Prototyping and Preliminary Evaluation of Sign Language Translation System in the Railway Domain

Cristina Battaglino[1(✉)], Carlo Geraci[2], Vincenzo Lombardo[1], and Alessandro Mazzei[1]

[1] CIRMA and Department of Informatics, University of Torino, Turin, Italy
battagli@di.unito.it
[2] CNRS, Institut Jean-Nicod, Paris, France

Abstract. This paper presents the prototype and the preliminary evaluation of an automatic translation system developed in the LIS4ALL project. The system domain is the corpus of railway station announcements in Italian. The output of the system is a 3D animated avatar that signs announcements in Italian Sign Language. The preliminary evaluation, which measures the accuracy of the translations at the sentence level, relies through the BLEU-RAC4 metric, a variant of the traditional BLEU metric used to evaluate Machine Translation, specifically designed for sentence level evaluation. The aim of the evaluation is to compare the LIS4ALL translation outputs with the human counterparts.

1 Introduction

Automatic translation from spoken into sign language (SL) is of growing interest for the scientific community. In fact, in addition to the traditional issues featured by the automatic translation for spoken languages, Sign Languages exhibit a new variety of challenges: dealing with under-studied languages (e.g., the absence of reference grammars), poorly understood linguistic phenomena (e.g., how to manage the signing space, where signs are performed), the lack of a suitable written form for SL that goes beyond the gloss level; the handling of the multichannel nature of SL articulators (namely, manual and non-manual articulators). Therefore, automatic translation into SL is an interdisciplinary research domain where linguistic, graphic and algorithmic skills are required.

Most of the current research on the automatic translation into sign languages features both symbolic [5,10] and statistical approaches [11]. Symbolic approaches adopt algorithms and knowledge bases that have a direct correspondence with traditional linguistics (grammar, vocabulary, etc.). Natural Language Processing tools are used for analysis and generation of morphological, syntactic and semantic features for both the spoken language input and the sign language output. Often, it is necessary to develop from scratch lexical resources, grammars and knowledge bases. In contrast, statistical approaches adopt algorithms based on alignment frequencies between texts in the source and target languages

M. Antona and C. Stephanidis (Eds.): UAHCI 2015, Part II, LNCS 9176, pp. 339–350, 2015.
DOI: 10.1007/978-3-319-20681-3_32

(sequences of glosses in the case of SL), respectively. Large resources (such as, e.g., parallel corpora) are needed to compute such frequencies. Both approaches have advantages and drawbacks in the specific context of the automatic translation into SL; both adopt avatar technology in order to visualize the translation output [5,7,11,15].

This paper presents a symbolic Italian-LIS translation system for the Italian Sign Language (called LIS - Lingua Italiana dei Segni, the language of the Italian Deaf community), with an avatar animation output, and its preliminary evaluation using the BLEU-RAC4 metric [17].

The paper is organized as follows: in Sect. 2 we present the LIS4ALL architecture and describe how the LIS output is generated from an Italian text. In Sect. 3 we describe the application domain, based on railway station announcements. Section 4 presents the results of the evaluation by using the BLEU-RAC4 metric and discusses the results. Section 5 concludes the paper.

2 The Architecture of LIS4ALL

Current research projects on the automatic translation into SL investigate relatively small domains in which avatars show a good performance, such as, e.g., post office announcements [2] and drivers license renewal [15]. Project LIS4ALL does not make an exception, and its domain is the corpus of announcements broadcast in Italian railway stations.

The project approach relies on the experience, knowledge, and resources of the previous ATLAS project [10], a pioneering project on the automatic translation from Italian into LIS that set up the complete pipeline and focused on the weather forecasting domain. The LIS4ALL project extends the coverage of syntactic constructions and the lexicon built for ATLAS (about 2350 signs), by adding the signs that are specific to the railway domain (about 120).

The major innovations of LIS4ALL are: (1) the account of new linguistics issues that are typical of the domain addressed, and (2) the translation architecture that is partially modified with a parser based on regular expressions. This choice is motivated by the fact that the railway station announcements are based on pre-determined templates and by the particular linguistic structure internal to railway station announcements (see Sect. 3). This allows us to build a parser based on regular expressions that recognizes the correct template for each specific announcement.

