Natural Interactivity Resources - Data, Annotation Schemes and Tools

Laila Dybkjær and Niels Ole Bernsen

Natural Interactive Systems Laboratory (NISLab)
University of Southern Denmark
Science Park 10, 5230 Odense M, Denmark
{laila, nob}@nis.sdu.dk

Abstract

This paper presents results of three surveys of natural interactivity and multimodal resources carried out by a Working Group in the ISLE project on International Standards for Language Engineering. Information has been collected on a large number of corpora, coding schemes and coding tools world-wide. The paper presents the information collection process, the description and validation methods used, the surveyed resources, and brief conclusions for each of the three resource areas reviewed. Observations on user profiles, user needs and best practices are briefly presented.

1. International Standards for Language Engineering surveys

The long-term vision of natural interactivity envisions that humans communicate, or exchange information, with machines (or systems) in the same ways in which humans communicate with one another, using thoroughly co-ordinated speech, gesture, gaze, facial expression, head movement, bodily posture, and object manipulation [Bernsen 2001]. The idea of multimodality is to improve human-system interaction in various ways by using novel combinations of (unimodal) input/output modalities [Bernsen 2002]. Natural interactivity is by nature (mostly) multimodal.

Across the world, researchers and companies are beginning to focus on investigating and exploiting the potential of natural interactive and multimodal systems. An important foundation for work on such systems is resources, i.e. data (corpora), corpus annotation schemes and annotation tools. A good starting-point for those working in the field therefore is information about which resources are already there, how they might be accessed, what they might be used for, etc., so that fewer people try to re-invent the wheel than might otherwise be the case.

This paper presents substantial and, to our knowledge, unprecedented groundwork on resources carried out in the European Natural Interaction and Multimodality (NIMM) Working Group of the EU-HLT/US-NSF project International Standards for Language Engineering (ISLE). ISLE is the successor of EAGLES (European Advisory Group for Language Engineering Standards) I and II and includes working groups on lexicons, machine translation evaluation, and NIMM, respectively. The NIMM working group (isle.nis.sdu.dk) began its work in early 2000 and has now completed three comprehensive surveys. The surveys address NIMM data, annotation schemes, and annotation tools, respectively. Focus has been on producing resource descriptions which are systematic, follow standard formats, and are sufficient for providing interested parties in research and industry with the information they need to decide if a particular resource matches their interests. Each resource comes with contact information on its creator(s). The three surveys are available in html and pdf format at the ISLE NIMM website isle.nis.sdu.dk.

The report on NIMM data resources [Knudsen et al. 2002a] reviews a total of 64 resources world-wide, 36 of which are facial resources and 28 are gesture resources. Several corpora combine speech with facial expression and/or gesture. The report also includes a survey of market and user needs produced by ELRA (the European Language Resources Agency) and 28 filled questionnaires collected at the Dagstuhl workshop on Coordination and Fusion in Multimodal Interaction held in late 2001.

The survey of NIMM corpus annotation schemes [Knudsen et al. 2002b] reviews 7 descriptions of coding schemes for facial expression and speech, and 11 descriptions of annotation schemes for gesture and speech.

The survey of NIMM corpus coding tools [Dybkjær et al. 2001] describes 12 annotation tools and tool projects most of which support speech annotation combined with gesture annotation, facial expression annotation, or both.

In the following, we first describe the three surveys in more detail (Sections 2-4). We then present conclusions on users and resources (Section 5).

2. Data resources

The approach adopted for producing the NIMM data resources survey [Knudsen et al. 2002a] was to (i) first identify a common set of criteria for selecting the data resources to be described and decide upon issues concerning quality of content as well as of presentation; then (ii) establish a common template for describing each data resource; (iii) identify relevant data resources world-wide based on the web, literature, networking contacts among researchers in the field, etc.; and, finally (iv), interact with the data resource creators to the extent possible in order to gather information on their resources and ask them to verify the data resource descriptions produced. These four steps are described in the following.

2.1. Criteria and quality

The first step taken was to identify the following set of selection criteria and decide on guidelines for ensuring quality of contents and presentation.

Accessibility: The data resource must be accessible for research and/or industrial purposes. An indication should be included of whether a resource is free or if there is a fee to be paid.

Annotation: If a data resource has been marked up this is considered an advantage. The coding scheme used should be included in the ISLE NIMM report on coding schemes if it satisfies the requirements for inclusion in

that report (Section 3.1). If the coding scheme does not satisfy those requirements, a short informal coding scheme description should be included in the data resource report alongside the corpus description. If not marked up, the corpus should be highly suitable for markup and the types of phenomena which could be marked up should be indicated.

