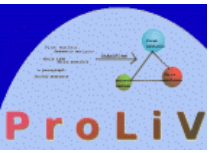
Universität Hamburg

MULTIMEDIA KONTOR HAMBURG



ProLiV - Animated Models for Visualizing Complex Linguistic Theories

Irina Aleksenko, Christina v.Bremen, Monica Gavrila, Walther v.Hahn, Angelika Redder, Cristina Vertan

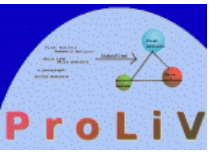
LSP, August 2007

Contents

- General Overview of the Project - W. v.Hahn
- System Architecture - M. Gavrila, C. Vertan
- Modules:
  - LSA Module, Lexicon Formats Module – M. Gavrila
  - Question Modelling Module (*German*) – A. Redder, C v.Bremen
  - Topic-Focus Module (*German*) – I. Aleksenko
- Demo

Universität Hamburg

MULTIMEDIA KONTOR HAMBURG



PROLIV - General Overview of the Project

Walther v.Hahn

University of Hamburg • Computer Science Department

Natural Language Systems Group

WWW: http://nats-www.informatik.uni-hamburg.de/view/User/WaltherVHahn

E-Mail: vhahn@informatik.uni-hamburg.de

Aim

- The PROLIV project aims at increasing the mutual knowledge of relevant, but complex theories and technologies by efficient learning-oriented information.
- Humanities students should understand and use some relevant formal methods, and
- Computer science students should acquire knowledge about complex linguistic interpretation of human communication.

Efficient learning-oriented information?

- Alternatives to text-oriented learning,
- Multimedia presentation,
- Exercises,
- Visualization,
- Animated examples.

Technical Goals

- The development of interactive multi-media e-Learning modules for first-year courses in (computational) linguistics and literary studies.
- A platform-independent implementation
- The integration of such tools and their e-Learning contents into a homogeneous surface within "WebCT“ (the e-Learniung system of Hamburg University)

Modules

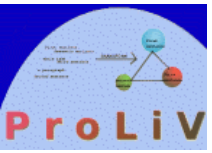
- Latent Semantic Analysis (LSA) → linguists
- Lexicon encoding and formats → linguists, computer scientists
- Question-Answer-Modelling → computer scientists
- Topic-Focus-Articulation → computer scientists
- Hidden Markov Models→ linguists, computer scientists

Integration

- The e-learning environment as well as the single tools have been tested and evaluated in real courses,
- Can be attached to other e-learning environments,
- Will be used in future courses in linguistics and computer science.

Universität Hamburg

MULTIMEDIA KONTOR HAMBURG



PROLIV - System Architecture Modules: LSA, Lexicon formats

Monica Gavrila, Cristina Vertan

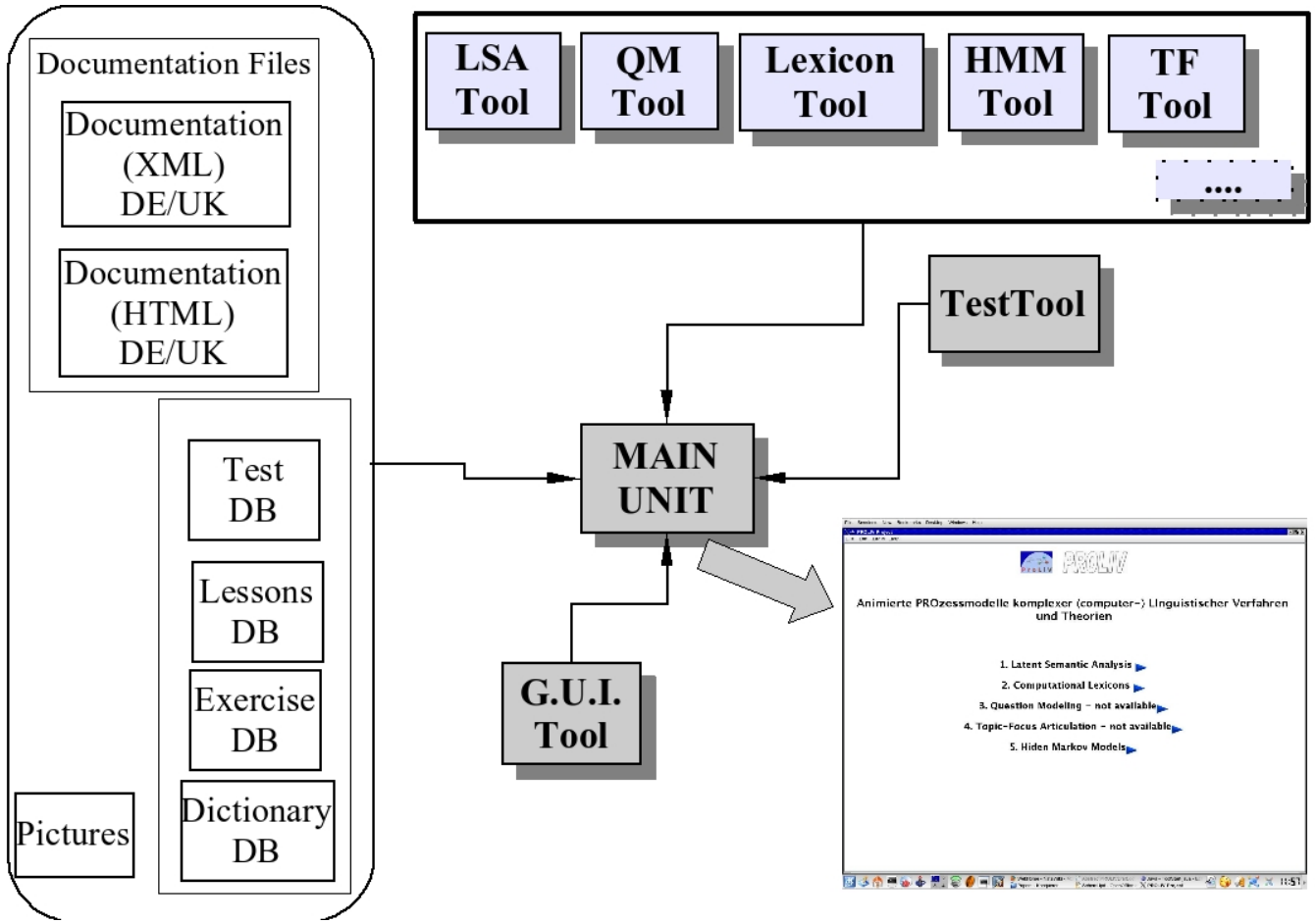
University of Hamburg • Computer Science Department

