POLITECNICO DI TORINO
Repository ISTITUZIONALE

An Editor for Assisted Translation of Italian Sign Language

Original
An Editor for Assisted Translation of Italian Sign Language / HATAMI MAZINANI, Nadereh; Prinetto, Paolo Ernesto; Tiotto, Gabriele. - STAMPA. - (2009), pp. 415-418. (Intervento presentato al convegno EWDTS 2009: 7th IEEE East-West Design & Test Symposium tenutosi a Moscow (Russian Federation) nel Sept 18-21. 2009).

Availability:
This version is available at: 11583/2462976 since:

Publisher:
SPD FL Stepanov V.V.Ukraine, 61168, Kharkov, Ak. Pavlova st., 311

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 April 2024

An Editor for Assisted Translation of Italian Sign Language

Nadereh Hatami
Politecnico di Torino
Dipartimento di Automatica e Informatica
Torino, Italy
Nadereh.hatami@polito.it

Paolo Prinetto
Politecnico di Torino
Dipartimento di Automatica e Informatica
Torino, Italy
Paolo.prinetto@polito.it

Gabriele Tiotto
Politecnico di Torino
Dipartimento di Automatica e Informatica
Torino, Italy
gabriele.tiotto@polito.it

ABSTRACT[1]

This paper discusses the design and implementation issues of ALEA: an editor for assisted translation. This editor constitutes a complete environment that allow the translation from Italian Language to Italian Sign Language of deaf people. It is interfaced with a database that stores information about source and target language and with the Italian Wordnet for disambiguation. The tool has been developed within the Automatic TransLation into Sign languages (ATLAS) project.

1. Introduction

Italian Sign Language is the official language of the deaf community in Italy. Though what is commonly considered, Sign Language is a true language with its grammar and morphology [1]. Despite of the great achievements in research, the task of automatically translate a source language into another is hard to be performed and sign language is not an exception. These advances in linguistic research and in new statistical methods for machine translation offer the possibility to develop a system that could help deaf people accessing services, improving their integration in the hearing community. In this scenario the assisted translation approach can provide for the problems that rise with automatic translation by manually solving them. This can ensure us a correct translation performed with the user intervention.

2. Related Works

In the recent years several research groups have worked on machine translation of sign languages. Basically two approaches are considered: automatic and computer-assisted translation. In the automatic translation field Bauer et al. [2] proposed a framework for statistical based sign language translation. Safar and Marshall proposed their work on rule based translation for English to British sign language [3]. Huenerfault worked on Rule based American to American Sign Language translation [4]. In his works he argues that statistical translation is not suited for sign language translation as it is difficult to create huge corpora. In recent evaluations like Chinese to English and Arabian to English translations, it was found that statistical approaches were comparable or superior to conventional systems corpora [5]. In general recent works demonstrated that statistical and rule based translation can't completely translate into sign languages as they are error prone. Computer-Assisted translation is a complex process involving specific tools that help the user to perform operations on the sentences. This process can be completely manual or can be partially automated suggesting the next word to be used in the sentence [6] [7]. The tools for computer assisted translation can be smart enough to suggest to the user the whole sequence of words in the sentence [8] [9]. To our knowledge no work is reported that consider user assisted approach for the translation into sign languages. In this paper we describe a tool called ALEA that allows a complete translation to be used on text content such as subtitles and books. In section…

3. ISL Formal Representation (AELIS)

The goal of the ALEA tool is to support the manual translation form Italian to Italian Sign Language (ISL).

[1] 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 : Cog-nitive Science and ICT).

The target language is thought to be subsequently visualized by means of a virtual character. A fundamental difference between sign languages and spoken languages is that sign languages are visual languages. A sentence is signed in the same way as describing a scene by putting the sentence elements in the space in front of the signer. This means that no written ISL is defined. As a consequence ALEA will not provide written sentences in the ISL grammar but we have to specify a formalism expressing the language content and the parameters needed by the avatar for the movements. For this purpose we defined a formal representation that contains both linguistic information and the relationship between the elements of the sentence. This formalism is called AELIS

For example if we have the input sentence: "The mum gave a big peach ice cream to the child". In ISL we have the signer that performs the sign of mum that is put in a certain position in the space, then the sign of child that is put in another position in the space, then the sign of ice cream performed in a "big" fashion by adding facial expression and keeping the hand in a way that refers to the dimension of the ice cream. This sign is then followed by the one of peach. Then the sign of ice cream that goes from the mum to the child. In the sign of ice cream we have an incorporation because the sign of "to give" is incorporated in the sign of ice cream. The movement from the location of mum to the location of child gives use the information that the mum gives the ice cream to the child and not vice-versa. In the end the sign of "done" is performed. Summarizing we have to translate the written sentence in input in:

Mother+Child+Ice-Cream(BIG)+Peach+ (movement from mother to child)+Done.

