Using Signing Space as a Representation for Sign Language Processing

Boris Lenseigne and Patrice Dalle

IRIT, Université Paul Sabatier,118, route de Narbonne 31062 TOULOUSE cedex 9, France

Abstract. Sign language processing is often performed by processing each individual sign. Such an approach relies on an exhaustive description of the signs and does not take in account the spatial structure of the sentence. In this paper, we will present a general model of sign language sentences that uses the construction of the signing space as a representation of both the meaning and the realisation of the sentence. We will propose a computational model of this construction and explain how it can be attached to a sign language grammar model to help both analysis and generation of sign language utterances.

1 Introduction

Sign languages, such as the French sign language, use gestures instead of sounds to convey a meaning. They are deaf peoples' natural languages. Unlike oral languages, sign languages are characterized by a great multilinearity due to the fact that the signer can simultaneously use several body parts to communicate: hand configuration, localisation and motion, facial expression, body motion,...

Most of the time, one considers two levels of language: standard utterances that only use standard signs, the ones that can be found in dictionaries, and iconic utterances, so-called "classifier predicates", where most of the meaning relies on iconic structures. Iconic structures are widely used in spontaneous sign language so that they need to be taken in account in automatic sign language processing systems.

Works in French Sign Language (FSL) linguistics [4][3] have shown that, in both standard and iconic utterances, the meaning of a sign language production could be accessed by considering the construction of the signing space. The signing space is the space surrounding the signer and where the signs are produced. During this production, the signer will use that space to position the entities that are evoked in the sentence and to materialize their semanting relationships, so that the resulting construction can be considered as a representation of the meaning of the discourse.

In this paper, we propose a computational representation of this organisation, and describe how this representation can be used to help both automatic interpretation and generation of sign language.

2 Previous work

Most of existing works on the description of sign language focus on describing, as precisely as possible, the gestures that are used to produce the signs. Linguistic sign description systems such as W.S. Stokoe's [5] or, more recently, HamNoSys [20] led to descriptions based on discrete sets of parameters and values to describe signs (hand configurations, elementary motions,...).

In the field of sign language automatic processing, those description systems have been used as templates to define the primitives that have to be characterized for sign recognition. For instance, C. Vogler and D. Metaxas [6][7] use the Liddel&Johnson description system [21] and R.H. Liang, M. Ouhyoung [18] the Stokoe notation, to define the primitives to identify. In the field of sign language uterrances generation, the VisiCast project is based on SigML which is an XML implementation of the HamNoSys notation system [10][11].

Many other works, especially in the field of vision-based sign recognition, use specific feature vectors depending on the data available for recognition, as in [1] or [17]. Only a few one take in account the spatial structure of the utterance. In the field of automatic translation, H. Sagawa et al. [19] propose a vision-based sign recognition system that handles directional verbs and A. Braffort's sign language traduction system, Argo [2], is able to translate iconic sentences that have a fixed structure, by using a single dataglove. The sign language generation system by S. Gibet and T. Lebourque [22] allows the generation of spatially arranged sentences by including the notion of targets to specify gestures which enables the use of explicit designations and directional verbs. Finally, M. Huenerfauth [13] proposes a classifier predicate generation system based on a virtual reality system that is used to specify the relative locations of the entities evoked in the predicate.

But, for the moment, none of these works led to a global model of sign language utterances spatial structure.

3 A computational signing space representation

The initial goal of this modelisation was to provide a internal representation of a sign language sentence for vision-based sign language analysis purposes. To avoid the risk of an abusive simplification implied by an incomplete description, it focuses on representing a subset of the sign language. This subset concerns sentences produced in the context of a timetable, that means sentences that brings on play persons, places, dates and actions.

3.1 What to represent in that model ?

The construction of the signing space is mainly useful to represent the relationships between the entities evoked in the discourse, which can be done without knowing the exact kind of those entities. From this point of view, the signing space representation does not have to represent exactly every notion evoked in

the signed production. It is limited to those that may be involved in some relationship, so that the entities can be distinguished by the relationships that may link them.

From a cognitive point of view [23][14][9], there are only a few possible relationships that can be evoked in a sentence :

- *Temporal relationships* that can be either absolute or relative to the current time of the production.
- *Spatial locations* between two entities.
- *Actions* that can link two or more entities.

