

Transcription and Annotation

Working Group Report

July 11-13, 2003

Electronic Metastructures for Endangered Language Data 2003 Workshop
East Lansing, Michigan

report by Arienne M. Dwyer

Working Group:

Bernard Comrie, Max Planck Institute - Leipzig, Co-Chair

Edward Garrett, Eastern Michigan U, Co-Chair

Arienne Dwyer, U of Kansas, Co-Chair

Sebastian Drude, Freie U Berlin/Museu Goeldi

Brenda Farnell, U of Illinois, Urbana-Champaign

D. Terence Langendoen, U of Arizona

Philippe Martin, Université de Paris 7

Rob Vann, Western Michigan U

Ljuba Veselinova, Stockholm U

Dietmar Zäfferer, Ludwig-Maximilians-Universität München

- Priorities
- Definitions
- Recommendations
- Desiderata
- Existing Tools + evaluation
- Presentation Formats

1. Priorities

Though it at first may seem counterintuitive, we suggested that the first priority of language documentation should be **translation**.

(As Bernard so famously noted, “We don’t want to create any more varieties of Linear A.”) The second priority, transcription, constitutes an entryway to documentation. From it, we can make inferences (and other forms of annotation) on meaning, grammatical structure, and content structure.

2. Definitions

Transcriptional Annotation

A transcription is a representation of the perceivable dimensions (form) of the sign appropriate to a description of the modality (gesture, speech, writing).

1. A transcription is information needed for a speaker or machine to reproduce the linguistic form including co-linguistic (paralinguistic [e.g. tone of voice] / situational [e.g. pauses, self-corrections]).
2. Forms of transcription, which are always interpretive and therefore may be considered *rendered material*, may include but are not limited to: orthographic (graphemic/graphetic), phonetic, phonemic, and kinemic/kinetic.
3. A body of European work (see e.g. Zäfferer, EMELD 2003) differentiates annotation tiers into “positive” and “negative” tiers, whereby the positive tiers represent the signifier or perceivable forms (e.g. the A/V signal, phonetic, phonemic, kinemic/kinetic, prosodic, and morphophonemic tiers), while negative tiers represent the signified or inferable content (such as morphological, syntactic, and meaning structure tiers, as well as translation tiers).

Nontranscriptional annotation:

non-linguistic annotation: metadata, format, comment

linguistic annotation: morphological, syntactic, semantic, discourse, pragmatics

Nontranscriptional annotation constitutes forms of representing form structure and content of types of linguistic signs.

It is at a different level of interpretation than transcription, as it can be *recursive* or *iterative* (it can annotate itself or a transcription).

Furthermore, it itself does not require a transcription. (One can annotate a video without transcribing it, for example with closed captions or with a translation into a major language.)

3. Recommendations

- a. **GOA**: (General Ontology for Annotation):

We recommend establishing a common structured ontology of annotation types parallel to the GOLD ontology. This could later be expanded to include special-purpose annotation ontologies for e.g. lexicons.

A first stab at such an ontology would be:

“positive” [rendered text]

- orthographic
 - graphemic
 - graphetic
- phonic
 - phonetic
 - phonemic
- prosodic
 - tone...
- kinesthetic
 - kinemic
 - kinetic

“negative”

- morphological
- syntactic
 - syntactic structure, syntactic relations...
- translation
 - literal
 - English
 - Russian
 - free
 - English
 - Russian...

Dwyer comments: The above ontology would have to be greatly expanded by the working group and then vetted by specialists. For this the Working Group should resign itself to the establishment of a subgroup (i.e. a *work* Working Group) to accomplish these goals.

In establishing such an ontology, one major challenge is how to deal with multipurpose annotation, especially the “classic” parts-of-speech (POS) tagging. This (beloved but uncool in some circles) form of annotation collapses formal, functional, morphological, and syntactic tiers into one all-purpose tier. One solution might be to simply have the abstract term “morphosyntax” as an upper tier of the negative tier hierarchy. Another related issue is to ensure that formal tiers are delineated from functional ones (drawing on the ideas of Lieb & Drude).

b. Tools evaluation

The Working Group stated general requirements for tools development, and established the beginnings of a framework to evaluate existing tools. These evaluations should be incorporated into the EMELD web site’s Tool shed. Again, the working group perhaps unwittingly assigned itself future work, which would include the following steps: solicit Working Group and EMELD group input for the evaluations; incorporate similar evaluations from other sources; include a field for user reviews (both to obtain the widest degree of input possible, as well as to keep the information up to date).