Figure 1 illustrates the pipeline of the LIS4ALL architecture, which includes four modules (for further details about the system and the translation process see [4]):

1. Regular expression parser for Italian;
2. Filler/slot based semantic interpreter;
3. Generator for the LIS grammar;
4. Avatar performing the synthesis of the sequence of signs (i.e., the final LIS sentence).

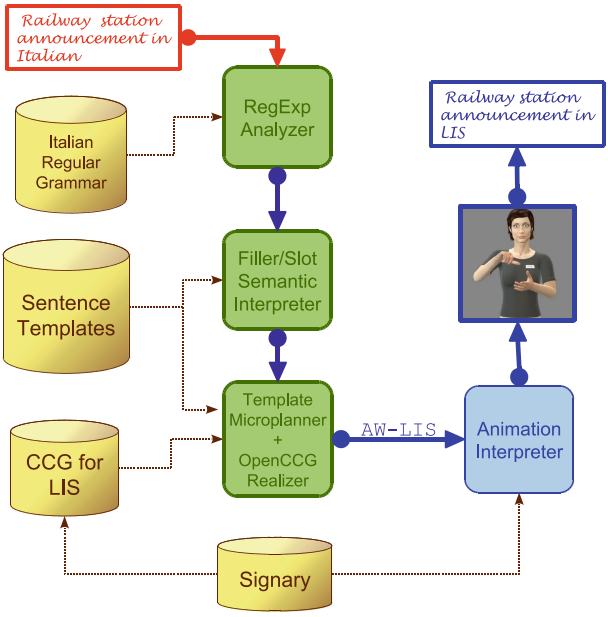

Fig. 1. LIS4ALL Translation Architecture.

The architecture of the LIS4ALL project employs a regular expression-based analyzer that produces a simple (non recursive) filler/slot based semantics to parse the Italian input. This has proven to be more effective because of the large number of complex noun phrases, with several prepositional phrases and nominal modifiers, resulting in degraded parser performance due to multiple attachment options (see Sect. 3).

The LIS4ALL generator consists of two sub-modules: a microplanner and a realizer [14]. The microplanner decides about the syntactic organization of the LIS sentence and about the signs to use in the generation. Following [3], the microplanner is based on templates, which exploit the filler/slot structure produced by the semantic analyzer. The output of the microplanner is a hybrid logic formula in a tree structure (XML), that encodes an abstract syntactic tree. Extending the *Combinatory Categorical Grammar* (CCG) grammar [18] designed in the ATLAS project [10] and using the parallel Italian-LIS corpus produced in LIS4ALL, we implemented a new CCG grammar for LIS that can be used by the OpenCCG realizer to produce LIS sentences in the railway domain [19]. The output of the realizer is an XML file specified with the AWLIS (Atlas Written LIS) language, i.e., a sequence of *lemmata*, accompanied by a description of the meaning of each lemma, its syntactic number and the link to the corresponding sign. The AWLIS language is an XML based language and is used for communication between the generator and the avatar. The Animation Interpreter (see Fig. 1) takes as input the AWLIS representation of the sentence and generates the animation of the virtual signer.

In order to display the translation using a virtual avatar, the following operations are necessary. The signs are collected (through motion capture or key-frame animation techniques) and stored in a repository, the "signary". The signs that create a sentence are then retrieved, concatenated, and synthesized, so that the animation player can guide the virtual avatar in the realization of the translation. The concatenation of the signs that form the LIS sentence is expressed through an animation language [8] that encodes the animation curves into tracks associated with the body parts engaged.

3 LIS4ALL Application Domain: Railway Station Announcements

Railway station announcements are the domain of application of the LIS4ALL project. The structure and the templates for these announcements are described in the Manuale degli Annunci Sonori (MAS – Manual of the Spoken Announcements), filled out by Rete Ferroviaria Italiana (RFI – Italian Railway Network company) [1]. MAS specifies 39 templates that RFI uses to automatically produce the messages announced in all Italian railway stations: 15 templates concern departures, 13 templates concern arrivals, 11 templates concern special situations, such as, e.g., strikes.

