Multimodal Annotations in Gesture and Sign Language Studies

P. Wittenburg, St. Levinson, S. Kita, H. Brugman

Max-Planck-Institute for Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
peter.wittenburg@mpi.nl

Abstract

For multimodal annotations an exhaustive encoding system for gestures was developed to facilitate research. The structural requirements of multimodal annotations were analyzed to develop an Abstract Corpus Model which is the basis for a powerful annotation and exploitation tool for multimedia recordings and the definition of the XML-based EUDICO Annotation Format. Finally, a metadata-based data management environment has been setup to facilitate resource discovery and especially corpus management. Bt means of an appropriate digitization policy and their online availability researchers have been able to build up a large corpus covering gesture and sign language data.

1. Introduction

The MPI for Psycholinguistics has a long history of research on the synchronization between different modalities in human communication. In the 1980s eyetracking signals and signals about pointing gestures produced important information about the mental processes responsible for speech production [1, 2]. Such signals were typically recorded in relation to spoken utterances. The equipment used was designed to make automatic fine grained temporal analysis possible. For gesture registration IR-light based methods were used. Mort recently, ultrasonic equipment was used for this purpose identifying the location of maximally 8 sources. This tradition is still continued in the "baby labs" where eye tracking is recorded to study, for example, the focus of childrens' attention during linguistic tasks. In recent years brain imaging methods (EEG, MEG, PET, MRI) have often been added to get online information about brain activities during speech production and perception task.

In the last few years, research using multimodality shifted towards observational methods in communicative situations of various sorts. Child-caretaker interaction is studied with the help of extensive video recordings to better understand how childrens' language learning is influenced by 'input' and environmental factors. The use of various types of gestures (pointing, iconic and emblematic) is studied in different situations. The following studies should be mentioned in particular: (1) ethnography of pointing gestures; (2) gestural facilitation of speaking or understanding; (3) gestural expression of motion events; (4) speech dysfluencies and gestures; (5) influence of gestures on recipients' gaze movement; (6) hemispheric specialization of types of gestures [3, 4, 5, 6, 7]. In addition, studies about sign language and their comparison to gestural patterns were carried out [8]. The goal of these recordings is fundamental research about the relation between language and thought and the role of gesture in human communication. Since gestures are very much dependent on language and culture, most of the recordings are cross-linguistic, i.e. various countries and cultures are included.

Nowadays the study of multimodal communication based on video recordings is much easier. Information technology allows science to work with digitized video greatly facilitating the analysis work. For the last two years, all recordings at the MPI have been digitized. yielding an online multimedia corpus consisting of more than 7000 sessions (units of linguistic analysis). Gesture studies form a substantial part of these recordings. Powerful corpus management with the help of metadata descriptions and multimodal annotation tools were developed at the institute to enable the type of research explained. Annotations are stored in well-documented formats well adapted to capturing the complexity of the annotations which are typical of multimodal studies.

2. Multimodality Research

Multi-modal records allow us not only to approach old research problems in new ways, but also open up entirely new avenues of research. An old issue, for example, is just how 'modular' language processing is, that is to what extent non-linguistic processes can intervene in the course of linguistic processing. This can be studied by looking at the interaction between two entirely different behaviour streams, gesture and speech. A large multi-media corpus of natural dialogue shows, for example, that when speakers self-edit speech, gesture inhibition actually occurs earlier, suggesting interaction between the speech and gesture execution systems. Similarly, in the comprehension process it can be shown that gesture content is incorporated into the immediate 'message'. Eye-tracking shows that speakers can manipulate the likelihood of this by looking at their own gestures, which are then more often fixated by listeners. More fundamentally, we can look at the role of the two cerebral hemispheres in the production of the two behaviour streams, speech and gesture. Careful studies of the gestures of split-brain patients show that gesture production is largely driven from the right hemisphere, while language of course is normally processed in the left.

In addition to contributing to such long-standing theoretical issues, annotated multimedia records also make possible entirely new lines of research. For example, we have been interested in whether the semantic character of a specific language leads to a special construal of a scene to be described. The study of gesture during online

production shows that the way a language 'packages' information has a demonstrable effect on the depiction of a scene in gestures. Turkish for example packages movement with direction in a single clause but puts manner of motion into a separate adverbial clause ('The ball descended, rolling') – while English allows manner and direction to occur in the same simple clause ('The ball rolled down'). Turkish speakers tend to produce separate gestures for direction and manner, while English speakers tend to fuse them. In a similar way, we have been able to study spatial thinking as it occurs in non-spatial domains, by examining the gestures of speakers talking about e.g. kinship relations.

