

Resources for Computational Linguistics

Annotation Tools: RSTTool & MMAX

Presentation by Ekaterina Biehl



In the Last Session

- Corpus Linguistic;
- Corpus;
- Unannotated;
- Annotated;
- **Annotation;**
 - Levels of Annotation;
 - POS- Tagging;
 - Grammatical Parsing;
 - Semantic Tagging;
 - Discoursal and Text Annotation
(RST, Discourse Tags, Anaphoric Annotation);
 - Prosodic Annotation....

So,



We need annotation tools!

Annotation Tools

What is important?

- should be able to do your task;
- speed, stability, and practical usability;
- ready and easy to use;
- standardized input/output format (*XML*).

Today' Session

1. Text analysis => **RSTTool**;
2. Multi-level annotation => **MMAX**;
3. Examples.

Questions

1. What is to be annotated?
2. What are the *markables*?
3. What is the guideline and *annotation scheme*?

RSTTool



Michael O'Donnell

RSTTool

Graphical interface to facilitate the marking up of the RST structure of text =>

- segmentation of text;
- graphical linking of the segments into an RST Tree.

RST = Rhetorical Structure Theory

- offers an explanation to COHERENT texts
 - i.e. with no gaps and non-sequiturs;
- describes text structure by means of „BUILDING BLOCKS“ at
 - principal level
 - „nuclearity“ <> „relations“;
 - second level schemas.

„Nuclearity“

➤ Mononuclear: Nucleus => Satellite

Relation Name	Nucleus	Satellite
Background	text whose understanding is being facilitated	text for facilitating understanding
Elaboration	basic information	additional information
Preparation	text to be presented	text which prepares the reader to expect and interpret the text to be presented.

