

Description de la juxtaposition en Langue des Signes Française à partir d'une grammaire réursive

Mohamed Hadjadj^{1,2} Michael Filhol

(1) (2) LIMSI, Rue John Von Neumann, 91400, France

mohamed.hadjadj@limsi.fr, michael.filhol@limsi.fr

RÉSUMÉ

La présente communication s'inscrit dans le cadre du développement d'une grammaire formelle pour la langue des signes française (LSF). Générer automatiquement des énoncés en LSF implique la définition de certaines règles de production pour synchroniser les différents articulateurs du corps, signes, mouvements, etc. Cet article présente dans sa première partie notre méthodologie pour définir des règles de production à partir d'une étude de corpus. Dans la deuxième partie nous présenterons notre étude qui portera sur deux règles de production pour juxtaposer quelques types de structures en LSF. Nous finissons par une discussion sur la nature et l'apport de notre démarche par rapport aux approches existantes.

ABSTRACT

The present communication tackles formal grammar development of French Sign Language (LSF). To generate automatically the statements in LSF it is necessary to have a production rules to synchronize different articulators and signs, etc. In this paper, we describe our methodology to define the link between an observable form and linguistic function. A second section presents the study on the links of juxtaposition between two items from LSF corpus. A discussion of the contribution of our approach compared with other existing approaches will conclude this paper.

MOTS-CLÉS: Langue des signes, grammaire formelle, séquence.

KEYWORDS: Sign language, formal grammar, sequence.

1 Introduction

Afin de générer automatiquement des énoncés en langue des signes française (LSF) à l'aide d'un signeur virtuel à partir d'une grammaire, il est indispensable de définir des règles de production pour synchroniser les différents articulateurs du corps, signes, mouvements, etc. Dans une visée universaliste qui tendrait à utiliser les mêmes outils pour décrire toutes les langues, la dimension séquentielle est essentielle pour définir le niveau syntaxique (Neidle et al, 2000). La LSF est considérée alors comme une langue d'ordre canonique OSV (De Langhe et al, 2003). Cependant, ce rapprochement avec la description des langues vocales ne rend pas systématiquement compte des spécificités que l'on observe dans les langues des signes (LS), liées à leur modalité visuo-gestuelle. Contrairement aux langues écrites dont l'agencement des mots en séquence est pertinent pour définir un ordre linéaire canonique, la multilinéarité des LS (des mouvements simultanés d'articulateurs manuels et non manuels) ainsi que leur aspect iconique rendent un tel ordre moins évident (cf. fig.1). D'autres études présentées essentiellement par Cuxac (1996, 2000) et ses successeurs (Sallandre,

2005) ont mis l'utilisation de l'espace par le signeur et l'icongicité au centre de la description linguistique de la LSF. Leur démarche a permis de mieux appréhender les particularités du fonctionnement des LS, cependant leurs descriptions ne sont pas formelles au sens d'une modélisation utilisable pour générer automatiquement des énoncés à l'aide d'un signeur virtuel.



FIGURE 1: Mouvement de plusieurs articulateurs simultanément

Dans l'objectif de développer une approche qui prend en compte les caractéristiques de la LSF et permettrait sa modélisation, nous avons proposé un modèle formel dans (Filhol et al, 2014). La première partie de cet article présente la méthodologie que nous utilisons pour définir des règles de production. Dans une seconde partie, nous présenterons les résultats d'une étude basée sur un critère de forme, la juxtaposition entre deux items interprétables et successives, qui nous permettra de comparer notre approche avec des démarches basées sur la séquence pour définir la syntaxe des LS.

