From Computer Assisted Language Learning (CALL) to Sign Language Processing: the design of e-LIS, an Electronic Bilingual Dictionary of Italian Sign Language and Italian

Chiara Vettori, Oliver Streiter, Judith Knapp

Language and Law
EURAC; European Academy of Bolzano
Viale Druso/Drususallee 1, 39100 Bolzano/Bozen, Italy
{cvettori;ostreiter;jknapp}@eurac.edu

Abstract

This paper presents the design of e-LIS (Electronic Bilingual Dictionary of Italian Sign Language (LIS) and Italian), an ongoing research project at the European Academy of Bolzano. We will argue that an electronic sign language dictionary has to fulfil the function of a reference dictionary as well as the function of a learner's dictionary. We therefore provide an analysis of CALL approaches and technologies, taking as example the CALL systems ELDIT and GYMN@ZILLA developed at the European Academy of Bolzano too. We will show in how far these approaches or techniques can be ported to create an electronic dictionary of sign languages, for which system components new solutions have to be found and whether specific modules for the processing of sign languages have to be integrated.

1. Introduction: Dictionaries of LIS

Around 50.000 people in Italy are deaf. The first language of the majority of them is LIS, Lingua Italiana dei Segni (Italian Sign Language), but there is also an undetermined percentage of oralist deaf people. LIS is also acquired as a second or third language by hearing family members, teachers, interpreters and logopedics, amounting to about 170.000 people using LIS, in various degrees of language competence. Unfortunately, the quality and accessibility of LIS-courses and supporting material (dictionaries, text books, and videos) lack behind the actual need. Moreover, the official support does not meet the high standards of other countries and does not comply with international recommendations, e.g. Recommendation 1598 (Council of Europe 2003), which advice, among others, to broadcast television programs in sign language, to utilize new technologies for teaching sign languages and to include sign languages as a valid academic qualification. It is most likely that such status quo also depends on the position of the Italian government which has not yet officially recognized LIS.

As for LIS dictionaries, the vast majority of them are paper based ones, e.g. Radutzky 1992 (752 signs, 2500 sign meanings); Angelini et al. 1991 (400 signs). The paper format, however, cannot obviously account for the possibility of describing the three-dimensional complexity of each sign. A first, significant attempt in Italy to exploit new technologies to approach sign languages in an innovative and more proficient way, was made by the team of Cooperativa Alba. Its members have opened an Internet portal for LIS (DIZLIS) that now features more than 1000 video-filmed signs, which represent a respectable size for a sign language dictionary, cfr. Sternberg 1987 (3300 signs), Stewart et al. (2500 signs). Italian serves as vehicular language and dictionary index.[1] The advantage of this presentation of signs over the schematic and static drawing in paper dictionaries is evident and has motivated similar projects in other countries.

2. Towards e-LIS

Most sign language dictionaries form a hybrid between a reference dictionary and a learner's dictionary. This often occurs because sign language is implicitly considered as the second language of a "learner's dictionary" de-facto created for the needs of hearing people. At the same time these lexicographic works pretend to fulfil the function of a reference dictionary of the involved sign language, only in virtue of the presence of drawings and photos representing different signs. "*A major feature of such dictionaries is the absence of definitions, it being assumed that each sign would have exactly the same meaning(s) as the written word with which it is linked*" (Brien 1997). This sort of production treats signs as equivalents of the words of a spoken language and neglects the complexity, the dignity of sign language and its peculiarities in semantics and syntax. Lexical units in a sign language differ in a number of important features from the translational equivalents in the spoken language. These are:

- the referential extension, i.e. which objects, states and events are referred to by a word,
- the conceptualization, e.g. as *Abendstern*, *Morgenstern* or *Venus* (Frege 1892),
- the micro-syntax, e.g. the derivational history of a word from its bases via compounding or derivation to its final form,
- the stability with which they belong to a word class (nouns vs. verbs),
- the lexical relations they maintain, e.g. expressed as lexical functions (Melc'uk 1974) and
- the affiliation of a word to a word class which does not exist in the other language, e.g. classifiers in sign language or functional prepositions in the spoken language.