Natural Language Systems Group

E-Mail: gavrila@informatik.uni-hamburg.de

PROLIV Architecture

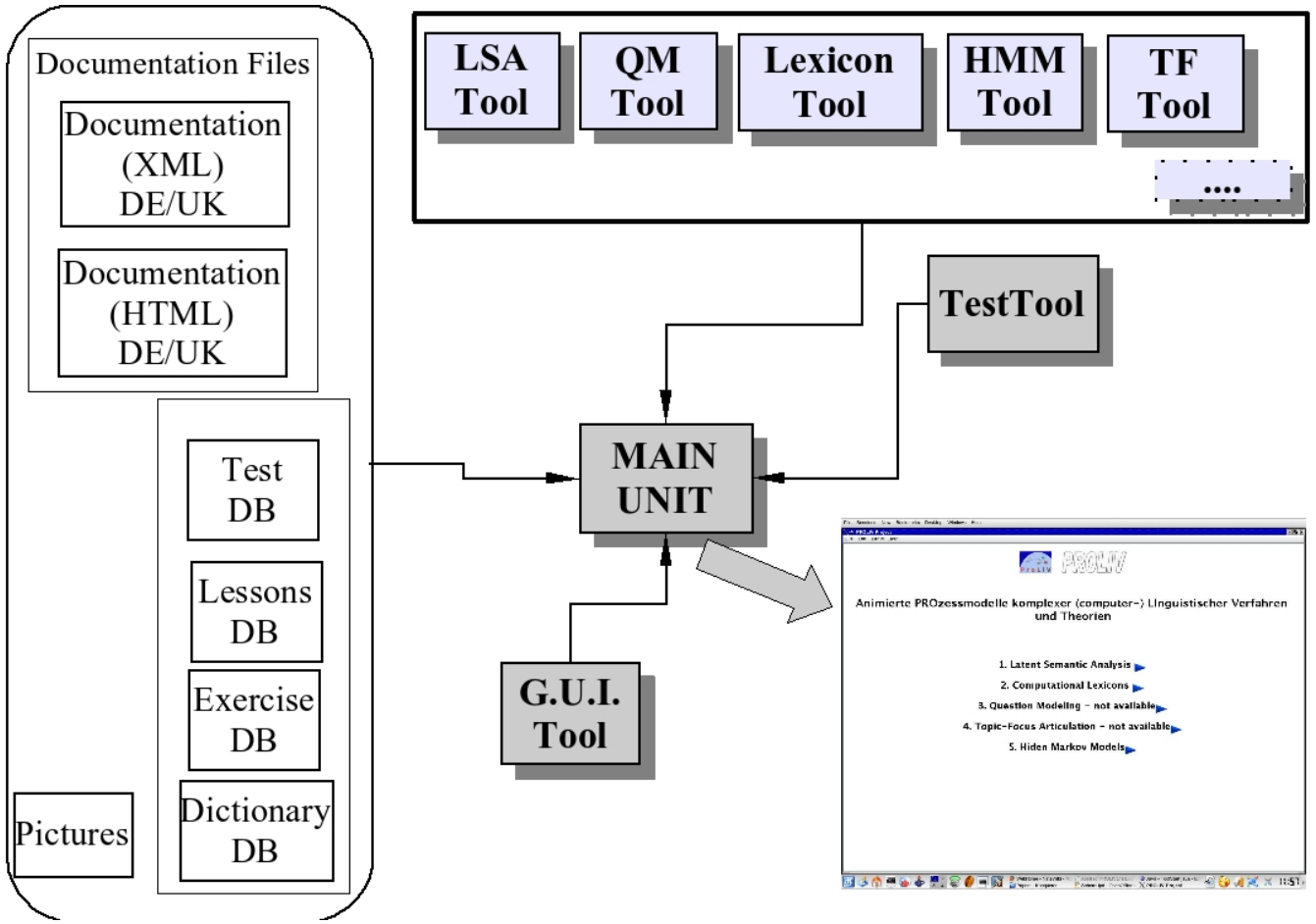
- **Modules:**
  - **GUI**
  - **MAIN**
  - **LEXT**
  - **QMT**
  - **LSAT**
- **Files:**
  - **Lesson content**
  - **Animations**
  - **Exercises**
  - **System documentation**



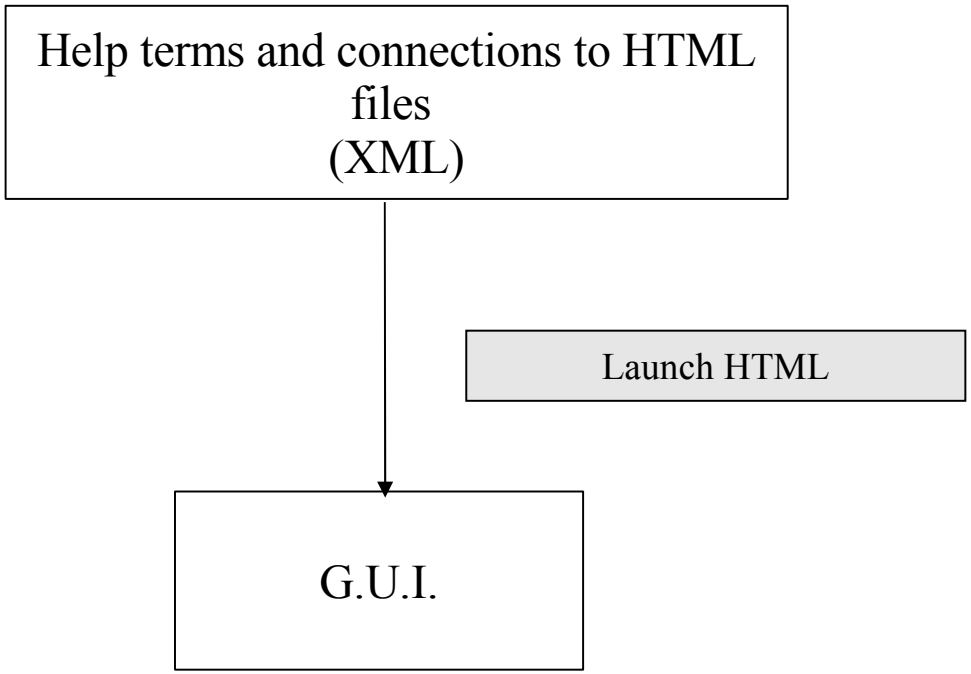
Languages considered: German and English

PROLIV Architecture - 2

- The system has a flexible structure, that is:
- ***modular,***
  - ***minimum interaction, and***
- ***extensible***
  - ***new tools***
  - ***new lessons***
  - ***.....***
- ***consistent***



Generation of the G.U.I. for the System Documentation

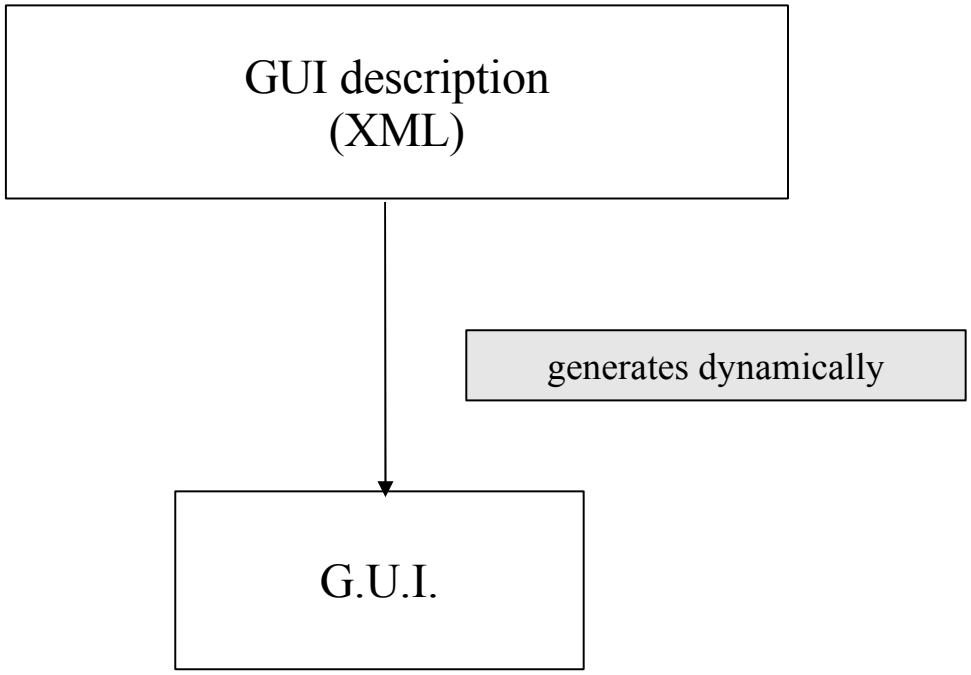



```
<?xml version="1.0"
encoding="ISO8859-1" ?>
<help lang="english">
 <topic name="Help and Support
Center,,>
  <text>HelpIndexUK.html</text>
  <search>help general</search>
  <topic name="Using Help">
   <text>UsingHelpUK.html</text>
   <search>using help window</search>
   <topic name="Searching Help,,>
    <text>SearchHelpUK.html</text>
    <search>search help</search>
   </topic>
  </topic>
</topic>
```