The templates have been designed by a group of linguists to yield concise and direct messages in Italian. Full relative clauses, sentential coordination and complex structures (e.g., ellipses) at the sentential level are avoided. As a consequence, the language domain is a controlled language. However, while the syntactic complexity is kept simple at the sentential level, the level of complexity of nominal expressions is considerably high. Consider the following example:

1. "Il treno straordinario Frecciabianca 9764, di Trenitalia proveniente da Roma Termini e diretto a Torino Porta Nuova, delle ore 13:57 è in arrivo al binario 5." (*"Trenitalia Frecciabianca 9764 special train, from Roma Termini, directed to Torino Porta Nuova, with scheduled arrival at 1:57pm is arriving at platform 5."*)

The syntactic structure of the entire clause simply involves a nominal subject ("il treno" / *"the train"*), an unaccusative predicate ("è in arrivo" / *"is arriving"*), and a prepositional complement ("al binario" / *"at platform"*). However, the internal structure of the subject is incredibly complex, involving the following six components:

1. an intersective adjective (e.g., "speciale" / *"special"*);
2. an appositive nominal modifier encoding the category of the train (e.g., "Frecciabianca");
3. an appositive nominal modifier encoding the number of the train (e.g., "9764");
4. a prepositional phrase encoding the enterprise that owns the train (e.g., "di Trenitalia" / *"Trenitalia"*);

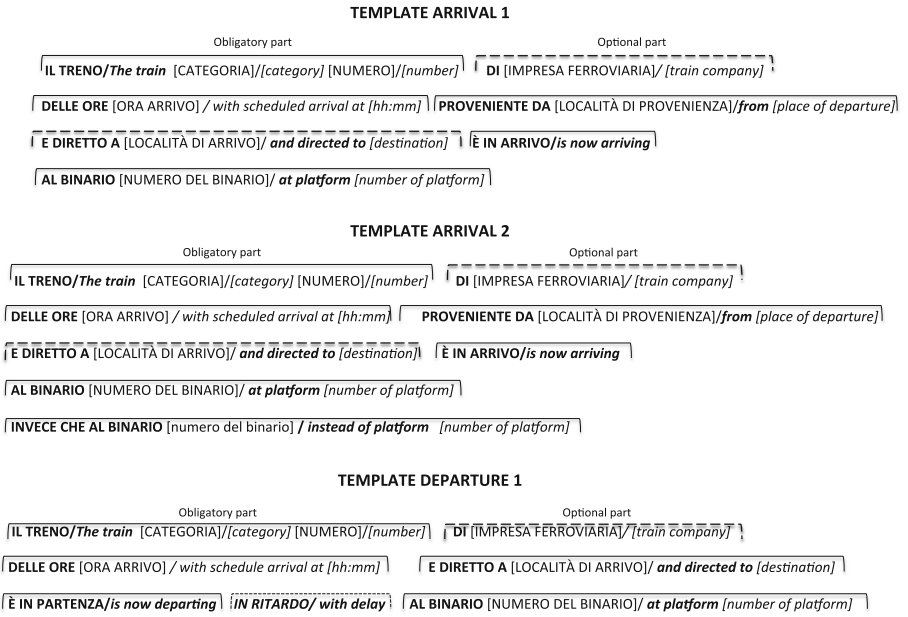

Fig. 2. The templates *Arrival 1*, *Arrival 2* and *Departure 1*. Fixed lexical entries are indicated in bold. The square parenthesis indicate variable lexical entries. The dotted lines indicate the optional parts, while solid lines indicate the mandatory parts.

5. a coordination of two reduced relative clauses encoding origin and final destination of the train (e.g., "proveniente da Roma Termini"/ *"from Roma Termini"* and "diretto a Torino Porta Nuova"/ *"directed to Torino Porta Nuova"*);
6. a prepositional phrase encoding the scheduled time (e.g., "delle ore 13:57"/ *"with scheduled arrival at 13:57"*).

The MAS manual specifies what parts of the template are obligatory or optional, respectively. The optional parts are the first intersective adjective, the name of the company, and the final destination of the train. Both obligatory and optional parts are composed of fixed parts, invariable lexical items, and variable parts that depend upon specific features of the train (e.g., the name of the final destination of the train). For example, in the template *Arrival 1* (see Fig. 2), "Il treno'/ *"The train"* is a mandatory part composed of fixed lexical items ("Il" + "treno"), while"diretto a [località di arrivo]"/ *"directed to [destination]"* is an optional part composed of fixed lexical items (e.g., "diretto" + "a") and variable lexical items (e.g., "località di arrivo""/ *"destination"*, in square brackets).

