



HAL
open science

Vers la traduction automatique d'adverbiaux temporels du français vers la langue des signes française

Sandra Bellato

► **To cite this version:**

Sandra Bellato. Vers la traduction automatique d'adverbiaux temporels du français vers la langue des signes française. 26e Conférence sur le Traitement Automatique des Langues Naturelles, 2019, Toulouse, France. pp.605-616. hal-02611215

HAL Id: hal-02611215

<https://hal.science/hal-02611215>

Submitted on 18 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Vers la traduction automatique d'adverbiaux temporels du français en langue des signes française

Sandra Bellato¹

(1) MoDyCo, 200 Avenue de la République, 92001 Nanterre, France
bellato.sandra@parisnanterre.fr

RÉSUMÉ

Nous présentons ici de premiers travaux abordant la question de règles de passage entre deux formalismes décrivant la sémantique d'adverbiaux temporels respectivement pour le français et pour la Langue des Signes Française (LSF). Ces travaux prennent place dans une visée de traduction automatique d'une langue vers l'autre. Nous nous appuyons sur un corpus rassemblant 95 adverbiaux temporels du français traduits par trois locuteurs de la LSF.

ABSTRACT

Towards the automatic translation of temporal adverbials from French to French sign language.

The present paper is an initial study addressing the question of rules of passage between two formalisms describing the semantics of temporal adverbials respectively for French and for French Sign Language (FSL). This research takes place in an objective of automatic translation from one language to the other. We rely on a corpus of 95 French temporal adverbials translated by three FSL speakers.

MOTS-CLÉS : adverbiaux temporels, français, LSF, traduction

KEYWORDS : temporal adverbials, French, French Sign Language, translation

1 Introduction

Dans cet article, nous abordons la problématique de la traduction automatique des adverbiaux temporels du français vers la langue des signes française (dorénavant LSF). Dans le domaine de la traduction automatique, deux types de méthodes sont principalement utilisées : la méthode fondée sur des règles dites expertes qui modélisent un savoir linguistique permettant le traitement de données d'une source vers des représentations plus abstraites puis vers des données d'une langue cible ; la méthode par apprentissage automatique fondée sur des récurrences observées dans les données source et dans les données cible. Le manque de données accessibles pour la LSF, et en particulier de données bilingues de bonne qualité français-LSF qui sont essentielles pour un apprentissage statistique, nous a conduits à choisir pour cette étude une approche de traduction automatique fondée sur des règles expertes. Dans cet article, nous revenons dans un premier temps sur ce à quoi renvoie la notion d'adverbial temporel en français ainsi que sur les modes d'expression de la temporalité en LSF, et présentons quelques avancées actuelles en matière de traduction. Dans un second temps, après avoir expliqué notre méthodologie et présenté les différents formalismes que nous avons utilisés lors de

cette étude, nous présentons le corpus choisi pour cette recherche. Pour finir, nous détaillons les résultats obtenus grâce aux différentes étapes de formalisation et proposons en particulier trois règles de généralisation permettant la traduction automatique d'adverbiaux temporels du français vers la LSF.

2 Problématique

La temporalité est un champ d'analyse linguistique complexe, qui nécessite de s'intéresser à de nombreux marqueurs entrant en interaction (verbes, morphèmes de temps verbaux, connecteurs temporels, semi-auxiliaires, adverbes temporels). Dans le domaine du TAL, les adverbes temporels ont reçu une attention toute particulière au cours des quinze dernières années en réponse à des attentes pour les systèmes d'extraction d'information (Pustejovsky et al., 2003).

