Additions to the IMDI metadata set for sign language corpora

Agreements at an ECHO workshop
May 8 + 9, 2003, Nijmegen University


Onno Crasborn  Department of Linguistics, University of Nijmegen
o.crasborn@let.kun.nl

Thomas Hanke  Institute for German Sign Language, University of Hamburg
Thomas.Hanke@sign-lang.uni-hamburg.de

Version: Tuesday, April 7, 2009, 10:18

A set of key-value pairs to store sign language specific properties

The IMDI standard provides a simple mechanism for extending its scope: key-value pairs. Every major section of the session description includes a sub-schema for key-value pairs. To store information here, you add a pair, give it a key comparable to schema elements and choose a value from a vocabulary to be assigned to the key. This results in a slightly less-structured representation than what could be achieved with an extension of the standard, but it does allow for a quick start, with the chance to later formally propose an extension of the standard promoting the keys to scheme elements. In order to simulate the concept of element groups, we suggest key values with dots for the moment, such as Hearing.Status.Hearing for Actor.keys.

The most important future development of the IMDI tools from the perspective of this proposal concerns the creation of ‘profiles’. Profiles will contain sets of key-value pairs specific for different subgroups of users, such as the sign language community. At this moment we can simulate and share a sort of profile by making and sharing a ‘master document’ as the one that can be found on the workshop web page (from early May on). Ideally, one would be able to choose one or more profiles from a list within the IMDI editor. This will be developed in the near future.

If the set of sign language extensions is seen as a useful and stable set by the sign language community, the ‘sign language profile’ can perhaps be given a more flexible layout in the IMDI editor and browser. In order to simulate the concept of element groups, we suggest key values with dots, such as ‘Deafness.DeafnessStatus’ for ‘Actor.keys’.

Numbers refer to paragraphs in the IMDI 3.02 proposal that already provide space for additional information in the form of ‘keys’.
3.3 Content
3.3.6 Content . Languages
3.3.6.2 Content . Languages . Description

Space for describing code mixing, sign supported speech, etc. used in this session in prose.

Do we need separate keys for describing code mixing and code switching between different languages or modalities?

3.3.7 Content . keys

Language Variety
Definition: Description of the language variety used in the session.
Encoding: string
Comments: Space for more constrained description of language variety used in this session. Information about language skills of the individuals should be entered in the actor’s description (cf. 3.4.2.15 Actor . keys).

Elicitation Method
Definition: A characterization of specific prompts used for eliciting language production.
Comments: Use ‘no prompt’ for spontaneous language. When working on the influence of German on DGS compounding, for example, it is essential to know if the spoken language competence has been activated by the elicitation situation. Content . Task might be appropriate for this purpose, but the open vocabulary seems to suggest different levels of detail: While Wizard of Oz certainly is not related to the utterance’s topic, some others are, such as room reservation. “Frog story” could already have a (TM), it is well known to name both contents and elicitation method. Content . Involvement would be a good place, if it were open vocabulary.

Interpreting
Group
Definition: Properties of interpreting appearing in the session.
Encoding: Interpreting . Source
Interpreting . Target
Interpreting . Visibility
Interpreting . Audience
Comments:

Interpreting . Source
Definition: Source modality and language type.
Encoding: OVL: sign language, speech / sign supported speech / text / fingerspelling / unknown / unspecified
Comments:

Interpreting . Target
Definition: Target modality and language type.
Encoding: OVL: sign language / speech / sign supported speech / text (subtitling) / fingerspelling / unknown / unspecified
Comments:
**Interpreting . Visibility**
Definition: Visibility of the interpreter in the video recordings.
Encoding: CCV: not visible / in view during whole session / in view during part of session, unknown, unspecified
Comments:

**Interpreting . Audience**
Definition: Presence and nature of an audience that the interpreter is signing for.
Encoding: CCV: audience not present (signing to camera) / audience known to the interpreter / heterogeneous group partly known to the interpreter / anonymous audience (e.g. theatre) / unknown / unspecified
Comments: If Interpreting . Target = subtitling, leave field empty.

3.4 Actors
3.4.2 Actor
3.4.2.15 Actor . keys

We propose to add a number of keys describing different aspects of the actors, mainly to characterize the language background. All of these keys refer to relatively stable properties (skills) of the actors, not to their actual behaviour in the specific session at hand.

Note: descriptions of groups of keys are aligned with the left margin; description of elements are all indented. The other formatting of the descriptions follows the IMDI documents. Keys that are further specified by a set of keys are followed by “(sub)” in the lists.

*General comment:* most of the subjective data could be paralleled with “objective” data, such as ‘db left’ and ‘db right’ for the item ‘hearing’, scores in a language competence tests etc. Is this needed? Does anyone have suggestions for specific field and values that are often measured in your corpus?

**Actor keys**
Group:
Encoding: Deafness (sub)
   Sign Language experience (sub)
   Family (sub)
   Education (sub)
Comments: Stable properties (skills) of the actor, not their actual use in a given session.