In the FSL, entities are evoked through signs and located in the signing space so that their relative position will correspond with the spatial relationships between those entities in the real world. Temporal relationships are evoked through entities that are located on "time lines" (fig. 1). Finally binary actions are evoked through directional verbs and more complex ones by grammatical structures called "transfers" in [3].

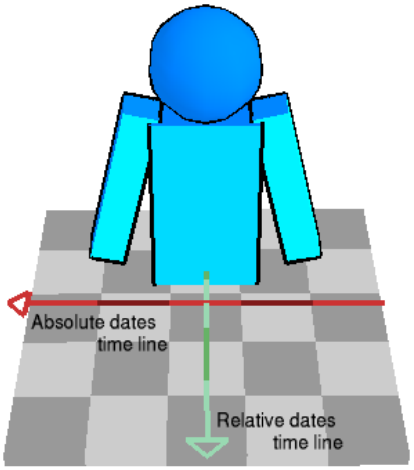

Fig. 1. Symbolic view of the signing space showing the time lines where the dates may be located.

Different kinds of entities depend on the kind of the relationships in which they may be involved :

- *dates* can be involved in temporal relationships ;
- *places* in spatial relationships ;
- *animateds* can perform an action or be located in relation to another ;
- finally *actions* can be referenced as a moment in time or as a protagonist of an action.

The specificities of the French sign language grammar require to consider some additional kind of entities: one needs to make a distinction between entities that whenever involved in a complex action are evoked by the signer taking their role(*persons*[1]) and the entities that cannot be evoked that way (*objects*). Finally, due to the temporal ordering of the signs, one needs to take in account the case of actions that are evoked before one of their protagonists. This entity has an *implicit Type*.

Table 1 gives an overview of the different kinds of entities that can be evoked depending on the relationships that may link them.

Entity	Potential relationships			
	Relative temporal location	Absolute temporal location	Spatial location	Action
Date	×	×		
Place			×	
Animate			×	×
Person			×	×
Action		×		×
Object			×	×
Implicit			×	×

Table 1. Different kinds of entities that may be evoked in a signed sentence and relationships that can exist between them.

3.2 General structure of the model

The symbolic representation of the signing space consists in a cube surrounding the signer, regularly divided into *Site*(s)[2]. Each location may contain a single *Entity*, each *Entity* having a *Referent*. A *Referent* is a semantic notion that can be found in the discourse. Once it has been placed in the signing space, it becomes an *Entity* and has a role in the sentence. So that, building a representation of a sign language sentence consists in creating a set of *Entities* in the *SigningSpace*. The meaning contained in this signing space construction is represented in terms of *Entities*(s) whose *Referent*(s) can have successively different *function*(s) during the construction of the sentence (*locative, agent,...*). A set of rules maintains the consistency of the representation by verifying that enough and coherent information has been provided when one needs to create a new entity in the signing space. The figure (fig. 2) describes the global architecture of the model in the UML notation standard.

[1] In French sign language, persons are not necessarily humans, they can be assimilated to animals or even objects of the real world in humorous stories for example.

[2] Terms written using a *slanted* font are elements of the model.