Pour le français, les travaux de Battistelli et al. (2008) ou Ehrmann et Hagège (2009) ont montré l'intérêt d'une approche fine de cette catégorie de marqueurs temporels qui s'avère plus complexe qu'elle n'en a l'air au premier abord, notamment parce qu'elle renvoie à une définition qui ne fait pas consensus en linguistique. La catégorisation de ces adverbiaux, souvent par une fonction et non par une nature grammaticale bien définie (Wilmet, 2003), ainsi que leur mobilité dans la phrase pouvant impliquer des différences sémantiques (Le Goffic, 1994), sont des exemples de cette complexité. Suivant les principes décrits dans Teissèdre (2012) à la suite de Battistelli et al. (2008), nous retiendrons le terme d'adverbiaux de localisation temporelle (dorénavant ALT) pour désigner les éléments de cette catégorie, laquelle sera entendue comme regroupant aussi bien des adverbes de temps (*depuis*), des syntagmes prépositionnels (*en été 2018*), des subordinées circonstancielles de temps (*depuis qu'elle est arrivée*), que des syntagmes nominaux se comportant comme des compléments adverbiaux (*la veille*).

Si les travaux sur les adverbiaux en français sont nombreux, il n'en existe pas à notre connaissance en LSF, ou alors ils sont intégrés à une analyse plus générale de l'expression du temps en LSF.

Différents travaux (Maeder, 1994 ; Ivani Fusellier-Souza et Jimmy Leix, 2005 ; Sinte, 2015) ont montré que les signeurs, s'exprimant dans différentes langues des signes, possèdent plusieurs moyens à leur disposition pour rendre compte de notions temporelles. Ces moyens renvoient à :

- De nombreux signes lexicaux : ils peuvent être considérés comme équivalents à des noms communs (par exemple des signes désignant les jours de la semaine, les mois de l'année, ...), à des adverbes ou prépositions (par exemple les signes AVANT, APRÈS, DEPUIS, ...), ou encore comme pouvant traduire des notions aspectuelles (par exemple les signes FINI, ENCORE, ...).



Figure 1 : Captures d'image d'une traduction observée de *lundi* dans une vidéo du corpus

- Des lignes de temps : elles permettent de donner une signification particulière à des signes lexicaux ayant à la base une forme dite neutre, mais qui sont flexibles dans l'espace (Sinte, 2015). La LSF utilise quatre lignes de temps, les lignes A et B sont porteuses de notions de temps (comme

le passé ou le futur) et sont situées sur le côté du corps, sur un axe avant-arrière, tandis que la ligne C est porteuse de notions d'aspect (comme la régularité, la continuité ou la durée) et est située devant le corps sur un axe droite-gauche. La ligne D quant à elle est utilisée pour indiquer une idée de croissance, qu'elle soit physique ou au niveau temporel, et n'est pas toujours attestée comme une ligne de temps (Maeder, 1994).

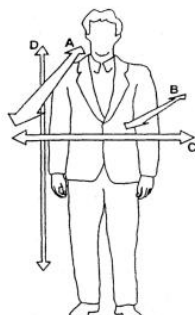


Figure 2 : Schéma des lignes de temps en LSF (Maeder, 1994)

- Des composants non manuels (clignements des yeux, expressions faciales, positions du visage et du corps, ...) : ils servent à affiner le sens, de nuancer certains signes temporels (Filhol et al., 2014). Une ouverture des yeux peut par exemple signifier une accentuation du signe effectué.
- Un ordre dans lequel les événements et les signes sont signés. Les événements, par exemple, sont généralement signés dans l'ordre chronologique dans lequel ils se sont déroulés (Maeder, 1994).

Soulignons que tous ces moyens ne sont pas exclusifs les uns des autres mais apparaissent de manière combinée.

Il est à noter que les langues des signes, dont la LSF, sont des langues encore très peu outillées. Dans le domaine de la traduction automatique quelques efforts ont cependant été initiés (Morrissey, 2008 ; Kervajan, 2011), et les nouvelles avancées technologiques, notamment concernant l'animation d'avatars (Wolfe et al., 2016), permettent également d'envisager de grandes avancées dans ce domaine.

3 Méthodologie

Afin d'établir des règles de traduction automatique, il nous a fallu tout d'abord analyser des ALT en français ainsi que leur traduction en LSF. Pour ce faire, nous nous sommes appuyés sur deux formalismes : l'un dédié aux ALT en français (Teissèdre, 2012), et l'autre à la formalisation de la LSF (Filhol & Hadjadj, 2016). Nous avons appliqué ces formalismes sur un corpus français-LSF en nous concentrant sur les parties correspondant aux ALT en français. Nous avons ainsi pu observer certaines régularités dans les traductions en LSF selon différents critères sémantiques et syntaxiques. À partir de ces observations, nous avons ensuite proposé des règles générales de traduction qui permettraient le passage d'un formalisme vers l'autre.

