Where should I put my hands? Planning hand location in sign languages

Alice Ruggeri[1], Cristina Battaglino[1], Gabriele Tiotto[2], Carlo Geraci[1], Daniele Radicioni[1], Alessandro Mazzei[1], Rossana Damiano[1], Leonardo Lesmo[1]

[1] Dipartimento di Informatica, Univeristà degli Studi di Torino, Italy
[2] Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy
lesmo@di.unito.it

Abstract. This paper describes a corpus-based study addressing the spatial positioning of hands in Italian Sign Language (LIS). The corpus includes LIS sentences extracted from TV news concerning weather forecasting. The results shows that spatial relations depend on contextual and geographical features. These results have been exploited in the implementation of a planning module that generates a sequence of commands that drive the generator of a virtual character. The planner is part of an Italian to LIS translation system whose goal is the automatic translation of weather forecast programs.

Keywords: Sign languages, automatic translation, signs place of articulation.

1 Introduction

This paper describes the module of an Italian to LIS (Italian Sign Language) translator that takes care of hand positioning in the animation of the virtual character. The overall organization of the translator, the ATLAS project (Automatic Translation into the Language of Signs), includes various software modules taking care of the linguistic translation and of the virtual character animation. Most of these modules, in particular the ones devoted to language interpretation, have been used in various projects and aim at a wide coverage of linguistic structures. However, the structure of the planner, which which covers hands position, is based on the results of a corpus-based study on spatial relations in LIS, and takes care of the general principles of hand positioning in sign languages.

Sign languages (SL), namely the languages used by Deaf people for everyday exchanges, are visuo-spatial languages in the sense that they use the visual apparatus to perceive the linguistic input and use space as a crucial component to transmit linguistic meaning. Signers (like speakers) may adopt two different perspectives to convey spatial information: either they use a route perspective where the viewpoint is within the scene, or they use a survey perspective where the viewpoint is outside the environment [1]. In the route perspective the signing space reflects the individual's perspective of the environment in a 3-D like representation, while in the survey viewpoint it takes a fixed bird's eye view over a horizontal plane. Weather forecasting provides an interesting case study of the use of space in SLs. Weather conditions always refer to geographic areas and a common way to conceptualize them, in the SL use, is by adopting survey perspective. The domain of weather forecasting offers a slightly different survey perspective from the bird's eye view. Specifically, the perspective adopted in this situation is mapped onto imaginary geographical maps displayed along the vertical axis, reflecting the standard way of illustrating forecasting on TV news. For instance, the sign ALPS[1] in LIS (i.e. the mountain chain in the north of Italy) is iconically realized in the upper part of the signing space (i.e. in the north of the imaginary map), as illustrated in (1a). Accordingly, meteorological events happening in that part of the country would be articulated in the same spatial location. For instance, this could be the case of the sign CLOUD-GATHERING, as in (1b). If the cloud gathering had happened in the south of Italy, say in Sicily, the place of articulation of that sign would have changed accordingly.

In the next section we describe the architecture of the translator; section 3 presents the corpus analysis, section 4 sketches the plan-based implementation, while section 5 concludes the paper.

(1) a. ALPS b. CLOUD-GATHERING

2 System architecture

The translator is based on a traditional rule-based approach, where the input sentences are interpreted in terms of an ontology-based logical representation, which acts as input to a linguistic generator that produces the corresponding LIS "sentence" in a language we called AWLIS (Atlas Written Italian Sign Language). A LIS sentence is a sequence of glosses, annotated with some syntactic pieces of information; the sequence is sent to a planner that takes a decision about the position where a sign must be articulated and its output is analyzed by a character animator that visualizes the movements of the virtual character on different media (see Fig.1).

The syntactic analysis is carried out by the Turin University Parser (TUP), which returns syntactic structures in dependency format. Dependency grammars represent syntactic relations by means of labeled binary relations between pairs of words (a Head and a Dependent). The parser includes a chunking step, where chunks (usually nominal groups) are collected and a verbal analysis step, where the chunks are connected to verbs on the basis information about verbal subcategorization (e.g. transitive vs. intransitive). Semantic interpretation is based on an ontology describing

[1] Throughout the paper, we use all-capital words to refer to the glosses of the LIS signs.

the concepts of the weather forecasting domain. The interpreter analyses the syntactic tree and, by accessing the ontology, builds a logical formula that takes advantage of thematic relations in order for assessing the semantic role of the verbal dependents. This formula is used as input for a CCG (Combinatory Categorial Grammar) linguistic generator [2]. This module uses rules describing the AWLIS grammar that have been designed by exploiting the straightforward syntax-semantics interface that is one of the main features of CCG grammars. These rules account for various linguistic phenomena, as the morphological realization of plural, and coordination.