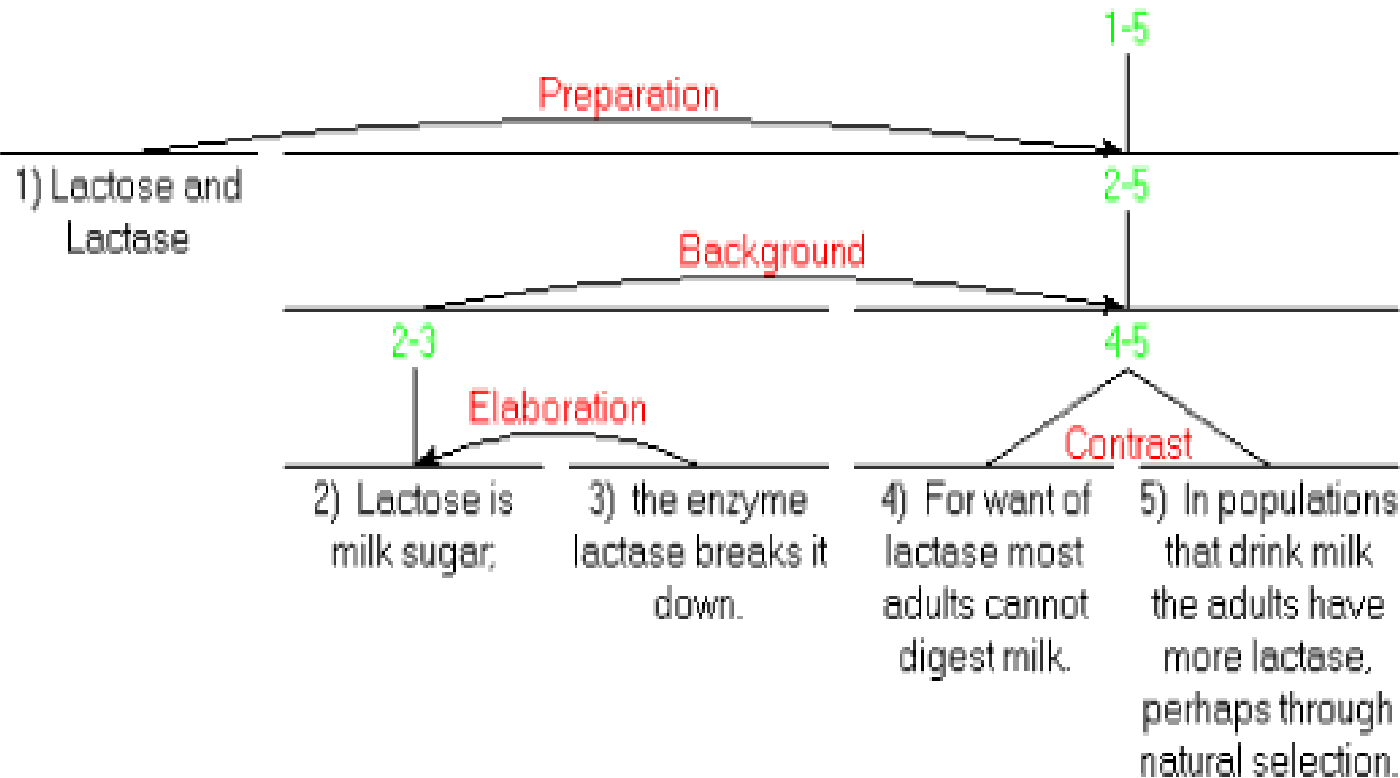
➤ Multi-nuclear: Span = Other Span

Relation Name	Span	Other Span
Contrast	one alternate	the other alternate

Rhetorical Structure Theory

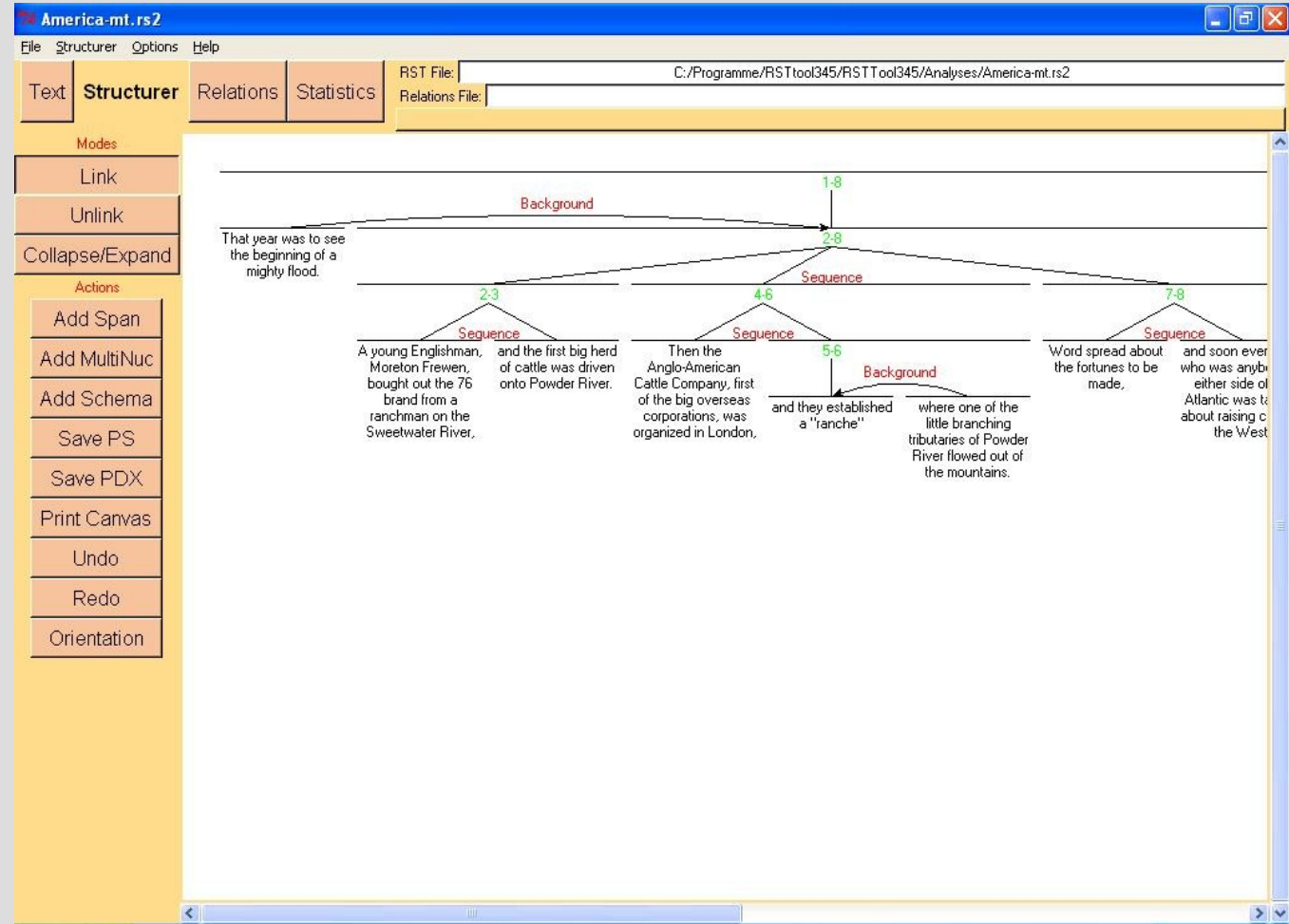
Example of the analysis tree:

Scientific
American,
October 1972.



RSTTool

Example of
the
annotation



RSTTool: Summary

1. Annotation tool for particular purpose;
2. Tree visualisation;
3. Graphical interface;
4. Analysis time reduction;
5. Statistics.

MMAX



Dr. Michael Strube

was developed at
EML Reserch,
Heidelberg

Christoph Müller

MMAX

- „light-weight and highly customizable annotation tool“ (Müller & Strube (2001a, 2001b, 2003);
- supports the multi-level annotation of (potentially multi-modal) corpora;
- based on the concept of *markables* carrying *attributes* and standing in certain *relations* to each other.

Concepts: Markable

- carries the annotation information;
- can be defined on arbitrary levels of linguistics annotation;
- is an entity that can consists of arbitrary sets of elements from the data base;

Concepts: Markable II

- can represent multiple levels of linguistic description;
- can be overlapping or discontinuous.
- the principle of STAND-OFF annotation

Concepts: Attributes

- markables can have arbitrarily many attributes (name-value pairs);
- *nominal* attributes which have a closed set of possible values.

Concepts: Relations

➤ relations between markables:

- **member-relation** : markables having the same value in an attribute;
- **pointer-relation**: directed relations between a source markable and arbitrarily many target markables.

Guideline => Annotation Scheme

➤ **Guideline=Instruction**

➤ **Annotation scheme = formal guideline:**

- describes which phenomena are to be annotated using which set of attributes;
- defines all attributes for a linguistic level.

MMAX: the Tool

- is written in Java;
- XML;
- consists of main annotation window, Search window, attribute window.

EXtensible Markup Language

- XML is a **markup language** much like HTML;
- XML was designed to **describe data** and to focus on **what data is**;
- XML tags are not predefined. You must **define your own tags**

```
<note>  
<to>Tove</to>  
<from>Jani</from>  
<heading>Reminder</heading>  
<body>Don't forget me this weekend!</body>  
</note>
```

- XML does not **DO** anything. XML was created to **structure, store and to send** information.

MMAX: the Tool

Example of
the
annotation

The screenshot displays the MMAX V0.94 build 77 interface. The main window shows a text document titled "Leben und Wirken von Georg Philipp Schmitt." with a paragraph of German text. The text is annotated with various grammatical roles and agreement markers. A search panel on the right allows for filtering annotations based on criteria like comment, np_form, grammatical_role, agreement, member, pointer, Pointer Target, exp_type, ante_sub_anaphoric, and ante_sub_bridging. Below the main text, a detailed annotation window for "Eva Kaysser und Georg Philipp Schmitt (markable_60)" is visible, showing a list of grammatical roles and their corresponding radio button options. The roles include comment, np_form, grammatical_role, agreement, member, pointer, and exp_type. The np_form options are none, ne, defnp, indefnp, pper, ppos, and pds. The grammatical_role options are none, sbj, and obj (with a value_3 dropdown). The agreement options are none, 3m, 3f, 3n, 3p, 1s, 2s, 1p, and 2p. The exp_type options are none, anaphoric, and bridging. The interface also includes buttons for "Apply", "Undo changes", "Search", "Reset", "AND/OR", "Previous match", "Clear", and "Next match". The status bar at the bottom indicates "AutoApply is OFF!".