Sign languages are another domain which has been opened up by multi-media technology. Sign languages are fully-expressive languages which utilize not only the hands, but also the face, gaze and even body-posture to construct complex utterances with phonology, morphology, syntax and 'prosody'. These different 'articulators' express different distinctions in overlapping time windows, where the offset can indicate e.g. the scope of a question. Even the simplest description of a signed utterance therefore requires a multi-tiered annotation of a video-record, and the development of such annotation tools make possible systematic databases for sign language research for the first time. Fascinating questions can now be pursued about effects of modality on language – for example does the spatial nature of the visual-gestural channel have profound effects on the nature of sign languages, andgive sign languages an underlying commonality? Most deaf signers are exposed to the gestural systems of the surrounding spoken language, and we can also ask to what extent these gestural systems are recruited into the sign language. Preliminary results from the study of a sign language in the process of standardization (Nicaraguan sign language) suggests that there is such an interaction.

These examples should serve to indicate just what a revolution in our understanding of language and its relation to other aspects of cognition is being made possible by the new technologies. There are also fundamental advantages to archiving multi-media records for all branches of the language sciences. For example, studies of the acquisition of language are hugely enriched by having available the very scene available to the infant language user – we now know for example that unexpressed arguments (e.g. subjects and objects) in Inuit care-takers' speech are often recoverable by the child just because they are most likely in the child's field of view at the moment of utterance. Similarly, records of dying or endangered languages are greatly enhanced by having visual information correlate with the language use. In all these cases, richly annotated multi-media records make possible the extraction of systematic information about the correlation of linguistic and non-linguistic events.

3. Gesture Encoding Schemes

General

This variety of studies all based on observational methods (i.e. audio and video, sometimes also gaze) required many different gesture encoding schemes on the different linguistic levels, efficient procedures and powerful tools. Since our researchers are involved in international projects broad agreements on the methods for encoding multimodal behavior are very important. Yet for international standards it seems to be too early, the discipline is too young, although it would facilitate integrating and comparing the data of all the scholarly work.

Most of the studies require careful encoding of the articulator movements and their global timing pattern. Naturally, we are faced with similar problems to those for identifying the articulator movements in the case of speech production. The articulator movements form a continuum, are overlapping and have tolerances dependent on the situation. Therefore, it is not only difficult to make proper time segmentation, but also to classify them. The articulators are different for gestures[1] and Sign Language[2]. In the case of sign language we can identify the same linguistic levels compared to verbal utterances (phonemes, words, morphology, syntax, semantics). In general, the properties of the encoding schemes and requirements are comparable for Sign Language and spoken languages. They are discussed in detail in [9,10]. Also the differences between and within families of Sign Languages were the subject of studies indicating differences in the requirements of encoding them. However, more detailed studies have yet to be carried out on this subject.

For gestures which are movements of the arms and its parts accompanying verbal communication acts, it is sufficient to annotate their type and meaning in addition to the articulators. The type of a gesture is a taxonomic classification of its principle purpose and role in communication. It is widely accepted to separate between pointing, iconic and emblematic gestures. Pointing gestures refer to a spatial point or a movement. They appear either as isolated gestures where the meaning is obvious to the listener or mostly in overlap with verbal utterances where the gestures are much more simple to generate and interpret than verbal descriptions. Their meaning is easy to describe by the object they refer to and their intrinsic purpose. Also iconic gestures appear spontaneously as co-speech activities while emblematic gestures stand alone. Iconic gestures have a culturally bound meaning since they are widely accepted within an area.

[1] For gestures we have as articulators the arms and its parts up to the fingers. Characteristic movements of the head and the eyes in communicative situations are not treated as part of the gesture although they have similar purposes.

[2] Also for Sign Language the movements of the arms are most important. However, other body parts such as eyes, facial expressions, in particular the movement of the lips, orientations of the body and the head are also being used.