Exceptions: Exception to the above is only to be made if a data resource is so rare or innovative for its domain that its very existence might be of interest to researchers in the field.

NIMM data resources are not always easy to get access to. We have adopted the following guidelines for contents inspection, validation and presentation:

Access: It is highly desirable that the describer of a certain data resource has actually had access to that resource. If this has not been possible, it should be clearly indicated in the description.

Validation: All descriptions should be validated by someone other than the describer, if at all possible with the data resource creator in the loop, either as describer or as validator.

Examples: Whenever permissible, a short example of the data resource should be presented in the ISLE NIMM survey. If, for whatever reason, it has not been possible to access and inspect a resource example first-hand, this should be stated clearly in the description.

2.2. Description template

In order to help describers and providers of data resource information document the data resources, facilitate the reading of the ISLE NIMM survey, and allow some measure of easy comparison among the data resources presented, resource descriptions have a common structure which, to the extent possible, provides the same kinds of information about all data resources. The common structure includes eight main entries in addition to the resource name as header, as shown in Figure 1. Each main entry subsumes a number of more specific information items.

The common description template went through several revisions as work on the survey proceeded, for instance in order to take into account types of information which it would be useful to include but which had not been anticipated from the outset. One example is the concluding question in Figure 1 about creators' regrets. It turned out that several data resource creators, being first-time creators, were keen to share their experience on pitfalls in data resource creation in order to help others avoid the errors made.

Reference (specify resource by project name, main authors or laboratory)
Description header
Main actor (name and email)
Verifying actor (name and email)
Date of last modification of the description
References
Web site(s) (make a short description of what can be found on the site.)
Short description
Illustrative sample picture or video file
References to additional information on the reviewed resource (journal or conference paper, white paper, …)
Recorded human behaviour
How many different humans have been recorded in the whole resource (none, one, two, more than two)?
How many humans are recorded at the same time (visible in the same frame)?
What is their profile (age, gender, profession…)?
Which human body parts are visible in the resource (face, arms, hand, whole body, …)?
Which modalities are annotated (speech, hand gesture, arm gesture, body posture, facial expression, …)?
Which other modalities are available/visible in the resource but have not been annotated (speech, hand gesture, arm gesture, body posture, facial expression, …)?
Recorded computer behaviour
Which interactive media are visible/audible in the resource and are used by the humans (none, graphical screen, computer pen, tactile screen, data glove, loudspeakers, …)?
Recording
What are the file types included in the resource? Are they organised in a database structure?
How much data does the resource contain (measured in duration, number of dialogues, or Mb)?
Who created the resource and when?
How was the resource created?
What is the application area (none, tourism, education, arts, …)?
What was the original purpose of creating the resource?
Accessibility
How does one get access to the resource?
Is the resource available for free or how much does it cost?
Has value been added to the original resource in terms of, e.g., transcriptions, annotations and/or tools, which are now available along with the original resource or otherwise?
Did the reviewer have access to the resource?
Usage
Which purpose(s) can the resource be used for/has the resource been used for?
Who used the resource so far/who are the target users of the resource?
Is the resource language dependent (which language(s)) or language independent?
Conclusion
How interesting/important/high quality is the resource?
What do the authors regret, if anything, (not) to have done while building the resource?

Figure 1. Common data resource description template.

2.3. Surveyed data resources

Relevant data resources were identified primarily via ISLE NIMM participants, the web, and published proceedings in the NIMM area. Figure 2 lists the reviewed data resources. The overall division is into data resources which have their main focus on facial expression, possibly including speech and other modalities, and data resources which have their main focus on gesture, possibly combined with speech and other modalities.

2.4. Interaction with corpus creators

Close interaction with the creators of data resources has been sought throughout the writing of the report on

NIMM data resources. We did that, firstly, of course, to seek their permission to publicly describe their data resource, and secondly to invite their collaboration in producing as useful and accurate information about the data resource as possible. Creators of the data resources reviewed were invited to comment on the description of their data resource and to validate the final description, resulting in feedback on, and validations of, more than half of the descriptions made (including the "lesser known" resources which are data resources for which we have not been able to find very much information). Many data resource creators pointed out the potential value of the data resource survey. In a few cases, data resource creators had already answered a questionnaire from COCOSDA (The International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques for Speech Input/Output) and did not want to repeat a similar exercise.

