A Multilanguage Database for supporting Sign Language Translation and Synthesis

Roberto Borgotallo[(2)], Carmen Marino[(2)], Elio Piccolo[(1)], Paolo Prinetto[(1)], Gabriele Tiotto[(1)], Mauro Rossini[(2)]

(1) Politecnico di Torino, Department of Control and Computer Engineering
Corso Duca degli Abruzzi 24, Torino, Italy
(2) RAI Radiotelevisione Italiana, Centre for Research and Technological Innovation
Corso Giambone 68, Torino, Italy
r.borgotallo@rai.it, carmen.marino@rai.it, elio.piccolo@polito.it,
paolo.prinetto@polito.it, m.rossini@rai.it, gabriele.tiotto@polito.it

Abstract

The design of a language database is an important task within projects targeting sign language research. In this paper is presented a database structure that supports both linguistic information and visualisation oriented data to assist a final publication of services for deaf people. The database has been designed within the Automatic Translation into sign LAnguageS (ATLAS) project that takes aim at getting the automatic translation from written Italian to Italian Sign Language (LIS). The final step of the overall process is the enrichment of the original video with a superimposed virtual character realised by 3D animated computer graphics. The top element within the database is the A_Product defined as the main primitive element managed by the ATLAS platform under which all the other data, from input sources to the final publication modalities and attributes lay. The A_Product includes the reference to the original content and all the intermediate elaborations results towards the final publication comprehensive of the virtual character animations. Among the others, the most important transformation is the automatic translation from a written Italian text to the intermediate language AEWLIS (ATLAS Extended Written LIS), formalized within the ATLAS project.

1. Introduction

The automatic translation among national languages represents one of the greatest challenges undertaken by computer science. The automatic translation from Italian language into Italian Sign language, the mother tongue for signing deaf people, may be regarded as a venture even more difficult because the syntax, the grammatical structure and the lexical heritage of the two languages are very different. However, the request for Italian Sign Language (LIS) interpretation is increasing in different contexts nowadays, such as educational, legal, healthcare, entertainment and cultural environments.

In this paper is presented the structure of a Multilanguage database that supports the translation from Italian Language into Italian Sign Language. The adopted methodology and its structure can be extended to support the translation of each national language into the corresponding sign language. The database has been designed within the Automatic TransLAtion into italian Sign language (ATLAS) project that aims at provide the LIS translation of different typologies of contents, such as audio/video, subtitles, teletext pages, web pages, texts, and displays it through a virtual interpreter realized by 3D animated computer graphics. The ATLAS architecture allows the retention of linguistic information, including lexical and animation data into a unique database named Atlas MultiMedia Archive (AMMA).

The first section of the document is dedicated to the description of the basic process that transforms, through several intermediate steps, a generic source content into a final virtual actor animation.

The second part concerns the structure of AMMA and contains the description of all included databases.

2. Language Translation Process

The translation process from Italian language into Italian Sign language is articulated into three main steps:

- Generic content ingestion
- Text Translation into AEWLIS
- Virtual Actor commands generation and rendering

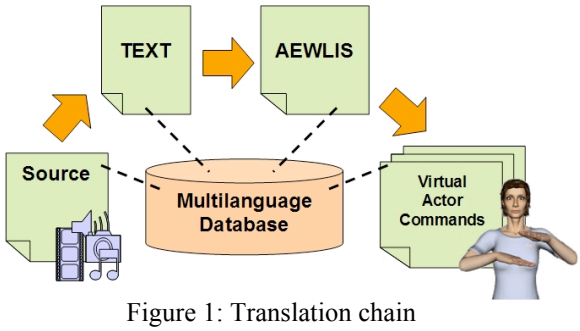

Figure 1: Translation chain

2.1 Generic Content Ingestion

The source content ingestion is the first phase of the translation chain that consists in the submission of different typologies of contents: audio, video, teletext pages, web pages, texts and subtitles. Whatever the format, the text is the fundamental part because it is

