Glossary compilation of LSP including a signed language: a corpus-based approach

Cynthia Kellett Bidoli

Scuola Superiore di Lingue Moderne per Interpreti e Traduttori – Università degli Studi di Trieste
Via Filzi 14, 34127 Trieste, Italy
jkellett@units.it

Abstract

Sign language interpreters not only work in a ‘community’ context but also are called to conferences on deafness-related issues containing language for special purposes (LSP). In Trieste, within an Italian national research project, one particular area of research has been centred on investigating textual recasting that may take place during English to Italian Sign Language (LIS) interpretation based on the compilation of parallel multimodal corpora in English, Italian and LIS. Electronic analysis of the corpora enabled the collecting and concordancing of specialized terminology and the development of a pilot version of a trilingual electronic terminological dictionary (on CD-ROM). The glossary will contain dynamic imagery of LIS and will provide a useful and innovative tool for future interpreter trainees.

1. Introduction

Since 2002, joint research has been conducted by teachers of English at the universities of Turin and Trieste within two Italian national research projects on how and to what extent the English language influences cultural and linguistic communication in contact with Italian\(^1\). Italian society, as all societies throughout the world, conceals an invisible ‘community of practice’ within its confines composed of Deaf people\(^2\), which (amongst other social categories identified by the research unit as belonging to different discursive domains\(^3\)) was targeted by the Turin unit as representing a very intriguing area of interlinguistic/cultural contact to study wherein socially and ideologically marked identity traits are not only discursive characteristics but are negotiated in the very choice of communicative code: sign language - in this case Italian Sign Language (Lingua dei Segni Italiana - LIS). The access by Deaf cultures to international communication, has long been ignored in research in Italy in the field of English language/linguistics, and is considered in this particular research project as a relevant intercultural situation to investigate.

Research was first focussed on Intercultural Practices and Strategies of Textual Recasting to verify if and to what extent the production/reception of written and oral English discourse within a number of different domains leads to a propensity for cultural and linguistic intrusion from English into Italian and hence also into LIS (cf. Kellett Bidoli 2004, 2005b, forthcoming a; Ochse 2004a, b, 2005). Research has now turned to Construction of Identity in Socio-political Discourse to investigate the discoursal processes of construction, manifestation and negotiation of social identity, namely in intercultural situations such as the teaching of English to the Deaf or during interpretation from English to LIS.

In order to explain the link between the research projects, corpus linguistics and trilingual dictionary compilation, there follows a brief outline on contact between the English-speaking world and the Italian Deaf community.

2. The English language within the Italian Deaf community

Information about the English-speaking world reaches the Italian Deaf almost exclusively through written Italian sources: newspapers, magazines, translated books and articles, subtitled films and Italian websites on British or American issues. A minor visual source is provided by TV world news through simultaneously interpreted sign language at set times during the day (Kellett Bidoli 2004: 129). But there is also a good deal of direct exposure to the English language at school or university (Ochse 2001, 2004a), through the Internet, in the workplace, during periods of study abroad on cultural exchanges (Socrates Erasmus or Fulbright Scholarships), at public conferences on Deaf issues in the presence of English native speakers, and to a lesser extent at home (Kellett Bidoli forthcoming b). Deaf people are able to read and write in English if given adequate instruction at school, but where contact involves spoken English, professional interpreters are required to enable communication to take place. In 2003, a survey was conducted among professional Italian sign-language interpreters to determine the extent of English to LIS interpretation and discover which genres are commonly involved (Kellett Bidoli 2005a). It was found that interpreters with an active knowledge of English are more numerous than expected, but unfortunately only a handful are willing or able to mediate directly from English to LIS, and thus interpretation from English is normally always filtered in relay through Italian to a second interpreter who transfers the received message into a gestural/visual mode (LIS) for the deaf audience. The
survey uncovered a number of genres within the context of conference interpreting and in particular the field of linguistics (conferences on various linguistic aspects of sign language and interpretation). This finding led to closer investigation of interpreted discourses and more specifically to the selection of an LSP corpus of representative, authentic, English discourses aimed at discovering to what extent the English language influences interpreted LIS (either directly or filtered through Italian) and hence the message received by a deaf end-user, and to what extent divergences may arise due to basic cultural or linguistic distinctions (Kellett Bidoli 2004, 2005b, forthcoming a).

