Sharing Multimodal Data: A Novel Metadata Session Profile for Multimodal Corpora

Farina Freigang[1], Matthias A. Priesters[1,2], Rie Nishio[3], Kirsten Bergmann[1]
[1]Faculty of Technology, Center of Excellence "Cognitive Interaction Technology" (CITEC)
Bielefeld University, P.O. Box 100 131, 33501 Bielefeld, Germany
{firstname.lastname}@uni-bielefeld.de
[2]Human Technology Centre (HumTec), Natural Media Lab
RWTH Aachen University, Theaterplatz 14, 52056 Aachen, Germany
priesters@humtec.rwth-aachen.de
[3]Institute of German Sign Language and Communication of the Deaf
University of Hamburg, Binderstr. 34, 20146 Hamburg, Germany
rie.nishio@sign-lang.uni-hamburg.de

Abstract

In the natural sciences and humanities, scientific data management and in particular the categorisation of data and the publication of (meta)data becomes ever more relevant. A new focus in corpus-based research are *multimodal* data. However, metadata profiles for multimodal data are rare and do not fit the needs of researchers who are searching for particular data. In this paper, we present a novel metadata session profile for describing data collections which contain other modalities beyond text and speech. The profile is based on experiences gained during the work on three different corpora comprising communicative speech-gestural behaviour as well as sign language data. The profile is aimed at creating metadata for individual recording sessions and is technically implemented in the CMDI format. Furthermore, it is designed to be paired with an existing profile for media corpora, which was extended for multimodal data.

Keywords: Metadata profile, multimodal data, multimodal corpora, gesture, sign language, CMDI, ISOcat, CLARIN

1 Introduction

The production of high-quality multimodal corpora is extremely expensive and hence it is of major importance to manage these resources in a way that they are easily searchable and reusable for other researchers. In fact, the reuse of resources is an issue strongly promoted by research funding organizations, for example, by the European Union in terms of their "open data strategy".[1] In the field of corpus linguistics and language resources it is widely agreed that the ever-expanding number and growth of corpora needs *metadata* for the purpose of corpus management. For linguistic resources there already exist a large number of metadata schemes, but so far not much effort has been put into the development of metadata schemes for the particular structure of *multimodal* corpora. This is, at least in parts, due to the fact that multimodal corpora are highly heterogeneous. They might include different modalities or communication channels of natural communicative behaviour such as gestures, facial expressions, body posture or eye gaze for which no standardised coding schemes exist. Moreover, multimodal corpora might comprise multiple synchronous data streams, such as video, audio, time series data (e.g., motion capture or eye tracking) and annotation data. These aspects have previously not been captured by metadata profiles.

In CLARIN-D, the discipline-specific working group on "Speech and Other Modalities"[2] has initiated a discussion on these issues (cf. Freigang and Bergmann, 2013) which has led to the proposal of a novel

[1] http://ec.europa.eu/digital-agenda/en/open-data-0

[2] Indicated by the group name *SpeechAndOtherModalities* in the CLARIN Component Registry: http://catalog.clarin.eu/ds/ComponentRegistry

metadata session profile for multimodal data: the `MultimodalSessionProfile`.[3] The profile is based on a detailed evaluation of three different multimodal corpora: the Speech and Gesture Alignment ("SaGA") Corpus from Bielefeld University (Lücking et al., 2013), the Dicta-Sign DGS Corpus from University of Hamburg (Matthes et al., 2012) and the Natural Media Motion Capture ("NM-MoCap") Corpus from RWTH Aachen University (Hassemer, 2015). The profile has been developed according to the CMDI[4] standard (Broeder et al., 2012; de Vriend et al., 2013) including unique ISOcat[5] definitions within and for (but by no means exclusively for) the CLARIN infrastructure. It offers a wide variety of corpus descriptions especially designed for, but not limited to, multimodal data. Furthermore, it has been used for the integration (and publication) of the three mentioned corpora into the CLARIN-D infrastructure.[6]

This paper aims to present the new metadata session profile for multimodal data. First, we briefly introduce necessary technical terms in section 2. Subsequently, in section 3, we review existing metadata profiles for multimodal data and discuss why they were not sufficient for the requirements of our corpora. In section 4 we introduce the novel session profile with its modality components and other specifically developed components, and in section 5 the accompanying corpus profile is described. We conclude with a discussion in section 6.

2 Terminology

In this section, we briefly introduce the basic technical terminology pertaining to the CMDI metadata standard insofar as it is necessary for the understanding of the following sections. For more detailed information on CMDI and the underlying infrastructure, see Broeder et al. (2008), Broeder et al. (2012) and de Vriend et al. (2013).