MMAX V0.94 build 77 C:\MMAX094\samples\text\texsample.anno

File Tools Settings Help Tahoma 14 wrap display

Leben und Wirken von Georg Philipp Schmitt.

Am 28. Oktober 1808 wurde Georg Philipp Schmitt in Spesbach geboren. Den entscheidenden künstlerischen Anstoß für seine Entwicklung fand Schmitt mit noch nicht siebzehn Jahren, als er durch die Vermittlung Xellers, dessen Jugendfreund Peter von Cornelius vorgestellt wurde, der Gefallen an dem Talent des Jungen fand und ihn mit nach München nahm. Dort wurde Schmitt sein Schüler an der Akademie der bildenden Künste. In den Selbstbildnissen, die zwischen 1827 und 1828 entstanden kündigte sich bereits das große Talent Georg Philipp Schmitts an. Nach fünfjährigem Akademiebesuch ließ sich Georg Philipp Schmitt 1830 als "freier Maler" in Heidelberg nieder und wirkte hier bis zu seinem Tod im Jahr 1873 über vier Jahrzehnte. Den Künstler verbanden Freundschaften zu den Malern Christian Philipp Koester, Eduard von Steinle und zum Dichter Justinus Kerner. Im Dezember 1833 heirateten Eva Kaysser und Georg Philipp Schmitt. In jene Zeit fiel auch die Freundschaft mit dem Grafen Charles de Graimberg, der die "Kurfälzische Alterthümerhalle" gründete. Schmitt war als Mitsammler und Restaurator für die Graimbergsche Sammlung tätig und wurde 1844 zu deren hauptamtlichem Experten bestimmt. Auch vermittelte Schmitt Werke in andere Sammlungen. Im Jahre 1863 konnte Georg Philipp durch die Unterstützung seines Sohnes Guido in ein eigenes Haus am Klingenteich 6

Current Markable File: markables.xml

Eva Kaysser und Georg Philipp Schmitt (markable_60)

comment

np_form none ne defnp indefnp pper ppos pds

grammatical_role none sbj obj {value_3}

agreement none 3m 3f 3n 3p 1s 2s 1p 2p

member

pointer

exp_type none anaphoric bridging

Apply Undo changes

to front suppress check warn on extra attributes

AutoApply is OFF!

comment not <none selected>

np_form not <none selected>

grammatical_role not <none selected>

agreement not <none selected>

member any set not set

pointer any set not set

Pointer Target any set not set

exp_type not <none selected>

ante_sub_anaphoric not <none selected>

ante_sub_bridging not <none selected>

Search Reset AND OR

Previous match Clear Next match

defnp

definite noun phrase

MMAX: Summary

- annotation as set of simple concepts based on the notion of „markable“;
- almost any kind of annotation can be done;
- multiple levels;
- stand-off annotation;
- can express highly customizable annotation schemes;
- is compatible with ISO standard (ISO TC37 SC4).

Conclusion:

- ✓ RSTTool:
 - ✓ Rhetorical Structure Theory;
 - ✓ tree visualization;
 - ✓ dedicated annotation tool.
- ✓ MMAX:
 - ✓ flexible tool for almost any kind of annotation;
 - ✓ annotation refers to markables;
 - ✓ simple and customisable.

References:

Corpus Linguistics- Annotation:

- <http://www.coli.uni-saarland.de/courses/korbay/Complingres/Slides/corpora.pdf>
- <http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/contents.htm>
- <http://coli.lili.uni-bielefeld.de/forschung/xbrac/pdf/xbrac-dipperetal-sfb.pdf> - search='MMAX%2

RSTTool:

- <http://www.wagsoft.com/RSTTool>
- <http://www.sfu.ca/rst>

MMAX:

- <http://www.eml-research.de/english/research/nlp/download/sigdial03.pdf>
- <http://www.eml-research.de/english/research/nlp/download/mmax.php>

XML:

- <http://www.w3schools.com/>

ISO:

- <http://www.tc37sc4.org>

TEI

- <http://www.tei-c.org/>