3. From corpus to glossary

Video-recorded speeches were selected4 to create a small corpus of 12,616 English tokens of which there were 3,075 types. The original video recordings in VHS were transformed into a digital corpus for electronic analysis and viewing of the interpreted signed language using Code-A-Text Integrated System for the Analysis of Interviews and Dialogues software (C-I-SAID: Scolari, Sage Publications) capable of handling multimodal source data in the form of media files and plain text (sound, video and written text). The original sound files of the speeches in English were transcribed, together with the visual signed discourses to provide aligned, multimodal, parallel corpora to work on in order to reveal intercultural and linguistic aspects of textual recasting. The parallel corpora were composed of:

- a written transcription in English of the original spoken discourses;
- Italian glosses of the LIS signs (6,643 tokens and 1,819 types) transcribed with the help of an interpreter;
- a written ‘interpreted’ version in Italian of the signed corpus checked by a deaf teacher of LIS;
- a written ‘interpreted’ version in English of the signed corpus.

The horizontally aligned discourses in English and LIS were compared providing clear evidence of occasional disparity (from lexical items to whole chunks) leading to several instances of intercultural or interlinguistic communicative failure through semantic misrepresentation, distortion or omission. An example is given below which describes a preliminary exercise used in simultaneous interpretation training and illustrates corpora alignment:

**ENGLISH:** Students will listen to a fairy-tale that they know and they will..., they're asked to render this fairy-tale in their own words. They usually know the ideas of Little Red Riding Hood, let's say, anything

**ITALIAN:** Gli studenti ascoltano una favola che conoscono bene. Dopo non chiedo loro di ripetere le parole, ma di raccontare quello che conoscono della favola. Non devono solo fermarsi sulle singole parole, devono capire i concetti.

4 Such a system is SignStream, see: www.bu.edu/asllrp/signstream/contact.html.

glossed LIS version was thus filled with numerous bracketed annotations (rather than conventional straight lines with symbols placed above the glosses), as C-I-SAID automatically ignores all text in brackets, enabling rapid word counts and concordances to be performed.

On completion of the lengthy transcription of the LIS glosses, the text was segmented into small meaningful units, using the customary ‘musical score’ format for the transcription of bilingual mediation (as illustrated above). The addition of punctuation was also necessary in order to respect the requirements of C-I-SAID which is programmed to parse ‘segments’ of dialogue chronologically and horizontally, according to punctuation markers. Short segmentation was preferable where possible.

During comparative analysis of the parallel corpora the deaf expert uncovered several instances of omission or unclear, ambiguous signing of technical phraseology and deaf expert uncovered several instances of omission or unclear, ambiguous signing of technical phraseology and lexical items related to the semantic field of linguistics. As the source language of the corpus was English it was decided to manually select specialized terminology from the English word count which initially resulted in around 500 items. Concordances were run for each one using Wordsmith Tools and according to the degree of relevance and frequency of use a further selection was made. Thus, the potential pool of specialized English terminology was reduced to approximately 200 items. Rather than compiling a trilingual terminological glossary containing all the items, a pilot version was produced in electronic format on CD-ROM to be tested on students at the SSLMIT7 (Kellett Bidoli 2005c). Ten lexical items were selected for the pilot version which produced over 60 entries (including synonyms and cross-references) across the three languages. Compilation has continued beyond the pilot version, at present standing at around 100 English entries, and is expected to terminate by mid 2007.