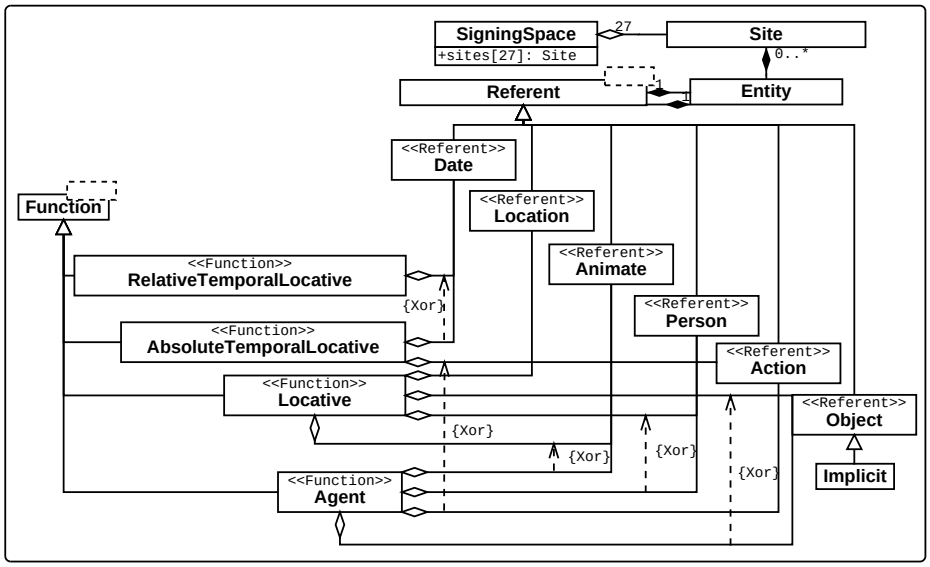

Fig. 2. UML class diagram of the semantic representation of the *SigningSpace*. The *SigningSpace* is regularity divided into *Sites*. Each *Site* can contain a single *Entity* whose *Referent* can have several *Function*(s) during the realisation of the sequence.

3.3 Creating a new entity in the signing space

Every time a new entity is created in the signing space, specific mechanisms are used to ensure the consistency of the new signing space instanciation. Those mechanisms depend on the type of that entity.

Creating a generic entity Generic entities, that are neither *Date*(s) nor *Action*(s), can have two *Function*(s) : *Locative* or *Agent*. Their default function is *Locative*. When such an entity is created in a given *Site*, a new *Referent* of the given *Type* is created. In the case of an automatic analysis system that doesn't take in account the lexicon, it not possible to determine the exact *Type* of the entity so that it remains *unknown* and can potentially be implied in every kind of relationship. The exact *Type* of such an entity will be later changed depending on its successive *Function*(s) during the production of the utterance. The mecanisms that lead to the creation of such an entity are described as an UML sequence diagram in figure 3.

Creating a *Date* The modifications of the *SigningSpace* in the case of the creation of a *Date* are quite the same as those used to create a generic *Entity*. Their default *Function*, which can be either an *Absolute* or a *Relative* temporal locative, depends on its location on one of the time lines (fig. 1). The details of those mechanisms are given in figure 4.

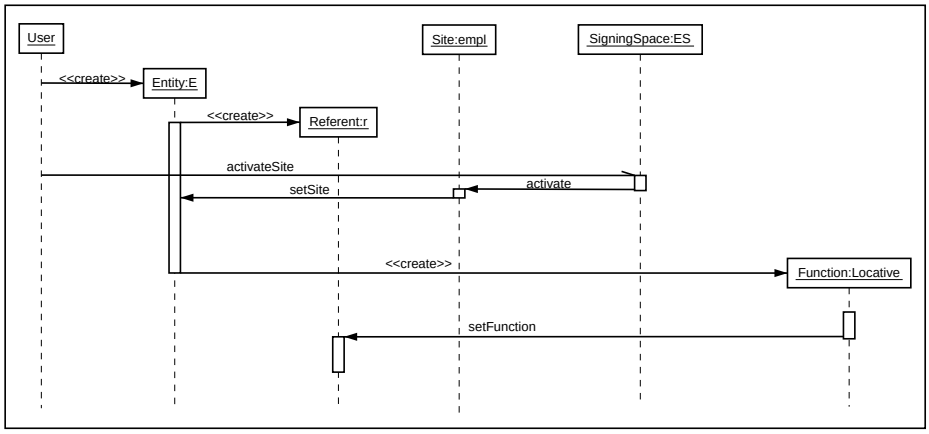

Fig. 3. UML sequence diagram describing the modications of the signing space resulting of the creation of a generic *Entity* in that space.