Gestures often correlate with emotional state, are used to facilitate the planning of speech production and to facilitate speech perception due to their disambiguation capability. Emotional state can be described, although there are no clear conventions yet.

Articulators in Gestures

The basis of all scientific work when studying gestures is an encoding scheme for the articulator movements. It was soon perceived that an exhaustive gesture encoding including all relevant characteristics would be ideal but impossible (except for small segments). On the other hand the recordings were perceived as so valuable that re-usage for various research questions was anticipated. To cope with this contradiction it was realised that only an iterative encoding approach would suffice where the needs of primary research projects do not hinder the addition of gesture encodings dedicated to completely different research interests. To support research, the underlying scheme should be exhaustive to define a grid allowing easy computational comparison. Therefore, for a number of recordings focused on in the Institute's gesture project, a thorough study was carried out to attain a general gesture encoding scheme that would allow comparative analysis to be made easily.

Based on Kendon's work a more accurate scheme was developed by v. Gijn, vd Hulst and Kita [11] to separate various phases in a gesture. A *MovementUnit* therefore can exist of several *MovementPhrases*. Basically, each of these can be seen as a sequence of a *Preparation* phase, an *ExpressivePhase* and a *Retraction* phase. An *ExpressivePhase* which covers the meaningful nucleus of a gesture is either an *IndependentHold* or a sequence of a *DependentHold*, a *Stroke*, and another *DependentHold*.

*MovementUnit = MovementPhrase**
MovementPhrase = (Preparation) => ExpressivePhase => (Retraction)
ExpressivePhase = IndependentHold
ExpressivePhase = (DependentHold) => Stroke => (DependentHold)
Preparation = (LiberatingMovement) => LocationPreparation >> HandInternalPreparation
Retraction (if subsequent movement) = PartialRetraction

*= consists of, * one or several, => discrete transition, () optional, >> normally blended out, occasionally discrete transition*

The authors developed a set of descriptive criteria to identify the phases and their usefulness was shown in several studies which were successfully annotated by student assistants.

v. Gijn, vd Hulst and Kita also developed an encoding scheme to describe mainly the articulator movements in the *ExpressivePhase* [12]. It is this phase where annotators are confronted with all the about 60 degrees of freedom and where not only the location and shape has to be described but also for example changes in motion and direction. The following aspects are described: *PathMovementShape (straight, circle, round, iconic, 7-form, ?-form, x-form, +-form, z-form), PathMovement Direction ([up|down], [front|back], [ipsilateral| contralateral]), HandOrien-tationChange ([supination| pronation], rotation, [flexion | extension], nodding, [ulnar flexion|radial flexion], lateral flexion), HandShape Change ([opening | closing], [abduction |adduction], [hinging |dehinging], [clawing |declawing], wiggling, opening wave, closing wave, rubbing, cutting), HandOrientation ([up| down], [front | back], [ipsilateral| contralateral]), and HandShape*. For the latter basically the HamNoSys scheme was re-used.

This scheme was also used to describe the movements of the arms and its parts in Sign Language. Due to its exhaustiveness it turned out to be very useful, although it cannot account for as a complete system for Sign Language since other articulators than the arm are included also.

To support the various gesture related research activities simple encoding schemes are most often derived from this exhaustive scheme. The reference back to the unified exhaustive scheme together with the online availability of the annotated multimedia document allows easy re-usage and an enhancement of the annotations. This can either be corrections of the existing or the addition of new tiers. When encoding gestures it is of great importance to understand the exact time relationships with the verbal utterances. This is not part of the gesture annotation scheme, but the annotation structure scheme has to provide adequate mechanisms.

4. Annotation Structures

While the encoding scheme describes how to encode the linguistic phenomena (a close handshape in gestures is encoded as "*close*"), the annotation structure scheme describes the expressive power in structural respects. It has to provide mechanisms for all possible structural phenomena. From our long experience with gesture and sign language studies we know that the annotations can become very complex. There are projects which try to solve this complexity by merging the annotations associated with different linguistic levels into one tier. This method, which is known especially from traditional annotation schemes such as CHAT [13], is also used in new projects. The resulting annotation includes many relations implicitly, i.e. it is the tool which has to include all the knowledge. At the MPI this method was not seen as useful for the future. Different linguistic levels should be separated and all relations such as interruptions, parallelism, semantic correlation should be made explicit. This is the only way to easily modify the coding later.