3.1 Les formalismes d'analyse sémantique des ALT

3.1.1 Formalisation des ALT en français

C'est l'approche de Teissèdre (2012), déployée à partir de Battistelli et al. (2008), que nous avons décidé d'adopter pour décrire la sémantique des ALT en français. Cette approche est particulièrement intéressante pour notre étude car elle permet de décrire et de formaliser tous les ALT quelles que soient leur nature grammaticale et leur type d'ancrage – aussi appelé 'base' - sur un référentiel temporel. Cette base peut ainsi être de nature calendaire (e.g. *le 1^{er} février 2003*), déictique (e.g. *la semaine prochaine*), anaphorique (e.g. *la veille*) ou relative à un événement (e.g. *dès l'élection présidentielle*). Ce formalisme distingue trois types d'opérateurs :

- L'opérateur de régionalisation : il permet de spécifier la région pointée par un ALT sur un axe temporel par rapport à une base. Elle est souvent dénotée en français à l'aide de prépositions ou de locutions prépositionnelles (vers, depuis, avant, bien après, ...). L'opération de régionalisation peut prendre comme valeur ID (e.g. *en 2003*), DEPUIS (e.g. *depuis peu*), JUSQUE (e.g. *jusqu'à aujourd'hui*), AVANT (e.g. *bien avant*), APRÈS (e.g. *postérieurement à 2003*) ou VERS (e.g. *aux alentours du 1^{er} février 2003*).
- L'opérateur de focalisation permet, quant à lui, d'effectuer un zoom par rapport à un repère temporel noyau. La focalisation peut être de différentes natures. Elle peut tout d'abord être granulaire, c'est-à-dire que l'on se focalise sur la granularité décrite par le repère temporel noyau (e.g. *ce jour-là*). La focalisation peut aussi se faire par subdivision, ce qui permet de zoomer sur une portion spécifique de la zone temporelle désignée (par exemple le début de la semaine dans *en début de semaine*). Cette focalisation par subdivision peut être numérique (quart, tiers, moitié) ou qualitative (début, milieu, fin). Il peut enfin s'agir d'une focalisation par sélection lorsqu'il s'agit de sélectionner une partie de la zone temporelle désignée (e.g. *dès le premier jour de l'hiver*). Cette focalisation par sélection peut être formée d'un rang et d'une unité calendaire.
- L'opérateur de déplacement : il permet d'effectuer un déplacement par rapport à une zone temporelle. Cet opérateur peut être formé d'un rang (e.g. *pendant le dernier mois*), d'une durée (e.g. *deux mois avant la fin*) ou d'une unité calendaire ordinale (e.g. *jusqu'au vendredi qui suivra son arrivée*). L'opérateur de déplacement a également une valeur liée à l'orientation du déplacement à effectuer, qu'elle soit égale à avant, après ou à ID si l'opération de déplacement est nulle.

Ces opérateurs s'appliquent dans l'ordre suivant :

OpRégionalisation (OpFocalisation / OpDéplacement (Base))

- (1) *Quelque 50.000 personnes se sont rassemblées jeudi soir à Podgorica, la capitale du Monténégro, pour exprimer leur volonté que cette république devienne indépendante, trois jours avant la tenue d'un référendum historique sur cette question.* (phrase 1G)

Pour exemple, le groupe nominal « *jeudi soir* » présent dans la phrase (1) extraite de notre corpus serait ainsi formalisé :

Base : Déictique (dépend de la situation d'énonciation)

Régionalisation : ID

Focalisation : Soir (focalisation par subdivision)

Déplacement : Orientation : avant (non explicite dans l'ALT mais déduit grâce à la conjugaison du verbe *rassembler*) ; Unité calendaire ordinale : jeudi

Formalisation : ID (Soir / Orientation : avant ; Unité Calendaire Ordinale : jeudi (Déictique))

