

Simple Annotation Tools for Complex Annotation Tasks: an Evaluation

Stefanie Dipper
Michael Götze
Manfred Stede

University of Potsdam

XML-based richly annotated corpora

LREC-2004 workshop
May 29, 2004

Outline

1. Project context
2. Annotation tools
3. Evaluation criteria
4. Results of evaluation

Project Context

- SFB 632 (collaborative research center)
 - “Information Structure: the linguistic means for structuring utterances, sentences and texts”
 - Potsdam + Humboldt University, started autumn 2003
- Objective: determine factors of information structure
- Individual projects
 - collect a lot of data of typologically different languages
 - and annotate them on various levels (manually):
phonetics, morphology, syntax, semantics

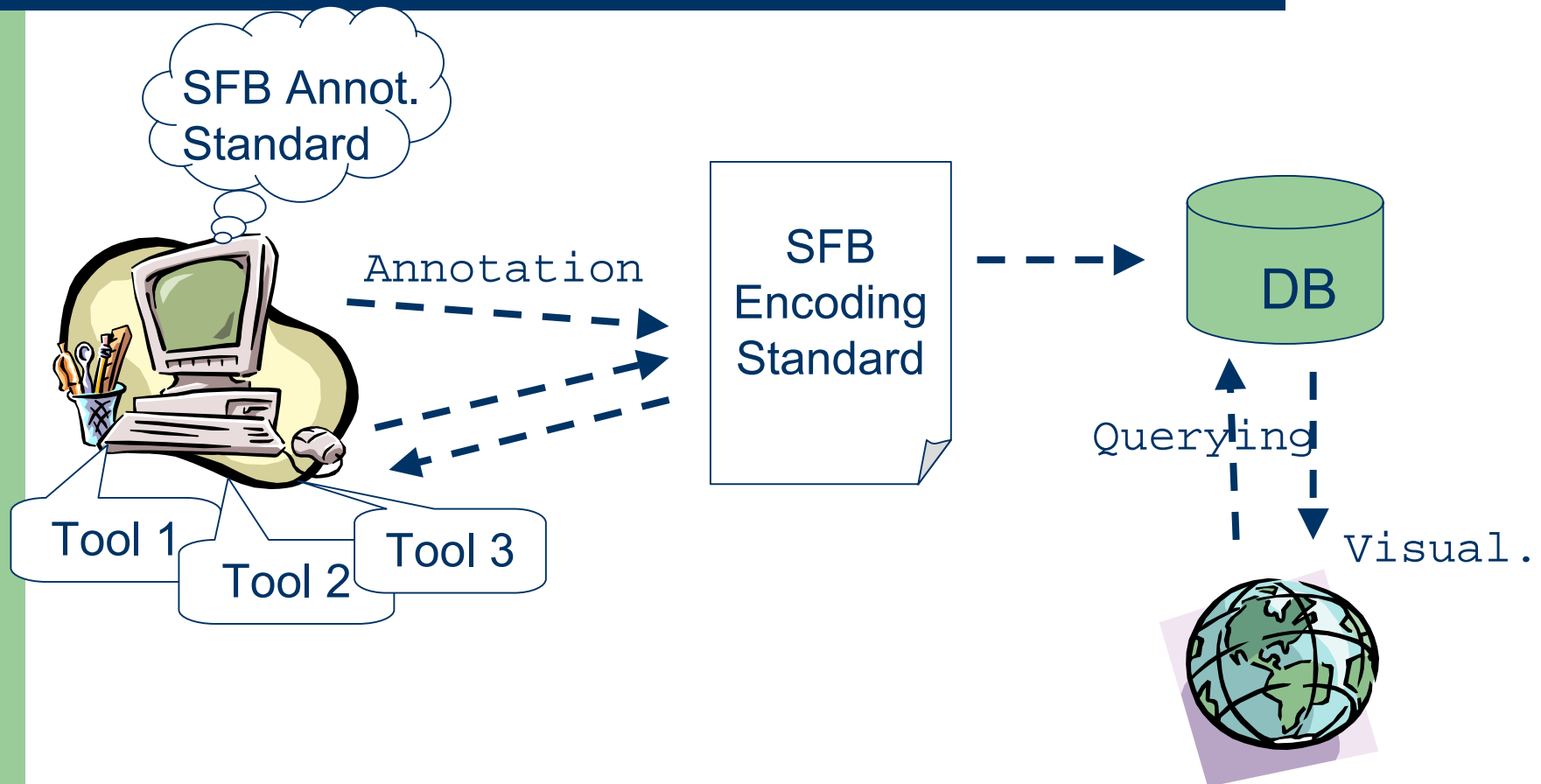
Project Context: Data and Annotation

- Semantics: quantifier scope, definiteness
- Discourse: rhetorical and prosodic structure
- Focus in African Languages
- Diachronic data
- Questionnaire for typologically diverse languages

Project Context: Standards

- Each project should profit from data of other projects
- Hence:
 - standardized annotation formats
(common tagsets and annotation criteria)
 - standardized encoding format (XML-based)
as the SFB-internal exchange format
 - (both under development)
- Database offers visualization and search facilities

Project Scenario



Requirements for Annotation Tools

- Diversity of data and annotation
 - written vs. spoken language, sentence vs. discourse
 - attribute-value pairs vs. pointers vs. graphs ...
 - multi-level annotation
- Convertibility
 - converters from/to other tools
 - standardized input/output format (XML)
 - > standardization, quality assurance

Requirements for Annotation Tools (cont'd)

- **Simplicity:**

Tools must be **ready and easy to use**

- annotators have no/few programming skills
- limited resources: annotating data is only one aspect of the project work
