Project ref. no.	IST-1999-10647
Project title	ISLE Natural Interactivity and Multimodality Working Group

Deliverable status	Public
Contractual date of delivery	31 December 2002
Actual date of delivery	February 2003
Deliverable number	D7.1
Deliverable title	Introduction to and Summary of the Final NIMM WG Guidelines
Type	Report
Status & version	Final
Number of pages	17
WP contributing to the deliverable	WP7
WP / Task responsible	NISLab
Editors	
Authors	Laila Dybkjær, Niels Ole Bernsen, Daan Broeder and Peter Wittenburg
EC Project Officer	Philippe Gelin
Keywords	Natural interactivity, data resources, annotation schemes, annotation tool, meta-data, guidelines.
Abstract (for dissemination)	This Deliverable D7.1 from the ISLE Natural Interactivity and Multimodality (NIMM) Working Group provides an overview of the work done by the group.

ISLE Natural Interactivity and Multimodality Working Group Deliverable D7.1

Introduction to and Summary of the Final NIMM WG Guidelines

February 2003

Authors

Laila Dybkjær[1], Niels Ole Bernsen[1], Daan Broeder[2], Peter Wittenburg[2]

1: NISLab, University of Southern Denmark. 2: MPI, Nijmegen, The Netherlands.

Contents

1. Introduction

This report provides an introduction to the work done in the ISLE NIMM (Natural Interaction and Multi-Modality) WG (Working Group). The ISLE NIMM WG began its work in early 2000 and has involved participation from the following sites:

- NISLab (Odense, Denmark) (WG leader)
- CNRS-LIMSI (Paris, France)
- DFE (Barcelona, Spain)
- DFKI (Saarbrücken, Germany)
- ELRA (Paris, France)
- HCRC (Edinburgh, UK)
- ILC (Pisa, Italy)
- IMS (Stuttgart, Germany)
- MPI (Nijmegen, the Netherlands)
- UROME (Rome, Italy)

The working group has produced 9 reports all of which represent pioneering work in the emerging field of natural interactivity and multimodal resources. More specifically, the reports address:

- NIMM data resources [Knudsen et al. 2002b, Knudsen et al. 2003]
- NIMM annotation schemes [Knudsen et al. 2002b, Dybkjær et al. 2003]
- NIMM annotation tools [Dybkjær et al. 2001a, Dybkjær et al. 2001b]
- NIMM meta-data [Broeder et al. 2000, Broeder et al. 2002a, Broeder et al. 2002b]

In parallel with the work carried out in the European ISLE NIMM WG, an ISLE NIMM WG has also been active in the US. There has been rather limited cross-Atlantic collaboration between the two groups because they came to address the general area of natural interactivity and multimodal resources in somewhat different ways and contexts, and to focus on sub-areas which only overlap to a limited extent. Whereas the European ISLE NIMM WG has implemented the programme of work outlined in the WG's ISLE NIMM contract with the EC, the US NIMM WG has focused on the following NIMM resources sub-areas:

- spoken language
- gesture, and
- discourse.

The US spoken language group has focused on the OLAC (Open Language Archives Community) initiative and on annotation graphs for internal data representation. The group's topical overlap with the European ISLE NIMM WG has mainly been on meta-data (OLAC) on which there has been interaction with MPI. Otherwise, there has been no overlap between the work of the US spoken language group and the work in Europe, since the European ISLE NIMM WG has focused on multimodal resources rather than on spoken language per se.

The US gesture group has focused on the development of a new gesture annotation scheme called FORM and on the production of a data resource annotated using this coding scheme. The focus of

European work on gesture and annotation schemes has been different. This work has addressed a larger number of natural interactive modalities than just gesture, collecting information world-wide on existing coding schemes as a basis for drafting guidelines for natural interactivity coding schemes.

Like the US spoken language group, the US discourse group has focused primarily on spoken dialogue rather than on natural interactive and multimodal dialogue more generally. Many of the European NIMM WG participants had already collected, in the MATE project (mate.nis.sdu.dk), information on coding schemes for (unimodal) spoken dialogue at several different levels, cf. [Klein et al 1998], and it was therefore decided from the very beginning that unimodal spoken dialogue would not be focal to the work ahead. This having been said, it is also important to point out that, in a more general sense, the US and European ISLE NIMM WG participants have many research interests in common. This is reflected in, e.g., the ISLE workshop on Dialogue Tagging for Multi-Modal Human Computer Interaction held in December 2002 in Edinburgh at the initiative of the US discourse group, which had participation from the European NIMM WG as well.