Profile CMDI profiles are templates used to create metadata records (XML files that describe a specific corpus or data set). They are implemented as XSD schemas which define the structure of the actual CMDI records. Profiles consist of components and elements. Profiles are published in the CLARIN Component Registry, which can be used to validate metadata records.

Component A CMDI component is an independent part of a CMDI profile, which groups subordinate components and elements, usually thematically related ones. Components are also published in the Component Registry and can be reused by including existing components into a newly created profile.

Element Elements are the basic building blocks in CMDI, which hold the actual metadata. They are embedded in a component or directly in a profile and have the form of *key–value* pairs. The values are typed and can contain strings, numbers, boolean values, URLs or different date formats. Furthermore, the user can define the values of elements in terms of regular expressions or controlled vocabularies (predefined sets of possible entries).

Cardinality Components and elements in a profile are marked with *number of occurrences* constraints. These constraints specify lower and upper boundaries of how often the component or element can or must occur in an instance of the CMDI profile. Typically, the lower boundary is 0 or 1, and the upper boundary 1 or *unbounded*. In this paper, cardinalities are given in the figures in square brackets (e.g., [0–1]).

Attribute Additional data fields which can be attached to components and elements are called attributes. They can contain additional information about the respective elements or components, for example, attributes can indicate a component's language or link components to other components (cf. section 4.4 and Figure 6).

[3] Monospaced font for designations denotes names of CMDI profiles/components/elements/attributes as they appear in the CLARIN Component Registry.

[4] CMDI: Component Metadata Infrastructure (Broeder et al., 2012; de Vriend et al., 2013); which is compatible with other standards such as Dublin Core (DC), Open Language Archives Community's metadata set (OLAC), and Isle Metadata Initiative (IMDI).

[5] `http://www.isocat.org`

[6] Two corpora were ingested into the repository of the Bavarian Archive for Speech Signals:
`https://clarin.phonetik.uni-muenchen.de/BASRepository/index.php`

3 Related work and problem description

In Freigang and Bergmann (2013), we compared relevant CMDI metadata profiles for multimodal data from the CLARIN Component Registry: `media-corpus-profile` and `media-session-profile` by BAS[7], NaLiDa's[8] `MultimodalCorpus` profile, and `BamdesMultimodalCorpus` used for harvesting purposes by the Harvesting Day initiative. These profiles already included some aspects of multimodality, since simple modality components such as `cmdi-modality` and `ModalityInfo` exist (cf. the modality list in Figure 4), however, various other aspects were missing (for a detailed discussion, see Freigang and Bergmann (2013)). We are not aware of other related work concerning metadata for multimodal corpora.

Basically, we identified two major problems with existing components and profiles. The first problem occurred when generating metadata descriptions for the previously mentioned multimodal corpora (SaGA, Dicta-Sign DGS, NM-MoCap) from existing metadata profiles: the *granularity* in which modality (or multimodal) metadata descriptions were possible was not fine enough. So far, it was not possible to specify, for example, the handedness of an actor[9], the modalities of a stimulus, or that iconic gestures were annotated in the data. Hence, from the corpus user's perspective, it was not possible to search for detailed features of multimodal corpora. Therefore, we focused particularly on the following:

- the development of detailed descriptions of the two modalities *gesture* and *sign language*, among others

- descriptions of how actors relate to different modalities (text, speech, gesture, sign language, etc.): for example, what is their written German language proficiency, or do they gesture a lot, or do they have experience with sign language, or do signers have regular contact to non-signers?

- multimodal descriptions for the study design and the data collection (environment, content, elicitation phase, etc.): for example, did the recording take place in a studio, or what modalities were involved in explaining the task, or did someone gesture in a stimulus video?

- descriptions of the annotation scheme (e.g., according to a gesture or sign language researcher) which is used to analyse the multimodal data.

A second problem with existing metadata profiles was that *technical descriptions* for media files and annotation files were missing. With the recording of multimodal data, novel technical devices typical for gesture or sign language studies are used. For example, one of our reference corpora includes motion capture recordings of gestures. Compared to other profiles, the `media-session-profile` is rather advanced and already includes components for time series data and stereo video (3D) recordings. However, describing a marker setup as used in the NM-MoCap corpus in this metadata structure proved cumbersome and unintuitive. Therefore, a more meaningful way for describing motion capture data among others was one of the requirements for a new profile. Furthermore, descriptions for technologies used in recordings, as for example HD videos, were not elaborate enough and needed extension.