[1] http://www.dizlis.it

As LIS is an autonomous language and not a mere visual representation of Italian, we designed a dictionary which describes two systems at the same time, the Italian and the LIS one, and which can also build a bridge between them through a sort of "translating interface". In this perspective, accepting Stokoe's description of what he calls "serious dictionaries" (Brien 1998), we are greatly motivated to focus on the definition of sign meanings that could reveal much of the deaf culture.

This accommodates for two distinct user groups. (a) Hearing Italian people who study LIS and who will start with an Italian query term in an Italian environment (Italian definitions, explanations etc.); (b) LIS-signers looking for a sign and who should have the possibility to formulate query terms in LIS and have a LIS environment.

In order to assure the description of sign language in the sign language itself, therefore accounting for the specificity of this linguistic code[2], appropriate modes of rendering it into a Web-interface are required. One unexplored way of providing signs' definitions could be realized through the adoption of SignWriting (Rocha Costa & Pereira Dimuro 2003). In contrast to filmed definitions, in fact, SignWriting renders the definitions, explanations and menu buttons searchable (Aerts et al. 2004, Rocha Costa et al. 2004) and hyperlinkable. Words contained in a definition may thus be linked to lexical entries, which feature, as main component, the filmed sign.

3. ELDIT

One of the tools we already count on and from which we intend to develop the e-LIS dictionary is ELDIT, an electronic learners' dictionary for German and Italian. Inspired by the lexicographic research started in the '50s and according to recent psycholinguistic and didactic theories (Aitchison 94, Kielhöfer 96), it covers a limited vocabulary consisting of approximately 3.000 words for each language. It also stores a large set of information for each word entry and highly interlinked information pieces.

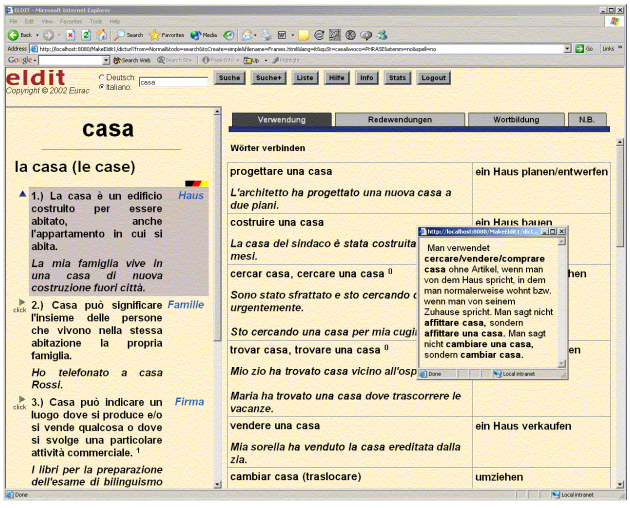

Figure 1: Dictionary entry for the Italian word "casa" (house) in ELDIT.

Figure 1 shows a screenshot of the dictionary entry for the Italian word "casa" (Engl. "house"). The left-hand frame shows the different word meanings. Each meaning is described by a definition, a translation, and an example. The right-hand frame shows additional information, which depends on the selected word meaning. The collocation tab lists the most frequent collocations along with their translation and an illustrative example. In the semantic field tab word relations (such as synonymy, antonymy, etc.) are illustrated in graphs for the learner. Verb valency is explained using colours and movable elements. Adopting a comparative approach, ELDIT also stresses specific differences between the Italian and the German language. Such differences are indicated by footnote numbers. Last but not least, each word used in the system (e.g. in the definitions or in the example sentences) is annotated with lemma and part-of-speech and is linked to the corresponding dictionary entry, which facilitates a quick dictionary access for unknown words.

4. GYMN@ZILLA

A further interesting way of facing LIS and Italian is represented by Gymn@zilla, a browser-like application which integrates existing educational and non-educational modules in a new didactic environment. Gymn@zilla allows to access documents from the Internet and to convert its text into an easy reader text, a glossary and a completion exercise.