In many cases different linguistic interpretations of a gesture or sign are possible. The annotation scheme has to take this into account. Essentially, we follow the indicated way: add another tier which can be used by a new annotator. If only adaptations of the existing annotations are intended, a copy action may be useful for bootstrapping the tier.

Number of Tiers

In multimodal annotations one can easily have more than 40 description tiers, since many articulators will be described in addition to the well-known speech description layers. This differs from earlier annotations where one maximally had up to 10 tiers. This requires the support tool to be flexible enough to allow the definition of new tiers without limitation, to do intelligent tier visualization due to the limited pixel space and to store complex tier setups.

Temporal Relations

In multimodal communication and Sign Language the temporal relations of the participating articulators (including the speech track) are of great importance. All sort of relations can occur as indicated in figure 1 a and b.

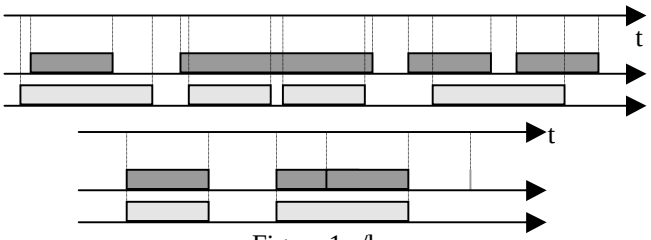

Figure 1 a/b

The first two examples (1a) show types of inclusions, the next (1a) left or right types of overlap, the third and fourth (1b) full matches of movements, and the last (1b) a unitary event which does not have an extension in time and can itself be in relation with movements. The remaining type is that there is no timing overlap between two movements. It is widely agreed to annotate a unitary event as an annotation with a time extension of 1 to simplify the handling. While traditional formats mix linguistic encoding with time marking, it seems to be widely accepted that time information and linguistic encoding should be encoded separately. For many research purposes the elapsed time for a particular type of movement is relevant, but mostly one wants to calculate the temporal relationships (type and time). The most obvious way to do this is to store the time references for all independent annotations. With queries which contain the encoding pattern and the time relationship all corresponding hits can be found. It should be noted that often certain temporal relationships coincide with dependencies (see below).

Spatial Relations

Gesture related research is mostly based on recordings of natural communicative situations such as requesting, for example, a route description. In such cases there is no fixed spatial reference system which could be used for calibration purposes. This is the reason why numerical position detection in general does not help. The information "the left hand hits the left ear" may be relevant to detect meaning. The specification of the coordinates of the left hand only makes sense if we have an environment which allows to automatically infer that the hand hit the ear. For this reason spatial annotations within a coordinate system and beyond the encoding scheme mentioned above are seldom used. However, to remain generic we agree with the suggestions of Bird and Liberman who introduced "regions" in their extended Annotation Graph formalism, not only to represent time regions but also for spatial regions (i.e. locations on the screen). At MPI a similar annotation technique was used to trace the movement of a certain articulator. The location of an articulator was identified with mouse clicks in subsequent frames. Given that the main axis of the camera coincides with the normal vector of the movement, such quantitative annotation can be useful to describe, for example, typical movement patterns.

Hierarchical Relations

In many cases annotations have dependency relations to other annotations. We have indicated above the type of temporal relations which can exist. It was described that these relations can occur incidentally - token based. Often, however, relations are type based, i.e. they are part of the definition of the tier. Three examples are given to demonstrate the problems: (1) In gesture annotation we separate phases. A "MovementPhrase" can exist of a "PreparationPhase", an "ExpressivePhase" and a "RetractionPhase". By definition the start of the "MovementPhrase" coincides with the start of the "PreparationPhase". The same is true for the end of the "MovementPhrase" and the end of the "RetractionPhrase". We can therefore define a time relationship between the types mentioned as indicated in figure 2.