By analyzing a corpus of messages produced within 24 h of a random day at the Torino Porta Nuova Station (5014 messages total), we found that a small number of templates cover the majority of announcements, while others are virtually absent. The three most frequent templates are *Arrival 1*, which covers

Table 1. Example of a railway station announcement from Arrival 1 in: Italian, simplified Italian, human LIS translation, and human LIS translation of the simplified version and LIS4ALL automatic translation.

Italian	ITA	Il treno Frecciabianca, 9764 di Trenitalia, proveniente da Roma Termini, e diretto a Torino Porta Nuova, delle ore 13 : 57, è in arrivo al binario 5
Simplified Italian	ITA'	Il treno Frecciabianca, 9764, proveniente da Roma Termini, delle ore 13 : 57, è in arrivo al binario 5
Human translation LIS	H_{LIS}	TRENO FRECCIABIANCA NUMERO 9764 TRENITALIA POSSESSIVO ORA 1.57 POMERIGGIO ROMA TERMINI VENIRE, TORINO PORTA NUOVA ANDARE, BINARIO NUMERO 5 LUI ARRIVARE FUT_PROG
Simplified Human Translation LIS	H'_{LIS}	TRENO FRECCIABIANCA NUMERO 9764 ORA 1.57 POMERIGGIO ROMA TERMINI VENIRE, TORINO PORTA NUOVA ANDARE, BINARIO NUMERO 5 LUI ARRIVARE FUT_PROG
System Translation	$LIS4ALL$	TRENO FRECCIABIANCA NUMERO 9764 ORA 1.57 POMERIGGIO ROMA TERMINI VENIRE_PROVENIRE BINARIO NUMERO 5 ARRIVARE FUT_PROG

36 %, *Departure 1* that covers 26 %, and *Arrival 2* that covers 14 %; altogether, they cover about 80 % of the total number of announcements. Therefore, we focused on the translations of the railway station announcements that feature these three templates. All these templates are exemplified in Fig. 2.

Analyzing the corpora of the announcements, we built three regular expressions that match the three templates above. Specifically, for each template, we designed a sequence of semantic slots that are filled, during the translation process, with lexical elements (e.g., scheduled time, platform, station name, destination, place of departure, train category). Each slot corresponds to a variable part of the template. However, considering the high complexity of the nominal subjects in the source language and the fact that nominal modification is highly understudied of the LIS grammar [9], we could not address all the types of nominal modifiers omitted in the templates. So, we limited the development of the automatic translation to the mandatory components of the templates (including both fixed and variable parts, i.e. bold and square parenthesis parts in Fig. 2), by introducing a pre–processing module that simplifies a sentence by deleting the optional components.

Table 1 reports an example of a railway station announcement belonging to the *Arrival 1* template. The first row reports the original announcement in Italian (ITA), the second row reports the simplified announcement in Italian (ITA'), the third row reports the LIS human translation (H_{LIS}), the fourth row reports the human translation of the simplified announcement (H'_{LIS}), the (last) fifth row reports the machine translation output (LIS4ALL). Specifically, the name of the train enterprise and the second conjunct of the reduced relative clause (the one specifying the final destination of the train) have been removed.

Without entering the details of the syntactic structure of the human translation, one important aspect to notice is that the human translation includes a pronominal pointing (i.e., Italian third person singular pronoun "LUI"/ "IT"), that is missing from the automatic translation. This pronominal pointing corresponds to a sort of subject clitic doubling, which is required by the LIS grammar when the subject is too complex. In general, a number of relevant aspects of the LIS grammar are not accounted for by the LIS4ALL project, namely: non-manual "articulators", classifier constructions, grammatical use of the signing space, and prosodic structuring of the message. We are planning a thorough evaluation to identify the priority of each construct to be addressed; in the rest of this paper, we describe a preliminary evaluation that takes into account the components implemented so far through accuracy measures that allow to compare the human, with respect to the automatic Italian–to–LIS translation.