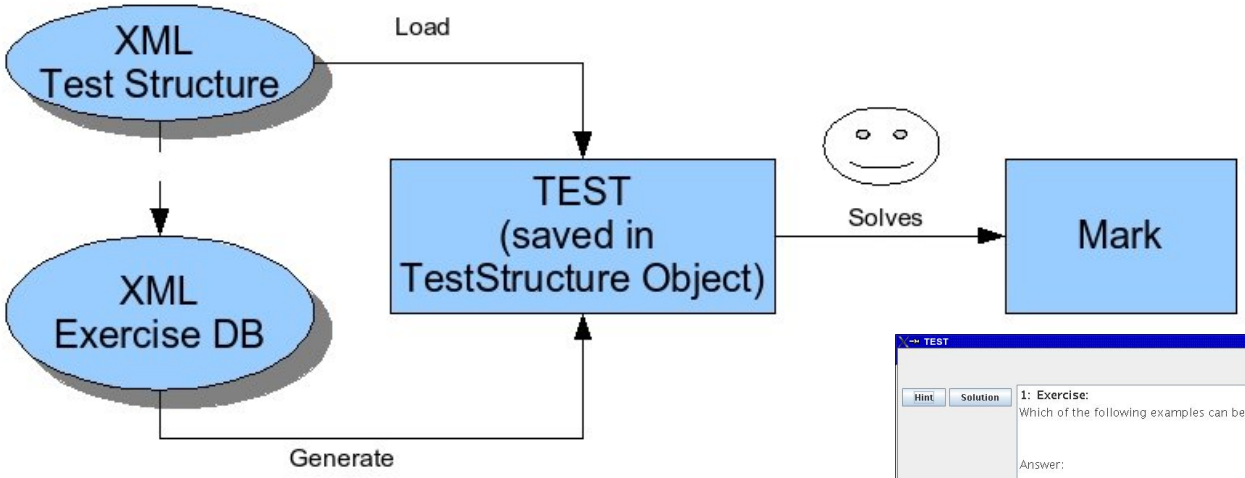
Generation of the G.U.I. for the Lessons

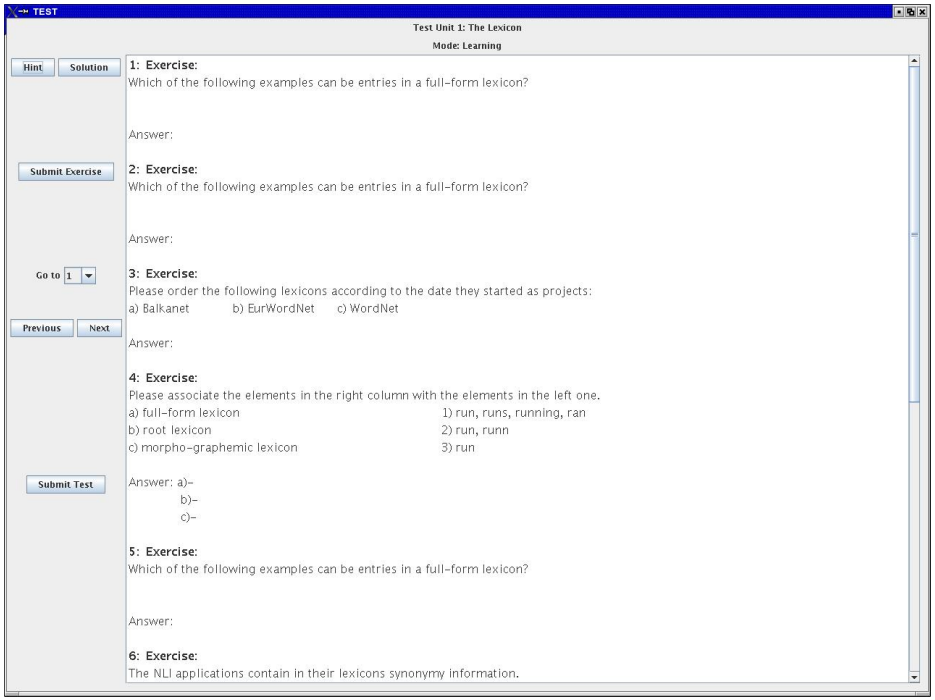
Advantages

- Information well structured
- Easy to update
- GUI generation work done once
- Extendible