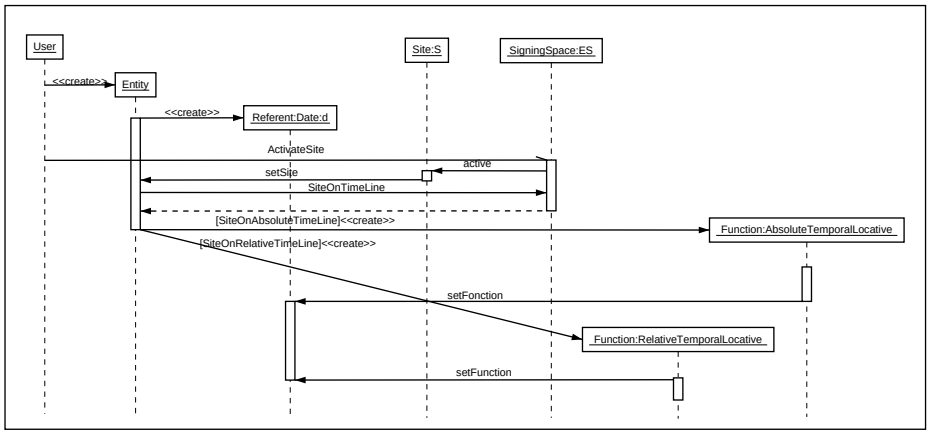

Fig. 4. UML sequence diagrams describing the modications of the signing space resulting of the creation of a new *Date* in that space.

Creating an *Action* *Action*(s) don't have their own *Location*. They link several entities depending on the *Arity* of the *Action*. Those entities are *Protagonist*(s) of the *Action* and their new *Function* is *Agent*. *Protagonists* are defined thru the *Site*(s) that are activated when the signer evokes the *Action*. In the case of a *Site* being empty, an *Implicit Entity* is created in this *Site* (fig. 5).

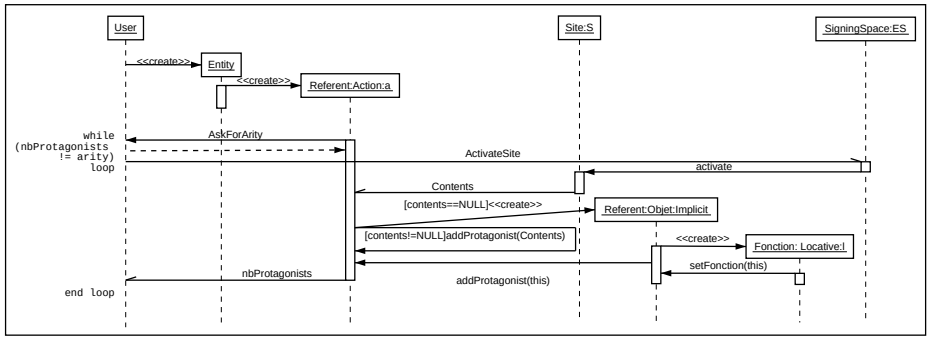

Fig. 5. UML sequence diagrams describing the modications of the signing space resulting of the creation of a new *Action* in that space.

3.4 An example of the construction of the signing space

As an illustration of the use of that model, we will now describe the construction of a *SigningSpace* corresponding to an utterance that concerns a question on a cinema program : *"Who is the director of the movie that plays at* 9.30 *pm on thursday february the* 26^{th} *in Toulouse at the Utopia"*. The sign-to-word translation of that sentence is : *"In Toulouse (1) - In the movie theatre called Utopia (2) - The movie (3) - In this theater (4) - Thursday february the* 26^{th} *at* $9:30$ *pm (5) - the one who (6) - made (7) - That movie (8) - Who is it ? (9)*. The successive configurations of the *SigningSpace* are represented by $3D$ scenes in figure 6. In this *SigningSpace* representation, each kind of *Entity* corresponds to a geometric primitive as detailled in table 2. This transcription was made using an interactive application that allows to manually build the signing space and that implements our model. During the production of the utterance, entities are successively created in the signing space by the mean of signs or specific grammatical structures in the following order :

1. The first *Entity* to be created is a *Place* that is located on the left of the *SigningSpace* and that corresponds to the city of Toulouse.
2. A second *place* is created in the same *Site* of the *SigningSpace* this means that the movie theater is located in Toulouse.
3. The movie that plays in that theater is prepresented by an *Object* that is located in the same *Site* as the movie theater, thus meaning that the movie plays in that movie theater.

Type of the entity	Geometric primitive
Date	white horizontal cross
Place	yellow horizontal disc
Personne	red vertical cylinder
Action	arrow
Object	green cube

Table 2. Geometric primitives that are used to represent different kinds of entities in the $3D$ representation of the *SigningSpace*.