To express this sort of raw written ISL in a formalized way we need to add some attributes such as we have the following elements:

- Mother(1)
- Child(2)
- Ice-Cream(3, scaled_of=BIG)
- Peach$1(4, attr_of=3)
- Give(src=1, dest=2, obj=3, temp_attr=done)

As shown above we have the Lemma, such as Mother, Child, Ice-Cream and Peach. Each lemma is tagged with a unique progressive identifier (the number). Then we have some properties of the lemmas. In this case the scale (scaled_of) whose value is BIG and attr_of, which means that the peach is to be referred to the Ice-Cream. Give is a verb which is incorporated in the Ice-Cream and is expressed with a movement that pass through some waypoints. They are expressed as source and destination attributes while the sign to be moved is the object (3). The last attribute is the temporal attribute which follow the verb (done). Note that the lemma "peach" is completed with the $1 label. This is because the Italian word for peach is ambiguous as it has different meanings:

pesca[1]noun, feminine= (fruit) peach
pesca[2]noun, feminine= fishing
andare a pesca verb = go fishing.
For this reason we have disambiguate it. The label $1 means that this is the lemma "pesca" referred to the disambiguated meaning of "peach". To turn the input sentence in a sentence of the form specified above, several modules are needed. In the following section we discuss the ALEA workflow and how it interoperates with the modules.

4. The ALEA Editor

ALEA is used by following some steps listed below:

1. Italian Text importing
2. Segmentation
   - Automatic
   - Manual corrections when needed
3. LIS sentence creation in the AELIS format
   - Word manual shifting inside the sentence (Drag & Drop)
   - Single lemma translation from Italian to LIS sign:
     - Automatic (when no ambiguity: e.g., MAMMA)
     - Manual disambiguation (e.g., PESCA$1)
   - Adding semantic attributes (e.g., MAMMA(1), DARE(source=1, destination=2, object=3))
   - Sentence cleaning (articles & proposition cancellation)
   - Information adding
4. Exporting

ALEA allows the user to import Italian text for the translation. The text is then automatically segmented into sentences. The user can also select the sentence to be

Figure 1. Segmentation window

segmented by simply selecting and acquiring it manually. (see figure 1).

For the translation process ALEA takes information from a set of modules integrated in the whole environment:

- Lexical, Syntactical and Semantic Analyzers
- Italian Dictionary

For each sentence ALEA retrieves information from these modules and gives the user the possibility to perform some important operations such as word shifting, word syntactical role specification and disambiguation of lemmas. Figure 2 shows a portion of the main window in which it is possible to perform these operations.

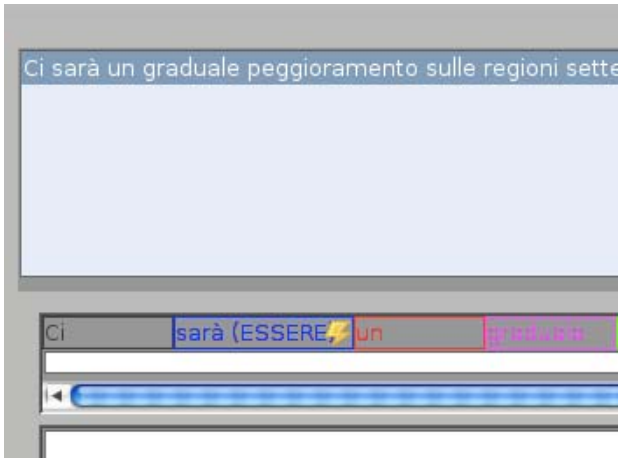

Figure 2. Main Window

On the top of the windows the segmented sentence is visualized. On the bottom of the picture the words are visualized on the basis of their semantic roles (between parentesis is the verb in infinite form). The red words are the ones identified for deletion. From this window is possible to shift the word by simply drag and dropping them.