4 Introducing the `MultimodalSessionProfile`

Based on the identified problems, we developed various new components covering different modality aspects and technical descriptions. As discussed in Freigang and Bergmann (2013), there are several

[7] `http://www.phonetik.uni-muenchen.de/Bas`

[8] `http://www.sfs.uni-tuebingen.de/nalida/en`

[9] We chose to use the term *actor* throughout our profiles and in this paper, firstly, because it is most accurate. *Participant* and *subject* are terms which imply an arranged setting such as in studies, which is not always the case since some corpora are collections of data, such as the Dicta-Sign DGS Corpus or a collection of news broadcasts. Secondly, *actor* is the most neutral term available: the terms *speaker* or *signer* would exclude users of signed or spoken languages, respectively. Therefore, we used the term *actor* in newly created components of our session (and corpus) profile. The terms *subject* and *participant* occur rarely and only where components were reused.

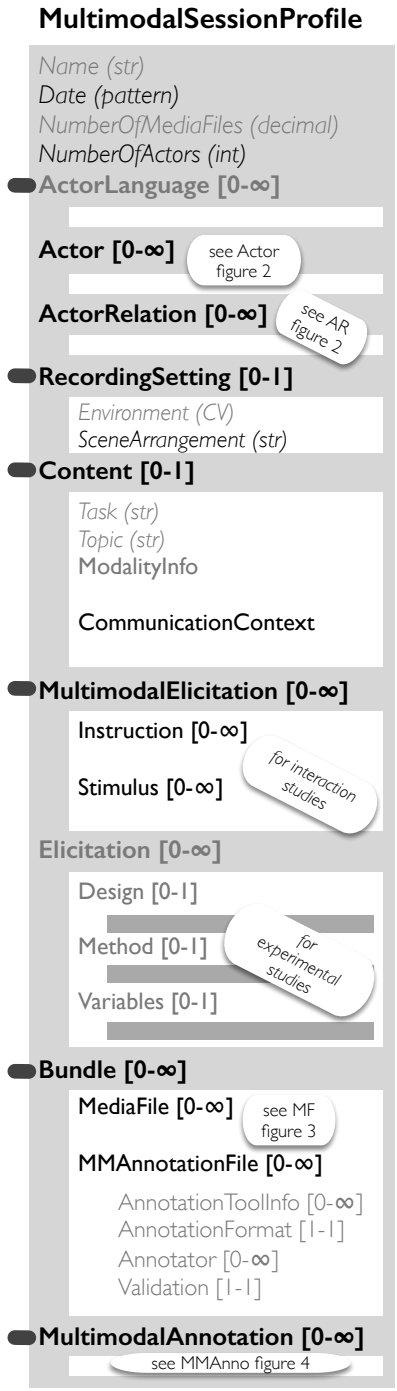

Figure 1: An overview of the session profile with six main thematic parts. Two elicitation components cover interaction studies and experimental studies.

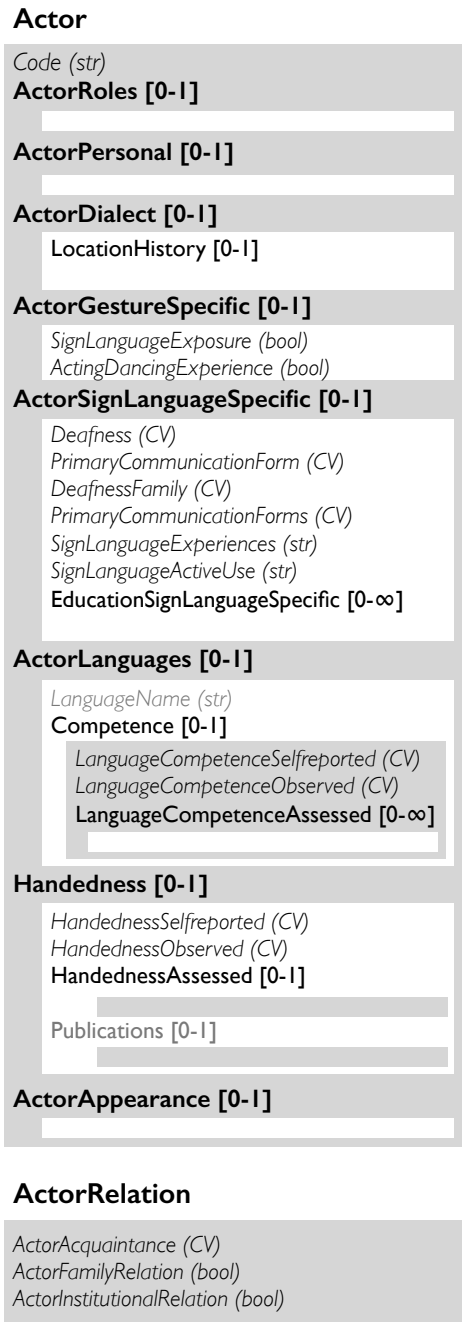

Figure 2: The actor and actor relation components introduced in Figure 1. All components have been newly created for the purpose of a fully multimodal description of actors.