4. The *Date* is located in front of the signer in a *Site* that is located on the time line.
5. An *Entity* whose *Type* is *Person* is created and located on the right of the *SigningSpace*.
6. The person is linked to the movie by the mean of an *Action* so that both the person and the movie become *Protagonist*(s) of that *Action*. Their *Function* is changed into *Agent*.
7. Finally, the question that concerns the person is evoked in the same *Site* as the person. Note that entities whose *Type* is *question* are specific to the interactive implementation of our model.

4 Using the model

The representation of the signing space can be linked to the meaning of the discourse by giving access to the relationships between entities that were evoked and referenced. On the other hand, the iconicity theory by C. Cuxac [4][3] provides a description of the grammar of the sign language in terms of gesture sequences that leads to creating a new entity in the signing space so that it permits to link this representation to the gestures that where used to create the current signing space instantiation [15][16]. Such a predictive model can be used for both analysis and generation of sign language sentences.

Using the signing space for sign language analysis Using that model for sign language analysis leads to two classes of tools : interactive tools intended for linguists to evaluate the model and automatic analysis tools that can be used in many fields of application (linguistic analysis, automatic interpretation,...).

At present time, an interactive tool has been developed in order to represent the construction of the signing space during the production of the utterance. This tool consists in a transcription software that allows to synchroneously link the different steps of the construction of the signing space and the video sequence that is transcripted. This application was designed to evaluate the model on several kinds of utterances and to determine how this model can be considered as a generic representation of sign language utterances.