The word shifting is necessary, as shown in the previous section, to put the words in the correct order. ALEA provides the sentence with each word colored on the base of its syntactic role in the sentence. This is made possible because the sentence is analyzed by the Lexical, Syntactical and Semantic Analyzers called Tule and TextPro.

TULE [10] is a:

–Morphological analyzer

–Tokenizer

–Chunk-rule based dependency Parser

developed at the University of Turin. It produces a dependency tree describing the Natural Language syntactic structure of the sentence and is used in ALEA to perform syntactical & semantic analysis

TextPro [11] is a suite of tools for Natural Language Processing developed by the Foundation Bruno Kessler of Trento and is used in ALEA to perform lexical analysis.

It performs morphological analysis for each input word & provides info such as:

–Lemma

–Lexical Category

–Other features (gender, number, …)

The disambiguation is performed by using an Italian language dictionary. ALEA uses Italian Wordnet in such a way.

WordNet [12] [13] is a lexical database targeting:

–Information Retrieval

–Semantic tagging

–Disambiguation

–Ontologies

–Terminologies

It was originally developed by the Princeton University and further extensions were developed by institutions all over the world to provide support for different languages.

In Wordnet groups of synonyms (synsets) are used to identify lexical concepts.

By retrieving information from these modules ALEA is able to provide the sentence with disambiguated lemmas and verbs in their infinite form. If some syntactic role is not correctly recognized by the analyzer it is possible to specify the roles manually. With ALEA is also possible to automatically delete the articles and prepositions since they are not used in ISL. Finally it is possible to add words such as “done” in the previous example. The editor gives in output an XML file that contains all the information about the sentence in the AELIS format. The use of an XML format provide us the possibility to directly send the information to the avatar and give us more flexibility.

5. Conclusion and Future development

In this paper we presented a tool for computer-assisted translation from Italian to Italian Sign Language of deaf people. It provides an output that is oriented to the sign visualization through a virtual avatar. Future development aims at adding necessary information on the output sentence. This is relationship between words in the form of tags and properties of the lemmas to support signs incorporation and modification.

6. References

[1] Stokoe, W. C.: Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. Studies in Linguistics: Occasional Papers 8, (1960).

[2] Bauer, S. Nießen, and H. Hienz. 1999. Towards an automatic Sign Language translation system. In *Proc. of the Int. Workshop on Physicality and Tangibility in Interaction*, Siena, Italy.

[3] F. Brown, S. A. Della Pietra, M. J. Della Pietra, and R. L. Mercer. 1993. Mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.http://www.interfacebus.com/MIL_STD_462_Description.html

[4] Huenerfauth. 2004. A multi-path architecture for ma- chine translation of English text into American Sign language animation. In *Proc. Student Workshop at Human Language Technolgies conference HLT-NAACL*, Boston, MA, USA.

[5] P. Langlais, G. Foster, and G. Lapalme, "TransType: A computer-aided translation typing system," in *Proc. Workshop on Embedded Machine Translation Systems (NAACL/ANLP2000)*, Seattle, WA, May 2000, pp.46–52.

[6] L. Nepveu, G. Lapalme, P. Langlais, and G. Foster, "Adaptive language and translation models for interactive machine translation," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP04)*, D. Lin and D. Wu, Eds., Barcelona, Spain, Jul. 2004, pp. 190–197.

[7] F. Och, R. Zens, and H. Ney, "Efficient search for interactive statistical machine translation," in *Proc. 10th Conf. Eur. Chapter of the Association for Computational Linguistics (EACL03)*, Budapest, Hungary, Apr. 2003, pp. 387–393.

[8] J. Civera, J. Vilar, E. Cubel, A. Lagarda, F. Casacuberta, E. Vidal, D. Picó, and J. González, "A syntactic pattern recognition approach to computer assisted translation," in *Proc. Advances in Statistical, Structural and Syntactical Pattern Recognition—Joint IAPR Imternational Workshops on Syntactical and Structural Pattern Recognition (SSPR 2004) and Statistical Pattern Recognition (SPR 2004)*, Lecture Notes in Computer Science,