Dynamic Facial Data Resources with Audio

1. Advanced Multimedia Processing Lab
2. ATR Database for bimodal speech recognition
3. The BT DAVID Database
4. Data resources from the SmartKom project
5. FaceWorks
6. M2VTS Multimodal Face Database
7. M2VTS Extended Multimodal Face Database – (XM2VTSDB)
8. Multi-talker database
9. NITE Floorplan Corpus (Natural Interactivity Tools Engineering)
10. Scan MMC (Score Analysed MultiModal Communication)
11. VIDAS (VIDeo ASsisted with audio coding and representation)
12. /'VCV/ database

Dynamic Facial Data Resources without Audio

1. LIMSI Gaze Corpus (CAPRE)

Static Facial Data Resources

2. 3D_RMA: 3D database
3. AR Face Database
4. AT&T Laboratories Database of Faces
5. CMU Pose, Illumination, and Expression (PIE) database
6. Cohn-Kanade AU-Coded Facial Expression Database
7. FERET Database Demo
8. Psychological Image Collection at Stirling (PICS)
9. TULIPS 1.0
10. UMIST Face Database
11. University of Oulu Physics-Based Face Database
12. VASC – CMU Face Detection Databases
13. Visible Human Project
14. Yale Face Database
15. Yale Face Database B

Lesser Known/Used Facial Data Resources

1. 3D Surface Imaging in Medical Applications
2. ATR Database for Talking Face
3. Audio-Visual Speech Processing Project
4. Facial Feature Recognition using Neural Networks
5. Image Database of Facial Actions and Expressions
6. JAFFE Facial Expression Image Database
7. Multi-modal dialogue corpus
8. Photobook
9. Video Rewrite

Gesture Data Resources

1. ATR Multimodal human-human interaction database
2. CHCC OGI Multimodal Real Estate Map
3. GRC Multimodal Dialogue during Work Meeting
4. LIMSI Multimodal Dialogues between Car Driver and Copilot Corpus
5. LIMSI Pointing Gesture Corpus (PoG)
6. McGill University, School of Communication Sciences & Disorders, Corpus of gesture production during stuttered speech
7. MPI Experiments with Partial and Complete Callosotomy Patients Corpus
8. MPI Historical Description of Local Environment Corpus
9. MPI Living Space Description Corpus
10. MPI Locally-situated Narratives Corpus
11. MPI Narrative Elicited by an Animated Cartoon "Canary Row" Corpus 1
12. MPI Narrative Elicited by an Animated Cartoon "Canary Row" Corpus 2
13. MPI Narrative Elicited by an Animated Cartoon "Maus" and "Canary Row" Corpus
14. MPI Natural Conversation Corpus
15. MPI Naturalistic Route Description Corpus 1
16. MPI Naturalistic Route Description Corpus 2
17. MPI Traditional Mythical Stories Corpus
18. MPI Traditional Mythical Stories with Sand Drawings Corpus
19. National Autonomous University of Mexico, DIME multimodal corpus
20. National Center for Sign Language and Gesture Resources
21. RWC Multimodal database of gestures and speech
22. University of Chicago Origami Multimodal corpus
23. VISLab Cross-Modal Analysis of Signal and Sense Data and Computational Resources for Gesture, Speech and Gaze Research

Lesser Known/Used Gesture Data Resources

1. ATR sign language gesture corpora
2. IRISA Georal Multimodal Corpus
3. LORIA Multimodal Dialogues Corpus
4. University of California Video Series on Nonverbal Communication
5. University of Venice Multimodal Transcription of a Television Advertisement

Figure 2. Data resources surveyed.

2.5. Conclusions on data resources

The reviewed data resources reflect a multitude of needs and purposes, including the following (in random order):

- automatic analysis and recognition of facial expressions, including lip movements;
- audio-visual speech recognition;
- study of emotions, communicative facial expressions, phonetics, multimodal behaviour, etc.;
- creation of synthetic characters, including, e.g., talking heads;
- automatic person identification;
- training of speech, gesture and emotion recognisers;
- multimodal and natural interactive systems specification and development.

In many cases, the people working with the data, in particular those working with static image analysis, have

created their own resource databases. Algorithms for image analysis are sometimes dependent on lighting conditions, picture size, subjects' face orientations, etc. Thus, computer vision research groups often have had to create their own image databases. Image analysis using computer vision techniques remains a difficult task, and this may be the reason why we have primarily found static image resources produced by workers in this field.

In other areas, video recordings - mostly including audio - are needed. For example, studies of lip movements during speech, co-articulation, audio-visual speech recognition, temporal correlations between speech and gesture, and relationships among gesture, facial expression and speech, all require video recordings with audio.

Significantly, across all the collected data resources, re-use is a rare phenomenon. If a data resource has been created for a specific application purpose, it has usually been tailored to satisfy the particular needs of its creators, highlighting, e.g., particular kinds of interaction or the use of particular modality combinations. However, the lack of re-use may also to some extent be due to the fact that existing resources may be difficult to locate. On the other hand, it should be mentioned that vendors of data resources exist (e.g. ELRA and LDC). See [Dybkjær and Bernsen 2002] for a more detailed description of the intended and actual use of the surveyed data resources.