4 Evaluation

The evaluation of the structural components of our Italian–to–LIS translation adopts the BLEU-RAC4 metric [17], a variant of BLEU (BiLingual Evaluation Understudy) [6,13], a common evaluation metrics in machine translation, also for the case of sign languages [12,15,16]. The BLEU-RAC4 score is a measure based on the correspondence of n-grams (sequence of adjacent lexical items) between a reference translation (in our case, the Italian–to–LIS human translation) and a candidate translation (the LIS4ALL automatic translation). The BLEU result is a measure of precision p_n that ranges from 0 to 1 (often reported as a percentage from 0 to 100 %). This measure reflects the accuracy of the candidate translation relative to the temporal order of the sequence of signs. While the classical BLEU metric considers the *precision* based on n-grams and combines each n-gram precision through a geometric mean, the BLEU-RAC4 considers *recall* to yield a better performance at the sentence level and relies on the arithmetic mean [17]. Similarly to BLEU, BLEU-RAC4 assigns a score between 0 and 1 as a measure of the quality of the machine translation. We adopted the BLEU-RAC4 metric rather than BLEU, because our domain of application is made of single sentences and not of concatenated sentences.

The aim of the experiment is to assess the correspondence between the LIS4ALL translation output, which does not account for the optional parts, and the human translation, which does account for the optional parts. In particular, given a fixed number of optional parts, we selected a sample of sentences that

uniformly contain such parts; then, we built a modified sample consisting of the same sentences lacking the optional parts; both samples were translated manually by the human interpreters (H_{LIS} and H'_{LIS} translation, see above); both samples were also translated through the LIS4ALL system (LIS4ALL translation, see above); for each pair of translated samples, the one with the optional parts and the one without the optional parts, we computed the BLEU-RAC4 score that measures the difference between the human and the system translation, respectively; finally, we applied a statistical t–test to measure the distance between the two scores.

Each sample of sentences in Italian contains 21 tokens for each of the three templates above, 63 announcements total (21 for Arrival 1, 21 for Arrival 2, and 21 for Departure 1, see Sect. 3). The number 21 comes out of a combinatory calculation that takes into account two specific optional components (the train company and destination/delay, respectively, see below) and the possible lexical gaps due to incompleteness of the sign repository (in turn due to uncertainty in the definition of the individual signs in such a niche domain). Tokens from the first sample contained a selection of the optional components, concerning the phrases corresponding to the train company (e.g., "di [Impresa ferroviaria]" / *"[train company]"*) and either the destination of the train (for arrivals only, e.g., "diretto a [località di arrivo]" / *"directed to [place of arrival]"*) or the amount of delay (for departures only, e.g., "in ritardo" / *"with delay"*). In addition to these two optional components, we included the problem of lexical gaps for the case of train categories missing in the lexicon (which numbered three). The combination of multiple optional parts together with lexical gaps leads to a sample of 21 sentences per template. These parts were removed from the second sample, which only consisted of sentences with components implemented in LIS4ALL, which could only contain accidental lexical gaps.

Then, on the one hand, the two samples were manually translated by following the set of rules elaborated by a team of interpreters and a linguist (one of the authors of this paper), on the other, the two samples were automatically translated by the LIS4ALL system (i.e., they were the output of the open CCG realizer – see Fig. 1). An example of the announcements with the optional parts, a simplified version, and their human and automatic translations in LIS, respectively, are given in Table 1. For the purpose of this paper, we only focus on the comparison between the sequences of glosses produced by the human and the automatic translations, respectively. Section 4.1 illustrates how the BLEU-RAC4 score is computed, Sect. 4.2 reports the discussion of the results.

4.1 Computing the BLEU-RAC4 Score

The *BLEU-RAC4* is defined as follows:

$$BLEU - RAC4 = \left(\frac{1}{4} \sum_{n=1}^{4} r_n \right) \tag{1}$$

where the recall r_n is defined as:

$$r_n = \frac{Shared}{Total} \tag{2}$$

Shared is the number of n-grams shared by the candidate translation and the reference translation, *Total* is the total number of n-grams in the reference translation. For example, given the $LIS4ALL$ translation compared to the H_{LIS} translation, the *2-gram* "Treno Frecciabianca"/*"Train Frecciabianca"* finds a match in the H_{LIS} translation, and the same is for the *3-gram* "Ora 1.57 pomeriggio" *"1.57 p.m."*. Since the 2-gram "Treno Frecciabianca" and the 3-gram "Ora 1.57 pomeriggio" appear both in the H_{LIS} and in the LIS4ALL translations, so both increase the *Shared* counter. The computation of the total score of the $LIS4ALL$ translation compared to the H_{LIS} translation is given in Fig. 3. Notice that this system does not penalize for lexical items that for some reason appear in the candidate but do not appear in the reference translation.

n-gram	Shared n-gram between H-LIS and LIS4ALL Translator announcement	Total n-gram in H-LIS announcement	$r_{n\text{-gram}}$
1	15	22	15/22
2	11	21	11/21
3	7	20	7/20
4	4	19	4/19
BLEU-RAC4 SCORE		44,15% (0, 4415)	

Fig. 3. An example for computing the BLEU-RAC4 score.