starting from it in order that the translation takes place. The multimedia components are as well important for the generation of an effective service over communication channels such as the digital television or mobile streams where the multimedia experience is mandatory.
In fact the ingested multimedia components just pass through the system when they are already suitable for the final service, otherwise they are adapted by mean of transcoding to satisfy the specific publication channel requirements (e.g. suppose that a video mpeg2 with standard definition is supplied, it is adequate for the DVBT channel but has to be transcoded to lower resolution and bit rates for mobile streaming).
As the text is a mandatory component of the ingestion phase, it is interesting to manage the creation of a text starting from the audiovisual content. Such an operation is possible with the adoption of an automatic speech recognition subsystem that derives the spoken words automatically from the audio signal. Usually after this stage, a manual revision is required because of the non-negligible error rate, both the automatic speech recognition and the manual revision are considered as *transformations*.

2.2 Translation into AEWLIS

Within the ATLAS project it has been formalized an intermediate language called AEWLIS that contains all necessary information to derive a good animation of the virtual interpreter.
The AEWLIS inherits the specific morphologic and syntactic structure of LIS and includes the so called Communication Channels that specify the position/direction of the hands, body, shoulders, head, gaze, labial and facial expressions that are very important in sign communications.
Translation from written Italian to AEWLIS text is the most tricky part of the overall process and it is based on very complex statistical and mathematical algorithms. There are basically two distinct approaches to fulfill this task: ruled based and statistical translation.
The first kind of translation is based on rules for mapping the grammar and syntactic structure of the input language to the output language (AEWLIS). Statistical translation is based instead on classical machine learning algorithms. In this case, after a preliminary learning phase, where a large number of manually translated phrases (corpora) are submitted, the machine translation system is expected to automatically generate the translation for new input sentences with a sufficient precision.

2.3 Virtual Interpreter Animation

Nowadays, through computer graphics, it is possible to model and animate a virtual character that reproduces the LIS movements. In addition to movements of fingers and hands, also arms and facial expressions can be reproduced in a detailed way. This is a crucial aspect because in LIS the mimic and gestural expressiveness of the body and face assumes an relevant importance. The animations are automatically derived from the AEWLIS by mean of an articulated engine for the planning of the movements along the timeline, followed by the creation of virtual actor commands that are finally used for rendering.

3. Database structure and utilization

The database is an essential component to support the above mentioned translation process. It constitutes the persistence layer for all the sub processes implied, for example the manual annotation aimed at the creation and assessment of the corpora for the automatic statistic translation. Moreover, on the database are stored all the elaboration results like the Italian and AEWLIS texts, the commands for the virtual actor, the pointers to external references like Wordnet and Radutzky dictionary.
Figure 2 shows an overview of the database with a leading concept of A_Product under which all the other databases is are linked: Multimedia, Corpora, Radutzky, New Signs, Wordnet, AEWLIS, Animation.

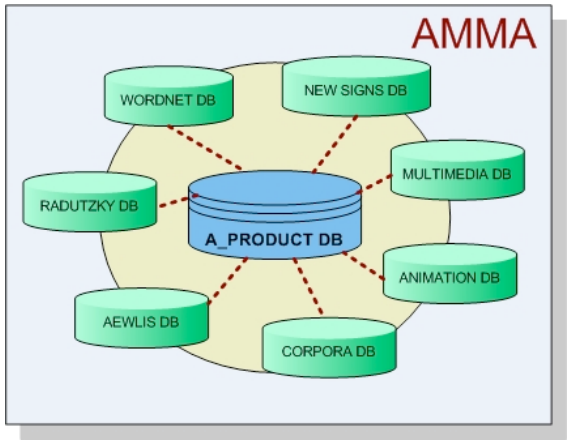

Figure 2:AMMA structure

Several other projects targeting sign language studies brought to the definition of a data storage and retrieval system, such as the Finnish Lexical Database (Savolainen & Leena, 2001), the Purdue American Sign Language Database for Sign Language recognition (Kak, 2002) and database for motion capture data retrieval (Award et al., 2009). These projects and know how constituted valid guidelines to organise the presented work.