3. Annotation schemes

The approach adopted for producing the ISLE NIMM annotation schemes survey [Knudsen et al. 2002b] was basically the same as the one reported for data resources in section 2. Thus, the four steps of (i) identifying selection criteria and deciding on issues concerning quality of content and of presentation, (ii) establishing a common template for describing each coding scheme, (iii) identifying relevant coding schemes, and (iv), interacting with the coding scheme creators, were also followed in the coding schemes description process. These steps are described in the following sections.

3.1. Criteria and quality

To keep the survey focused on reasonably well-documented and validated coding schemes, the following criteria were adopted for selecting the coding schemes to be included in the ISLE NIMM survey:

Documentation: It is interesting if the coding scheme is well-documented in the sense that it has a coding book which describes the purpose and the domain for which the coding scheme has been developed. Exception to this point is made if the coding scheme is rare in its domain or still under development. Moreover, the coding scheme should be substantiated by examples for better understanding and come with a contact address where one can get further information.

Usability: The coding scheme should have been used by a decent number of researchers. This criterion was adopted in view of the fact that coding schemes that have only been used by their developers tend to be too subjective and difficult to use. However, a coding scheme that has been used only by its developers can be included in the survey, provided that the scheme has been used to code a large data resource which is included in the ISLE NIMM survey of data resources. Moreover, to demonstrate its usability, the coding scheme must have been used to annotate a certain number of dialogues, or interactions, and it must be in recent use or have potential for future use.

Mark-up language: The coding scheme must come with a list of phenomena which have been annotated by using the coding scheme (tag set) in order to enable comparison with related coding schemes. Moreover, the markup language should be described.

Evaluation desirability: It is interesting if users outside the group of developers have evaluated the coding scheme on, e.g., matters of inter-coder agreement. It is interesting if the coding scheme has been used to code a certain number of the data resources included in the ISLE survey of NIMM data resources. Moreover, it is interesting if the coding scheme has tool support, and if the tool is included in the ISLE survey on NIMM coding tools.

Exceptions: Exception may be made to the above if a coding scheme is so rare or innovative for its domain that its very existence might be of interest to researchers in the field. However, it is still desirable that the coding scheme is generalisable to a certain degree, at least, so that it is not only suitable for a single, very particular and limited task.

NIMM coding schemes tend to be rather voluminous, and they are sometimes carefully protected against intruders in the sense that it costs time and money to become an approved-by-the-developers user of them. To realistically compromise among the above selection criteria, we have adopted the following guidelines for contents inspection, validation and presentation:

Hands-on: It is highly desirable that the describer of a certain coding scheme has actually tried to use the coding scheme.

Validation: All descriptions should be validated by someone other than the describer, if at all possible with the coding scheme creator in the loop, either as describer or as validator.

Examples: Whenever permissible, a short example of coding of a resource should be presented in the survey. If, for whatever reason, it has not been possible to access and inspect a coding example first-hand, this should be stated clearly in the description.

3.2. Description template

In order to help providers of coding scheme information to document their coding schemes, facilitate the reading of the ISLE NIMM survey, and allow some measure of easy comparison among the coding schemes presented, coding scheme descriptions have a common structure which, to the extent possible, provides the same kinds of information about all coding schemes. The common structure includes seven main entries in addition to the name of the coding scheme as header, as shown in Figure 3. Each main entry subsumes a number of more specific information items.

The common description template went through several revisions as work on the survey proceeded, for instance in order to take into account types of coding schemes which had not been anticipated from the outset. Other adjustments became necessary during the validation process where the close contact to the coding scheme creators often demonstrated that the creators took a critical approach to their own coding schemes, providing valuable

information on what they would do differently were they to create their coding schemes once more.

Reference (specify coding scheme by project name, main authors or laboratory)
Description header
Main actor (name and email)
Verifying actor (name and email)
Date of last modification of the description
References
Web site(s)
Short description
Illustrative example of coding
References to additional information on the coding scheme (journal or conference paper, white paper, …)
Coverage
Which types of raw data are referenced?
Which modalities is the coding scheme meant to code?
Which annotation level(s) does the coding scheme cover, such as facial expression and prosody?
Which coding tasks has the coding scheme been used for?
Detailed description of coding scheme
Which header file information is included (meta-data)?
Coding purpose of the coding scheme?
List and description of phenomena which can be annotated by the scheme
Description of markup language/markup declaration
Examples
Description of coding procedure, if any
Creation notes, i.e. who wrote the coding scheme, when, and in which context?
Usage
Origin of the coding scheme and reasons for creating it
How many people have used the coding scheme and for what purposes?
How many dialogues, or interactions, have been annotated using the coding scheme?
Has the coding scheme been evaluated?
Is the coding scheme language dependent or language independent?
Is there tools support for using the coding scheme or an API for editing/parsing coded descriptions? In which language?
Accessibility
How does one get access to the coding scheme?
Is the coding scheme available for free or how much does it cost?
Conclusion
How well described is the coding scheme?
How general and useful is the coding scheme?

