Modeling of geographical location in French sign language from a semantically compositional grammar

Mohamed Nassime Hadjadj
LIMSI, CNRS, Universite Paris Saclay
Rue John Von Neumann, 91400 Orsay, France
mohamed.hadjadj@limsi.fr

Abstract

The use of the specificities related to the visuo-gesual modality of SL, such as the use of the signing space and the simultaneous articulation of multiple channels allows the signer to express structures in a more illustrative way. The description of this structure goes beyond the linear linguistic organization initially applied to describe spoken languages. In this paper, we are interested in modeling structures that rely on the signing space to designate the location of one object relative to another. We are particularly interested in the study of location of one place in relation to another one in French Sign Language (LSF). After a presentation of the corpus and the methodology followed to analyze it, we present the study carried out as well as the results obtained.

Keywords: French sign language, computer modelling, formal grammar

1. Introduction

Many language studies on sign language (SL) have observed that some of the structures are not defined as a linear sequence of signed units (Woll, 2007), (Filhol et al., 2014). The specificities related to the visuo-gesual modality of SL, such as the use of the signing space and the simultaneous articulation of multiple channels, allow the signer to express structures in a more illustrative way. The description of this structure goes beyond the traditional linguistic organization initially applied to describe the spoken languages (Huenerfauth, 2006), (Cuxac and Sallandre, 2007). In this paper, we rely on an empirical approach to model structures that use signing space to designate the location of one object relative to another. We are particularly interested in the location of one place in relation to another in French Sign Language (LSF).We briefly present in this paper the methodology we used to model some structures from a corpus study. We then present the studied corpus and the results obtained.

2. Methodology

To model the localization structures present in corpora, we adopt an approach that moves away from the constraints imposed by linear models (Hadjadj et al., 2018). In other words, we do not suggest any linguistic organization of SLs beforehand. Our approach aims to identify, from LSF corpora, a set of "observable forms" that refers to the same "semantic function". The notion of observable form defines gestural articulations and the different synchronizations that can take place between them. If we take as an example the following articulations:

A: "an eye blink"

B: "move the index finger down"

Each of these two gestural articulations can be defined as a form. Also, a form can be composed simultaneously of several articulations (A and B for example). In this case, the different synchronizations that characterize them are considered as a form criterion. By the notion of semantic function, we mean the interpretation attributed to an observable form. The following examples are considered as possible semantic functions:

C: the concept of "House"

D: Negative expression on a variable element

E: location of an object (obj1) in relation to another object (obj2)

Any systematic association between a single group of observable forms and a semantic function is considered as a rule that participates in the linguistic organization of the language. This article is part of a project of LSF generation. Thus, we are interested in identifying production rules (function-to-form links). It should also be noted that the identification of a link between a single form group and a semantic function may require several function-to-form iterations. We define a production rule by the following triple:

- An identifier: usually the name of the semantic function of the rule
- Arguments of the rule: the set of parameters on which the rule may depend
- Associated form: the invariant forms of the rule and their possible dependencies on the arguments.

Box diagrams (figure 1) can illustrate a production rule where:

- The horizontal axis represents the production time.
- The boxes represent time intervals in which an articulation must take place. The articulators are set in bold; their positions are designated in italics
- The blue boxes are invariant specifications.
- The boxes in red represent the time intervals during which an argument is to be produced.

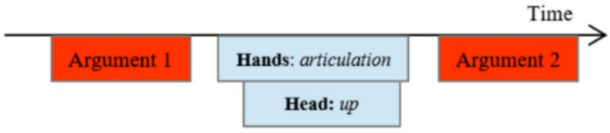

Figure 1: Example of a production rule

3. Study and results

In order to study LSF structures that do not necessarily respect a linear order, we start this study with a criterion L: the geographical location of one place in relation to another. The expression of a link between two places or objects may require a more complex representation than a linear sequence of signed units (Lejeune, 2004). We present below the corpus used and the study carried out

3.1 Corpus

The corpus "websourd AFP 2007" consists of 2000 short summaries of AFP newswire articles of the year 2007. It was signed throughout 2007 by the signers of the company Websourd and it covers various topics: economy, politics, health etc. The large number of newswire articles ensures a relevant number of occurrences of the same linguistic phenomenon as well as its production by several signers.