Due to the limited topical overlap among the two groups it was decided early on that the US ISLE NIMM group would not contribute to the European deliverables and vice versa. Instead, collaboration has been at the level of providing progress reports and other material of interest to the other part, and joint participation in meetings and workshops.

The results achieved by the European ISLE NIMM WG are summarised in the following sections. Section 2 provides a brief overview of the work on NIMM data resources. Section 3 summarises results on NIMM annotation schemes. Section 4 surveys the work on NIMM annotation tools. Section 5 summarises results of the meta-data activities.

2. NIMM data resources

The work on data resources has proceeded in two steps and results are documented in two corresponding reports.

The first step was to collect information on a large amount of existing NIMM data resources, far larger, in fact, than has been done before. The report on NIMM data resources [Knudsen et al. 2002b] 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 approach adopted for producing the NIMM data resources survey 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.

The reviewed data resources reflect a multitude of coding needs and purposes, notably: 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.

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, vendors of data resources exist, such as ELRA and LDC. Another important finding is that data resources are quite often poorly documented.

Based on the data resources survey, the second step was to propose guidelines for the creation of NIMM data resources. This work is documented in [Knudsen et al. 2003]. The report addresses guidelines for two main issues:

- general resource specifications, and
- data specifications.

The part on general resource specifications discusses and describes

- legal aspects,
- modalities used,
- organisation of the data, and
- data recordings in terms of procedures, equipment, recruitment, and assessment.

Following the structure of report D8.1 [Knudsen et al. 2002b], the part on data specifications addresses data presented in three different generic modalities, i.e.

- audio data,
- static image data, and

- video data.

In addition, the report provides an

- introduction to validation and validation criteria.

The report also describes a series of national and international bodies and organisations involved in data resource creation, dissemination, and documentation.

3. NIMM annotation schemes

The work on NIMM annotation schemes followed a process which is similar to the one described for data resources, i.e. it has been a two step approach leading to results documented in two corresponding reports.

The first step was to collect information on existing annotation schemes. The survey of NIMM corpus annotation schemes [Knudsen et al. 2002a] reviews 7 descriptions of coding schemes for facial expression and speech, and 11 descriptions of annotation schemes for gesture and speech.

The approach adopted for producing the ISLE NIMM annotation schemes survey 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.

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.

Based on the material collected, it may be concluded that there is still a 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 [Klein et al. 1998]. When we move beyond spoken dialogue annotation to considering facial coding, we do find a couple of frequently used and substantially evaluated coding schemes for different aspects of the facial expression of information (eyes, facial muscles), i.e. MPEG-4 and FACS. 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.

Based on the coding schemes survey, the second step was to develop and propose guidelines for the development of NIMM annotation schemes. This work is documented in [Dybkjær et al. 2003]. The report discusses work on standardisation by describing a series of activities world-wide which share the aim of influencing NIMM coding scheme-related standardisation, and presents recommendations and guidelines for NIMM coding scheme development.

Existing (de facto) standards in the NIMM area, or activities seeking to create standards in the area, include, e.g., TEI, ToBI, SAMPA, ISO TC37/SC4, MPEG-4, MPEG-7, ATLAS, and NITE, all of which are explained and described in [Dybkjær et al. 2003].

The recommendations and guidelines on how to create NIMM coding schemes cover the following aspects:

- how to create NIMM coding schemes;
- how to document NIMM coding schemes;
- how to represent NIMM coding schemes and annotations in a computer readable format;

- how to locate and select an appropriate existing coding scheme;
- how to adapt an appropriate existing coding scheme.

The report discusses each of these issues in detail.

4. NIMM annotation tools

The work on NIMM annotation tools has followed an approach which, in a general sense, is similar to the one followed for data resources and annotation schemes, i.e. it has been a two step approach leading to results documented in two corresponding reports.

The first step was to collect information on existing NIMM annotation tools. The survey of NIMM corpus coding tools [Dybkjær et al. 2001b] describes 12 annotation tools and tool projects which support natural interactivity and multimodal data annotation, i.e. tools which support annotation of spoken dialogue, facial expression, gesture, bodily posture, or cross-modality issues.

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 have 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.

Our initial expectations as to the scarcity of coding tools for natural and multimodal interactive behaviour were confirmed. 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 supporting one particular project's needs. When inspecting the properties of the tools in more detail, it becomes quite clear that it is far from easy to build adequate and robust, general-purpose NIMM coding tools. Moreover, tools originating from research projects are usually research demonstrators with what that entails in terms of fragile and buggy software. Today's users are aware that the tools which are currently available are far from optimal. It is up to the research and development community to try to meet their needs.