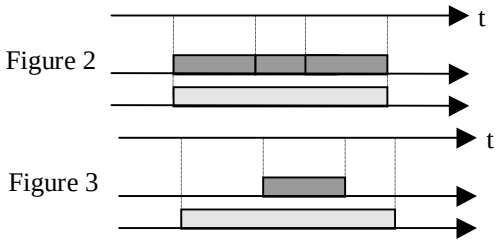

Figure 2

Figure 3

(2) Also in gesture annotations we can identify another time relation. If the handshape of the left hand, for example, is annotated as a part of a "MovementPhrase" then we can infer that the handshape annotation has to be within the boundaries of the "MovementPhrase". This is a constraint which can be implemented as part of the type definition as indicated in figure 3. (3) Another form of dependency can be seen in morphosyntactical and syntactical annotations. The result is the well-known tree structures. Here reference to time no longer makes sense. The dependencies are established between items of annotations on one tier and those of another tier as shown in figure 4.

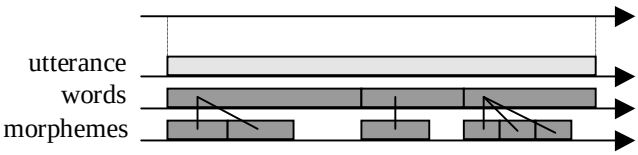

Figure 4

Dependencies imply that if a time was changed all dependent tiers share the same changes if the relation is based on time.

Cross-References

In many instances researchers want to annotate relationships which are beyond the strict hierarchy between annotations. There are many examples of such cross-references. On morphosyntactical level we know for example German words like "auflaufen" which can appear so that the past tense form can be split into "*lief* auf etwas *auf*". Here the first and fourth word together form the verb. Linguists want to mark this relationship within the annotation, i.e. the result is a cross-reference within a tier as indicated in figure 5.

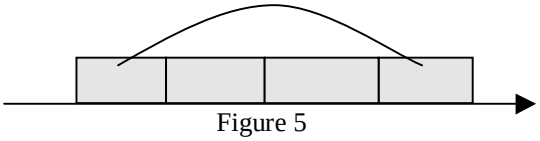

Figure 5

Very often in gesture annotation one wants to encode semantic relations between an element of a verbal utterance and, for example, a gesture as shown in figure 6. There are many other examples in language resources for such cross-references. Often they are related to the encoding of semantic phenomena.

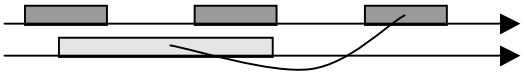

Figure 6

The usage of such cross-references can theoretically amount in recursive pointer structures. This phenomenon was already identified by Bird and Liberman when they extended their first acyclic annotation graph model [14] which was too restricted.

The relations can be of very different natures which requires them to be handled similarly to annotations, i.e. they have a label, type and can themselves be complex structured. An annotation structure scheme has to account for this.

Comments

Comments are a special form of an annotation linked to a linguistic unit which can be everything which is either defined on tier level or as, for example, words in an annotation on utterance level. Comments therefore do not raise new requirements.

Flexibility

The annotation structure scheme has to account for flexibility in several senses. It must be possible to add new annotations onto each tier, to add new tiers and tier types, and to modify existing tier setups and annotations. Tier types specify the tiers which includes, for example, the controlled vocabularies which can be used or the constraints defined for them. A tier type setup is the definition of all tier types and tiers for a given study. Given the complexity of tier setups in multimodal studies it must be possible to re-use tier type setups, i.e. they must be stored in a persistent form.

5. Abstract Corpus Model, Interchange Format and Tools

To provide researchers with an efficient annotation and analysis environment, the Institute began early on to setup digitization lines and to build true multimedia tools. The first was the MAC-based MediaTagger annotation tool [15] built in 1994. MediaTagger's implicit data model was setup to allow the user to define a large number of annotation tiers, his tier setup and tier types including closed vocabularies and tier dependencies. This implicit datamodel was soon extended to a relational database format including the annotations from many studies. This step allowed users to carry out various analyses on the whole or part of the included corpus. MediaTagger soon offered the opportunity to do incremental encoding and the great potential of digital methods and the limitations of the MediaTagger tool and its data model soon became apparent.

Consequently, the Institute decided to fully rely on all-digital techniques, i.e. all video and audio signals were digitized. For video it was decided to rely on MPEG1 (after an initial phase of using MJPEG and CINEPAK). Due to its limited resolution, for example, to identify facial expressions in field recordings, it was then decided to change to MPEG2 as a basis for the multimedia archive which has a factor of about 3 more data and bandwidth.