Gymn@zilla is used like any browser. The program accesses a web-page, identifies its language and encoding and performs a simple word-form stemming of text. The stemmed words and expressions are then linked to their respective translations in external dictionaries. The linked lemma is marked up as html tool-tip to provide an immediate translation aid even without following the external link.

Clicking on a word triggers two actions. First, the complete explanations of the external lexicon are opened. Second, the word, its context and its translation are added to a personal glossary. The learner can edit the vocabulary in his personal dictionary and use it for intentional vocabulary acquisition, as opposed to incidental vocabulary acquisition by annotated reading of the web-page. Last, the learner can create interactive quizzes from the personal glossary, for which Gymn@zilla automatically offers inflected, uninflected and misspelled forms to fill the gaps. Gymn@zilla handles a number of language pairs, some going from a spoken language to a Sign Languages (e.g. English=>ASL, c.f. Figure 2). Through a triangulation of the translation dictionaries (e.g. Italan => English => ASL) we will give Gymn@zilla new dimensions of usage.

[2] Cfr: Les Signes de Mano
http://www.ivtcscs.org/media/mano.htm

Figure 2: Annotated reading with Gymn@zilla.

5. e-LIS Architecture

Hence it becomes obvious, even after this schematic analysis, that an electronic dictionary of sign language can be much more than a list of search indices, each hyperlinked to a video file. The search will start with an Italian key word or a LIS key word entered in SignWriting yielding a parallel list of all matching lemmas and collocations in Italian-LIS (SignWriting), similar to *LEO*[3], developed by the Munich University of technology, and *bistro*[4], developed at the European Academy Bolzano. Clicking on a word or an expression makes this a search term, possibly inverting the direction of search. As in *bistro*, additional links will lead to the monolingual lexical entries.

The Italian entry will be close to its current form in ELDIT, which might be profitably reused for developing e-LIS (c.f. Figure 1). Link texts to related entries in LIS will be rendered in SignWriting. The LIS entry will feature the filmed representation of the LIS sign. All definitions and explanation in the LIS entry will be in LIS, rendered in SignWriting. As in the Italian entry, each sign will be hyper-linked to the corresponding LIS entry. Lexical functions, e.g. classifiers, collective nouns (antelope => herd, ant => army) etc. will be realized as hyperlinks to entries as well, as well as the backward relation. Example sentences, collocations, idioms in LIS which do not have a proper lexical entry will be directly linked to the filmed sign presentation. As for the video approach, we will draw on the materials already developed for the site DIZLIS by the Cooperativa Alba.

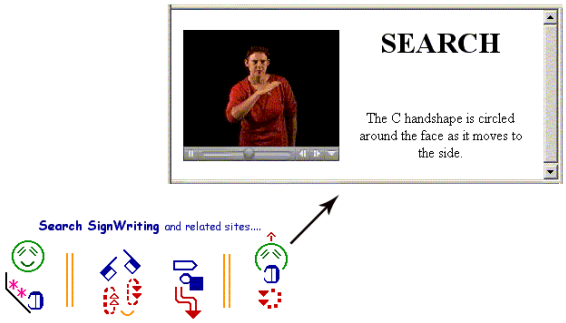

Figure 3: SignWriting in combination with Sign Language, a vision for the e-LIS system.

Beside this kind of inner metalinguistic description, we won't forget the peculiar needs of Italian speaking learners of LIS who will presumably not be able to read SignWriting and prefer videos of signs. For these users, as well as for signers studying Italian, Gymn@zilla can be easily invoked with its habitual functions:

- Italian words will be rendered as easy reader through video films or SignWriting
- SignWriting will be rendered as easy reader through video films or Italian
- personal word lists can be constructed
- completion test can be started at any time (in Italian and SignWriting)
- Texts Italian and SignWriting located in the WWW can be smoothly integrated into e-LIS, with proposed or freely selected texts, in order to allow the first steps outside the e-LIS environment. In case of any doubt, Gymn@zilla will take the user always back to e-LIS to provide the necessary explanations.