Based on the information collected in [Dybkjær et al. 2001b] as well as on the broad experience in coding tools among the NIMM WG participants, the second step was to create and discuss a set of requirements for a general NIMM annotation tool. This work is documented in [Dybkjær et al. 2001a]. The report discusses overall functionality, interface, architecture and platform requirements to a toolset in support of transcription, annotation, information extraction and analysis of NIMM data. The report starts by outlining a minimum set of requirements to a general NIMM coding tool. These requirements are briefly listed below.

- A flexible and open architecture which allows for easy addition of new tool components (a modular workbench).

- Separation of user interface from application logic and internal data representation. The internal data representation should be separated from the user interface via an intermediate logical layer so that the former two layers can be modified separately.

- Transcription support and annotation support at different levels of abstraction, for different modalities, and for annotating relationships across levels and modalities.

- To the extent that it is possible to (semi-) automate NIMM annotation processes, this should be supported by the toolset. Similarly, to the extent that it is possible to (semi-) automate NIMM data analysis, this should be supported by the toolset. Automation should be supported in two ways: (i) via the possibility to add (through an API) additional components for automatic annotation and data analysis, and (ii) via the use, as far as possible, of standard(ised) data formats, allowing easy importing and exporting of (automatically created) annotations.

- Powerful functionality for query, retrieval and extraction of data from annotated corpora; tools for data analysis, possibly including statistical tools.

- Adequate support for viewing and listening to raw data.

- Adequate visual presentation of annotated data.

- Easy-to-use interface. In general, the tool interface should support the user as much as possible, be intuitive, and as far as possible be based on interface standards which the user can be expected to be familiar with.

- Support for easy addition and use of new coding schemes and for defining new visualisations of annotated data (e.g., presenting annotations based on new coding schemes).

- Possibility of importing and thus reusing existing data resources via conversion tools.

- Possibility of exporting by means of conversion tools, coded data resources for further processing by external tools.

- Most importantly, perhaps, the tool must be robust, stable and work in real time even with relatively large data resources and complex coding tasks.

On this background, the report discusses in more detail the issues involved in creating a powerful and useful general NIMM annotation tool. The report has formed the starting point for work in the NITE project (nite.nis.sdu.dk) which is currently developing several different general-purpose NIMM coding tools.

5. NIMM meta-data

The research field of linguistics is very much a data oriented one and access to the original recording data is an important function that should be made as easy as possible. Therefore there has been early recognition of the need to create completely computer-based frameworks for the storage and processing of all linguistic data including primary data such as audio and video recordings. The combination of web-accessible computer-stored linguistic data and the explosive increase in the quantity of especially multi-media resources introduces management problems similar to those connected to the data explosion on the Web in general. The "accepted" solution for these discovery and management problems is to create web-accessible metadata descriptions. The recognition of this similarity led to the inclusion of a linguistic metadata framework development initiative as part of the ISLE NIMM WG.

The ISLE metadata initiative, named IMDI, first created a white paper containing an analysis of the above-mentioned problems and stated a number of goals that IMDI was to fulfil. The white paper was presented at a special international workshop at LREC 2000 devoted to metadata aspects of language resources. During the LREC 2000 conference also the IMDI organisational structure was set up. Meanwhile, all goals defined in the White Paper have been realised and the created metadata framework is now being further developed on the basis of new ideas such as "the semantic web". Central to IMDI's efforts was the design of a metadata vocabulary suited to the needs of general linguistics and language engineering. This vocabulary was developed after having made an overview of earlier initiatives and existing practices regarding this subject, and discussions with many specialists. Much interaction took place between the IMDI work in WP10 and the work in other ISLE work packages devoted to discussing NIMM aspects.

The resulting metadata set for the description of "sessions", the basic unit of linguistic analysis, contains many metadata elements important for resource management and, in particular, for resource discovery and exploitation. As parts of the metadata set definition, there are also several controlled vocabularies that constrain the metadata element values. Later, a separate proposal for a metadata set for the description of lexica was added that shares the general elements of the metadata set for sessions. This proposal emerged from strong interactions between the IMDI specialists and the ISLE work on lexicon standardisations. Important for the linguists is flexibility in the form of user definable elements and value sets. This is especially so since the way of describing language resources by metadata suitable for web-discovery and management is comparatively new and the existing set cannot be considered as final.

The format of the metadata descriptions is defined by an XML-Schema as has become standard practice for metadata descriptions. A special editor tool was developed that guides the user in creating metadata descriptions that conform to the Schema and the controlled vocabularies. The editor supports transparent downloading of controlled vocabulary definitions that can be stored on (remote) web-servers. This separate storage architecture of the definitions for the controlled vocabularies provides an extra degree of flexibility for linguistic sub-domains. For off-line work there is a controlled vocabulary caching mechanism so that linguists may use this tool under field conditions.