A semasiological approach was chosen leading to an alphabetical ordering of three separate indices: English, Italian and LIS. Headwords and corresponding articles in each of the three languages were first colour-coded, retained on the CD-ROM and enhanced by two national flags and a LIS label. Below is a monochrome example of trilingual lexical information as well as correct word order below the definition. Trainees will thus be able to obtain a dynamic image of the correctly signed lexical items related to family, feelings, food, health, weather, social events and so on. Deaf people may be acquainted with LSP in written form at work, but they rarely need to use it beyond the workplace. Standard signs therefore do not exist in LIS for numerous technical and complex concepts found in spoken Italian or English. Interpreters however, are expected to find a rapid adequate solution and do so by joining together existing signs or inventing new ones. Signed neologisms become established only if transparent enough to convey meaning to the Deaf and if frequently used by other interpreters the same way. If ‘technical’ signs differ in their configuration from one interpreter to another, this may cause perplexity among the Deaf, as was discovered on analysing the corpus of conference speeches.

4. The glossary

The digital conversion of the corpus into aligned parallel corpora permitted rapid word counts, calculation of word frequencies and the running of concordances of the English, Italian and LIS (glosses) to detect lexical items related to the field in question: linguistics. As the source language of the corpus was English it was decided to manually select specialized terminology from the English word count which initially resulted in around 500 items. Concordances were run for each one using Wordsmith Tools and according to the degree of relevance and frequency of use a further selection was made. Thus, the potential pool of specialized English terminology was reduced to approximately 200 items. Rather than compiling a trilingual terminological glossary containing all the items, a pilot version was produced in electronic format on CD-ROM to be tested on students at the SSLMIT7 (Kellett Bidoli 2005c). Ten lexical items were selected for the pilot version which produced over 60 entries (including synonyms and cross-references) across the three languages. Compilation has continued beyond the pilot version, at present standing at around 100 English entries, and is expected to terminate by mid 2007.

A semasiological approach was chosen leading to an alphabetical ordering of three separate indices: English, Italian and LIS. Headwords and corresponding articles in each of the three languages were first colour-coded, ordered vertically and alphabetically, irrespective of language, before being sent to Turin for transfer onto an HTML application (at the Piccola Cooperativa Sociale “Alba” a r.l – O.n.l.u.s). A ‘cross-browser’ approach was chosen that will permit access to the completed glossary through a wide choice of browsers and operative systems.

Colour coding of the trilingual entries permitted rapid, visual identification during compilation and will be retained on the CD-ROM and enhanced by two national flags and a LIS label. Below is a monochrome example of the trilingual articles for the lemma community interpreting. Where the word IMAGE appears, trainee interpreters will find an icon on which to click in order to obtain a dynamic image of the correctly signed lexical item, or a fully signed version of the example provided below the definition. Trainees will thus be able to obtain trilingual lexical information as well as correct word order sequences and collocations at the click of a mouse.

*****

7 An on-line dictionary of LIS being compiled at the Istituto di Comunicazione Specialistica e Phirilinguismo, Accademia Europea di Bolzano, Italy (http://elix.eurac.edu).

**community interpreting** noun/uncountable [kəˈmjuːnɪtɪ ɪnˈtɜːprɪtɪŋ – inˈtɜːprɪtɪŋ]

**interpretazione in campo sociale** [Italian index]
**interpretação em campo social** [Portuguese index]

*Definition:* Interpreting in two language directions consecutively and without notes, principally to assist migrants who cannot speak their host country’s language, in order to enable them to gain full access to legal, health, educational and social services.

*Example:* Over the past decade the field of community interpreting has increasingly attracted the attention of scholars worldwide.

*Synonyms:* public service interpreting, public sector interpreting.

*Note:* ‘Community interpreting’ is a form of ‘liaison interpreting’ which has long been practised but until recently largely ignored as a scholarly subject if compared to ‘simultaneous’ and ‘consecutive’ (conference interpreting). There is some debate as to whether ‘legal interpreting’ and ‘court interpreting’ can be considered as belonging to ‘community interpreting’. ‘Sign language interpreting’ is considered a form of ‘community interpreting’ as Deaf people within our societies are often in need of language assistance in social and institutional settings.