- tagsets will change from time to time
- annotation may be done during fieldwork
(no external support possible)

Requirements for Annotation Tools (cont'd)

Tools must

- run on any platform (Windows, Unix)
- be free of charge
- be maintained/actively supported
- be XML-based

-> selection criteria

Outline

1. Project context
2. Annotation tools
3. Evaluation criteria
4. Results of evaluation

2 Types of Annotation Tools

1. “Simple” tools
 - developed for special purposes
 - tuned
2. Complex tool kits
 - general-purpose tools
 - flexible, user-adaptable
 - tool offers platform, user defines application

“Simple” Tools: Specialized Tools

- Speech: Praat, TASX
- Discourse: MMAX, PALinkA, Systemic Coder
- Syntax: annotate
- ...

Complex Tool Kits

- NITE XML Toolkit
- AGTK (Annotation Graph Toolkit)
- CLaRK
- ...

SFB requirement 'ready and easy to use'

-> **simple tools, no tool kits**

(tool kits might be considered in future when SFB standards and annotation procedures have matured)

Annotation Tools: Tiers vs. Markables

1. Tier-based tools:
 - annotated information is represented by tiers
 - annotation is based on segments (“events”) that refer to common timeline
2. “Focus-based” tools:
 - annotation refers to markables
 - annotated information is visible for the currently active markable

Tier-based Tools

	0	1	2	3	4	5	6
[SOURCE]	Meenéenee	Tanko	yá	yíi	?	Bíya-n	h
[GLOSS]	what	T.	3sg.m.pf.dep	do		pay\N-of	ta
[CAT]	NP	NP	TAM	V			
[GF]	DO	SUBJ		PRED			
[IS]	FOC	BG					
[TRANS]	<i>What did Tanko do?</i>					<i>Tanko [paid t</i>	

“Focus-based” Tools

TRANS

GLOBB what

IS null foc bg

COMMENT

CAT null np pp vp v adjp advp ip cp foc tam unspec

GF null argument modifier other

GFarg null sbj do

to front

X -M MMAX V0.94 build 77 /home/sfb/tutorium/daten_annot/mmax/mn

File Tools Settings Help Bitstream Charter 14 wrap display

Meenéenee Tanko yá yíi ?

Bíya-n haráaji née Tanko yá yíi.

