



HAL
open science

Linguistic Resources for Automatic Natural Sign Language Generation

Francesco Terrone, Ritamaria Bucciarelli, Andrea Rodrigo, Javier Julian
Enriquez, Francesco Saverio Tortoriello, Ilaria Veronesi

► **To cite this version:**

Francesco Terrone, Ritamaria Bucciarelli, Andrea Rodrigo, Javier Julian Enriquez, Francesco Saverio Tortoriello, et al.. Linguistic Resources for Automatic Natural Sign Language Generation. 2024. hal-04478688

HAL Id: hal-04478688

<https://hal.science/hal-04478688v1>

Preprint submitted on 27 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linguistic Resources for Automatic Natural Sign Language Generation

F. Terrone¹; R. Bucciarelli²; A Rodrigo³; J. Julian Enriquez⁴; F.S. Tortoriello⁵; I. Veronesi⁵

¹ *Fondazione F. Terrone. Università degli studi di Napoli Federico II*

² *Università di Venezia*

³ *Universidad Nacional de Rosario [Santa Fe]*

⁴ *Universitat Politècnica de València*

⁵ *Università degli Studi di Salerno*

Abstract

Work-Tools (W. T) is a Linguistic Resources for Automatic Natural Sign Language Generation software for automatic textual analysis that describes and transforms from language to morphs, lexemes, and fixed phrases, to dialogue and build a communicative model of switching non-verbal nature languages L1 to verb L2. The W.T. The software is designed to handle complex tasks in natural language, such as parsing, to recognize and generate texts. The production, questioning, and evaluation of text construction involve interaction between humans and machines. It assists in the transformation of language and is utilized for didactic purposes. Work-Tools consists of: 1) a search engine, a database for data entry, in which the data described for cognitive areas and transformed into acronyms are indexed; 2) a writing corpus in which man/woman organizes the text in free sentences and fixed phrases, in accordance with the grammatical rules and the syntactic relations of the natural language that we propose with the data transfer from L₁ to L₂; 3) a transcoding corpus and faithful translation of text or textual parts from L₁ to L₂; A scroll bar where new text is transmitted in real time.

Keywords: Corporal phraseology; Dictionary for consultation; Text former.

1 Introduction

During an ultra-decennial experimentation conducted at the Department of Communication Sciences of the University of Salerno, in collaboration with other Research Centers and with the *Laboratoire d'Automatique et Linguistique* (CNRS - Paris 7), it was developed new methods for linguistic research.

Syntactic vocabularies were constructed with the goal of providing the most complete and formalized description of a language through data processing. The research is part of the Grammar Program of the Italian Language and LS (L.G.L.I.). The theoretical reference model is represented by the grammar "to operators and arguments" (Harris, 1970). The aim of this activity is to improve linguistic research methods and discover ways to use modern glottodidactic techniques in curricular settings.

The inclusion of didactics is proposed in Work-Tools, which is Linguistic Resources for Automatic Natural Language. The user uses linear-dominant interactive models, ITS (Intelligent Tutoring System) models, and becomes a builder of their own knowledge. It acquires the potential of semantic and syntactic dimensions and the possibility of subjective expansion of the linguistic heritage, which will allow it to be unusual mobility in the linguistic

space. The aim is to carry out a scientific description and production of a natural language to the analysis of the lemma and syntax speech, for a correct translation and trans codification production, which has a linguistic coherence found in applications in the sciences of information and communication.

2 Operating hypothesis

It needed to build a software that could solve the production of written texts; with the correct transcode from one natural code to the other, that is, the L1 native language to L2 native language of reception. It involved breaking the link between these two moments in the reformulation of the syntagmatic axes. The process involved manipulating international changes during the transition phase from L1 (which had socio-linguistic consequences for the speaker) to L2 with a faithful transcoding in the switching in how and with which reference models.

The reference works of the study were:

- Max Silberztein for a detailed description of the sentence and for electronic versions, to whom I worked in the POLO DSN team at the University of Salerno.
- G. Aston for the usefulness of the fixed phrase in the production of dialogues.
- Rosalee Wolfe project with scientifically comprehensible corpus for ASL description and switching in American Sign Language.