Figure 3. Common coding scheme description template.

3.3. Surveyed annotation schemes

As in the case of the data resources survey, relevant coding schemes were identified primarily via ISLE NIMM participants, the web, and published proceedings in the NIMM area. Figure 4 lists the reviewed coding schemes. The overall division is into facial coding schemes and gesture coding schemes. It should be noted that four of the entries in Figure 4 (number 2 under lesser know facial coding schemes, number 11 under gesture schemes, and numbers 2 and 3 under lesser known gesture schemes) are not coding schemes proper but rather more general descriptions of coding schemes for particular modalities.

Facial Coding Schemes
1. The Alphabet of eyes: formational parameters of gaze
2. Facial Action Coding System – FACS
3. The Maximally Discriminative Facial Movement Coding System (MAX)
4. MPEG-4 SNHC (Moving Pictures Expert Group, Synthetic/Natural Hybrid Coding)
5. ToonFace

Lesser Known/Used Facial Coding Schemes
1. BABYFACS – Facial Action Coding System for Baby Faces
2. General description of coding schemes for hand annotation of mouth and lip movements and speech

Gesture Coding Schemes
1. DIME: National Autonomous University of Mexico, Multimodal extension of DAMSL
2. HamNoSys - Hamburg Notation System for Sign Languages
3. HIAT -- Halbinterpretative Arbeitstranskriptionen
4. LIMSI Coding Scheme for Multimodal Dialogues between Car Driver and Copilot
5. MPI GesturePhone
6. MPI Movement Phase Coding Scheme
7. MPML - A Multimodal Presentation Markup Language with Character Agent Control Functions
8. SmartKom Coding scheme
9. SWML (SignWriting Markup Language)
10. TUSNELDA Corpus Annotation standard
11. General description of coding schemes for prosody, gestures and speech

Lesser Known/Used Gesture Coding Schemes
1. LIMSI TYCOON scheme for analysing cooperation between modalities
2. W3C Working Draft on Multimodal Requirements for Voice Markup Languages
3. The New England Regional Leadership Non-Verbal Coding scheme

Figure 4. Annotation schemes surveyed.

3.4. Interaction with coding scheme creators

Close interaction with the creators of the reviewed coding schemes was sought throughout the description process. Each creator was invited to comment on the description of that creator's coding scheme and to validate the final description. In this way, we managed to get feedback on, and validations of, most descriptions. Many coding schemes creators pointed out the potential value of the ISLE NIMM survey, arguing that had the survey been available when they first needed coding schemes for their own research, this might have made their work easier because they might have been able to use an already existing coding scheme instead of going through the laborious process of creating their own, or they might have been in a better position to learn from other researchers' experience with coding scheme creation.

3.5. Conclusions on annotation schemes

There probably exists a wealth of NIMM annotation schemes out there, far more than those which are presented in the current edition of the ISLE NIMM

survey. Most of them are tailored to a particular purpose and used solely by their creators or at the creators' site. Such coding schemes tend not to be very well described and they tend to be hard to find. The survey on annotation schemes includes a number of such coding schemes, several of which have been created by the ISLE NIMM participants themselves or by people known to ISLE NIMM participants, this being the main reason why we were aware of them. Other coding schemes included in the survey are fairly general-purpose ones, in frequent use, or even considered standards in their field.

Nearly all the reviewed coding schemes are aimed at markup of video, sometimes including audio. A couple of schemes are used for static image markup. See [Dybkjær and Bernsen 2002] for a more detailed description of the intended and actual use of the surveyed coding schemes.