2 Méthodologie

La méthodologie que nous proposons consiste à définir le lien entre:

- Des «formes» observables dans les corpus, et annotables dans un logiciel d'annotation (cf. fig.2). Nos pistes d'annotation prennent tous les articulateurs et mouvements au même niveau, elles peuvent porter sur les mouvements des mains, la direction du regard, le mouvement du buste, la durée de transition entre les signes...
- Et des interprétations sémantiques de formes que nous nommons «fonctions», par exemple l'énumération (exhaustive, non exhaustive) (Filhol et al, 2013), la temporalité (séquence ou précédence d'événements) (Hadjadj, 2014)

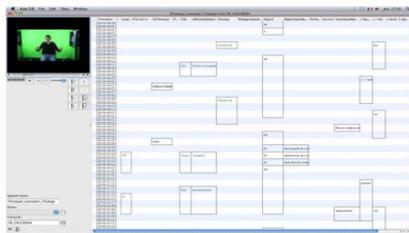


FIGURE 2: Pistes d'annotation (le temps s'écoule ici verticalement avec le logiciel *iLex*)

Notre objectif pour une synthèse de la LSF est de distinguer un groupe invariant d'occurrences de formes pour la même fonction linguistique. Une fois un tel groupe déterminé, une règle de

production est créée, dont le nom est la fonction identifiée et qui comporte la spécification et la synchronisation de toutes les formes paramétrées à générer à l'aide de notre modèle descriptif nommé "Azee" (Filhol et al, 2014)

Pour ce faire, nous appliquons une méthodologie de recherche en corpus alternant les recherches:

- D'occurrences de formes données pour trouver des groupes de fonctions identiques
- D'occurrences de fonctions données pour trouver des groupes de formes identiques

Un groupe identique trouvé devient le point de départ de l'itération suivante. Une fois stabilisé, un critère de fonction peut faire l'objet d'une règle de production. Par exemple, dans une étude antérieure portant sur l'expression de la durée en LSF (Filhol et al 2014), à partir d'un critère de fonction (un événement qui se passe entre deux dates), nous avons appliqué la démarche :

Notre critère de départ T: « événement entre deux dates»

- Exemple: un événement qui se passe entre 1900 et 1905.
Liste d'occurrences: nous avons annoté 70 occurrences de formes présentant cette fonction.
Observations: présence de différentes formes manuelles, clignement des yeux, mouvements du buste..etc.
Analyse: deux groupes de formes se dégagent : à savoir A et B (cf. fig.3)



(A)

(B)

Figure 3: Les deux groupes invariants de formes

Cette première itération a donné lieu à deux autres itérations T.1 et T.2 que nous présentons ci-dessous:

T.1 « Figure 3 (B)»

- Exemple: ce pays chaotique depuis son intervention de 1992 à 1994.
Liste d'occurrences 33 occurrences de formes correspondent à la même fonction sémantique.
Fonction stable: Règle «événement entre deux bornes temporelles»

T.2 « Figure 3 (A)»

- Exemple: d'ici la semaine prochaine.
Liste d'occurrences: 22 occurrences de formes correspondent à la même fonction sémantique.

Fonction stable: Règle «événement partant du temps de l'énonciation à une autre borne temporelle» (cf. Fig.4)

Ces règles de production sont décrites par le formalisme "Azee". Leurs formes peuvent être représentées graphiquement comme sur le diagramme ci-dessous (cf. fig.4) où:

- L'axe horizontal est l'axe du temps de la production.
- Les rectangles limitent la durée de la forme de l'articulateur sur l'axe du temps.
- Les éléments en italiques sont des arguments de la règle.

A titre d'exemple les constituants de la règle dont le critère de départ est T.2 sont respectivement:

- L'événement (Ev): l'argument de la règle.
- "esp.sg": signifie la direction du regard dirigée vers l'espace de signation où les mains sont placées.
- "Mdo; fwd-mw" et "Mdé": signifient respectivement le mouvement de la main dominante vers l'avant et la main dominée (cf. Fig.3 A)
- "Mtn:vb": est une petite rotation de la tête qui porte le menton vers le bas

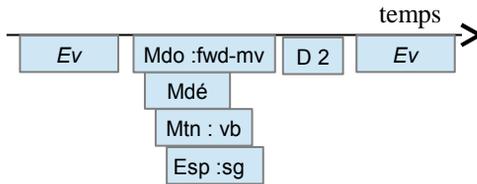


FIGURE 4: Règle «événement partant du temps de l'énonciation à une autre borne temporelle»