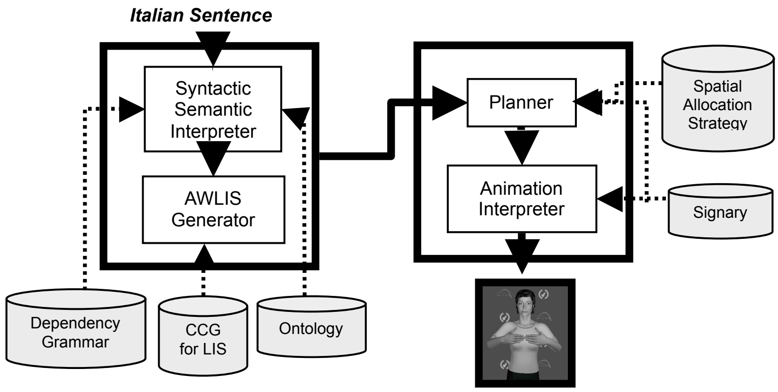

Figure 1: Architecture of the Atlas system

The resulting AWLIS sentence is fed to a hierarchical planner that produces a representation of the signed LIS where various visual features (as facial expression, hand movements, body position, etc.) are explicitly encoded (see section 4). Finally, the planner output is used by a character animation module that generates the actual character movements on the screen.

3 Corpus analysis

Within the ATLAS project, a corpus has been collected, consisting in the translation from Italian into LIS of 40 weather forecasting news recorded from the national broadcasting network. The corpus has been annotated with ALEA, a web application developed for the annotation of video content [3]. The following steps compose the annotation process:

- Sentence segmentation and tokenization (sequential ID association and linking of each token with a lemma entry in the LIS signary)
- Phonological tagging (manual and non-manual components)
- Morphological tagging (part of speech and lexical field)
- Syntactic tagging (syntactic dependencies)
- Discourse tagging (topic and focus identification)

In particular, the annotators indicated Sign Spatial Location (SSL). If the sign is in its standard location (as encoded in the signary) the SSL is left empty. In the case of a location diverging from the citation form, two options are possible: either the sign is relocated in the neighborhood of another (preceding or following) sign (attractor), or the sign is relocated absolutely, i.e. without an overt attractor. The former case is illustrated by the example in (1b) above, where the sign for CLOUD-GATHERING is relocated in the same spatial position where the sign for ‘Alps’ is produced. In presence of an attractor, the annotator encoded the sequential ID number of the attractor, and the relative position of the relocated sign with respect to the attractor (top, bottom, left, etc.). Absolute relocation is illustrated by the examples in (2a-b). In (2a), the sign for LIGURIA, a region in the northwest of the country, is produced in its citation form, i.e. the one you find in standard LIS dictionaries. Its place of articulation is in the middle of the signing space in front of the signer. However, the variant in (2b) is relocated on the left side of the signer, positioning the region on the northwest part of an imaginary map of Italy. In this case, there is no sign working as an attractor. In the case of absolute relocation, the notation simply indicates the position toward which the sign location has shifted (up, down, left, etc.).

(2) a. LIGURIA (citation form)   b. LIGURIA (absolute relocation)

Two studies on sign spatial location based on 10 out of 40 weather forecasting news have been conducted. In the first study we investigated the potential causes triggering relocation, while in the second study we addressed the issue of what triggers assimilation-style relocation or absolute relocation.

For the first study, we excluded from the 1027 annotated items those signs that never undergo the relocation process. Among the 513 remaining items (“relocatable”) the most frequent option is to maintain the standard place of articulation (61% of the cases), while in the 39% of the cases signs are produced in a relocated position. Several multivariate analyses, considering five potential predictors for the distribution of relocated vs. non-relocated signs: part of speech, lexical field, absence/presence of non-manual components, one/two-handed sign, and absence/presence of second-hand activity (i.e. whether the second hand was anticipating or perseverating the handshape of another signs in the utterance). The results reported in table 1 show the two predictors emerged as significant ($p<0.05$) in the VARBRUL analysis [4]: lexical field and number of hands.

Table 1. Input value: 0.381, total chi-square = 4.6363 (chi-square/cell = 0.5795), Log likelihood = –326.326.