The metadata descriptions that are created in this fashion can be bound into a distributed tree of resources by simply linking to the metadata description’s file URLs from other metadata descriptions. These “parent” metadata descriptions represent corpus and sub-corpus abstraction concepts. This way a complex (physically distributed) universe of linked IMDI metadata descriptions is created that is used for resource discovery and exploitation. A special browser tool was developed to navigate in this domain of XML-based IMDI metadata descriptions. It also offers a metadata search interface specifically tailored for the IMDI set and vocabularies. The different functions of browsing and searching can be combined and intermingled to find specific sessions of interest that are then displayed in the browser together with links to the original language resources.

The browser supports the starting and execution of resource type-specific applications by clicking on these resource links. Users can configure the resource type to application mapping in a separate configuration file. The editor tool is one of the tools that can be started by the browser to modify existing metadata descriptions or to create new corpus and sub-corpus nodes. In this way, a user can create a separate tree of IMDI descriptions that links to the user's private language resources and also add links to other interesting IMDI (sub-) corpora available on the Web.

For interoperability reasons, a mapping of the IMDI elements onto the standard DC metadata set was defined and implemented in a DC/IMDI bridge that allows metadata service providers that use the OAI protocol to harvest records from the IMDI universe, although, of course, this occurs with information loss.

A few major events can be mentioned that guided the IMDI work:

- May 2000: A dedicated workshop at LREC 2000 to discuss the principles and goals.
- September 2000: Launch of a broad overview about header and "metadata" initiatives.
- February 2001: Launch of the first distributed IMDI demonstration including 6 European institutions.
- March 2001: Presentation and discussion of a first proposal at an international workshop about the IMDI set.
- Summer 2001: Availability of the first editor and browser versions for users and availability of a registry service.
- May 2002: Presentation of the complete metadata infrastructure at the LREC 2002 conference.
- May 2002: Founding of ISO TC37/SC4 with strong participation of IMDI.
- June 2002: WP10 internal workshop about the use of databases for large domains.
- October 2002: Interoperability with OLAC presented.
- November 2002: Final IMDI workshop to discuss about experiences and new requirements.
- December 2002: Start of the ECHO project that will make use of the IMDI infrastructure.
- January 2003: Start of the INTERA project that will establish a European language resource area based on the IMDI set.

In summary, within two years the IMDI team developed a complete metadata infrastructure for the discovery and management of (multimedia) language resources based on state-of-the-art XML technology including a professional editor, a browser, a search tool and efficiency tools. Due to new projects and the installed base of about 15.000 openly available metadata descriptions, the comparatively small investment turns out to be justified. In the long term we see a merging of the current metadata concept towards the ideas of the Semantic Web. The participation of the IMDI team in the ISO TC37/SC4 guarantees that there will be a mutual fertilization.

6. References

Broeder, D., Offenga, F., Willems, D., Wittenburg, P.: Metadata Set for Multimedia/Multimodal Language Resources. ISLE Deliverable D10.2, August 2002a.

Broeder, D., Offenga, F., Willems, D., Wittenburg, P., Heid, U., Vögele, A. and Popescu-Belis, A.: IMDI Showcase. ISLE Deliverable D10.3, September 2002b.

Broeder, D., Offenga, F. and Wittenburg, P.: Overview of Metadata Initiatives and Corpus Metadata in Language Engineering and Linguistics. ISLE Deliverable D10.1, October 2000.

Dybkjær, L., Berman, S., Bernsen, N.O., Carletta, J., Heid, U. and Llisterri, J.: Requirements Specification for a Tool in Support of Annotation of Natural Interaction and Multimodal Data. ISLE Deliverable D11.2, July 2001a.

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, January 2001b.

Dybkjær, L., Bernsen, N.O., Knudsen, M.W., Llisterri, J., Machuca, M., Martin, J.-C., Pelachaud, C., Riera, M. and Wittenburg, P.: Guidelines for the Creation of NIMM Annotation Schemes. ISLE Deliverable D9.2, February 2003.

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

Knudsen, M. W., Bernsen, N.O., Dybkjær, L., Hansen, T., Mapelli, V., Martin, J.-C., Paulsson, N., Pelachaud, C., and Wittenburg, P.: Guidelines for the Creation of NIMM Data Resources. ISLE Deliverable D8.2, February 2003.

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, February 2002a.

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, February 2002b.