3.1 A_Product Database

The A_Product is a data structure that contains all the information associated to the source content and its transformations, aimed at the generation of the virtual character animation. The A_Product includes the identification and process metadata that allow to keep trace of it during the different elaboration steps..
In figure 3 is shown in detail the structure of the A_Product.
It contains a reference to the multimedia archive and to the text resources, the provenence of these components is

tracked as well by mean of the source block in the figure 3. For example, in this way is possible to know that a text is coming from the Teletext of a certain channel and date or a multimedia is coming from a specific archive.
An A_Product instance is usually composed by a multimedia (e.g. a weather forecast edition) and a text representing what is said in the programme. In the above mentioned example the text could be derived automatically from the audio by an ASR (Automatic Speech Recognition) tool, in the picture this is represented by the "MM to TEXT TRANSF". In other cases the A_product is formed only by text for example coming from Teletext or Web.
Both multimedia and text can be transformed respectively in order to get a suitable version of the multimedia for publication and to improve the automatic translation into LIS. This transformations are represented in the figure with a looping relation.
The main multimedia transformations are: audio track extraction, text to speech conversion, speech to text conversion, video transcoding and audio transcoding. As far as textual transformations are concerned, the most important are: text synchronization, subtitles adaptation , text manual revision and validation.
One of the most important step is the translation of the text into AEWLIS, better detailed in the paragraph 3.5. Starting from the AEWLIS text and the corresponding LIS Signs the animation process will produce the commands for the final virtual interpreter representation.
The publication is the final step that allows to deliver a profitable service to the final user over different communication channels (DTV, Mobile, Web). A typical multimedia service is composed by a video with a superimposed virtual character.
It has been considered important to store all the elements and the results of the different transformations in order to be able to improve the translation and use this data in other projects and research activities.

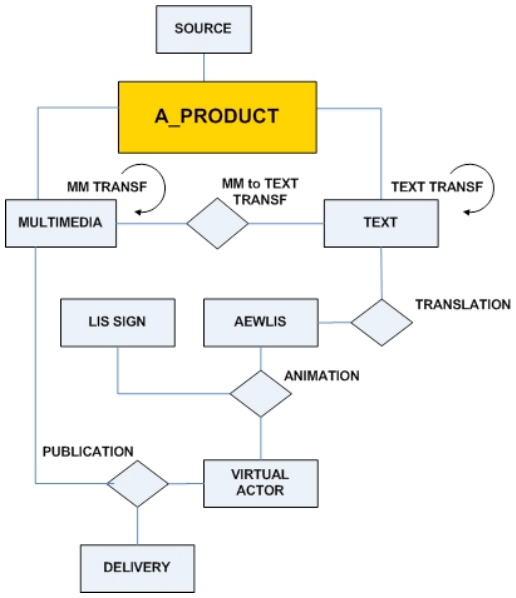

Figure 3:A_Product

3.2 Multimedia Database

This is the section where multimedia with different formats and resolutions are stored. Multimedia are saved on a common file system storage while they are referenced by pointers inside the database. The virtual actor animation usually does not belong to this section as it is rendered over the multimedia stream in real-time. Nevertheless in some cases the virtual actor could be superimposed when for example the device is not enough effective.

3.3 Corpora Databases

Here are preserved the annotations and manual translations that constitute the ground truth for the learning phase of the translation statistic engine. As the annotation is quite hard and long to carry out this is a very precious information.
Each enter of the Corpora Database contains:

- A pointer to a text file
- A pointer to the corresponding AEWLIS file
- A pointer to the associated movie with a LIS interpreter
- A pointer to the Virtual Character Commands file

A large set of metadata are collected resorting to the ECHO project regarding corpora definition metadata. For a deeper view on used metadata formats it has been considered the document "Metadata for sign language corpora" (Crasborn & Hanke, 2003).