Based on the collected material, it may safely be concluded that there is still a very long way to go before we will be able to code, on a scientifically sound basis, natural interactive communication and multimodal information exchange in all their forms, at any relevant level of analytic detail, and in all their cross-level and cross-modality forms. This observation is already true for the coding of spoken dialogue at several important levels of analysis, such as dialogue acts or co-reference, as shown in the MATE survey of spoken dialogue annotation schemes [Klein et al. 1998] which is available at the MATE website at mate.nis.sdu.dk. When we move beyond spoken dialogue annotation to considering facial coding, we do find a couple of general and substantially evaluated coding schemes for different aspects of the facial expression of information (eyes, facial muscles), cf. Figure 4. It seems clear, however, that we still need a number of higher-level facial coding schemes based on solid science for how the face manages to express cognitive properties, such as emotions, purposes, attitudes and character. In the general field of gesture, moreover, the state of the art is even further from the ideal described above. General coding schemes which go beyond the classification of gesture into few broad categories, and as opposed to coding schemes designed for the study of particular kinds of task-dependent gesture, are hard to find at all, the only exception being in the specialised field of sign languages. Also, the state of evaluation of particular gesture coding schemes is generally poor. Finally, when it comes to the most complex, and perhaps ultimately the most significant, of all areas of natural interactive behaviour annotation, i.e. that of cross-level and cross-modality coding, no coding scheme of a general-purpose nature would seem to exist at all. Even special-purpose coding schemes are hard to come by as yet in this area.

A key to progress, it would seem, is the availability of general-purpose coding tools for natural interactive and multimodal behaviour. Such tools do not yet exist, but their existence could mean a breakthrough in the scientific study of how humans express information through the intriguingly complex and massively coordinated use of multiple modalities and at multiple levels of abstraction within each modality involved.

4. Annotation tools

A number of tools in support of natural interactivity and multimodal data annotation, i.e. tools which support annotation of spoken dialogue, facial expression, gesture, bodily posture, or cross-modality issues, were reviewed in the ISLE NIMM coding tools survey [Dybkjær et al. 2001]. For this survey, and in view of the expected scarcity of NIMM coding tools world-wide, no particular selection criteria were set up except that it should be possible to somehow get access to the tools reviewed. With this exception, the same approach was taken as for the descriptions of NIMM data resources and coding schemes, i.e. (i) a common template was established for describing each coding tool, (ii) relevant coding tools were identified, and (iii) coding tool creators were contacted.

Name of tool
Introduction
Aim of the tool
Which domain does the tool cover?
Which task(s) does is solve?
What was the reason for creating the tool (user needs, own needs, curiosity, ...)?
Which version of the tool is being reviewed?
Is a demo available?
If yes, is the demo freely available or is there a fee?
Is the review based on hands-on experience with the tool, written descriptions, or other sources (include them in the reference list)?
Software design
Architecture
Implementation language(s)
Other software-related issues
Platform requirements
Operating system(s) on which the tool runs
Special hardware requirements, if any
Interface
Description of interface design and usability of the tool
Include screen shots if possible
Advantages and disadvantages of the interface
Functionality
Description of each main functionality with example(s) for each
Does the tool offer the functionality it promises?
How useful is it?
Advantages and disadvantages
Conclusion
General assessment of usability and functionality of the tool
How interesting is the tool (whole tool, particular aspects) for the long-term purpose of creating a tool in support of annotation of natural interactivity and multimodal data?
References

Figure 5. Common coding tool description template.

4.1. Description template

The tool descriptions share a common structure agreed upon early in the writing process. The purpose of this structure is to facilitate comparison across tools by providing similar information about each tool to the extent possible. The common structure has seven main entries in addition to the name of the tool, as shown in Figure 5. Entries may be merged or left out in the individual reviews, depending on the information available and how it is best presented. Under each main entry, a set of more specific information items have been added to serve as

guidelines for which information it would be desirable to include under each entry.

4.2. Surveyed annotation tools

Figure 6 lists the reviewed NIMM coding tools and tool projects. A couple of tools had not yet been implemented at the time of the review. However, the tool concepts presented by the projects developing the tools were found sufficiently interesting for including a project description, e.g., because the project has standardisation among its goals. Two tools are professional (commercial). The rest are research tools (or projects). MATE is a limiting case in another sense, because the MATE Workbench only supports spoken dialogue and text annotation. The tool is included because of its advanced properties for multi-level and cross-level annotation, which may show the way towards building a general-purpose natural interactivity coding tool. The CLSU Toolkit coding tool is for output generation. Finally, so far, at least, the SmartKom project is a user rather than a provider of NIMM coding tools.