Generation of the Tests

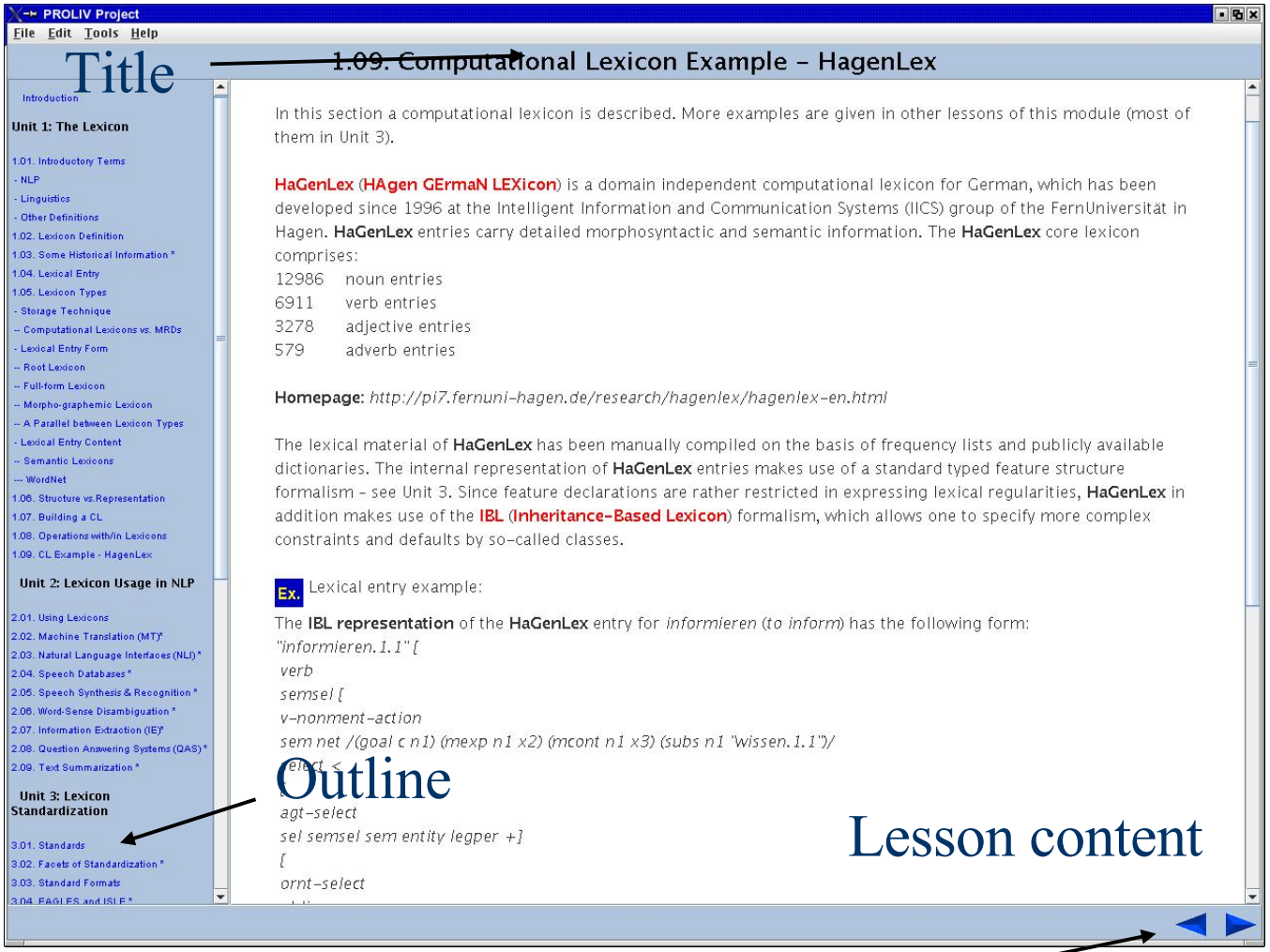




Two ways of accessing a test:
- learning mode
- testing mode

Types of „Screens“

- Course/lesson “screen”
  - Title
  - Content
  - Outline
  - Forward / Backward
- Dictionary “screen”
- Test “screen”
- Help “screen”
- Other “screens”



PROLIV Implementation

- The Software is implemented in Java. There are used java 1.5 and jdom-1.0. The tool used for implementation is Eclipse 3.1 . Everything is used is freeware and can be downloaded from the following websites:
  - JAVA: java.sun.com,
  - JDOM: www.jdom.org, and
  - ECLIPSE: www.eclipse.org.

PROLIV Tools

- Modules / Tools:
  - Main Unit
  - G.U.I. (DE/UK)
  - LSAT (Latent Semantic Analysis Tool)
  - LexT (Lexicon Tool)
  - QMT (Question Modelling Tool) *detailed*: A. Redder
  - TFT (Topic-Focus Tool) *detailed*: I. Aleksenko
  - HMMT (Hidden Markov Models Module)

Main Unit

- Gathers all modules
- Makes connections between modules – if necessary

- Makes the connection between a module and:
  - XML description files (lessons, etc.)
  - HTML description files (system documentation)

G.U.I.

- Generates the Graphical User Interface

LSA Module

- Lessons:
  - Utility and comparison with other similar algorithms (theoretical level)
  - Basic definitions and matrix algorithms (SVD)
  - General LSA algorithm description
  - Examples – algorithm explained
  - References (books, articles, links)
- Tests
- Dictionary
- Tool: possibility of seeing the results of the LSA algorithm on a very small corpus; possibility of choosing between some parameters