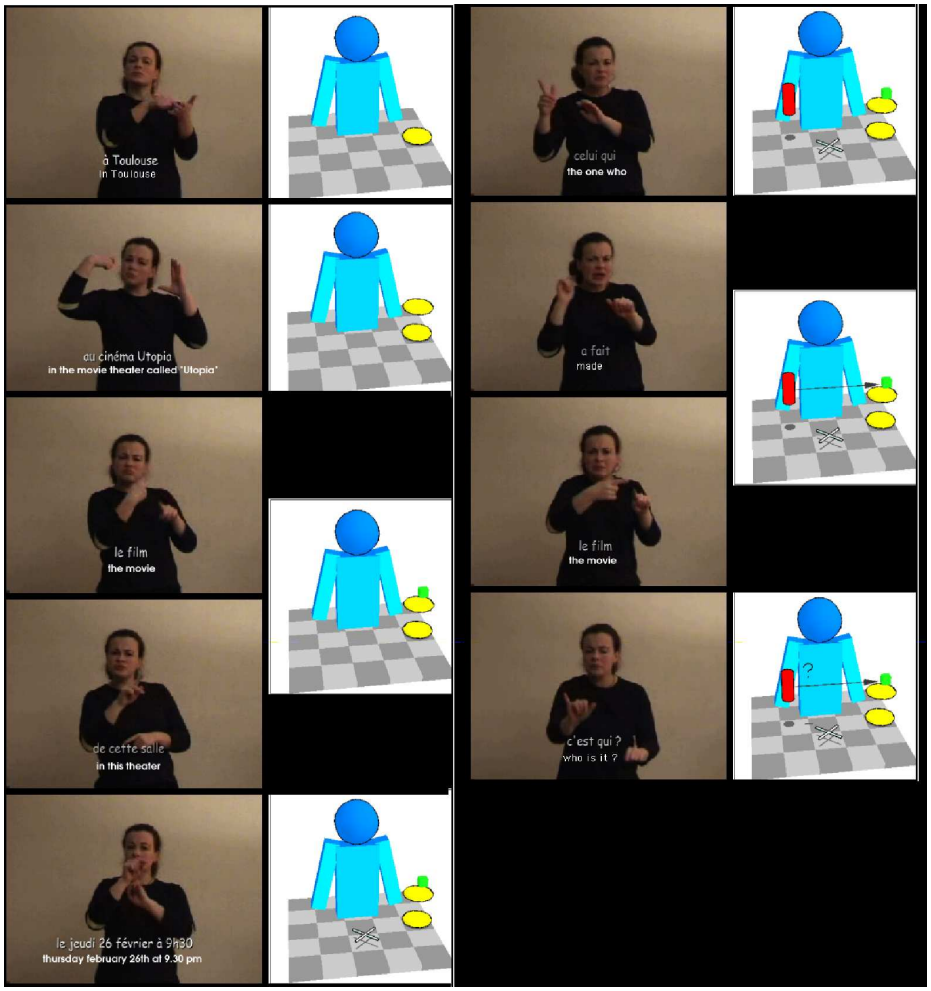

Fig. 6. An example of the construction of the signing space during the realisation of a sign language sentence built with the interactive video sequences transcription tool 4. The sign-to-word translation of that sentence is : *"In Toulouse (1) - In the movie theatre called Utopia (2) - The movie (3) - In this theater (4) - Thursday february the* 26^{th} *at* $9:30$ *pm (5) - the one who (6) - made (7) - That movie (8) - Who is it ? (9).*

In the field of automatic analysis, due to the fact that it is not possible, using a single camera, to build an exhaustive description of the gestures that are used, for automatic vision-based sign language analysis, the model of the signing space is used as a general representation of the structure of the sentence that allows simultaneously to access the meaning of the discourse. The grammar of the sign language that can be attached to that construction allows the use of a prediction/verification approach [16][8]: being given an hypothesis on the meaning of the discourse in terms of a signing space modification, it is possible to infer the gestures that where used to create the new entity in the signing space. Analysing the utterance is then reduced to verify whenever the data corroborates this prediction or not. Such an analysis can be performed without taking in account the lexicon, so that the gestures' descriptions that can be used need to be less precise that the ones required for exhaustive sign recognition. This makes the analysis of low resolution images possible.

However, in a reduced context, the spatial structure of the sentence may be a interesting guideline to identify the signs as it can be done by only considering discriminative aspects of the signs. For instance, by requesting a database concerning airline travels, *places* will be evoked by the name of the towns that can be identified by only considering hand positions.

The three different elements of such automatic tool (signing space representation, grammatical model, low level image processing) have been evaluated separately. It has been shown that in a reduced context, the prediction/verification approach that is described above was relevant and allowed to use simple $2D$ image processing operators instead of complex gesture reconstruction alorithms to performs the identification of the different kinds of entities that where used in the utterance.

Using the model for sign language generation For sign language synthesis, signing space modelling may be used to describe the general structure of the sentence to generate : the sentence is described as a temporally ordered sequence of the entities' creation. So that it is possible to attach to each entity's creation a preliminary description of the underlying gestures that will be used to constrain the generation of the signs properly speaking. This approach provides an easy-to-use way to describe the sentence to generate and will lead to produce spatially organized sentences that are much closer to natural sign language productions than simple coarticulated sign sequences.

The use of the model for sign language generation purposes has been studied in several fields of applications, but the existing elements of the sign language model have not been included in any sign language generation system for now.

5 Conclusion

By looking for a sign language representation that could be used for sign language image analysis, we proposed a general model of the structure of sign language sentences that takes in account the spatial organisation of those utterances. As

this representation can be attached to the gestures that where used to produce the sentence, it constitutes a generic model of the sign language grammar that does not need a fine gesture description. The predictive nature of this model makes it useful for both analysis and generation of sign language.

Moreover one of the main interests of our signing space representation is the possibility to use it, eventually interactively, as a transcription system for sign language sentences. This aspect suggests a new approach to study a written form of the sign language as well as for linguistic studies on the sign language grammar. The integration of the sign language model in a sign language analysis system requires its formalisation and will enable the linguists to collate the linguistic assumptions on this grammar to its application through the interpretation of the sequence.

Finally, gesture descriptions that are used in this model rely on functionnal terms such as "pointing out a place" or "produce a sign in a given place" rather than perceptual terms [12], which point out the interest of this kind of approach for gesture interpretation.