- a. General requirements for future tools
 - i. Open Source
 - ii. Undo
 - iii. Revision control system
 - iv. Avoid slow bloated software
 - v. Cross-platform (Web-based/offline)
 - vi.
 - vii. Unicode-compliant, XML-based
 - viii. Customizable annotation interface based on GOA
 1. Highly delimited \leftrightarrow highly powerful
 2. Modular? For computers with ltd HD space, older platform versions, processing speed...
 - ix. “Confidence Ranking” (Degree of Reliability)
 - x. A/V: variable speed playback
 - xi. Ability to visualize underlying graphs in multiple dimensions (cf. Bird’s Hyperlex)
 1. Directed graphs; annotation graphs
 - xii. Theoretical flexibility
 - xiii. Ability to annotate specific kinds of comments, e.g. disagreements on grammaticality by multiple speakers

b. Specific functionality for future tools

There was some brainstorming on the feasibility a single umbrella tool that is modular, allowing for digitization/capturing of A/V stream, transcription, and can do any kind of transcription / annotation. It could include the following plug-ins:

- i. Shoebox functionality for an XML-based, Unicode-compliant tool
 1. (Lookup (compare entry to lexicon), Jumping (hyperlinking), Interlinearization (semi-automatic fill-in of annotation))
 2. Requires a theoretical model that can be limiting
- ii. Playback with nondestructive signal modification
- iii. Praat-like acoustic analyzer with assisted alignment, spectrogram, F0 analysis
- iv. Spatial annotation of video (cf. MPEG-7 stds.; MPEG-4 wrt facial movement)
 1. for: movement path of signs; deictics; pn placement; gaze
 2. as e.g.: highlighting or circling a few frames
 3. also for: 3D spatial annotation

c. Lists of existing Transcription and Annotation Tools

We started a rating system in which recommendations were tied with the specific users and their requirements; for example, one tool might be adequate for text processing but hopeless at querying; one may be excellent for work in field or by untrained students, but poor for work in a computer science department.

| type | tool | URI/Ref | description | Unicode-compl? | export formats | platform | advantages | disadvantages | specific purpose | rating | your comments here |
|--------------------|--------------------|---------------------------------------|--|----------------|----------------|----------|---------------------|---|-----------------------------|--------|--------------------|
| Gestural | Laban | | | | | | | | | | |
| Gestural | Stokoe | | | | | | | | | | |
| Gestural | sign language | Annika Nonhebel usw | | | | | | | | | |
| Gestural | word glosses alone | | | | | | | | | ⊗ | |
| Gestural | Signwriter™ | | | | | | | proprietary | | ⊗ | |
| Time-aligned audio | Transcriber | LDC | | | | | | | | | |
| Time-aligned audio | SoundIndex | LACITO | associates audio files with annotations | ? | | | xml data, efficient | xml editor interface, may not support Unicode | | | |
| Time-aligned audio | Winpitch | | slowed speech, text to speech alignment, Unicode | Y | XML, Excel | | | | | | |
| Time-aligned A/V | ELAN | www.mpi.nl/dobes/... | | Y | | | | | | | |
| Time-aligned A/V | TASX | www.lili.uni-bielefeld.de/...tasxs... | | Y | | | | | | | |
| Time-aligned A/V | | | | | | | | | | | |
| Text | Shoebox | | | | | | | | for semi-automated glossing | *** | |

| | | | | | | | | | | | |
|----------|--------------------------------|------|--|--|--|--|---|---|------------------------|---------|--|
| | Shoebox | | | | | | | | for Unicode text | ☹ ☹ ☹ ☹ | |
| Text | FIELD | | | | | | | | | | |
| Text | ATT | | | | | | | | | | |
| Text | IDD | | | | | | | | | | |
| Text | FM pro | | | | | | | | | | |
| Text | spreadsheets e.g. Excel | | | | | | structure; sorting ; cross- platform, widely available (if templates are made available) | inconsistency | | | |
| Text | wordprocessors e.g. MS Word | | | | | | easy to use, Unicode- compliant, (temporary) proprietary font tolerant, printer friendly, widely available | inconsistency, lack of structure; moderately difficult to extract data | | | |
| metadata | | IMDI | | | | | | | | | |
| | | OLAC | | | | | | | | | |

The Working Group should augment this sample table and other sources (e.g. the Linguistic Annotation page; Corpus Linguistics Page; MATE evaluations etc.) should be consulted. It must include user feedback field to keep information up to date.

Desiderata

d. Transcription and Annotation methodology

- i. Prioritization
 1. Depends also on community/researcher goals
 2. Community resources
 3. Budgetary restraints
 4. Temporal restraints
 5. Projected audience
 6. Researcher competence (lx/target lang)
 7. Are there any universal priorities,
 - a. Audio/video recordings?
 - b. The knowledge speakers need to know to learn the language
 8. Put examples of suggested BP on School web site
 - a. Example 1: two remaining speakers, dying of cancer...
 - b. Example 2: plenty o' speakers, narrow ts.
 - c. Example 3: plenty o' speakers, lots o' texts
- ii. MS annotation: representing discontinuous/nonconcatenative morphology
- iii. Intonation transcription system: generic, minimal, extensible
 1. problem: prosodic boundaries
- iv. Standards for discourse markup
 1. Overlapping
- v. Standardization of (or at least conversion tables) gestural transcription systems

4. Presentation format:

Presentation formats are simply ways of rendering structured data. Examples include academic journal style sheets and the Leipzig Glossing Rules. Best Practice calls for the development of style sheets for articles and print dictionaries. The schoolhouse should include in the future XSLT stylesheets for these presentation formats.