4.2 Results and Discussion

For each announcement, we computed the BLEU-RAC4 score, comparing the $LIS4ALL$ translation against the human translation of the full announcement, H_{LIS}, and its simplified version, H'_{LIS} (see Sect. 3). The prediction is that the $LIS4ALL$ automatic translations have a better performance, compared with the human translation of the simplified announcements than compared with the translation of non-simplified announcements. Mean and standard deviation for each template are given in Table 2.

Paired sample t-tests reveal that the difference between the two series of scores is significant (Arrival 1: $t_{20} = -5.72$, $p < .001$; Arrival 2: $t_{21} = -4.30$, $p < .001$, Departure 1: $t_{21} = -6.90$, $p < .001$). Significance is also maintained at the global level ($t_{63} = -9.35$, $p < .001$). As expected, $LIS4ALL$ translations better match H'_{LIS} than H_{LIS}. Despite the fact that the simplified Italian version of the announcements is better handled by our system, a degraded performance with respect to human translations is still observed (overall BLEU-RAC4

Table 2. Mean and the standard deviation for LIS4ALL translation of templates A1, A2 and P1.

Result	Arrival 1	Arrival 2	Departure 1	Mean
H_{LIS}	0.55 (sd = 0.09)	0.62 (sd = 0.07)	0.58 (sd = 0.09)	0.58 (sd = 0.09)
H'_{LIS}	0.66 (sd = 0.09)	0.70 (sd = 0.08)	0.66 (sd = 0.09)	0.67 (sd = 0.09)

score = 0.67). This is partly due to the fact that our system is currently not able to manage the subject pronominal doubling observed in the H'_{LIS} translations and partly to accidental lexical gaps.

In addition to this, lexical gaps have unexpected outcome orders on the output of the open CCG realizer. This can be shown by looking at the boldfaced constituents in the two examples below:

1. H'_{LIS}: treno/*train* [**intercity notte**/*intercity notte*] numero/*number* [9 6 1 0] ora/*with scheduled arrival at* [5.02] mattina/*a.m.* [**napoli centrale venire**]/ *directed to [napoli centrale]* [**binario numero 16**]/*platform number [16]* ix3 arrivare fut_prog/ */is arriving*;
2. *LIS4ALL*: treno/*train* [**napoli centrale venire**]/*directed to [napoli centrale]* [**binario numero 16**]/*platform number [16]* numero/*number* [9 6 1 0] ora/*with scheduled arrival at* [5.02] mattina / *a.m.* arrivare fut_prog. / *is arriving.*

The effect of lexical gap on the order of signs in the LIS4ALL automatic translation scores 0.71. The subject modifiers referring to train origin and platform number are displaced to second and third position, right after the subject in the LIS4ALL automatic translation. This error correlates with lexical gaps on the train category ("Intercity notte" is missing in the LIS4ALL translation). The result is that the order of higher level constituents (larger n-grams) is disrupted, and the final score of the automatic translation is lower than expected.

5 Conclusion

The LIS4ALL prototype is a system that translates railway station announcements from Italian into LIS. The paper described its architecture, the domain of application, and the preliminary evaluation of its output. Currently, the system has been developed to handle a simplified version of three templates used in Italian stations. Recognition is done by a parser based on regular expressions, while generation is left to a filler/slot based semantic interpreter and to an open CCG realizer. The output is then sent to an animation interpreter which produces the translation into sign language. In this paper, we evaluated the output of the open CCG realizer module by comparing the temporal order of the glosses of the signs as produced by human and automatic interpreter, respectively. The temporal sequence of the glosses for 63 announcements (21 for each template) has been evaluated by using the BLEU-RAC4 metric. Results showed a mean

score of 0.67. Three sources of errors have been identified: (1) the inability to handle subject doubling, (2) lexical gaps, (3) displacement of some subject modifiers (possibly due to lexical gaps in parts of the sentence). While the field of automatic translation into Sign Languages is still in its infancy and several aspects of the human sign language production are still to be implemented in the automatic translation pipeline (especially those concerning the non-manual component), projects such as LIS4ALL show that the automatic translation into sign languages is a worth endeavor.