1. *Anvil:* Annotation of Video and Language Data, is a Java-based tool for annotating digital video files. See www.dfki.de/~kipp/anvil
2. *ATLAS:* Architecture and Tools for Linguistic Analysis Systems. No tool was available at the time of the review. See www.itl.nist.gov/iaui/894.01/atlas
3. *CLAN:* Computerized Language Analysis, is a program designed specifically for analysing data transcribed in the format of the Child Language Data Exchange System (CHILDES). Transcriptions can be linked to audio or video files. See childes.psy.cmu.edu
4. *CSLU Toolkit:* Center for Spoken Language Understanding Toolkit, is a suite of tools including the Rapid Application Developer, BaldiSync (for facial animation), SpeechView, OGIsable (an annotation tool), speech recognition tools, and a programming environment (CSLUsh). OGIsable is the only annotation tool included. It allows the user to attach properties to a text before it is spoken (via the Festival TTS engine), e.g. to synthesise facial expression synchronised with speech output. See cslu.cse.ogi.edu/toolkit/
5. *MATE:* Multilevel Annotation Tools Engineering. The MATE Workbench is a Java-based tool in support of multi-level annotation of spoken dialogue corpora and information extraction from annotated corpora. See mate.nis.sdu.dk
6. *MPI tools:* CAVA and EUDICO/Computer Assisted Video Analysis and European Distributed Corpora. Both tools support annotation of audio-visual files and information extraction. See www.mpi.nl/world/tg/CAVA/CAVA.html, www.mpi.nl/world/tg/lapp/eudico/eudico.html
7. *MultiTool* was developed in a project on a Platform for Multimodal Spoken Language Corpora. It is a Java-based tool in support of the creation and use of multimodal spoken language corpora (audio and video). See www.ling.gu.se/multitool
8. *The Observer* is a professional system for the collection, analysis, presentation and management of video data. It can be used to record activities, postures, movements, positions, facial expressions, social interactions or any other aspect of human or animal behaviour as time series of tagged data. See www.noldus.com/products/index.html?observer/
9. *Signstream* was developed as part of the American Sign Language Linguistic Research Project. It is a database tool for analysis of linguistic data captured on video. See web.bu.edu/asllrp/SignStream
10. *SmartKom* is a large-scale project which aims to merge the advantages of spoken dialogue-based communication with the advantages of a mixture of graphical user interfaces and gesture and mimetic interaction. SmartKom uses tools developed elsewhere: in Verbmobil for audio annotation, Anvil for mimics and gesture coding. See www.smartkom.org
11. *SyncWriter* is a professional tool for transcription and annotation of synchronous "events" such as speech and video data. See www.sign-lang.uni-hamburg.de/software/software.html
12. *TalkBank* is a project which aims to provide standards and tools for creating, searching, and publishing primary materials via networked computers. No tool had been developed in the project at the time of the review. However, further development of Transcriber (for orthographic transcription) and CLAN (see above) was ongoing. See www.talkbank.org

Figure 6. Annotation tools surveyed.

4.3. Interaction with annotation tool creators

As for the ISLE NIMM data resources and annotation schemes surveys, close interaction has been sought with the creators of the reviewed annotation tools. All coding tool descriptions were reviewed by their creators.

4.4. Conclusions on annotation tools

First of all, Figure 6 confirms initial expectations as to the present scarcity of coding tools for natural and multimodal interactive behaviour. Current needs for more general-purpose NIMM annotation tools may be viewed as being reflected in the nature of the reviewed tools many of which are intended to be somewhat general-purpose rather than specifically supporting one particular project's needs. When one inspects the properties of the tools in more detail in [Dybkjær et al. 2001], it becomes quite clear that it is far from easy to build adequate and robust, general-purpose NIMM annotation tools. Moreover, tools originating from research projects are usually research demonstrators with what that entails in terms of fragile and buggy software.

The rapidly growing interest in natural interactive and multimodal behaviours and systems is likely to lead to a larger market for NIMM annotation and data analysis tools and thus also to increased commercial interest in such tools. Two of the reviewed tools (8 and 11 in Figure 6) are, in fact, commercial tools which in limited fashion enable coding for natural interactivity and multimodality. If industry becomes more heavily involved – also in research projects - in general tools development for NIMM data annotation and analysis, it is likely that the resulting software will be more stable than is the case with the majority of tools reviewed by the ISLE NIMM Working Group.

Among the reviewed tools or projects, three explicitly mention standardisation as a goal, i.e. 2, 5 and 12 in Figure 6. Standardisation of frameworks for annotation schemes and of data and coding representation formats could significantly facilitate the work of NIMM data annotators and system developers. We believe that standardisation requires a robust, flexible and powerful general-purpose NIMM coding tool (or toolset) which supports the standards aimed at. Given the current state of the art, the first site or project which succeeds in producing such a tool is likely to exert considerable influence on the general acceptance of the standards proposed. The need for NIMM annotation tools is clear and the users of annotation tools are out there in plenty. Today's users are aware that the tools which are currently available are far from optimal, and they are eager for improvements. They are willing to try new tools and are just waiting for something better than what is being offered today. It is up to the research and development community to try to meet their needs.