*See also:* liaison interpreting, court interpreting, legal interpreting

---

**Italian**

interpretação di comunità [Italian index]

*Definition:* Interpretazione bi-direzionale e consecutiva, senza l’ausilio di appunti, con l’obiettivo di assistere principalmente gli immigrati che non parlano la lingua ospitante o i sordi al fine di permettere loro di usufruire dei servizi legali, sanitari e sociali.

**IMAGE**

*Example:* Nell’ultimo decennio il campo dell’interpretazione di comunità ha attratto l’attenzione degli studiosi nel mondo.

**IMAGE**

*Synonyms:* interpretazione in campo sociale

*Note:* la ‘interpretação de comunidade’ è una forma d’interpretação de tratativa che si è sempre praticata ma che è stata a lungo ignorata come disciplina di ricerca rispetto alla ‘interpretazione di conferenza’. L’interpretação nella lingua dei segni rientra in questa categoria, in quanto i Sordi hanno spesso bisogno d’interpreti che lavorano in ambito sociale o istituzionale. Nell’italiano esiste il sintagma ‘interpretazione in campo sociale’ mentre nell’interpretazione vocale si evita di usare ‘interpretação de comunidade’ in quanto ci si può confondere con l’interpretazione praticata nell’ambito delle istituzioni europee. In LIS invece ‘interpretação de comunidade’ è il termine più usato.

---

**See also:** interpretazione di trattativa, interpretazione in tribunale, interpretazione giuridica

---

**British English and if necessary American English phonetics are provided followed by colour-coded bilingual translation equivalents of the headword. At first sight there seems little difference between the Italian and LIS equivalents, but by clicking on one or the other, bidirectional access can be obtained to separate articles which have the same definitions and examples but often different notes, with the additional advantage of imagery in the case of the LIS articles. The glossary is tri-directional, so that users can start from an index item or entry article in any of the three languages in order to access information in the other two.

All exemplification of definitions was obtained from concordances of the corpus lexis, run in order to find all occurrences of each item as illustrated below in the extract of a concordance for **chunk** used as a verb or noun:

- not glued to the original. They can **chunk** the information because they...
- hold more information because each **chunk** will in itself already contain...
- rently. You look at larger and larger **chunks** in the original rather than...

Concordances led to exemplification in context of each headword, as shown in the following two examples:

---

**Italian**

**segmento** [Italian index]

*Definition:* A portion of discourse of variable length.

*Example:* You look at larger and larger **chunks** in the original rather than just at one particular word.

*Note:* An interpreter trainee should not memorize single words in the source discourse but learn to listen to **chunks** of information before translating into the target language. The length of a **chunk** depends on individual mnemonic capacity which expands through training and experience.

---

**English**

**chunk** noun/countable [tʃ ank]

*Definition:* A portion of discourse of variable length.

*Example:* You look at larger and larger **chunks** in the original rather than just at one particular word.

*Note:* An interpreter trainee should not memorize single words in the source discourse but learn to listen to **chunks** of information before translating into the target language. The length of a **chunk** depends on individual mnemonic capacity which expands through training and experience.

---

**Italian**

**segmentare** [Italian index]

*Definition:* To mentally select segments of discourse of variable length.

*Example:* You as an interpreter or as a trainee begin to **chunk** the original differently.

---

**English**

**communication** noun/uncountable [ˈkəmjuːnɪkeɪʃən]

*Definition:* The ability to transfer thought and feelings beyond the self.

---

Information is provided on spelling variants and linguistic or semantic features of interest obtained from the patterns of language usage, connotation and collocation that were revealed by concordancing. For example in the case of the head word **communication** three separate observations are noted:

---

**English**

**communication, comunicazione** noun/uncountable [ˈkəmjuːnɪkeɪʃən, ˌkoʊˈmjuːnɪkəˈtʃiːn]

*Definition:* The ability to transfer thought and feelings beyond the self.
Example: It's hard to doubt that the early human species had language, not rudimentary gestural communication or grunts and cries, but the ability to make gestures into word signs.