In 1997 the Institute started developing an Abstract Corpus Model [16] which would encompass the necessary structural richness to represent the annotation phenomena described above and which was seen as the nucleus of the planned new EUDICO multimedia tool set [17]. Various existing and well-known annotation formats such as CHAT, Tipster [18], Shoebox [19], BAS [20], MediaTaggers rDBMS [21], and others were analyzed to attain a format powerful enough to cover the relevant phenomena. At the same time ACM was extended to also cover phenomena such as random cross-references between annotations on different tiers.

The development of the Java-based EUDICO Tool Set for annotating and exploiting multimedia signals was begun in 1998 and has now reached a flexibility and functionality which makes it one of the most advanced tools for multimodal work. Its nucleus is based on ACM, i.e. it has a comprehensive internal representation power. It has a flexible and easy-to-use annotation and time linking component which allows the user to define his tier setup, which can work with audio and/or video signals in the same way and which makes it possible to do the annotation in various writing systems. It has input methods, for example, for IPA, Chinese, Cyrillic, Hebrew and Arabic. Annotations can either be linked to moments in time in the media stream or to other annotations. It is possible to include hierarchical annotations which is necessary, for example, for an interlinearized representation of morphology.

The EUDICO tool set also provides various views on the multimedia data which can be sound, video, or annotation tracks or other types of signals such as eye tracking tracks.

There are a number of stereotypic views on the annotations scientists prefer, therefore EUDICO supports different views and more views can be added according to individual scientists' needs. An important feature is that researchers can easily select and arrange the data tracks they want to see. All viewers in EUDICO are synchronized, i.e. whenever the cursor in a viewer is set to a certain time or segment, all other viewers will move to that instance. The tool set also has a flexible search interface which allows the user to define patterns and associate them with annotation tiers (including all supported input methods) making it possible to enter complex patterns covering several tiers and distances between the patterns. The EUDICO tool set can work in a fully distributed environment where annotation and media tracks are at different locations and support media streaming of fragments. An XML-based generic interchange format was defined (EUDICO Annotation Format), but other formats such as rDBMS, CHAT and Shoebox are also supported. For EAF a schema was developed which is available via the web.

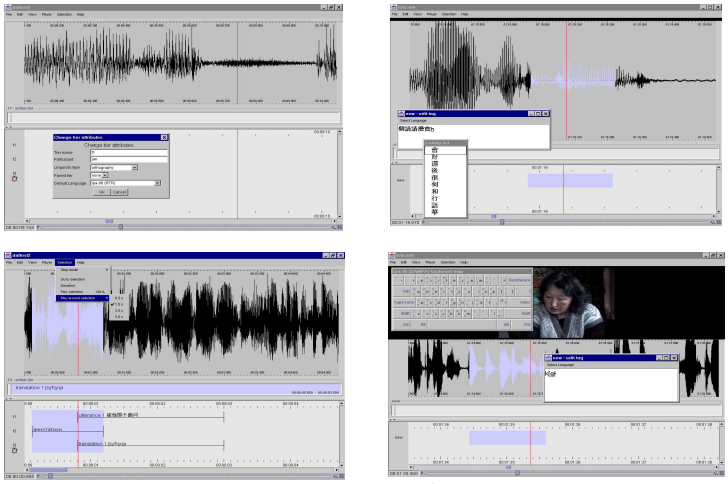

Figure 7 shows some typical screenshots of the annotation tool within EUDICO. It offers much support to easily specify and control segments and to do annotations in various writing systems. For other languages such as Chinese input methods are available.