Current Markable File: /home/sfb/tutorium/daten_annot/mmax/mmax_haus

Evaluated Tools:

1. Tier-based Tools

- EXMARaLDA (Hamburg)
 - annotation of multi-modal data
 - dialogue, multi-lingual
- TASX Annotator (Bielefeld)
 - multi-modal transcription: speech, video
 - dialogue, prosody, gesture

Evaluated Tools:

2. “Focus-based” Tools

- MMAX (Heidelberg)
 - discourse, dialogue
 - coreference, bridging relations
- PALinkA (Wolverhampton)
 - discourse
 - anaphora resolution, centering, summarization
- Systemic Coder (WagSoft)
 - discourse
 - register analysis

Outline

1. Project context
2. Annotation tools
3. Evaluation criteria
4. Results of evaluation

Evaluation Criteria

- Criteria based on ISO 9126-1
(Software engineering – product quality)
- Criteria concern
 1. Functionality
 - checks presence of task-relevant features
 - concerns relation *tool – task*
 2. Usability
 - evaluates effort needed for use
 - concerns relation *tool – user*

Functionality: Properties of Primary/Source Data

- Which input formats?
 - discourse (= sequence of sentences)
 - speech
 - ...
- Is preprocessing necessary? (e.g. tokenizing)
- Is Unicode supported?

Functionality: Properties of Secondary Data (= Annotation)

- Which data structures?
 - atomic features
 - relations, pointers, trees
 - conflicting hierarchies
 - ...
- Which metadata?
 - header information
 - comments

Functionality: Interoperability

- Export/import
- Converters
- Plug-ins
- ...

Usability

Operability

- customizability by specifying annotation levels and tagsets
- (semi-)automatic annotation
- visualization of annotated information

Usability

- Documentation
 - help, tutorial, example files, ...
- Compliance
 - Does the tool adhere to standards/conventions?
e.g. shortcuts, undo/redo, copy/paste, ...
- Learnability, attractiveness
 - People should as much as possible enjoy annotation

Outline

1. Project context
2. Annotation tools
3. Evaluation criteria
4. Results of evaluation

Selected Results

Criteria that measure aspects of:

- Functionality
- Ready and easy to use
- Quality assurance
- Learnability, attractiveness

Functionality: Primary Data

- all tools
TASX
discourse
audio, video
- all tools
Unicode

Functionality: Secondary Data

- all tools atomic features
- all but Coder multi-level annotation
- MMAX, PALinkA relations, pointers
- PALinkA bracketing
- MMAX conflicting hierarchies

Ready and Easy to Use: Preprocessing

- TASX, EXMARaLDA
 - + no preprocessing or tagset specification necessary
annotation can start immediately
- MMAX, PALinkA, Coder
 - preprocessing and tagset specification obligatory
- Coder
 - + tool-internal tagset specification

Ready and Easy to Use: Compliance, Documentation

- TASX, EXMARaLDA
 - + copy/paste, shortcuts, ...
- EXMARaLDA
 - + tutorial (detailed walkthrough)

Ready and Easy to Use: Visualization

- MMAX, PALinkA, Coder
 - + nice visualization of primary data
- TASX, EXMARaLDA
 - + nice visualization of annotated information

Quality Assurance

- MMAX, PALinkA, Coder
 - + predefined tagsets (customizable)
- MMAX, Coder
 - + structured tagsets

Learnability, Attractiveness

SFB tutorial

- annotation of lemma, morphology, part of speech, constituents, ...
- no discourse relations, no co-reference
- tools: EXMARaLDA, MMAX, PALinkA
- participants were asked to complete a questionnaire

Learnability, Attractiveness

Results

- EXMARaLDA offers most attractive visualization
- despite script files, preprocessing of data (tokenizing) is difficult
- customization of tagsets is difficult

Conclusion I

- “Simple” tools offer a lot
- Tool suitability depends on annotation scenario

	TASX, EXMAR.	MMAX	PALINKA	Coder
Immediate A.	+	–	–	–
Consistent A.	0	+	+	+
Guided A.	–	+	0	+

Conclusion II

“Wishlist” to tool developers

- suitable visualization of source data **and** annotated information
- tool-internal tokenizer
- tool-internal interface for tagset customization