2.1 Reference models: M. Silberztein

Silberztein's focus (1993) lies on the intersection of descriptive linguistics, formal linguistics, automatic language processing, formal informatics, and algorithmic texts. The NOOJ approach is used in this book to present a complete project on formal description of natural languages. It analyzes the representation of the basic language.

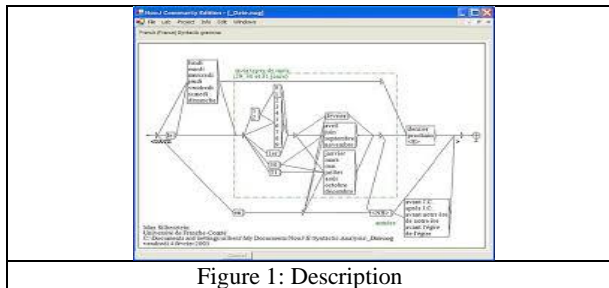


Figure 1: Description

2.2 Reference models: G. Aston

According to Aston (2001, 2015), the phrase phraseology of fixed phrases in the automatic construction of speech is necessary. Above all, it can be a valuable aid for individuals, who strive to integrate into social life when we use a natural language (L2) that is different from their own.

Phraseology is essential for fluency in speech production and reception, especially in cognitively/effectively demanding contexts like interpreting, there is substantial evidence. Phraseological items are limited in repertoire for most second language speakers because they lack knowledge of their lexical, functional, and prosodic aspects. Constructing speech corpora that align transcripts with audio from subtitled video materials is a simple task, and learners can use them to view and hear concordance data. Examples are provided for phraseological item documented in a one-million-word corpus of talks from the TED – Ideas worth WordSmith Tools (Scott, 2012). Activities for performance with and by learners are also suggested, aimed at increasing their phraseological awareness and expanding their repertoires.

Sinclair (1987) proposed that the 'idiom principle' is crucial to language production and comprehension, based on retrieving and elaborating phraseological items as wholes.

2.3 Reference model American Sign Language (ASL)

We developed an avatar that reflects sign language using reusable hand animation following Novogrodsky *et al.* (2014). Sign languages lack a universally accepted written form, unlike many spoken languages, and depicting them convincingly remains an open question. We aim to create a representation that could synthesize novel

utterances from American Sign Language (ASL) as realistic and...

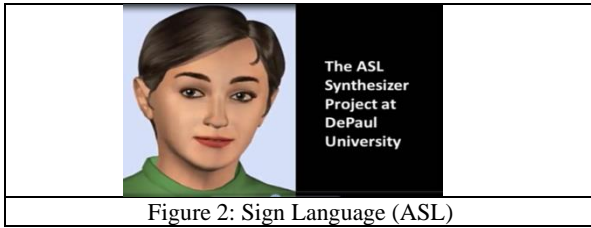


Figure 2: Sign Language (ASL)

3 The challenge

Structure an operating software that should have the following requirements:

- The validity conferred by experience and based on the reference models.
- Reproduce texts or fixed texts.
- Rephrase clear-cut and precise phrases, namely transferring and switching from one natural code to another.
- Translation in real time.

Thus, it was necessary to introduce DB-WT fixed phrases that are not compositional and have a production location where free phrases can also be included. The new digital-computational WT has several interesting moments from this experience. Enhancing the spheres of linguistic knowledge can improve IT output qualitatively. The software consists of the following steps.

3.1 Data storage

It allows for the development of a morphological parser called syntactic semantics that can make structural choices based on codes, such as the position of a particular component. Fixed structures and formulas are provided for analysis of different parts of the text, with the aim of starting transcription systems. So, you can get a taxonomy of phrases or textual parts that transforms and reduces the phrase. You can create a database for interfaces and

implementations of ADT stack and FIFO queues by obtaining fixed structures encoded in acronyms, which will treat very useful algorithms and serve as a basis for many of the analyzed implementations and taxonomic collections. For example, we can talk about merge (join) operations of queues to combine two queues into a single queue, a type of data that can exist multiple instances, and that we can assign to variables. In this phase of archiving, we aim to evaluate the various probabilities of converting morphemes and phrases into smaller elements. Phonological analysis of the significant elements of modern language reveals that the grammatical morpheme is the pivotal unit of morphology. Hence, constant reductions of the same can be obtained. Morphological is the term used to describe the form's encoding and gathering of information in taxonomies, such as prefixation, suffixing, and conjunctions. The more challenging it gets to synthesize the important elements of the word, in terms of the lexical morphemes, we are referring to words that convey the message we are trying to convey. The reduction in code extends to the fixed structures and textual parts. Finally, in the archiving phase you can insert objects, such as documents, image files.