Acknowledgments. This work has been partially supported by the project LIS4ALL partially funded by Regione Piemonte, Innovation Hub for ICT, 2011–2014, POR-FESR 07–13. Part of the research leading to these results also received funding from the European Research Council under the European Union's Seventh Framework Program (FP/2007–2013) / ERC Grant Agreement n. 324115-FRONTSEM (PI: Schlenker). Research was partially conducted at Institut d'Etudes Cognitives (ENS), which is supported by grants ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-0087 IEC.

References

1. Acoustic Announcements Manual - MAS, Rete Ferroviaria Italiana (RFI) (2011). http://www.rfi.it/cms-file/allegati/rfi/MAS.pdf (in Italian)
2. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: Tessa, a system to aid communication with deaf people. In: Proceedings of the Fifth International ACM Conference on Assistive Technologies, pp. 205–212. ACM (2002)
3. Foster, M.E., White, M.: Techniques for text planning with xslt. In: Proceedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, pp. 1–8. Association for Computational Linguistics (2004)
4. Geraci, C., Mazzei, A.: Last train to "rebaudengo fossano": the case of some names in avatar translation. In: Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages, pp. 63–66 (2013)
5. Huenerfauth, M.: Generating American Sign Language classifier predicates for English-to-ASL machine translation. Ph. D. thesis, University of Pennsylvania (2006)
6. Koehn, P., Monz, C.: Manual and automatic evaluation of machine translation between european languages. In: Proceedings of the Workshop on Statistical Machine Translation, pp. 102–121. Association for Computational Linguistics (2006)
7. Lombardo, V., Battaglino, C., Damiano, R., Nunnari, F.: An avatar-based interface for the Italian sign language. In: International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2011, June 30 - July 2, 2011, Korean Bible University, Seoul, Korea, pp. 589–594 (2011)
8. Lombardo, V., Nunnari, F., Damiano, R.: A virtual interpreter for the Italian sign language. In: Allbeck, A., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 201–207. Springer, Heidelberg (2010)
9. Mantovan, L., Geraci, C.: The syntax of cardinal numerals in LIS. In: Presented at the 45th Annual Meeting of the North Esta Linguistic Society (NELT), October 31 - November 2. MIT, Cambridge (2014)

10. Mazzei, A., Lesmo, L., Battaglino, C., Vendrame, M., Bucciarelli, M.: Deep natural language processing for Italian sign language translation. In: Baldoni, M., Baroglio, C., Boella, G., Micalizio, R. (eds.) AI*IA 2013. LNCS, vol. 8249, pp. 193–204. Springer, Heidelberg (2013)
11. Morrissey, S., Way, A., Stein, D., Bungeroth, J., Ney, H.: Combining data-driven mt systems for improved sign language translation. In: Proceedings of Machine Translation Summit XI (MT 2007) (2007)
12. Morrissey, S., Way, A.: Lost in translation: the problems of using mainstream mt evaluation metrics for sign language translation. In: Proceedings of the 5th SALTMIL Workshop on Minority Languages, May 23, 2006, Genoa, Italy (2006)
13. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
14. Reiter, E., Dale, R., Feng, Z.: Building Natural Language Generation Systems. MIT Press, Cambridge (2000)
15. San-Segundo, R., Montero, J.M., Córdoba, R., Sama, V., Fernández, F., D'Haro, L.F., López-Ludeña, V., Sánchez, D., García, A.: Design, development and field evaluation of a Spanish into sign language translation system. Pattern Anal. Appl. **15**(2), 203–224 (2012)
16. Segundo, R.S., Pérez, A., Ortiz, D., Fernando, D.L., Torres, M., Casacuberta, F.: Evaluation of alternatives on speech to sign language translation. In: INTERSPEECH, pp. 2529–2532. Citeseer (2007)
17. Song, X., Cohn, T., Specia, L.: Bleu deconstructed: designing a better mt evaluation metric. In: Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING) (2013)
18. Steedman, M.: The Syntactic Process. MIT Press, Cambridge (2000)
19. White, M.: Efficient realization of coordinate structures in combinatory categorial grammar. Res. Lang. Comput. **4**(1), 39–75 (2006)