5. Users and resources

Through conducting the surveys presented above, the European ISLE NIMM group has gathered information on user profiles, user needs and best practices in the fields addressed. Results on those issues are presented and discussed in the surveys. What follows is a brief summary.

As expected, the potential users of NIMM resources work with data and annotation in many different areas, from widely different perspectives and professional backgrounds, and with an unlimited number of different annotation purposes. User profiles include, e.g., people working in spoken (and sometimes written) dialogue systems development and evaluation, many kinds of multimodal systems, embodied agents, speech, face and/or gesture annotation, research in prosody, linguistics, psychology, anthropology, human behaviour, and human factors, documentation of disappearing languages and cultures, etc.

There is a felt, and growing, need in academia and industry for all the kinds of NIMM resources surveyed. For instance, the computer games industry and the interface agents community need resources for developing natural and human-like characters, and educational institutions need NIMM resources for learning purposes. Best practice and standards are fairly developed in some areas but more or less absent in others. For instance, MPEG-4 and FACS set the current best practice standard for facial expression annotation, whereas no standards are in sight for gesture markup. A number of tools already exist, each of which support coding and analysis of particular aspects of natural interactivity data according to one or several coding schemes. However, none of them even come close to covering the full range of aspects to be addressed, and many are research tools which have severe deficiencies, hampering their practical use. With the development of increasingly sophisticated applications of natural interactive technologies, there is a strong need for tools that can effectively support annotation and analysis of the full variety of natural interactivity phenomena and their interrelationships.

Although probably far from being exhaustive, we believe that the surveys presented above significantly contribute to our common knowledge of the state-of-the-art in data, coding schemes, and tools for natural interactivity and multimodal interaction. However, it is evident that the information collected will quickly get outdated if not regularly updated. Since the ISLE project has a limited duration (until 2003) and budget, we would like to kindly invite the NIMM community to help us and each other maintain an up-to-date and added-value collection of NIMM resource information. A web-based facility is being set up at isle.nis.sdu.dk which will enable any interested colleague to upload information about a NIMM resource which has not been included on the website already.

6. Acknowledgements

We gratefully acknowledge the support of the ISLE project by the European Commission's Human Language Technologies (HLT) Programme. We would also like to thank all European ISLE NIMM participants for their contributions to the surveys described in this paper.

7. References

Bernsen, N. O.: Multimodality in language and speech systems - from theory to design support tool. In Granström, B. (Ed.): Multimodality in Language and Speech Systems. Dordrecht: Kluwer Academic Publishers 2002 (to appear).

Bernsen, N. O.: Natural human-human-system interaction. In Earnshaw, R., Guedj, R., van Dam, A. and Vince, J. (Eds.): Frontiers of Human-Centred Computing, On-Line Communities and Virtual Environments. Berlin: Springer Verlag 2001, Chapter 24, 347-363.

Dybkjær, L., Berman, S., Kipp, M., Olsen, M. W., Pirrelli, V., Reithinger, N. and Soria, C.: Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data. ISLE Deliverable D11.1, 2001.

Dybkjær, L. and Bernsen, N. O.: Data Resources and Annotation Schemes for Natural Interactivity: Purposes and Needs. Proceedings of the LREC'2002 Workshop on Multimodal Resources and Multimodal Systems Evaluation, 2002.

Klein, M., Bernsen, N. O., Davies, S., Dybkjær, L., Garrido, J., Kasch, H., Mengel, A., Pirrelli, V., Poesio, M., Quazza, S. and Soria, S.: Supported Coding Schemes. MATE Deliverable D1.1, 1998.

Knudsen, M. W., Martin, J.-C., Dybkjær, L., Ayuso, M. J. M, N., Bernsen, N. O., Carletta, J., Kita, S., Heid, U., Llisterri, J., Pelachaud, C., Poggi, I., Reithinger, N., van ElsWijk, G. and Wittenburg, P.: Survey of Multimodal Annotation Schemes and Best Practice. ISLE Deliverable D9.1, 2002b.

Knudsen, M. W., Martin, J.-C., Dybkjær, L., Berman, S., Bernsen, N. O., Choukri, K., Heid, U., Mapelli, V., Pelachaud, C., Poggi, I., van ElsWijk, G. and Wittenburg, P.: Survey of NIMM Data Resources, Current and Future User Profiles, Markets and User Needs for NIMM Resources. ISLE Deliverable D8.1, 2002a.

The ISLE NIMM surveys presented in this paper are available at the website for the European ISLE NIMM Working Group at isle.nis.sdu.dk

MATE reports are available at mate.nis.sdu.dk