3.2 First iteration L

Starting this study with a first iteration of function form, we identified three groups of forms. Thus, in our approach, each identified group of occurrences becomes a starting criterion for a new iteration. We stop this process once we define a link between a semantic function and a single form group. We present below the different steps of this study

Iteration L: Location of a place 1 in relation to a place 2 (function criterion)

Example: "Tens of thousands of Shiites arrived on Monday in Najaf (160 km south of Baghdad)" (place 1: Najaf, place 2: Baghdad)

- Number of occurrences in the corpus (Nocc) = 147
- Number of groups identified (of form in this iteration) (Ngp) = 3
- Occurrences that do not fit into any group (Nout) = 15

Group L.1:

Figure 2: pointing sign Figure 3: Articulation of both hands

Figure 4: Location

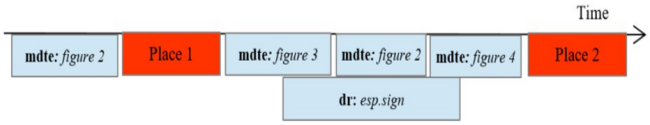

Figure 5: Form of group L.1

The form of group L.1 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands - (mdte): figure 3
- Articulation of the strong hand (mdte): figure 2
- Articulation of the weak (mde) and strong hands (mdte): figure 4
- Eye gaze directed to the signing space (dr: esp-sign)
- Argument 2: place 2

Group L.2:

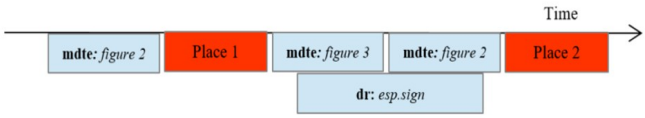

Figure 6: Form of group L.2

The form of group L.2 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands (mdte): figure 3
- Articulation of the strong hand (mdte): figure 2
- Eye gaze directed to the signing space (dr: esp-sign)
- Argument 2: place 2

Group L.3

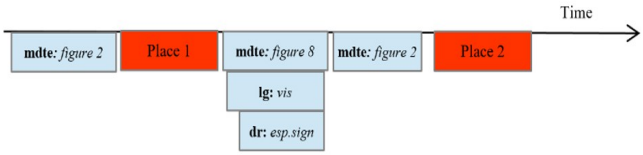

Figure 7: Form of group L.3

Figure 8: Near

The form of group L.3 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands (mdte): figure 8
- The tongue of signer: vsible (lg : vis)
- Eye gaze directed to the signing space (dr: esp-sign)
- Articulation of the strong hand: figure 2
- Argument 2: place 2

3.3 New iterations from L.x groups

We present in the following sections the different iterations made from the three form groups identified during the first iteration as well as the defined production rules.

Form criterion L.1: cf.fig.5

- Nocc = 70
- Ngp = 1
- Nout = 5

Single group: place 2 is a part with undefined borders within place1

Example: place 1 “France”, place2 “south of France”

Function criterion L.1.1. Place 2 is a part with undefined borders within place1

- Nocc = 65
- Ngp = 1
- Nout = 0

The condition of our methodology is verified, the iteration starts with a function criterion associated with a unique group of forms. This defines a production rule, specified as follows:

Production rule L1.1:

- **Identifier**: Place 2 is a part with undefined borders within place1
- **Arguments**: place1, place 2
- **Form**: see Figure 5

Form criterion L2: cf.fig.6

- Nocc = 33
- Ngp = 1
- Nout = 6

Single group: place 2 is a part with defined boundaries inside of place 1

Example: place 1 “ France ”, place 2 “ Marseille”

Function criterion L.2.1: Place 2 is a part with defined boundaries of inside place 1

- Nocc = 27
- Ngp = 1
- Nout = 0

Production rule L2.1:

- **Identifier**: Place 2 is a part with defined boundaries inside of place 1
- **Arguments**: place1, place2
- **Form**: see Figure 6

Form criterion L.3: cf.fig.7

- Nocc = 29
- Ngp = 1
- Nout = 4

Single group: place 1 is near place 2

Example: “Heads of G8 diplomacy meet on Wednesday in Potsdam near Berlin to prepare the international agenda for the Heiligendamm Summit (6-8 June)”. Place 1: Potsdam is near place 2: Berlin.