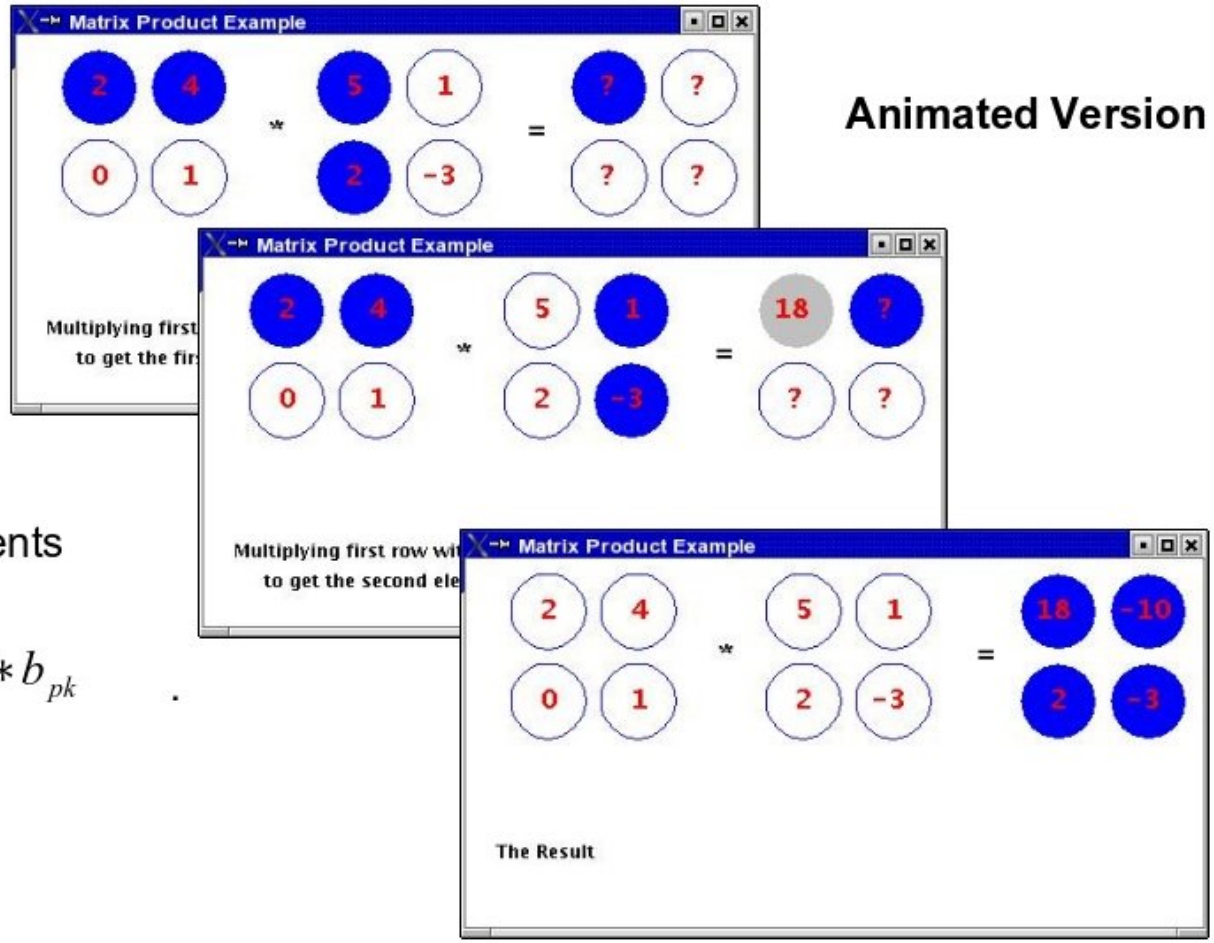
Images/Animations vs. Formulas

Text Version

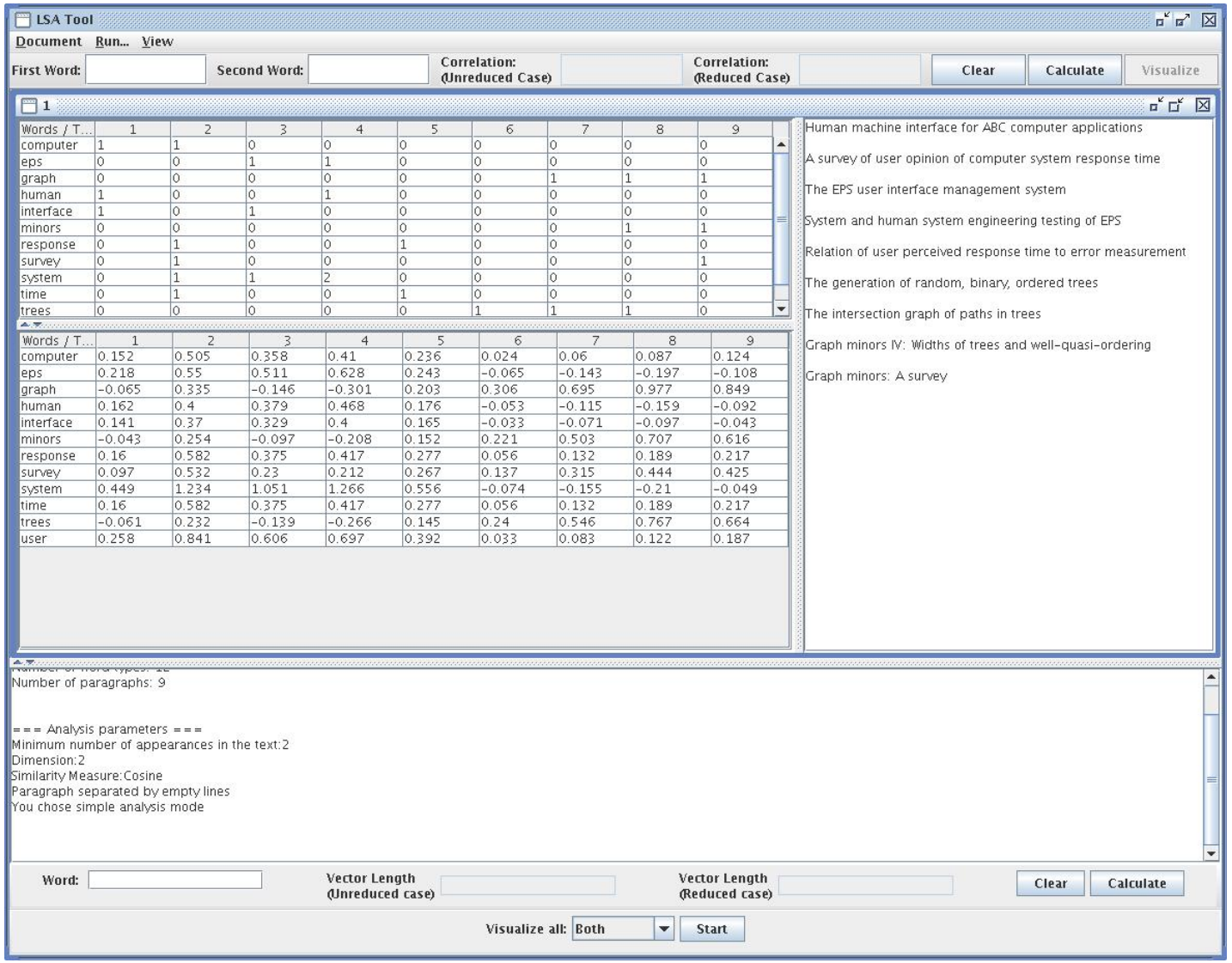
Let $A=(a_{ij})$ a matrix of dimension mxp and $B=(b_{jk})$ a matrix of dimension pxn . We want to calculate A*B . A*B is a matrix $C=(c_{ik})$ of dimension mxn , where the elements have the following form:

$$c_{ik}=a_{i1}*b_{1k}+a_{i2}*b_{2k}+\ldots+a_{ip}*b_{pk} \quad .$$

Animated Version

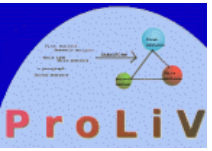


LSA Module


LSA Tool
Document Run... View
First Word:
Second Word:
Correlation: (Unreduced Case)
Correlation: (Reduced Case)
Clear
Calculate
Visualize
1
Words / T... 1 2 3 4 5 6 7 8 9
computer 1 1 0 0 0 0 0 0 0
eps 0 0 1 1 0 0 0 0 0
graph 0 0 0 0 0 0 1 1 1
human 1 0 0 1 0 0 0 0 0
interface 1 0 1 0 0 0 0 0 0
minors 0 0 0 0 0 0 0 1 1
response 0 1 0 0 1 0 0 0 0
survey 0 1 0 0 0 0 0 0 1
system 0 1 1 2 0 0 0 0 0
time 0 1 0 0 1 0 0 0 0
trees 0 0 0 0 0 1 1 1 0
Words / T... 1 2 3 4 5 6 7 8 9
computer 0.152 0.505 0.358 0.41 0.236 0.024 0.06 0.087 0.124
eps 0.218 0.55 0.511 0.628 0.243 -0.065 -0.143 -0.197 -0.108
graph -0.065 0.335 -0.146 -0.301 0.203 0.306 0.695 0.977 0.849
human 0.162 0.4 0.379 0.468 0.176 -0.053 -0.115 -0.159 -0.092
interface 0.141 0.37 0.329 0.4 0.165 -0.033 -0.071 -0.097 -0.043
minors -0.043 0.254 -0.097 -0.208 0.152 0.221 0.503 0.707 0.616
response 0.16 0.582 0.375 0.417 0.277 0.056 0.132 0.189 0.217
survey 0.097 0.532 0.23 0.212 0.267 0.137 0.315 0.444 0.425
system 0.449 1.234 1.051 1.266 0.556 -0.074 -0.155 -0.21 -0.049
time 0.16 0.582 0.375 0.417 0.277 0.056 0.132 0.189 0.217
trees -0.061 0.232 -0.139 -0.266 0.145 0.24 0.546 0.767 0.664
user 0.258 0.841 0.606 0.697 0.392 0.033 0.083 0.122 0.187
Human machine interface for ABC computer applications
A survey of user opinion of computer system response time
The EPS user interface management system
System and human system engineering testing of EPS
Relation of user perceived response time to error measurement
The generation of random, binary, ordered trees
The intersection graph of paths in trees
Graph minors IV: Widths of trees and well-quasi-ordering
Graph minors: A survey
Number of paragraphs: 9
=== Analysis parameters ===
Minimum number of appearances in the text:2
Dimension:2
Similarity Measure:Cosine
Paragraph separated by empty lines
You chose simple analysis mode
Word:
Vector Length (Unreduced case)
Vector Length (Reduced case)
Clear
Calculate
Visualize all:
Both
Start