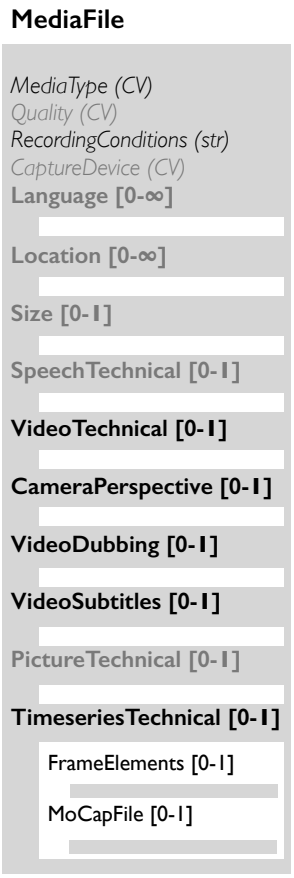

Figure 3: The Media file component introduced in Figure 1. The component is comprised of old and new components.

options of how to realise new metadata components. Many changes were necessary, so that the integration of the new components into an existing profile structure was not feasible. Therefore, we created the novel **MultimodalSessionProfile**[10] in a bottom-up fashion: We designed new components and subsequently compiled them together with relevant existing components from the CLARIN Component Registry into a large profile structure. All figures in this paper illustrate reused components and elements in grey font; all others are newly created. In some cases a component has been newly created which combines (almost) only existing components and elements in a novel fashion. The aim was to group components and elements thematically. Furthermore, the exact names of components and elements may have been changed in this paper for better readability (for the exact designations see the profiles online) and the figures illustrating the metadata components are depicted not to the full extent but are reduced to the most important parts. The profile construction is oriented at `media-session-profile` by BAS and NaLiDa's `MultimodalCorpus` profile, among others. In cooperation with BAS, we also created a multimodal version of the `media-corpus-profile`, discussed in section 5.

4.1 Session Profile Overview

The **MultimodalSessionProfile** (Figure 1) consists of six main thematic parts: metadata descriptions about the *actors*, the *recording setting*, the *content* of a study or a corpus, the *elicitation* methods, the accumulated *data* (media files and linked annotation files), and a description of the *annotation* design (`MultimodalAnnotation`). In some cases, the outline is reminiscent of the temporal development of a data set: for example, when conducting a study, one typically starts with the participants and the study design, continues with the study itself, the recorded data, the post-processing of the data, and the theory behind the used methods. On the top level, the profile contains a number of elements for basic metadata about the session: the date, the time and the place of a recording and the numbers of files and actors involved.

In the following, the main components and the various possibilities of description they provide are discussed. Note that the profile has been designed for flexibility, therefore all components and elements on the profile's top level are optional. If a component has been chosen for description of a data set, some elements or components are obligatory. For example, if `MultimodalAnnotation` has been chosen, we assume that the corpus contains multimodal data that needs to be described and, thus, the `ModalityInfo` component, reused from the META-SHARE metadata profile (Gavrilidou et al., 2012), is mandatory. Or as soon as an actor is involved in a data set it needs to be stated which role she took (cf. section 4.2.1). Since there is a lot of homogeneity among multimodal corpora, flexibility is of major importance to allow users to adapt the metadata profiles according to their needs.

4.2 Modality and multimodal components

Natural communication data include various modality aspects of which only a few are found in metadata descriptions. In order to get the full picture of multimodality, a first step was to define categories for gesture and sign language. In Figure 5, the newly created components are grouped according to three different categories: *speech*, *gesture* and *sign language*. We refined the granularity of the modality metadata descriptions in various ways: to depict the details of performed modalities by the *actors*, the influence of the modalities on the *data collection* design, the *content* of the material, and the *annotations* concerning the various modalities.

In the following, we will give a few examples of how certain components (which will be discussed in more detail below) fall into these categories. We developed category-specific descriptions, namely the `ActorGestureSpecific` and `ActorSignLanguageSpecific` components. Speech-specific descriptions are covered by reused components such as `cmdi-subjectlanguages`. Other components are kept general and allow for speech and sign language descriptions, among others, as does the component `ActorLanguages`. Other components that serve both of these categories are `ActorDialect` with its component `LocationHistory`. The component `Handedness` has been

[10] The `MultimodalSessionProfile` has been published on May 5th 2014 (and edited by CLARIN on July 2nd 2014): `http://catalog.clarin.eu/ds/ComponentRegistry?itemId=clarin.eu:cr1:p_1381926654659`

kept general for the description of both gesture and sign language. Finally, some components fall into all three categories: for example, the MultimodalElicitation component (which is supplemented by the reused Elicitation component for experimental research data) and the Content component comprising the study task, the modalities which are used during the study, and the communication context.