Function criterion L3.1: place 2 is near place1

- Nocc = 25
- Ngp = 1
- Nout = 0

Production rule L3.1:

- **Identifier**: Place 2 is near place1

- **Arguments**: place1, place2
- **Form**: see Figure 7

3.4 Synthesis of the study L

Starting with a function criterion L, the location of a place 1 in relation to a place 2, we defined after several iterations three production rules. The table 1 is a summary of all iterations performed as well as the production rules defined in this study.

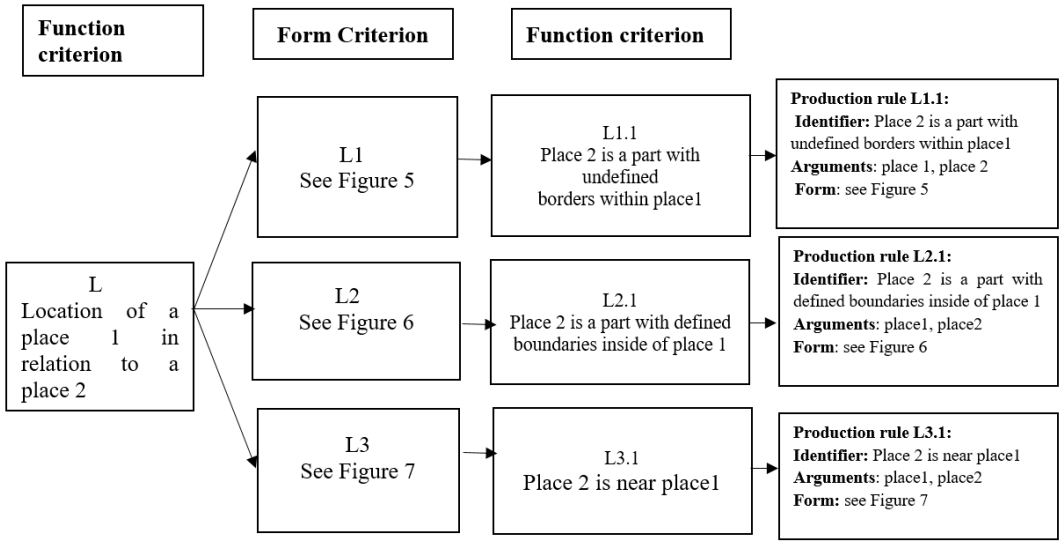

Table 1: synthesis of the study

4. Conclusion

This article has presented the description of some localization structures in LSF. To take into account specificities related to SL, in particular the multilinearity and the use of the signing space, we carried out a study of corpus using a semantic approach. It consists in identifying a systematic link between an observable group of forms and a semantic function. By applying it on the analysis of occurrences of geographical location in the corpus "websourd AFP 2007", we have identified three production rules relating to the location of a place 1 in relation to a place 2.

It should also be noted that some of the rules defined in this study have been merged with other more global production rules presented in Hadjadj et al. (2018). For example, L.1.1 rule and L.2.1 rule are merged with a rule named "addinfo". This production rule includes structures whose second item carries additional information to item 1. If we take the example of the production rule L.2.1, the second item gives additional information to item 1 (its geographical location). This important semantic coverage of the rules is interesting for describing different LSF structures, using a reduced number of production rules.

5. Bibliographical References

Cuxac, C. and Sallandre, M. A. (2007). Iconicity and arbitrariness in French Sign Language : Highly Iconic Structures, degenerated iconicity and diagrammatic iconicity. In A. Pizzuto, E. P. Pietrandrea, & R. Simone (eds.), *Verbal and Signed Languages : Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter. 13-33.

Filhol, M., Hadjadj, M. N., and Choisier, A. (2014). Nonmanual features: the right to indifference. In Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC), Reykjavik, Iceland.

Hadjadj, M. N., Filhol, M., and Braffort, A. (2018). Modeling French Sign Language: a proposal for a semantically compositional system. Language resource and evaluation conference (LREC), Miyazaki, Japan.

Huenerfauth, M. (2006). Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Lejeune, F. (2004). Analyse sémantico-cognitive d'énoncés en langue des signes française pour une génération automatique de séquences gestuelles, Thèse de doctorat, Université Paris 11 Orsay.

Woll, B. (2007). Perspectives on linearity and simultaneity. In M. Vermeerbergen, L. Leeson, & A.Onno (eds.), *Simultaneity in Signed Languages: Form and Function.* Amsterdam: John Benjamins. 103-125.