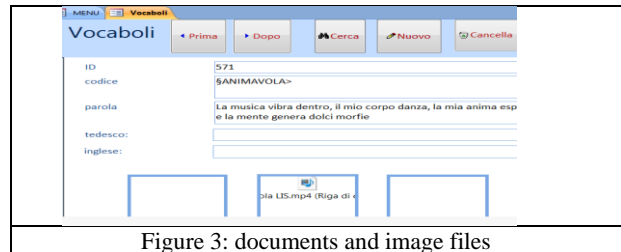


Figure 3: documents and image files

3.2 Production phase

At this point, the protagonist is responsible for guiding and building the text, which is composed using both fixed phrases (for recall) and free phrases. That is, he /she defines the content and manipulates the emitting language with that of receiving and opens the video object for comprehension in language, i.e. he /she writes the text composed of acronyms and phrases free from L_1 in L_2 .

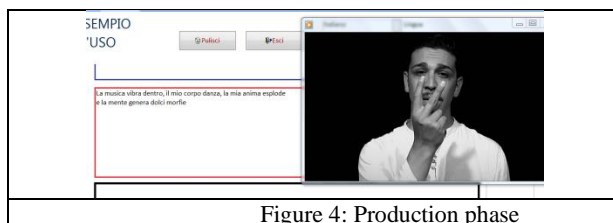


Figure 4: Production phase

3.3 Transmission phase

In real time, the text is made with a simple reminder: reformulated and reproduced in full on the scroll bar, with the advantage of reformulating the axes of the language in the transmission.

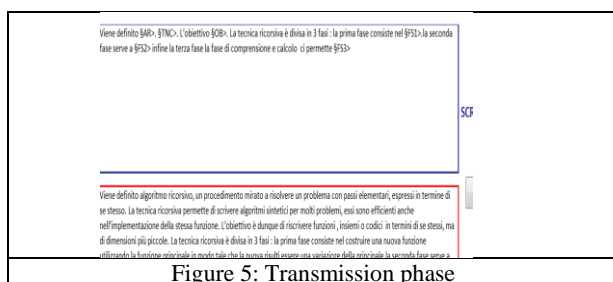


Figure 5: Transmission phase

4 Conclusions

It is a valuable tool for professionals, who want to work quickly on the electronic construction of a document, whether it is to translate an LIS language to written text or to reproduce spoken language in written language. As a support tool for students experiencing perceptual disorientation or absence, it can be useful in teaching. Not only does the software produce text but it also stores it in a short amount of time.

5. Author's Considerations

It is utopian to think that you can translate from one language to another by thinking of retaining the semantics in its entirety, because there is always a lack of information in the languages. Only the translation of fixed, i.e. professional languages can keep the content intact.

References

- Aston, G. (2015). Learning phraseology from speech corpora. *Multiple affordances of language corpora for data-driven learning*, 69, 65-84.
- Aston, G., & Gavioli, L. (2001). Enriching reality: Language corpora in language pedagogy. *ELT journal*, 55(3), 238-246.
- Harris, Z. S., & Harris, Z. S. (1970). The two systems of Grammar: Report and Paraphrase. *Papers in structural and transformational linguistics*, 612-692.
- Novogrodsky, R., Fish, S., & Hoffmeister, R. (2014). The acquisition of synonyms in American Sign Language (ASL) toward a further understanding of the components of ASL vocabulary knowledge. *Sign Language Studies*, 14(2), 225-249.
- Silberztein, M. (1993). *Dictionnaires électroniques et analyse automatique de textes le système INTEX*. Masson ed. Paris.
- Sinclair, J.M., 1987. Collocation: a progress report. In: Steele, R., Threadgold, T. (Eds.), *Language Topics: Essays in Honor of Michael Halliday*, II. John Benjamins, Amsterdam, pp.319-331.
- Scott, M. 2012. WordSmith Tools (Version 5.0) [Computer Software]. Available from <http://www.lexically.net/software/index.htm>.