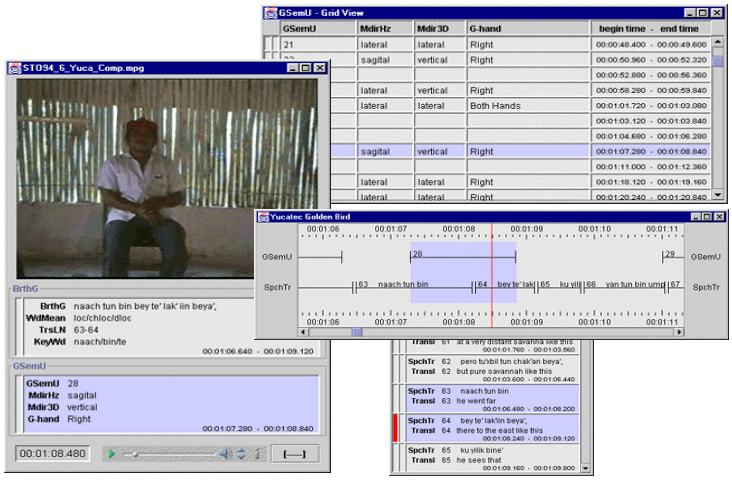

Figure 8 shows the visualization power of EUDICO. Dependent on the project different stereotypic visualizations of the material can be selected. The type of output, the tiers and the order of tiers can be selected by the user. The range of viewers covers dynamic subtitles, a time line view and text viewers with compressed texts.

Tier types can be defined including controlled vocabularies and constraints. Pixel management is very important when dealing with complex tier structures. The user can define the tiers he wants to see and specify the order of presentation. Currently, MPEG1 streaming is supported. MPEG2 is also supported, however downsizing of the video widget is absolutely necessary in order to see the annotations as well.

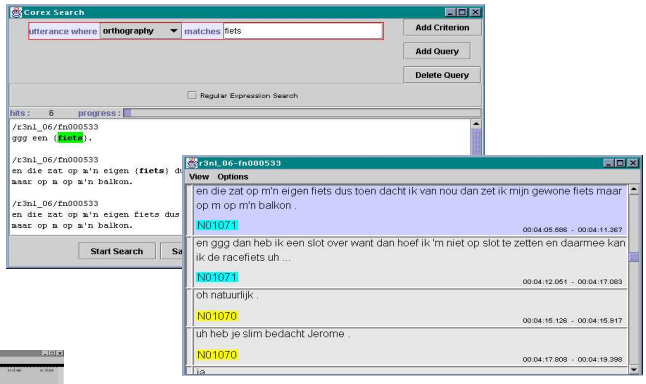

Figure 9 gives an impression of the search feature. It basically allows the user to define search patterns, associate them with tiers and logically combine these patterns to a complete query where also distances can be specified. The result is a list of hits which can be clicked to directly yield the corresponding fragment.

Further details to the EUDICO Tool Set can be seen on the web-page [22].

6. Corpus Management

A large subset of the Institute's multimedia corpus, including most of the gesture and sign language data, was recently described with the help of IMDI type metadata descriptions developed within the ISLE project [23]. All metadata descriptions were then integrated into a browsable and searchable hierarchy where the nodes are meaningful conceptual layers for scientists. Additional information such as project information was hooked up to the different nodes. Researchers can now browse or search in this metadata domain to find desired resources. The browser tool allows for the direct application of one of the possible operations to the resource(s) found such as starting, for example, the EUDICO viewer tool or starting a sound analysis program like PRAAT. Through this process, the MPI has solved the retrieval problem for its large multimedia/multimodal corpus and every researcher with access rights can easily access the resources. The metadata domain is also an excellent domain for corpus management, i.e. to integrate all relevant information.

The whole corpus is available to researchers in the local area network. After some important ethical and legal problems have been solved, the metadata domain will be made available on Internet for external researchers.

Currently, the Institute's multimedia archive contains more than 2 TB digitized recordings and more than 7000 sessions which are the linguistic units of analysis in multimedia recordings.

7. Conclusions

At the MPI for Psycholinguistic the study of human usage of multimodality has been a relevant topic for many years already for various purposes in better understanding speech comprehension, production and language acquisition. Also the relation between language and thought stimulated researchers to investigate especially gestures in various cultures. During the last years the institute shifted to using digitized multimedia recordings. The availability of powerful annotation and exploitation tools and the online availability stimulated the researchers to invest much time in developing exhaustive schemes for gesture encoding which were applied successfully. Tools for managing corpora allowed the researchers to collect and exploit a large gesture corpus.