Lexicon Module

- Lessons
  - General definitions, types, usage, examples
  - Standards
  - Lexicon structure / model (grammatical features) vs. encoding / format vs. content
  - References
- Other tools: ManageLex, GERL: for lexicon structure, entries, etc
- Tests
- Dictionary



Linguistische Analyse von Fragen - interaktive Modellierung

(Teilprojekt „Fragemodellierung“ im ELCH-Projekt PROLIV 2005-7)

Angelika Redder
& Christina von Bremen

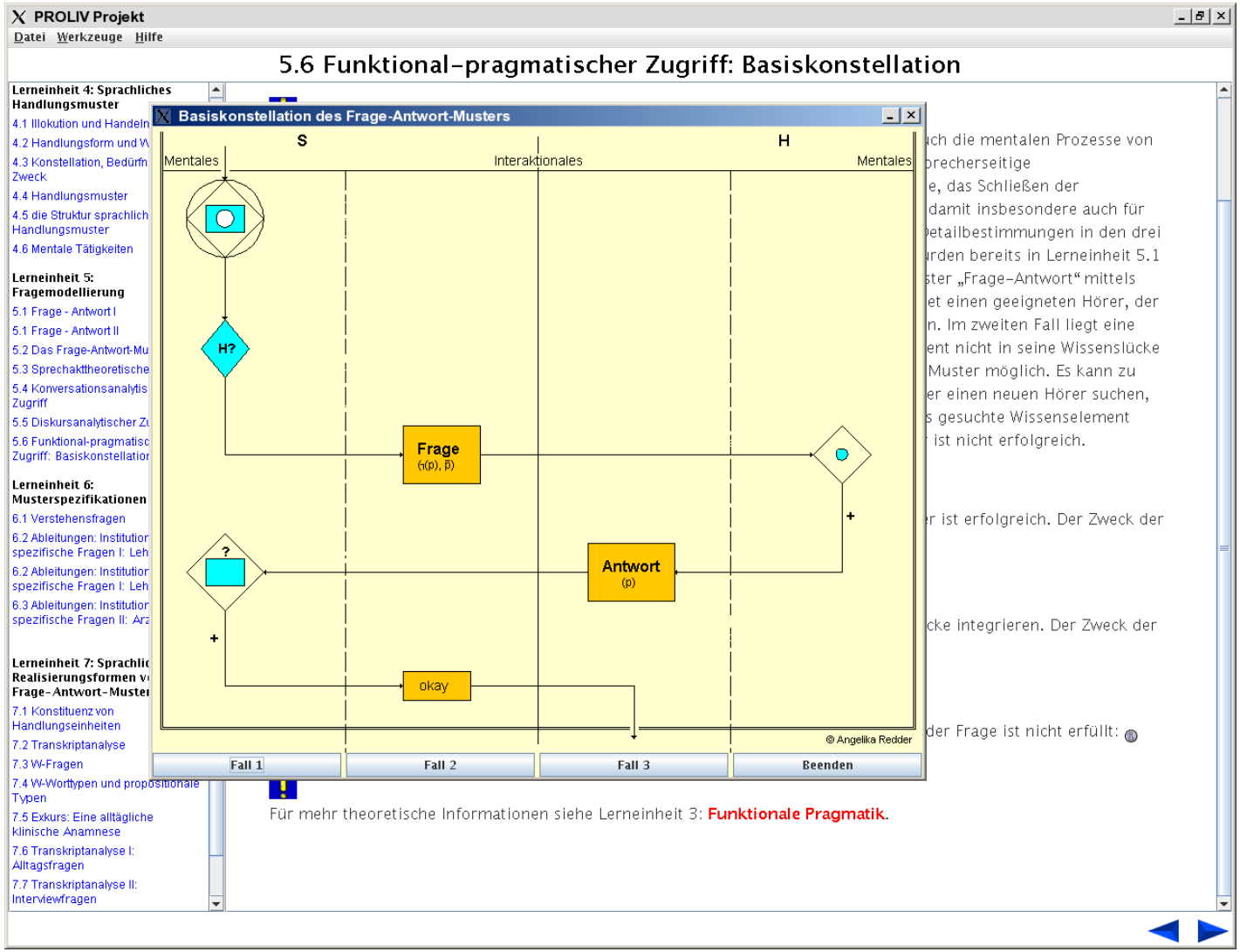
Linguistische Analyse von Fragen - interaktive Modellierung

- „Animierte PROzessmodelle komplexer (computer-) LInguistischer Verfahren und Theorien“
- Entwicklung von interaktiven Lernobjekten hauptsächlich für die Vertiefung von Grundstudiumsveranstaltungen in Linguistik und Informatik (Sprachverarbeitung)
- Visualisierung komplexer Verfahren und Theorien

Linguistische Analyse von Fragen - interaktive Modellierung

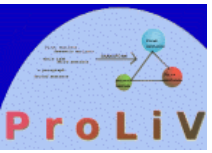
- Einführung in pragmatische Theorien mit Bezug auf die Sprechhandlung Frage
- Darstellung des Frage-Antwort-Musters in der Funktionalen Pragmatik
- Einzelzugriffe auf mentale und interaktionale Musterpositionen
- Vergleich mit Sprechakttheorie und Konversationsanalyse
- Abgeleitete Fragemuster (z.B. Lehrerfrage, Prüfungsfrage, Arztfrage)
- Exemplarischer Bezug zu einem Transkriptausschnitt

Linguistische Analyse von Fragen - intera



Universität Hamburg

MULTIMEDIA KONTOR HAMBURG



Topik-Fokus Annotation für ein Visualisierungsmodul im Projekt ProLiV

Iryna Aleksenko

Hamburg Universität • Department Informatics

Natürlichsprachliche Systeme

Email: Irina.Aleksenko@gmx.de

TFA-Einführung

- Topik und Fokus sind Begriffe, mit denen wesentliche Aspekte der Informationsstruktur eines Satzes beschrieben werden.
- Bei der Informationsstruktur einer Äußerung handelt es sich um die Art und Weise, in der die mit der Äußerung übermittelte Informationen sprachlich geglidert sind.
- Mit ihr wird auf das bereits vorhandene Wissen der jeweiligen Rezipienten Bezug genommen.

Quellen: PD Dr. Johannes Dölling
http://www.uni-leipzig.de/~doelling/veranstaltungen/topikfokus1.pdf

TFA-Einführung

- Grundlegend für die Analyse der Informationsstruktur sind die nur schwer definierbaren Begriffe der Bekanntheit oder Vorerwähntheit und der Neuheit. Zu den zentralen informationsstrukturellen Kategorien gehören:
  - Topik = der Gegenstand der Aussage und
  - Fokus = wichtige, oder neue Information

TFA-Einführung

a. Where will Mary drive tomorrow?

Mary will drive to Prágue tomorrow.

*Máry will drive to Prague tomorrow.

b. Who will drive to Prague tomorrow?

Máry will drive to Prague tomorrow.

*Mary will drive to Prágue tomorrow.

Quelle: http://www.uni-leipzig.de/%7Edoelling/veranstaltungen/topikfokus1.pdf

TFA-Prager Schule

- Topik-Fokus Annotation soll die Struktur des Satzes wiedergeben, d.h. zwei unterschiedlichen Realisationen eine Satzes liegen zwei unterschiedliche TFA, oder zwei unterschiedliche tektogrammatischen Bäume zugrunde.