4.2.1 Actors

Providing detailed information about the persons appearing in the corpus material was a major focus of our metadata profile. We realised this in two components: Actor for the properties of an individual person and ActorRelation for capturing relations between multiple persons taking part in the same session (Figure 2). The Actor component comprises multiple subcomponents which allow the metadata user to describe different aspects of participants in various use cases: ActorPersonal contains elements for basic data about participants, such as their name, sex, age, educational or professional status. ActorRoles captures the roles a person takes in a corpus, for example, *experimenter*, *subject* or *confederate*. ActorAppearance describes the actors' physical appearance, insofar it is relevant for the purposes of the recordings (e.g., if the actor is wearing glasses or clothing that could cause problems for image recognition or motion tracking techniques). Besides specifying the language used in the corpus, the component ActorLanguages allows for recording all languages spoken by an actor, including their self-reported or measured proficiency (Competence).

The requirements of researchers recording gesture and sign language corpora include metadata about participants usually not captured by metadata profiles. For theses purposes, we developed the ActorGestureSpecific and the ActorSignLanguageSpecific components. The focus of both types of corpora is set on communicative manual action and thus, our profile includes information about an actor's Handedness (either self-reported or assessed using a test). The gesture component records whether the actor has had previous exposure to sign language, acting or dancing, factors which could influence their gestural behaviour. For corpora including sign language, the personal history of actors, such as their educational background or the location where they grew up, are especially important, as these strongly influence sign language proficiency and the signed dialect (ActorDialect). Furthermore, the sign language component contains detailed elements for describing Deaf actors, such as the use of hearing aids, the deafness status of their family members and their degree of involvement in Deaf culture (e.g., through sign language teaching or using sign language in art). In developing these components, we built upon and extended a set of ISOcat data categories for describing signed language resources compiled and implemented by Crasborn and colleagues (Crasborn and Hanke, 2003a; Crasborn and Hanke, 2003b; Crasborn and Windhouwer, 2012).

4.2.2 Elicitation

The *Elicitation* components provide room for describing the design of the data collection, i.e. the methods which were applied to elicit communicative behaviour from the recorded participants. For studies employing experimental methodologies (e.g., from a psychological background), we included the Elicitation component from the NaLiDa metadata profile. Additionally, we designed a new component for interaction studies (MultimodalElicitation). It is mainly divided into a MultimodalInstruction and a MultimodalStimulus component, each framing a component named InformationChannel, stating how information was given to the actors during the instruction or stimulus phase: which kind of medium was used, how it was physically presented, and which modalities were involved. For stimuli which are common in a certain research community and which have been published, the respective publications can be attached using the documentInfo component, also developed by META-SHARE. Furthermore, it can be specified whether an instruction was recorded beforehand and if a stimulus was accessible to the actor during language or gesture production.

4.2.3 Context and content

For a full picture of the data collection design, information about the content of the language resources and the recording environment is necessary. The profile includes two components for these purposes,

Content and RecordingSetting. The latter includes facts such as the Environment (e.g., *studio* or *field*), the VisualBackground, the Weather or the SceneArrangement (a free description field).

Content offers various ways of specifying what the content of a data set is about, including the Task and Topic elements as well as a modality component and the CommunicationContext component. CommunicationContext provides space for specifics of the conversation, for example the SocialContext (e.g., *family* or *public*), the Channel (e.g., *face to face* or *telephone*), the ConversationType (e.g., *dialogue*), or whether there is an Audience watching the scene.

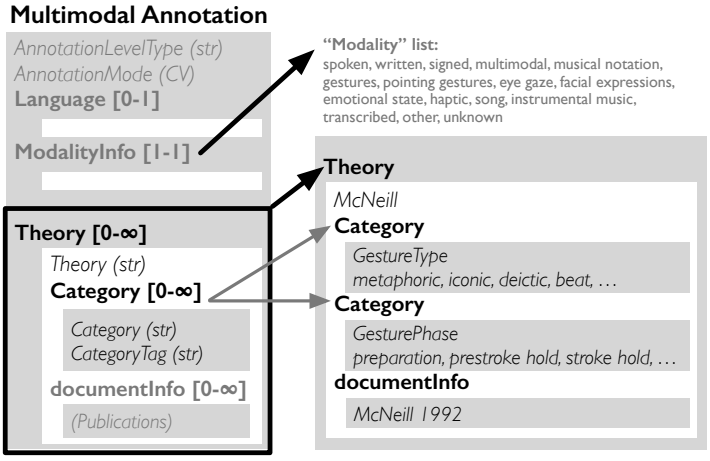

Figure 4: The MultimodalAnnotation component with an exemplary use of the Theory component for gesture categories by McNeill (1992). For the position within the MultimodalSessionProfile, cf. Figure 1.