International collaboration, however, is often prohibited or slowed down due to the poor degree of agreements amongst the researchers about good encoding schemes, open and powerful formats based on XML definitions. The institute investigated time to define such schemes. The EUDICO Annotation Format is based on the results of building the Abstract Corpus Model and fits to the needs. However, when the ATLAS Interchange Format will be mature enough the institute would like to turn over to facilitate the exchange and re-usability of data. Good tools for the annotation and exploitation of multimedia data have been developed which are based on the described ideas and the institute is happy to share them with others. In the framework of the ISLE project the institute could participate in building a metadata environment which turns out to be extremely useful not only for resource discovery, but also for managing large multimedia corpora with multimodal annotations.

8. References

[1] W.J.M. Levelt (1980). Online processing contraints on the properties of signed and spoken language. In Biological Constraints on linguistic form. U. Bellugi, M. Studdert-Kennedy (eds.). Vgl. Chemie, Weinheim.

[2] G. Richardson (1984). Word recognition under spatial transformation in retarded and normal readers. Journal of Experimental Child Psychology 38, 220-240.

[3] S. Kita, J. Essegbey (to appear). Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. Gesture.

[4] S. Kita (1998). Expressing a turn at an invisable location in route direction. In Ernest Hess-Lüttich, J.E. Müller & A. vanZoest (eds.), Signs & SPace. 159-172. Tübingen: Narr.

[5] A. Özyürek, S. Kita (1999). Expressing manner and path in English and Turkish: Differnces in speech, gestures, and conceptualization. In M. Hahn and C. Stones (eds.), Proceedings of the 21 st Annual Meeting of the Coginitive Science Society. 507-512. Amsterdam.

[6] M. Gullberg, K. Holmqvist (2001). Eye tracking and the perception of gestures in face-to-face interaction vs. on screen. In C. Cave, I. Guaitella, S. Santi (Eds.), Oralite et gesturalite: Interactions et comportemetns multimodaux dans la communication (pp. 381-384). Paris: L'Harmattan.

[7] H. Lausberg, S. Kita (2001). Hemispheric specialization in spontaneous gesticulation investigated in split-brain patients. In C. Cave, I. Guaitella, S. Santi (Eds.), Oralite et gesturalite: Interactions et comportemetns multimodaux dans la communication (pp. 431-434). Paris: L'Harmattan.

[8] M. Seyfeddinipur, S. Kita (2001). Gesture and dysfluency in speech. In C. Cave, I. Guaitella, S. Santi (Eds.), Oralite et gesturalite: Interactions et comportemetns multimodaux dans la communication (pp. 266-270). Paris: L'Harmattan.

[9] R. Sutton-Spence (1999). The linguistics of British Sign Language: an introduction. Cambridge University Press. Cambridge

[10] K. Emmorey (2001) Language, cognition and the brain: insights from sign language research. Erlbaum, Hilldale, NJ

[11] S. Kita, I. v. Gijn, H. vd. Hulst (1998). Movement Phases in Signs and Co-speech Gestures, and their Transcription by Human Coders. In I. Wachsmuth and Martin Frühlich (eds.), Gesture and Sign Language in Human-Computer Interaction, Vol. 1371: 23-35. Proceedings of the International Gesture Workshop Bielefeld, Lecture Notes in Artificial Intelligence. Berlin: Springer Verlag.

[12] S. Kita, I. v. Gijn, H. vd. Hulst (2000). Gesture Encoding. MPI Internal Report.

[13] B. MacWhinney (1999). The CHILDES Project: tools for analyzing Talk. Second ed. Hillsdale, NJ: Lawrence Erlbaum.

[14] S. Bird, M. Liberman (2001). A formal framework for linguistic annotation. Speech Communication 33 (1,2), pp 23-60

[15] www.mpi.nl/world/tg/lapp/mt/mt.html

[16] H. Brugman, P. Wittenburg (2001). The application of annotation models for the construction of databases and tools. In Proceed*ings of the Workshop on Linguistic Databases*. Philadelphia.

[17] www.mpi.nl/world/tg/lapp/eudico/eudico.html

[18] www.cs.nyu.edu/cs/faculty/grishman/tipster.html

[19] www.sil.org/computing/catalog/shoebox.html

[20] www.phonetik.unimuenchen.de/Bas/ BasHomedeu.html

[21] www.mpi.nl/world/tg/CAVA/CAVA.html

[22] www.mpi.nl/tools

[23] www.mpi.nl/ISLE