References

1. T. Starner A. Pentland. Real-time american sign language recognition from video using hidden markov models. Technical Report TR-375, M.I.T Media Laboratory Perceptual Computing Section, 1995.
2. A. Braffort. Argo : An architecture for sign language recognition and interpretation. In P. Harling and al., editors, *Progress in Gestual Interaction*, pages 17–30. Springer, 1996.
3. C. Cuxac. French sign language: proposition of a structural explanation by iconicity. In Springer: Berlin, editor, *Lecture Notes in Artificial Intelligence : Procs 3rd Gesture Workshop'99 on Gesture and Sign-Language in Human-Computer Interaction*, pages 165–184, Gif-sur-Yvette, France, march 17-19 1999. A. Braffort, R. Gherbi , S. Gibet , J. Richardson, D. Teil.
4. C. Cuxac. *La langue des Signes française. Les voies de l'iconicité.* ISBN 2-7080-0952-4. Faits de langue, Ophrys, Paris, 2000.
5. W. Stokoe D. Casterline, C. Croneberg. *A dictionnary of american sign langage.* Gallaudet College Press, 1965.
6. C. Vogler D. Metaxas. Toward scalability in asl recognition: Breaking down sign into phonemes. *Gesture Workshop'99, Gif-sur-Yvette, France*, mars 1999.
7. C. Vogler D. Metaxas. Handshapes and movements: Multiple-channel asl recognition. In G Volpe A. Camurri, editor, *Proceedings of the Gesture Workshop'03, Lecture Notes in Artificial Intelligence*, volume 2915 of *Lecture Notes in Computer Science*, pages 247–258, Genova, Italy, April 15-17 2003. Springer.
8. P. Dalle and B. Lenseigne. Vision-based sign language processing using a predictive approach ans linguistic knowledge. In *IAPR conference on Machine vision and application (MVA 2005)*, pages 510–513, Tsukuba Science City, May 2005.
9. J.P. Desclés. La prédication opérée par les langues. In *Langages*, pages 83–97. Larousse, Paris, 1991.
10. J.A. Bangham et al. An overview of ViSiCAST. In *IEEE Seminar on Speech and language processing for disabled and elderly people*, London, April 2000.

11. R. Elliot et al. An overview of the SiGML and SiGMLsigning software system. In *Workshop on Representation and Processing of Sign Language, 4th Internationnal Conference on Language Ressources and Evaluation (LREC 2004)*, pages 98–103, Lisbone Portugal, 30 May 2004.
12. B. Lenseigne F. Gianni, P. Dalle. A new gesture representation for sign language analysis. In *Workshop on Representation and Processing of Sign Language, 4^{th} Internationnal Conference on Language Ressources and Evaluation (LREC 2004)*, pages 85–90, Lisbon Portugal, 30 may 2004.
13. M. Huenerfauth. Spatial representation of classifier predicates for machine translation into american sign language. In *Workshop on Representation and Processing of Sign Language, 4th Internationnal Conference on Language Ressources and Evaluation (LREC 2004)*, pages 24–31, Lisbon Portugal, 30 May 2004.
14. R. Langacker. *Foundations of cognitive grammar: Theoretical Prerequisites.* Stanford University Press, Stanford, CA, 1987. Vol 1, 1987 (Hardcover), 1999 (Paperback).
15. B. Lenseigne. *Intégration de connaissances linguistiques dans un système de vision. Application à l'étude de la langue des Signes.* PhD thesis, Université Paul Sabatier, Toulouse, décembre 2004.
16. B. Lenseigne and P. Dalle. A model for sign language grammar. In *2^{nd} Language and technology Conference*, Poznan Poland, April 21-23 2005.
17. R. N. Whyte M. G. Somers. Hand posture matching for irish sign language interpretation. In *ACM Proceedings of the 1st international symposium on Information and communication technologies*, pages 439–444, Dublin Ireland, September 24-26 2003. Trinity College Dublin.
18. R.H. Liang M. Ouhyoung. A real-time continuous gesture recognition system for sign language. In 3^{rd} *International conference on automatic face and gesture recognition*, pages 558–565, Nara, Japan, 1998.
19. H. Sagawa M. Takeuchi, M. Ohki. Description and recognition methods for sign language based on gesture components. In *in Proceedings of IUI*, orlando, Florida, 1997.
20. S. Prillwitz and al. *HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide.* Hamburg : Signum, 1989.
21. S. Liddell R. E. Johnson. American sign language: The phonological base. *Sign Language Studies*, 64:195–227, 1989.
22. S. Gibet T. Lebourque. High level specification and control of communication gestures : the GESSYCA system. In *Computer Animation'99*, Geneva, 26-28 May 1999. Computer Graphics Society (CGS) and the IEEE Computer Society.
23. R. Thom. *Stabilité structurelle et morphogénèse.* Édisciences, Paris, 2^e edition, 1972-1977.