3 Étude et résultats

Dans cette étude, nous nous sommes appuyés sur un corpus vidéo LSF (Filhol, 2014). Il s'agit de traductions en LSF de brèves journalistiques AFP par plusieurs professionnels locuteurs natifs. Contrairement à l'étude précédente, nous avons cette fois commencé notre étude par un critère de forme, à savoir la juxtaposition entre deux items interprétables qui se succèdent. L'objectif est de définir un lien de forme à fonction qui pourrait caractériser la synchronisation de deux structures interprétables. Nous avons appliqué la méthodologie présentée ci-dessus à la forme «juxtaposition» d'items, dont nous résumons le déroulé ci-dessous:

Notre critère de départ J: « la juxtaposition entre deux items»

Exemple: item1 « Pays», item2 « Brésil»

Liste d'occurrences: 300 occurrences de formes ont été annotées

Interprétation: plusieurs liens sémantiques entre les deux items se dégagent, nous citons: l'item2 est la localisation géographique de l'item1, l'item1 présente le contexte de l'item2....etc.

Analyse: nous avons distingué essentiellement deux groupes de fonctions à savoir: l'item2 décrit l'état de l'item1 et l'item1 présente le contexte de l'item2.

Nous traitons dans les itérations suivantes uniquement des cas qui portent sur les occurrences où l'item 2 décrit l'état de l'item1.

J.1: « l'item2 décrit l'état de l'item1 »

Exemple: item1 « Le président de la Côte d'Ivoire », item2 « Gbagbo »

Liste d'occurrences: 110 occurrences de formes ont été annotées.

Interprétation: deux groupes invariants de fonctions se distinguent **J.1.1 et**

J.1.2

J.1.1 « item2 donne une information supplémentaire (non essentielle) sur l'item1 »

Exemple: « Aujourd'hui, mercredi » où l'item 2 « mercredi » est ajouté comme information supplémentaire sur « aujourd'hui ».

Liste d'occurrences: 70 occurrences annotées, sur 92% des cas on constate un mouvement du menton vers le haut (c:up) sur le deuxième item.

Forme stable: règle « info-sup » (cf. fig. 5).

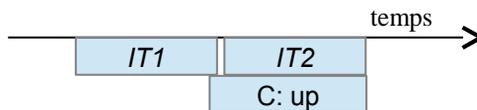


FIGURE 5: Règle de production « info-sup »

J.1.2 « item2 donne une catégorie (p.ex. hyperonyme) de l'item1 »

Exemple: « Pays, Brésil » l'item 2 est apposé pour donner la catégorie de l'item 1

Liste d'occurrences: 33 occurrences annotées, sur 83% des cas on constate un mouvement du menton vers le haut (c:up) sur le premier item.

Forme stable: règle « cat » (cf. Fig. 6)

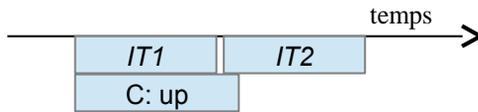


FIGURE 6: Règle de production «cat»

4 Discussion

Nous considérons les résultats de cette étude intéressants et ce pour deux points:

D'un point de vue informatique

Nous avons défini deux règles de production, générant chacune une forme pour une fonction linguistique. Par exemple, la forme décrite pour la fonction (info-supp) dépend des éléments en italique, ses arguments. Cette règle, à son tour, peut être incorporée dans une autre règle. Par exemple, dans notre corpus, nous trouvons des structures qu'on pourrait paraphraser de la manière suivante "Ce pays, localisé ici, est le Monténégro", cette structure est décrite dans notre formalisme avec l'imbrication suivante:

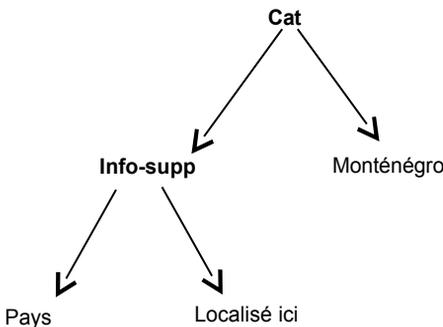


FIGURE 7: La récursivité des règles

Ce constat est pertinent dans la mesure où il nous permet de vérifier l'aspect récursif des règles, caractéristique sans laquelle on ne pourra pas avoir un langage généré infini. Cet aspect rejoint les paradigmes précédents comme celui de la grammaire générative. La différence dans notre modèle cependant est que l'ordre des feuilles ne correspond pas nécessairement en tout point à une séquence observée dans les formes de surface. Des feuilles pourront être générées simultanément selon la forme décrite par la règle mère.

D'un point de vue linguistique:

Influencées par les modèles descriptifs des langues vocales, de nombreuses études définissent le niveau syntaxique de la LSF à partir de l'ordre des signes. Dans notre étude, nous ne définissons pas de tels niveaux grammaticaux et nous appuyons uniquement sur le lien direct entre une forme observable et sa fonction linguistique (interprétée) pour définir des règles et agencer les composants qui ne sont pas nécessairement placés en séquence. Partir sur une démarche moins restrictive que celle

des approches considérées comme universelles et partagées par toutes les langues ne peut qu'enrichir les discussions sur la description linguistique dans son ensemble.

5 Conclusion

Cet article a présenté une étude sur la description de deux cas de juxtaposition en LSF dans le cadre du développement d'une grammaire formelle. La première partie a porté sur la présentation de la méthodologie que nous avons utilisée pour définir nos règles de production. Dans la deuxième partie, nous avons présenté nos résultats portant sur deux règles de production émergeant d'une étude sur la juxtaposition comme forme.

Aussi, l'étude présentée dans cet article peut servir de première étape pour définir la nature de la syntaxe en LS, dont le contour est encore largement débattu sur le terrain linguistique. Ce travail est à poursuivre afin de définir d'autres règles de production portant notamment sur la juxtaposition avec l'objectif de couvrir toutes les combinaisons présentes dans nos corpus.

Remerciements

La région Île-de-France

Références

- CUXAC C. (2000). La Langue des Signes Française (LSF) : Les voies de l'iconicité. *Faits de Langue*, 15-16. Ophrys
- CUXAC C. (1996). Fonctions et structures de l'iconicité des langues des signes. Analyse descriptive d'un idiolecte parisien de la langue des signes française. *Thèse de Doctorat d'Etat*. Université René Descartes - Paris V
- DE LANGHE O., GUITTENY P., PORTINE H., RETORE, C. (2003). À propos des structures OSV en langue des signes française.
- FILHOL M. (2014). Grammaire récursive non linéaire pour les langues des signes, Traitement automatique de la langue naturelle (*TALN*), Marseille, France.
- FILHOL M., HADJADJ M. N., CHOISIER A. (2014). Non-manual features : the right to indifference. In Language resource and evaluation conference (*LREC*), 6th workshop on the representation and processing of Sign Language : beyond the manual channel, Reykjavik, Islande.
- FILHOL M., HADJADJ M. N., TESTU B. (2013). A rule triggering system for automatic text-to-sign-translation. In Sign Language translation and avatar technology (*SLTAT*), Chicago, USA
- FILHOL M., TANNIER X. (2014), Construction of a French-LSF corpus, Building and Using Comparable Corpora, Language resource and evaluation conference (*LREC*), Iceland.
- HADJADJ M. (2014). Description de la durée en LSF à partir d'une grammaire formelle (*RECITAL*), Marseille, France.

NEIDLE C., KEGL J., MACLAUGHLIN D., BAHAN B., LEE RG (2000). The Syntax of American Sign Language: Functional Categories and Hierarchical Structure,, Cambridge, MA: *The MIT Press*.

SALLANDRE M A. (2003). Les unités du discours en Langue des Signes Française. Tentative de cgorisation dans le cadre d'une grammaire de l'iconicité. *Thèse de Doctorat*, Univ. Paris 8