3.4 New signs database

New signs are created if they are not present in the database. They are detected by means of a lexical frequency analysis. We linked the results of this analysis with the Radutzky Dictionary in order to find what signs are present within the domain in analysis, that are not present in the database yet. The new signs can be detected during the source texts analysis and during annotation. In this case a new sign has been agreed by the group of annotators including people of the deaf community. In the database is stored the new sign with its meaning and its video recording. The annotation tool used allows to easily detect the new signs by automatically searching into the database and retrieving the signs and comparing it with the signs being annotated. This tool called ALEA has been developed in the ATLAS project and represents a web based annotation editor.
To define a new sign is necessary to establish some parameters that indicate if it is:

- Size Modifiable
- Speed Modifiable
- Space Relocatable
- Compound Sign
- Symmetric Sign
- Static Sign
- Repeated Sign

Other additional information concerns the number of the involved hands, the movements of body, shoulder, head,

gaze and labial.

3.5 AEWLIS database

This is the section where the written LIS - the formal result of the language translation - is stored, subdivided into phrases. The visualization process starts from this formal representation with its conversion into virtual character commands. The stored data include the sentences written in AEWLIS and all the information that can be derived by the annotation in this formalism.

3.6 Radutzky and Wordnet databases

These are external databases, hooked to the system in order to increase the information. With Wordnet, each lemma is resolved with a better semantic representation while pointing to the Radutzky dictionary is possible to access to their basic sign representation. The Radutzky database contains the lemmas information resorting to the Radutzky Dictionary. Each sign is stored along with a coding that gives information about its parameters, such as:

- Hands Configuration
- Hands Orientation
- Place
- Movement

Each Radutzky sign is linked to the Wordnet Synset of each lemma to manage synonyms and perform disambiguation during manual translation. The association Radutzky sign-Wordnet synset allows to identify, for each new Italian word in the Italian source text, if it is present or not in the standard LIS dictionary. If it is absent, we can automatically find Italian synonyms that have a correspondence in the Radutzky dictionary. This process facilitates new sign creation and the use of other signs that are not standard but widely used in the deaf community.

3.7 Metadata

The database allows to store all the metadata related to the products created within the process. The metadata are always associated with a transformation (MM to text, textual transformations, etc…). They are used to store the information concerning the output of a specific transformation. A set of metadata is defined in order to store information on the product creation, date and time.

4. Conclusion

The presented database allows to share in a structured way information regarding the automatic translation process from Italian to LIS. The design of the database allowed to deeply investigate relations and dependencies between the major entities taking part into the translation process.

It supports all the operations from the content ingestion to the visualization by storing all the necessary data and supporting the transformations that are needed in the translation process. It provides the storage of metadata associated to each transformation that are useful to trace information about the creation of each product.

The modular structure of the proposed database and related processes allows to extend all the elaborations to other natural languages and other sign language dialects. We assume this is feasible with straightforward modifications and inclusion of additional databases connected to the A_Product.

The database data storage phase is still ongoing and future work will aim at the definition of a procedure in order to store critical data (i.e. new signs with associated movie and meaning) during manual processes, such as annotation and source text analysis.

5. Acknowledgements

The work presented in the present paper has been developed within the ATLAS (Automatic Translation into sign LAnguageS) Project , co-funded by Regione Piemonte within the "Converging Technologies - CIPE 2007" framework (Research Sector : Cognitive Science and ICT).

6. References

Savolainen, Leena, (2001). The database system used in the Finnish Sign Language Dictionary Project. *Sign Language and Linguistics*

Kak, (2002), 4th IEEE international Conference on Multimodal interfaces. *Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language.* http://dx.doi.org/10.1109/ICMI.2002.1166987

Awad, C., Courty, N., Duarte, K., Le Naour, T., Gibet, S.(2009). *A Combinated Semantic and Motion Capture database for Real-Time Sign Language Synthesis.* Universitè de Bretagne Sud, Vannes, France.

Crasborn, Hanke (2003). Background document for an ECHO workshop. *Metadata for sign language corpora*, Radboud University Nijmegen http://www.let.ru.nl/sign-lang/echo/docs/ECHO_Metadata_SL.pdf