TFA-Prager Schule

- TFA basiert auf zwei Phänomenen, nämlich

1. **Kontextgebundenheit**

   Kontextgebundene Einheiten: repräsentieren dem Hörer bereits bekannte oder vorhersehbare Informationen.

- ***Kontext:*** Tom kommt zusammen mit seinen Freunden.
- Meine Mutter hat nur IHN erkannt, aber keinen von SEINEN Freunden.

Quelle: http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/pdf/t-man-en.pdf

TFA-Prager Schule

Nicht kontextgebundene Elemente sind aus dem Kontext nicht bekannt.

c	the node represents a contrastive contextually bound expression
f	the node represents a contextually non-bound expression
t	the node represents a non-contrastive contextually bound expression

TFA-Prager Schule

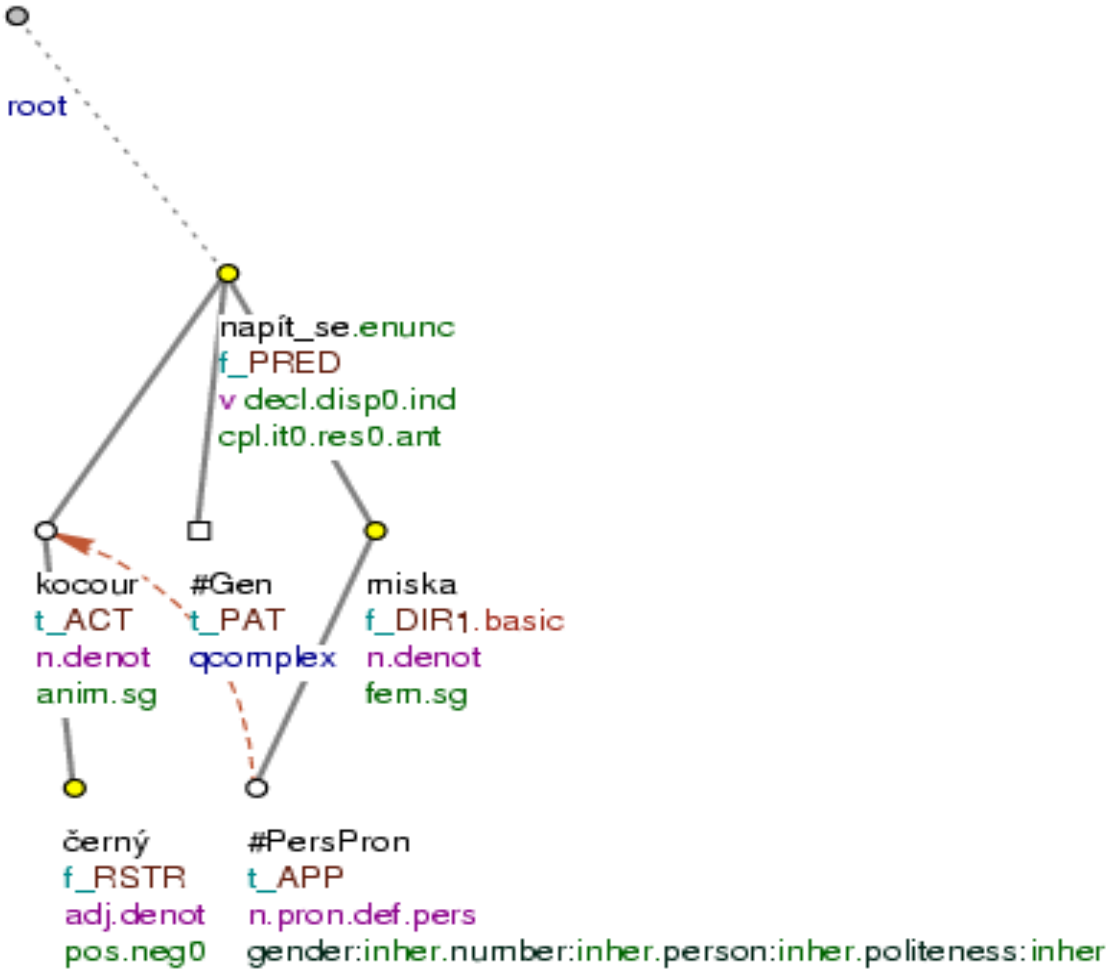
2. Kommunikativer Dynamismus (Knotenordnung)

„Communicative dynamism is a property of an expression that reflects its relative degree of importance in comparison with other expressions in the sentence attributed to it by the speaker; we consider contextually non-bound expressions to be more dynamic than expressions contextually bound (be they non-contrastive or contrastive).“

Quelle: Manfred Krifka :http://amor.rz.hu-berlin.de/%7Eh2816i3x/lehrstuhl.html

TFA-Prager Schule

Figure 10.4. Ordering of nodes in a tectogrammatical tree



Umsetzung im Projekt ProLiv. Wortordnung

- Durch die Wortstellung können „Akzente gesetzt werden“, d.h. sie leistet in der geschriebenen Sprache das, was in der gesprochenen Sprache durch den Akzent, die Betonung, ausgedrückt wird.
  - Kanonische Ordnung:
    - Ich habe meinem Freund das Auto geliehen.
  - mit Fokussierung:
    - Ich habe das Auto meinem Freund geliehen.
    - Meinem Freund habe ich das Auto geliehen

Quelle: Der kleine Duden. Deutsche Grammatik. 1997. S. 362-363

Umsetzung im Projekt ProLiv. Wortordnung

Englische Sätze:

Kanonische Ordnung: We went by car to a lake

We went to a lake by car

T F

Kanonische Ordnung : They moved from Chicago to Boston

They moved to Boston from Chicago

T F

Quelle: Eva Hajicovà, 2005, Information structure of the sentence and patterning of discourse

Annotierte Sätze. Wortordnung.

<SATZ ID="S1_001">

</KAN>

<TEXTE>Sie erhalten von Ihrem Diensteanbieter das Sperrkennwort.
</TEXTE>

</KAN>

<WO>

<BEISPIEL ID="1" ERKL="">

<**FOK TYP="Informativ" ID="1" ERKL="Die Verschiebung der Dativergänzung in die initiale Position führt zu Fokussierung dieses Elementes"**>Von Ihrem Diensteanbieter </FOK>

<TEXTE> erhalten Sie das Sperrkennwort.</TEXTE>

</BEISPIEL>

……..

Annotierte Sätze. Wortordnung.