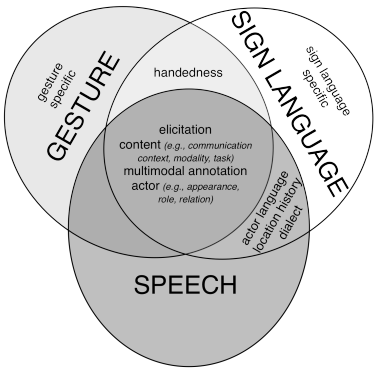

Figure 5: The keywords and names of newly created modality components classified into three main categories. Reused components are not shown.

4.2.4 Multimodal annotations

For multimodal communication data, the categorisation of observed phenomena is usually done by annotating recorded data according to predefined or emergent categories. One crucial aspect of our work was the development of a metadata component which is able to capture various categorisation systems. Our profile records this information in the description of the annotation schemes. In gesture studies, for example, gestures can be classified according to different criteria. One popular method follows McNeill (1992), who distinguishes between *iconic* (resembling the content of speech), *metaphoric* (image of abstract concept), *deictic* (pointing) and *beat* (marking the structure of the utterance) gesture categories. Furthermore, McNeill temporally segments gestures into phases such as *preparation*, *stroke*, *hold* and *retraction*.

In Figure 4, we have sketched how a gesture annotation scheme based on McNeill's categories could be captured with our MultimodalAnnotation component. Several Theory components can be added; each needs to be given a name (e.g., *McNeill*, cf. the example above) and may contain one or several Category components. A Category is also named (e.g., *GestureType* and *GesturePhase*) and contains one or several CategoryTag elements, which represent the individual annotation labels (e.g., *iconic* or *preparation*). Category and CategoryTag may be seen as an annotation category with one or more annotation labels possible. Each Theory component can be enriched with literature references in the documentInfo component. Additional information, for example, explanations about the exact meaning of annotation categories, can be stored in optional Description components. Overall, this Theory component is kept simple in its design and is still flexible enough to cover complex category systems, also those which may be developed in the future. We explicitly encourage metadata creators to use this component to also refer to their own theory or annotation frameworks.

The Theory component appears next to two other components and two elements in the MultimodalAnnotation component. One component is the modality list ModalityInfo mentioned in section 3. It provides modality-related keywords for characterizing the annotations performed on the corpus data. The difference between the Theory and the ModalityInfo components is that the former describes the annotation scheme and the theory behind it in detail, whereas the latter generally lists the specific modalities which were annotated. Thus, the ModalityInfo list allows for quick and shallow modality descriptions, if no particular framework has been used. Furthermore, the metadata creator can specify the AnnotationLevelType (e.g., *part of speech*, *gesture form*, etc.), the AnnotationMode (e.g., *manual*, *automatic*, etc.), and the language of the annotations.

4.3 Technical metadata

Besides capturing information about the recorded data on a conceptual and theoretical level, technical and organizational descriptions of the resulting data files are necessary. These metadata are collected in the Bundle component (Figure 1). Files are grouped in bundles if they belong to the same (usually synchronously recorded) data set. This can mean for example multiple simultaneously recorded video streams, together with motion capture and eye tracking data and the annotation files pertaining to these data.

4.3.1 Media files

With the MediaFile component (Figure 3), the profile includes fine-grained description categories for various types of media data, that is, *video*, *audio*, *image* and *time series* data. Most categories were reused from existing metadata profiles (most notably the media-session-profile), but some components were extended. Among the added features are information about camera perspectives, video dubbing/subtitling and the ability to describe multiple channels of a single video recording (needed for 3D stereo videos). The TimeseriesTechnical component was extended by components for marker sets used in optical motion capture systems and for kinematic data computed from raw motion capture data.

4.3.2 Annotation files

The treatment of annotation files differs from existing profiles in that *annotation files* (labels) are separated from *annotation schemes* (theories), the latter being realised in the Theory/MultimodalAnnotation component (cf. section 4.2.4). There are various established transcription and annotation tools, such as *Praat* (Boersma and Weenink, 2001), *ELAN* (Wittenburg et al., 2006), *Anvil* (Kipp, 2001), *iLex* (Hanke et al., 2010), *EXMARaLDA* (Schmidt, 2002), among others, each of which is based on different annotation file formats. The MultimodalAnnotationFile component (Figure 1) is limited to technical and organisational metadata. Each description instance of an annotation file is linked to the corresponding MultimodalAnnotation component, this way information about an annotation system or scheme only needs to be stored once in each session CMDI file.