Predictor	Level	Application value: relocated yes	
		Factor Weight[2]	%
Lexical field	Geographic	.666	52%
	Meteorological	.569	46%
	Functional	.477	35%
	Other	.307	23%
Hand-number	Two-handed	.534	40%
	One-handed	.422	37%

Lexical field had the strongest effect on the distribution of relocated and non-relocated signs (range = 0.359). Geographic and meteorological terms favors relocation (FW= .666 and FW= .569, respectively), while functional signs and the rest of the signs disfavor relocation (FW= .477 and FW= .307, respectively). A weaker but still significant effect is found for number of hands (range = 0.112): Two-handed signs slightly favor relocation (FW = .534); while one-handed signs disfavor it (FW= .422).

For the second study we concentrated only on those items showing relocation (i.e. on the 200 relocated signs). The results show that most relocated signs have an attractor preceding it in the clause, surfacing as cases of anticipatory assimilation (73%), while relocated signs with a following attractor are very rare (4%). A good portion of tokens shows absolute relocation (27%). For the purposes of the statistical analysis we only considered cases of perseverative assimilation and cases of absolute relocation and we found a significant effect of lexical field, as illustrated in table 2.

Table 2. Input value: 0.752. Total chi-square = 0.000 (chi-square/cell = 0.000). Log likelihood = –104.908.

Predictor	Level	Application value: anticipatory assimilation	
		Factor Weight	%
Lexical field	Functional	.804	93%
	Other	.531	77%
	Meteorological	.518	77%
	Geographic	.292	56%

The main effect of lexical field is due to the extreme behavior of functional and geographic signs. The former strongly favor anticipatory assimilation (FW= .804),

[2] Factor weights (also known as factor probabilities) are a numerical measure of the strength of each level of a predictor. Values that are close to 1 favor the application value (i.e. the level of the dependent variable used as 'baseline'), values close to 0 disfavor it, while values around 0.5 are neutral. In this case, values close to 1 favor relocation, while values close to 0 disfavor it. Factor weights can be easily converted into logits [7].

while the latter strongly disfavor anticipatory assimilation (FW= .292), therefore favoring absolute relocation. Other signs favor assimilation over absolute relocation (FW= .531), as well as meteorological terms, although only weakly (FW= .518).

The fact that geographic and meteorological signs favor relocation can be interpreted as the result of a mapping of spatial relations onto an imaginary geographical map stretched on the vertical plane, forcing displacement of signs somewhere on the map. In a sense, the domain of weather forecasting makes the iconic component of these signs extremely relevant forcing a schematic isomorphism between the aspects of the linguistic signal and aspects of the spatial scene [5]. However, geographic terms and meteorological signs crucially differ on the type of relocation in which they are involved. The former favor absolute relocation as the result of a direct link with the imaginary map, while the latter depend upon an already established spatial relation (e.g. the presence of a geographic sign, as in 'it's raining on the Alps'). Interestingly, absolute relocation is extremely unlikely in the case of functional signs. This is quite expected since the iconic component is virtually absent from these signs.[3]

4 The planner

The planner module has the specific role of organizing the sign flow in the signing space. Its input is the sequence of the lemmas and information on the semantic roles as produced by the generator module. The output returns the same sequence enriched with information on where hands has to be positioned in the signing space.

We use the SHOP2 planning system [6], which relies on the formalism of Hierarchical Task Networks (HTN). Given the sequence of signs, the planning component accounts for the use of the signing resources, namely hands, facial expression, torso, head, and the organization of the signing space in order to accomplish the animation of the given sign sequence. A library of linguistic plans describes how signs can be adapted to the context of a specific sentence (encoded in the AEWLIS input representation), given the constraints provided by the communicative situation and the interpreter's configuration (signing resources, availability of the resources, etc.) [8].

Figure 2 represents the top-level portion of the hierarchical task networks (HTN). The high-level task (*LIS-sign*) decomposes into the task of assigning the signs to the interpreter's hands (not shown), finding the location for each sign (*Localize* in the figure, achieved through the *Find-position* sub-task), then performing the sign (*Make-sign*). Also, the planner determines the initial and final location of the signs that have a parameterized trajectory, such as movement verbs (*Sign-relation* task).

---

3 Pointing signs, whose deictic nature might force absolute relocation, would represent a relevant exception within the class of functional signs. However, the number of pointing signs was extremely low in our corpus possibly due to the formal register used by the interpreters, therefore we couldn't test this prediction.