In addition, the analysis of possible difficulties LIS-signers encounter in studying Italian (Taeschner et al., 1988; Fabbretti 2000; etc.) suggests another usage a sign-language dictionary could be put to. We intend to supply the dictionary with an apparatus of contrastive grammatical macros in analogy to the footnote numbers in ELDIT. These macros are triggered whenever a lexical entry contains critical features, e.g. semantically weak prepositions such as "di" ("of") which cause translation difficulties for signers while writing in Italian, differences in word order etc. The lexical material of the entry and its parallel counterpart (in LIS or Italian) will be inserted into the macro and rendered from the point of view of the actual entry, yielding a comparative and synoptic descriptions of challenging grammatical aspects of the two languages compared with the lexemes of the current entry. Also in this perspective, the use of SignWriting could be particularly useful because it permits to parcel two equivalent strings in sign language and Italian and to interrelate the single syntagms/parts thus immediately showing the similarities and differences of the two systems with the aid of colours (for the corresponding elements) and explanations in sign language in case of differences.

[3] http://dict.leo.org/
[4] http://www.eurac.edu/bistro

6. Conclusions

We have presented so far the rationale of e-LIS, Electronic Bilingual Dictionary of Italian Sign Language (LIS) and Italian. A short analysis of existing Sign Language projects and of several CALL projects that have been carried out in the past years at the European Academy Bolzano has revealed that an electronic dictionary of sign language can be much more than a simple list of search indices, each hyperlinked to a video file.

While reusing some tools and options of the Italian-German dictionary ELDIT and enriching them through the many didactic functions provided by Gymn@zilla, a browser that converts Internet texts into easy reader ones, we will develop a new type of sign language dictionary.

We hope that our system might contribute to a research area that up to now has been quite neglected in Italy and that it could contribute to and accelerate the process which will lead Italian government to the official acknowledgement of LIS.

7. References

Aerts, S., Braem, B., van Mulders, K. De Weerdt, K., 2004. Searching SignWriting Signs*, this volume*.

Angelini, N., Borgioli, R., Folchi, A. and Mastromatteo, M., 1991. *I primi 400 segni. Piccolo dizionario della Lingua Italiana dei Segni per comunicare con i sordi.* Firenze: La Nuova Italia.

Brien, D. and Collins, J., 1997. The CD-ROM Dictionary of Deaf Community and Culture: Some Issues in the Creation of Sign Language Dictionaries. Paper presented at Empower'97: International Conference on Deaf Education. http://www.ssc.mhie.ac.uk/archive/proceedings/briencollins.html

Fabbretti, D., 2000. L'italiano scritto dai sordi: un'indagine sulle abilità di scrittura di sordi adulti segnanti nativi. *Rassegna di psicologia*, 1 (XVII):73-93.

Frege, G. 1892. *Über Sinn und Bedeutun,* Ztschr. f. Philosophie und philosophische Kritik, NF 100:25-50.

Melc'uk, I., 1974. *Opyt teorii lingvističeskix modelej 'Smysl-Tekst'*. Moscow, Nauka.

Radutzky, E. (ed), 1992. *Dizionario bilingue elementare della Lingua Italiana dei Segni*. Rome: Edizioni Kappa.

Rocha Costa, C. A. and Pereira Dimuro, G., 2003. SignWriting and SWML: Paving the Way to Sign Language Processing. Workshop *Traitement automatique des langues minoritaires et des petites languges, TALN 10e conference, tome2,* 193-202.

Rocha Costa, C. A. and Pereira Dimuro, G., Balestra de Freitas, J., 2004. A sign matching technique to support searches in sign language texts. *, this volume*.

Sternberg, M.L.A. (ed), 1987. *American Sign Language Dictionary*. New York: Harper & Row, New York.

Taeschner, T., Devescovi, A., Volterra, V., 1988. Affixes and function words in the written language of deaf children. *Applied Psycholinguistics*, 9:385-401.