4.4 Links between components

In order to better reflect the internal structure of a data set, many components can be linked to each other using attributes (Figure 6). This way, redundancy is kept at a minimum, since each piece of information has to be given only once. Components which can be linked to possess an 'ID' attribute, components which can link to other components possess a 'reference' attribute. The component Actor, for instance, has an attribute ActorID, which can be linked with components such as ActorRelation and MultimodalAnnotationFile through their ActorRef attributes. Those links can capture, for example, that the actor participated in different sessions of the same data set, that two actors are relatives or colleges from work, or that this specific actor in a video file is the same person whose interaction is labelled in an annotation file. Finally, session CMDI files and the corresponding corpus CMDI file are linked to each other.

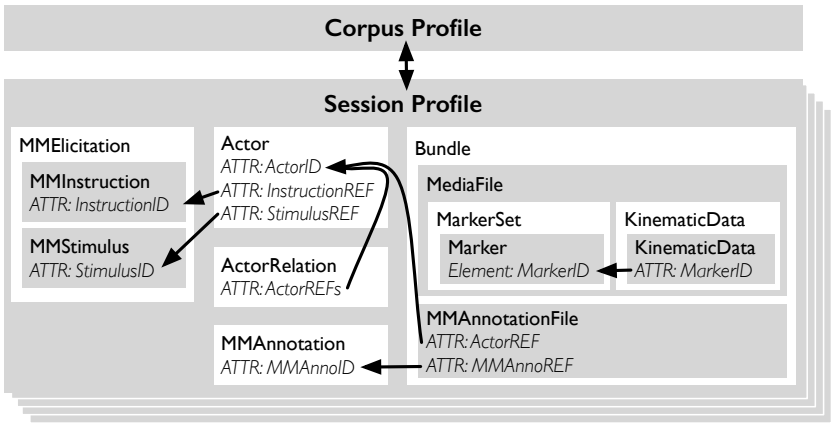

Figure 6: Links between components using attributes.

5 Corpus metadata

The `MultimodalSessionProfile` is designed to describe a single set of contiguous data, usually one recording session as part of a larger corpus. For the description of the corpus as a whole, an enclosing profile is needed, which 'frames' the session data. Therefore, in cooperation with BAS, we extended their `media-corpus-profile` with components for multimodal data (Figure 7; the components in grey font are the original BAS components of the profile). The first version of the profile was mostly geared towards speech corpora containing audio data. The extended version of the `media-corpus-profile`[11] (version 1.1) now contains a `MultimodalCorpus` component capturing information about modalities and an `AnnotationInfo` component with information about the annotated phenomena and the annotation tools and file formats used. The `MultimodalSessionProfile` and the extended version of the `media-corpus-profile` are designed to be used in combination in order to create a complete corpus metadata description. The usage is as follows: for each experiment or sub-study, one session CMDI file is created, whereby one actor can participate in several sub-studies. All sessions that belong to one data set are then linked to a single corpus CMDI file, which describes this data set (for links between components and profiles cf. section 4.4).

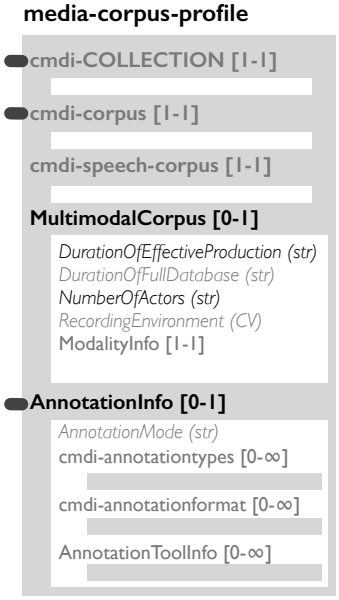

Figure 7: An overview of the enhanced BAS corpus profile with three main thematic parts.

6 Discussion and outlook

In this paper, we presented a metadata profile specifically addressing the needs of researchers working on multimodal communication data, which builds upon and expands earlier profiles. The presentation of our novel scheme and its realization have evoked fruitful discussions at conferences and workshops, both within the CLARIN community and in the relevant research communities. This shows a serious interest in the topic among potential users.

The development of our metadata profile has been driven by the requirements which resulted from work with specific corpora. Nevertheless, we aimed at developing a flexible profile universally applicable to multimodal data, in line with the philosophy behind CMDI: "The CMDI infrastructure encourages reuse of resources [...]. Therefore, metadata that are useful to any researcher [...] is especially valuable

[11] The `media-corpus-profile` has been published on May 5th 2014:
`http://catalog.clarin.eu/ds/ComponentRegistry?itemId=clarin.eu:cr1:p_1387365569699`

and should be focused on first." (de Vriend et al., 2013, 1320) Given the heterogeneity of multimodal corpora, further tests beyond our three corpora with other data collections are necessary to further improve the profile and make it as universally applicable as possible.