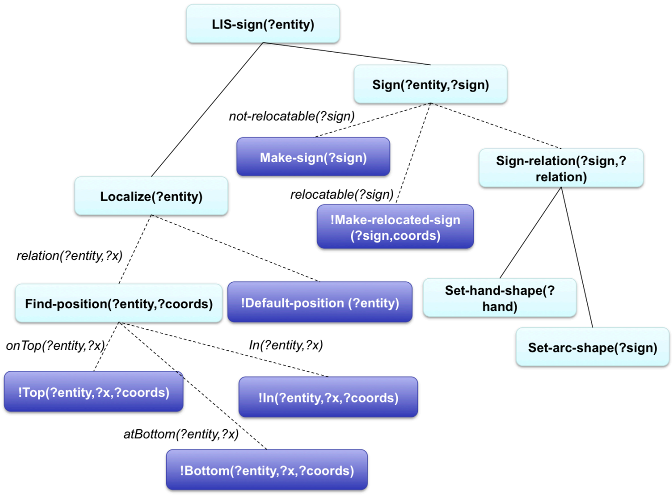

Figure 2: the top level of HTN encoding the signing strategy of the planning component. Dark boxes represent primitive actions.

For each sign the planner produces an ordered pair of spatial coordinates: the first one corresponds to the location of the sign in its citation form, the second one is generated by the planner itself and corresponds to the position in the signing space where the sign will be actually articulated. These latter coordinates are the result of the planning process, which currently take into account only spatial relations lexically expressed by specific lemmas (e.g. NORTH 'in the north', LEFT 'on the left', BOTTOM 'at the bottom of…', etc.), as provided by the generator module. This, however, fails to provide the correct spatial location coordinates for cases where signs are relocated without an overt indication of the spatial relation. These are precisely the situations of absolute relocation and assimilation of the place of articulation described in section 3. In order to fix this problem, we implement the planner by adding two further conditions on the generation of the final coordinates. The first condition introduces additional contextually salient information like the presence of a geographic map that in turns generates an imaginary map in the signing space. This will be used for the purposes of absolute relocation. The second condition capitalizes on the semantic information provided by the generator module. Specifically, we use the semantic role of location (in addition to lexically specified phonological constraints) to identify potential attractors that drive the assimilation of place of articulation. We illustrate here this latter case by considering the sign CLOUD-GATHERING, already introduced above. The first spatial coordinates produced for the sign CLOUD-GATHERING are those of its citation form. If no attractor is present in the utterance, the planner will use these coordinates as the final coordinates for the sign. If an attractor (e.g. ALPS) is present, the coordinates of the attractor are copied as the final coordinates for the sign CLOUD-GATHERING.

Conclusions

In this paper we presented the architecture of a generation module of the Italian into LIS automatic translator ATLAS. One of the crucial aspect of the grammar of SLs is the used of space in order to convey linguistic meaning. In the sign stream, signs can be articulated in positions different from their citation vocabulary-like position. We identified two modes of relocation: relative and absolute. A corpus study on weather forecasting revealed that both articulatory and grammatical factors are responsible for relocation. In particular, the mental representation of spatial relations is projected onto a virtual 'geographical' map along the vertical axis that is used as a guide for relocation. We implemented these findings as a set of rules that drive the behavior of a planner. The planning module gets as input a "symbolic" representation of the sentence in a written form (AWLIS) and takes the decision about where to put the hands, according to the principles identified in the corpus study.

References

1. Emmorey, K., Tversky, B., and Taylor, H.A. 2001. Using space to describe space: Perspective in speech, sign and gesture. *Spatial Cognition and Computation 2*, 157--180.
2. White, M. 2006 Efficient realization of coordinate structures in combinatory categorial grammar. Researh on Language and Computation, 39--75.
3. Barberis, D., Garazzino, N., Piccolo, E., Prinetto, P., and Tiotto, 2010. G. A Web Based Platform for Sign Language Corpus Creation. In proceedings of the International Conference in Computers Helping People with Special Needs, ICCHP 2010, Vol.2, 193--199.
4. Rand, D., and D. Sankoff. 1990. *Goldvarb 2.1: A Variable Rule Application for the Macintosh*. Montreal: Centre de Recherches Mathématiques, University of Montreal. Version 2.
5. Emmorey, K. 2007. The psycholinguistics of signed and spoken languages. In G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford University Press, 703--721.
6. D. Nau, T. Au, O. Ilghami, U. Kuter, J. Murdock, D.Wu, and F. Yaman, "SHOP2: An HTN planning system," Journal of Artificial Intelligence Research, vol. 20, no. 1, 379--404, 2003.
7. Morrison, G. 2005. Dat is what the PM said: a quantitative analysis of Prime Minister Crétien's pronunciation of English voiced dental fricatives, *Cahiers linguistiques d'Ottawa 33*, 1--21.
8. V. Lombardo, F. Nunnari, and R. Damiano. 2010. A Virtual Interpreter for the Italian Sign, in *Intelligent Virtual Agents: 10th International Conference*, IVA 2010, Philadelphia, PA, USA. Proceedings, Springer.