Note:

- Hearers associate communication with speech, but in parallel their feelings and emotions are conveyed through complex automatic gestures, facial expressions and body movements of which they are often unaware: Non-Verbal Communication (NVC). Signed languages, by excluding speech, emphasize and elaborate upon the NVC.
- Frequently related pre-modifiers are: child-, early-, gestural-, human-, visual-;
- Common collocations: - begins, - evolves, - of a language, - stage.

*****

The addition of numerous cross-reference entries (synonyms, related terms, compounds and derivations of article, but they guide the user to a headword with a necessary to provide ample information for trainees and interest), not all found in the original corpus, we're deemed required to produce ample information for trainees and are included as separate entries, often without a complete article, but they guide the user to a headword with a complete one. For example:

*****

English

ear-voice span noun/uncountable
[ɪə vɔɪs spæn - ɪə vɔɪs spæn]
décalage [Italian index]
décalage [LIS index]
See: décalage

*****

The most evident advantage of producing electronic sign language dictionaries is that dynamic images of signs provided by deaf signers versus the static ones of old can be and must be included. Without the inclusion of high quality images the principle aim of visually illustrating sign language is defeated. Terminological data collection and graphic representation take up a major part of the time required to produce sign language glossaries/dictionaries. However, visual representation is equally important. Great care and planning must go into the digital filming with optimal illumination. Details have to be taken into account like the contrast of the signer’s clothing with the background and agreement beforehand on how the headwords and examples in LIS should be signed. A major problem during filming remains the translation from voice to sign of technical words and neologisms as well as the fact that single words in English or Italian may not have a corresponding sign at all. Meaning in signed languages is conveyed not only through signs but also through classifiers, non manuals, or fingerspelling. For example the term ‘classifier’ used in sign language linguistics literature in English is translated by interpreters into Italian as ‘classificatore’ for want of an alternative linguistic term in Italian. ‘Classificatore’ in this case, is simply a convenient English loan because in Italian it normally means ‘loose-leaf file’, or ‘filing cabinet’ lacking any linguistic connotation. To further complicate matters, in the American literature there is also a distinction between different classifier types: entity classifiers, handling classifiers, tracing classifiers, quantity classifiers etc. In LIS no ready made distinction exists in sign; one simply signs CLASSIFIER because research in the field of LIS classifiers is at an early stage and distinctions have not yet been made. If and when identified they may not necessarily fit the American model. In the glossary, in the Italian and LIS indices, these examples and others are translated literally from English into Italian in order to locate them within the indices, but the signer had to find strategies to convey the full meaning without using single, equivalent, ready made signs.

5. Conclusion

Any form of electronic audio-visual support is an invaluable aid for anyone involved in sign language teaching or sign language interpreter training, not only from Italian to LIS, but also from English to LIS in view of the continuing spread of English as an international language of communication. There are numerous advantages in using electronic format in compiling dictionaries of spoken and signed languages singly or in combination:

- the possibility of including dynamic illustration of sign language terminology and its exemplification in context;
- the speed of instant access through hyperlinks to translation equivalents and related terms;
- unlimited space;
- creativity in the form of varied graphics, the use of colour, insets and numerous visual as well as acoustic devices.

Sign language discourse can be filmed, transformed by transcription into glosses and concordanced like any spoken language to study sign patterns and particular usages which should lead to a better understanding of sign language grammar. It is hoped that the methodology described above, which is essentially straightforward and simple (but time consuming), will encourage others to use corpus construction for the collection of samples of authentic discourses containing different LSP genres be they in Italian, or other signed languages in order to compile terminological dictionaries or specialist glossaries including a signed language.

6. References


Serena Corazza and Luigi Lerose – personal communication.