<BEISPIEL ID="2" ERKL="">

<TEXTE>Sie erhalten das Sperrkennwort </TEXTE>

<**FOK TYP="Informativ" ID="2" ERKL="Wortfolge Akkusativ vom Dativ kann zu Betonung der letzen Position führen"**>von Ihrem Diensteanbieter.</FOK>

</BEISPIEL>

<BEISPIEL ID="3" ERKL="">

<**FOK TYP="Informativ" ID="3" ERKL="Das Erscheinen des Akkusativ Objekts an erster Stelle hat seine Fokkusierung zu Folge"**>Das Sperrkennwort </FOK>

<TEXTE>erhalten Sie von Ihrem Diensteanbieter.</TEXTE>

</BEISPIEL>

</WO>

</SATZ>

Textannotation- Artikel

<SATZ ID="S1_001">

<**FOK TYP="Informativ" ID="F_001" ERKL="Neue informationen. Kein Kontext vorhanden."**> Junger, dynamischer Modernisierer gegen alten, gutherzigen Bewahrer.</FOK>

</SATZ>

Textannotation- Artikel

<SATZ ID="S2_002">

<**TOP TYP="Kontrastiv" ID="T_001" REF="S1T()F(1)"**
ERKL= "Bezug auf vorhandene Information 'Modernisierer' und Konfrontierung zu 'Bewahrer',,
>Michel Platini und </TOP>

<**TOP TYP="Kontrastiv" ID="T_002,, REF="S1T()F(1)"**
ERKL= "gleich zu T1"> Lennart Johansson </TOP>

<**FOK TYP="Informativ" ID="F_002" ERKL= "Neue Information über T1 und T2"**> die Bewerber </FOK>

<**FOK TYP="Informativ" ID="F_003" ERKL= "Neue Information"**> um den Präsidentenposten der Uefa, könnten unterschiedlicher kaum sein </FOK>

</SATZ>

Textannotation- Artikel

<SATZ ID="S3_003">

<**TOP TYP="Informativ" ID="T_003,, REF="S2T(2)F()" ERKL= ""**>Lennart Johansson ist </TOP>

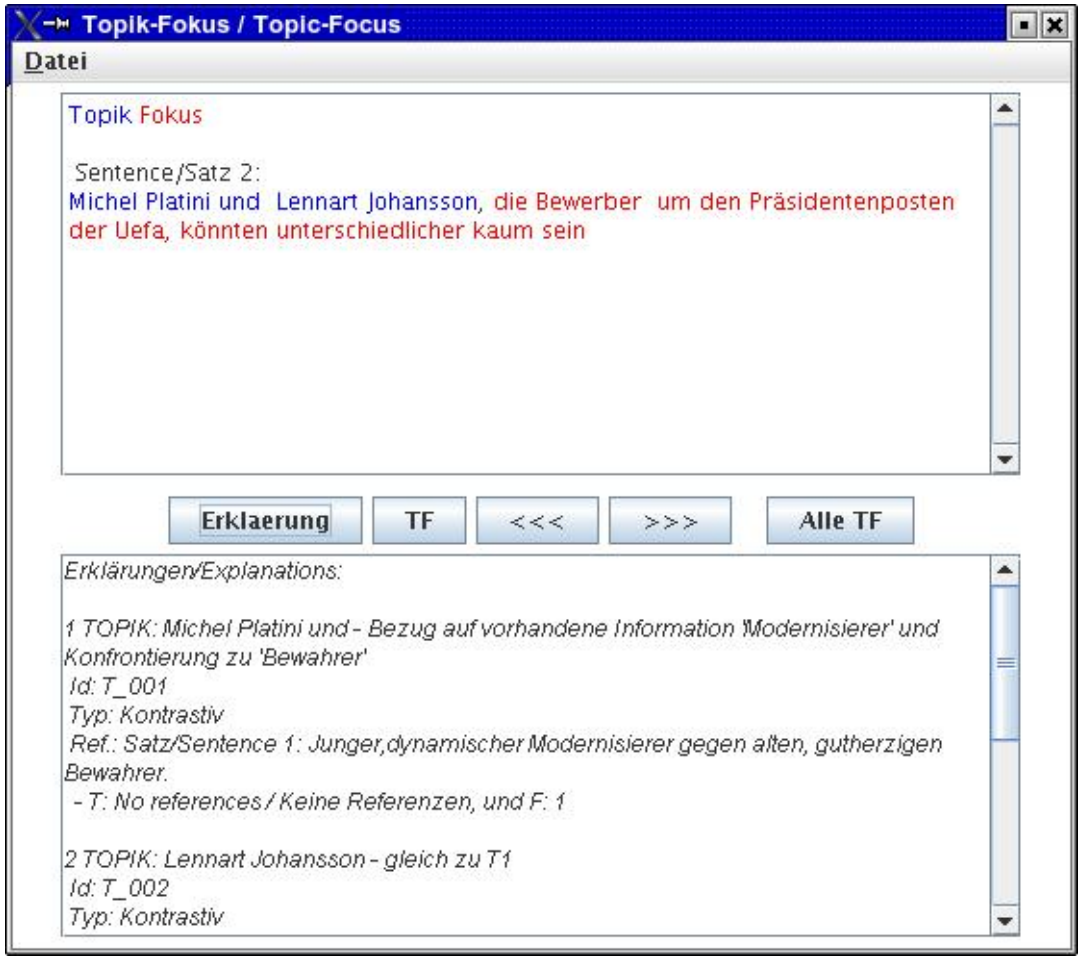
<**FOK TYP="Informativ" ID="F_004" ERKL= "Neue Information"**>kein besonders guter Wahlkämpfer -</FOK>

<**TOP TYP="Informativ" ID="T_0003" REF="S3T(3)F(4)" ERKL= ""**> er war es </TOP>

<**FOK TYP="Informativ" ID="F_0006" ERKL= ""**> noch nie.</FOK>

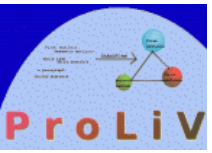
</SATZ>

Topik - Fokus


Topik-Fokus / Topic-Focus
Datei
Topik Fokus
Sentence/Satz 2:
Michel Platini und Lennart Johansson, die Bewerber um den Präsidentenposten der Uefa, könnten unterschiedlicher kaum sein
Erklaerung
TF
<<<
>>>
Alle TF
Erklärungen/Explanations:
1 TOPIK: Michel Platini und - Bezug auf vorhandene Information 'Modernisierer' und Konfrontierung zu 'Bewahrer'
Id: T_001
Typ: Kontrastiv
Ref.: Satz/Sentence 1: Junger,dynamischer Modernisierer gegen alten, gutherzigen Bewahrer.
- T: No references / Keine Referenzen, und F: 1
2 TOPIK: Lennart Johansson - gleich zu T1
Id: T_002
Typ: Kontrastiv

Universität Hamburg

MULTIMEDIA KONTOR HAMBURG



Demo and Project Information

Run Project

- **Multimedia Kontor Project, University of Hamburg**
- Beginning: 01.06.2005 / End: 31.12.2007
- Faculty of Mathematics, Informatics and Natural Sciences,Computer Science Department, Natural Language Systems Division &
- Faculty of Humanities, Department of Language, Literature and Media, Institute for German Studies I
- Homepage: https://nats-www.informatik.uni-hamburg.de/view/PROLIV/WebHome

Project Members

- **Walther v. Hahn**, Department of Informatics, Natural Language Systems Group, Department of Language, Linguistics and Media Studies
- **Angelika Redder**, Department of Language, Linguistics and Media Studies, Institute for German Studies I
- **Cristina Vertan**, Department of informatics, Natural Language Systems Group,
- **Shinichi Kameyama**, Department of Language, Linguistics and Media Studies, Institute for German Studies I
- **Monica Gavrila**, Department of Informatics, Natural Language Systems Group
- **Christina von Bremen**, Computer Science Department, Natural Language Systems Group
- **Olga Szczepanska**, Department of Informatics, Natural Language Systems Group
- **Irina Aleksenko**, Department of Informatics, Natural Language Systems Group
- **Svetla Boytcheva**, Academy of Sciences Sofia