To date, a user-friendly tool for the creation of CMDI files based on the `MultimodalSession-Profile` is not available. Some CMDI generation tools exist, as for example ARBIL[12] (Withers, 2012), developed at the Max Planck Institute for Psycholinguistics in Nijmegen, or custom-built CMDI generations scripts. However, these tools are either designed for a particular profile structure or they are not easy to use with complex profiles such as the `MultimodalSessionProfile`. The creation of actual CMDI files remains a challenge, as the profile's size and complexity makes the manual creation of larger numbers of CMDI files infeasible. Technical metadata can easily be extracted automatically from the data itself, but for content metadata, easy-to-use tools for researchers are required and remain future work. Therefore, we highly encourage further tool development for the automatic generation of CMDI files, which would be extremely helpful to create, use and share CMDI files and further improve metadata profiles. In future, such a tool may even be used for flexible and 'on-the-fly' profile creation: Thus, no complete CMDI profiles would need to be prepared as templates, but components from the Component Registry could be combined flexibly by the metadata creator while compiling metadata for a data collection.

Acknowledgments

We thank Florian Schiel, Menzo Windhouwer and Onno Crasborn for their support and cooperation of the multimodal corpus profile. This research was supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 "Alignment in Communication", the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277), Bielefeld University, and CLARIN-D, the German division of the "Common Language Resources and Technology Infrastructure". Additionally, we thank the anonymous reviewers for their comments and ideas.

References

Paul Boersma and David Weenink. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9–10):341–345.

Daan Broeder, Thierry Declerck, Erhard Hinrichs, Stelios Piperidis, Laurent Romary, Nicoletta Calzolari, and Peter Wittenburg. 2008. Foundation of a component-based flexible registry for language resources and technology. In *Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation*, pages 1433–1436.

Daan Broeder, Menzo Windhouwer, Dieter van Uytvanck, Twan Goosen, and Thorsten Trippel. 2012. CMDI: a Component Metadata Infrastructure. In *Proceedings of the workshop on Describing LRs with Metadata (LREC 2012)*.

Onno Crasborn and Thomas Hanke. 2003a. Additions to the IMDI metadata set for sign language corpora. Agreements at an ECHO workshop, May 8–9, 2003, Radboud University, Nijmegen. `http://www.ru.nl/publish/pages/522090/signmetadata_oct2003.pdf`.

Onno Crasborn and Thomas Hanke. 2003b. Metadata for sign language corpora. Background document for an ECHO workshop, May 8–9, 2003, Radboud University, Nijmegen. `http://sign-lang.ruhosting.nl/echo/docs/ECHO_Metadata_SL.pdf`.

Onno Crasborn and Menzo Windhouwer. 2012. ISOcat data categories for signed language resources. In Efthimiou, Eleni and Kouroupetroglou, Georgios and Fotinea, Stavroula-Evita, editor, *Gestures in embodied communication and human-computer interaction*, pages 118–128. Springer.

Folkert de Vriend, Daan Broeder, Griet Depoorter, Laura van Eerten, and Dieter van Uytvanck. 2013. Creating & testing CLARIN metadata components. *Language Resources and Evaluation*, 47(4):1315–1326.

[12] `http://tla.mpi.nl/tools/tla-tools/arbil`

Farina Freigang and Kirsten Bergmann. 2013. Towards metadata descriptions for multimodal corpora of natural communication data. In *Proceedings of the workshop on Multimodal Corpora: Beyond Audio and Video, (IVA 2013)*.

Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papageorgiou, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz, and Valerie Mapelli. 2012. The META-SHARE metadata schema for the description of language resources. In *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*.

Thomas Hanke, Lutz König, Sven Wagner, and Silke Matthes. 2010. DGS Corpus & Dicta-Sign: The Hamburg studio setup. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (LREC 2010)*, pages 106–110.

Julius Hassemer. 2015. *Towards a theory of Gesture Form Analysis: Principles of gesture conceptualisation, with empirical support from motion-capture data.* Ph.D. thesis, RWTH Aachen University.

Michael Kipp. 2001. Anvil – A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370.

Andy Lücking, Kirsten Bergmann, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based Analysis of Speech and Gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its Applications. *Journal on Multimodal User Interfaces*, 7(1–2):5–18.

Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 2012. Dicta-Sign – Building a Multilingual Sign Language Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC 2012)*.

David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, Chicago and London.

Thomas Schmidt. 2002. Exmaralda – ein System zur Diskurstranskription auf dem Computer. *Arbeiten zur Mehrsprachigkeit, Serie B (34). Hamburg: SFB Mehrsprachigkeit*.

Peter Withers. 2012. Metadata Management with Arbil. In *Proceedings of the workshop on Describing LRs with Metadata (LREC 2012)*.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.