**Deafness**
Group
Definition: Groups information about the deafness status of the actor. Only the first element is relevant for all actors, the other elements specify details about hearing loss.
Encoding: Deafness . Status
   Deafness . Aid Type
Comments:

**Deafness . Status**
Definition: Actor’s ability to hear.
Encoding: CCV: hearing / hard-of-hearing / deaf
Comments:
**Deafness . Aid Type**
Definition: Type of hearing aid the actor has.
Encoding: CCV: none / conventional / CI
Comments:

**Sign Language Experience**
Group
Definition: Groups (partly subjective) information on the actor’s experience with sign language.
Encoding: Sign Language Experience . Exposure Age
Sign Language Experience . Acquisition Location
Sign Language Experience . Sign Teaching
Comments:

**Sign Language Experience . Exposure Age**
Definition: Age at which exposure to sign language and sign language use started.
Encoding: c (years;months)
Comments: Nativeness can be expressed by Language . Mother Tongue.

**Sign Language Experience . Acquisition Location**
Definition: Place where sign language was learnt.
Encoding: OVL home from family/home from tutor/ preschool teachers / teachers / family beyond home / friends
Comments:

**Sign Language Experience . Sign Teaching**
Definition: Amount of experience with teaching sign language.
Encoding: OVL: none / some / extensive
Comments:

**Family**
Group
Definition: Describes deafness status of closest contact persons as well as preferred communication systems used.
Encoding: Family . Mother (sub)
Family . Father (sub)
Family . Partner (sub)

**Family . Mother**
Group
Definition: Characterises language input from actor's mother.
Encoding: Family . Mother . Deafness
Family . Mother . Primary Communication Form

**Family . Mother . Deafness**
Definition: Describes mother’s deafness status.
Comments: Where appropriate, describe deafness status of alternative primary caregiver.
Definition: Describes mother’s language input towards the actor.
Encoding: OVL: sign / sign-supported speech / gesture / mix between signing and speaking / speech only / writing
Comments: Where appropriate, describe primary communication form of alternative primary caregiver.

Family . Father
Group
Definition: Characterises language input from actor's father.
Encoding: Family . Father . Deafness
Family . Father . Primary Communication Form

Family . Father . Deafness
Definition: Describes father’s deafness.
Comments: Where appropriate, describe deafness status of alternative primary caregiver.

Family . Father . Primary Communication Form
Definition: Describes father’s language input towards the actor.
Encoding: OVL: sign / sign-supported speech / gesture / mix between signing and speaking / speech only / writing
Comments: Where appropriate, describe primary communication form of alternative primary caregiver.

Family . Partner
Group
Definition: Characterises language input from actor's partner.
Encoding: Family . Partner . Deafness
Family . Partner . Primary Communication Form

Family . Partner . Deafness
Definition: Describes partner’s deafness status.
Comments: Describe situation at the time of the recording.

Family . Partner . Primary Communication Form
Definition: Describes partner’s language input towards the actor.
Encoding: OVL: sign / sign-supported speech / gesture / mix between signing and speaking / speech only / writing
Comments:

Education
Group
Definition: Describes where the actor was educated.
Encoding: Education . Age
Education . School Type
Education . Class Kind
Education . Education Model
Education . Location
Education . Boarding School
Comments: It should become possible in the editor to specify this whole set of elements repeatedly for each school the actor has attended. Currently, this is not possible, and it will need to be
determined in the future how this can be done. In the mean time, it is recommended that users specify values for multiple schools in each field, separated by commas.

**Education . Age**
Definition: Describes the age during which the school was attended.
Encoding: string
Comments: Formatting: start age, dash, end age
For example: 3-6, 6;3-12;2, etc

**Education . School Type**
Definition: Describes the type of school.
Encoding: OV: bilingual home programme / kindergarten / preschool / primary school / vocational training / college / university
Comments:

**Education . Class Kind**
Definition: Describes the kind of class in the school.
Encoding: OV: deaf / hard-of-hearing / deaf class in hearing school / individually integrated
Comments:

**Education . Education Model**
Definition: Describes the education model used at the school.
Encoding: OV: bilingual / oral / mixed / sign monolingual / oral with interpreter
Comments: For combinations of oral education with cued speech, use ‘oral’, combinations with fingerspelling, use ‘mixed’.

**Education . Location**
Definition: Describes where (town or region) the institution was located.
Encoding: string
Comments:

**Education . Boarding School**
Definition: Is the school a boarding school?
Encoding: CCV: yes / no
Comments:
Links

Workshop home page  
http://www.let.kun.nl/sign-lang/echo/events.html

The background document  
http://www.let.kun.nl/sign-lang/echo/docs/Metadata_SL.doc

Sign language master files for IMDI  
http://www.let.kun.nl/sign-lang/IMDI

ECHO project, home page  
http://echo.mpiwg-berlin.mpg.de/

ECHO project, case study 4  
http://www.let.kun.nl/sign-lang/echo

ECHO project, technology  
http://www.mpi.nl/echo

ECHO project, state of the art  
http://www.ling.lu.se/projects/echo/contributors/

IMDI standard  
http://www.mpi.nl/IMDI

IMDI tools  
http://www.mpi.nl/IMDI/tools

ISLE metadata glossary  
http://www.mpi.nl/ISLE/glossary/glossary_frame.html

ELAN annotation software  
http://www.mpi.nl/tools/elan.html

References

Warning: the document and tools available online refer to version 2.5-2.8! Updated tools are available from September 2003.


http://www.mpi.nl/IMDI/documents/Proposals/ISLE_Lexicon_1.0.pdf

http://www.mpi.nl/IMDI/tools/IMDI_Editor_Manual_2_0.doc

http://www.mpi.nl/IMDI/tools/IMDI_Browser_Manual-02-09-08.doc