3.1.2 Formalisation en LSF

Afin de formaliser les traductions en LSF du corpus, nous nous sommes appuyés sur les travaux de Filhol & Hadjadj (2016). Le formalisme AZee décrit dans ces recherches a comme avantage la prise en considération de facteurs non manuels, comme les expressions du visage, la spatialisation, les durées entre éléments séquentiels et les synchronisations entre ceux simultanés, qui sont très importants dans les langues des signes. Filhol et al. (2016) partent du postulat qu'une production langagière est un ensemble de formes observables et de fonctions sémantiques interprétables. La modélisation d'un lien systématique entre une forme (articulée, ou à articuler) et une interprétation (sémantique) donne lieu à une règle de production, et un ensemble (système) de telles règles pouvant couvrir tous les phénomènes linguistiques en LSF est postulé. Nous avons par exemple utilisé les règles suivantes :

– règle Context :

Fonction : L'item 1 décrit le temps, la situation, l'endroit, pendant laquelle l'item 2 se déroule ou est vrai.

Forme : juxtaposition des signes.

Exemple : item 1 = « *aujourd'hui, le 3 mai* » ; item 2 = « *élection présidentielle* » ; interprétation combinée = « *élections présidentielles d'aujourd'hui, le 3 mai* »

– règle Category :

Fonction : L'item 1 indique comment il faut interpréter l'item 2.

Forme : Le menton et/ou les sourcils sont levés au début de l'item 1, avec la tête sensiblement penchée. Il y a également une séparation courte (approximativement 100ms) entre l'item 1 et l'item 2.

Exemple : item 1 = « *pays* » ; item 2 = « *Monténégro* » ; interprétation combinée = « *Monténégro* »

– règle Side-info :

Fonction : L'item 2 donne des informations supplémentaires sur l'item 1, mais le focus de l'information est sur l'item 1.

Forme : Le menton et/ou les sourcils sont levés au début de l'item 2. La transition est minimale entre l'item 1 et l'item 2.

Exemple : item 1 = « *ville* » ; item 2 = « *le pouvoir est le tourisme* » ; interprétation combinée = « *Ville touristique* »

– règle Info-about :

Fonction : L'item 2 est l'information donnée sur l'item 1. Il y a cette fois un focus sur l'item 2.

Forme : Séparation plus longue entre les deux items (environ 340ms) et un clignement des yeux à la fin de l'item 2.

Exemple : item 1 = « *pouvoir* » ; item 2 = « *tourisme* » ; interprétation combinée = « *le pouvoir c'est le tourisme* »

– règle Séquence temporelle (SeqTemp) :

Fonction : Permet de formaliser une chronologie d'événements. Cette fonction n'est pas nécessairement binaire.

Forme : Juxtaposition des événements.

Exemple : item 1 = « *un homme entre dans un bureau de poste* » ; item 2 = « *5 personnes prises en otage pendant 3 heures* » ; interprétation combinée = « *Un homme entre dans un bureau de poste puis 5 personnes sont prises en otage pendant 3h* »

– règle SeqRes :

Fonction : Cette fonction est binaire et prend en argument deux événements E1 et E2. Cette règle est utilisée quand l'événement E2 suit immédiatement l'événement E1, ou quand l'événement E1 a été le déclencheur de l'événement E2. Il y a ici une relation de cause à conséquence implicite.

Forme : l'événement E2 est signé immédiatement derrière l'événement E1, il y a un rythme rapide. De plus, il y a un abaissement durant E2 de la tête et des sourcils, suspendus durant E1.

Exemple : item 1 = « *pluies abondantes* » ; item 2 = « *éboulement de terrain* » ; interprétation combinée = « *Glissement de terrain provoqué par des pluies abondantes* »

– règle Finger-spell :

Fonction : Dénommer un endroit, une personne ou un concept en épelant son nom lettre par lettre.

Forme : Il y a un clignement des yeux avant le début de l'épellation, et le regard est fixé sur la main qui épelle.

Exemple : item 1 = P ; item 2 = O ; item 3 = D ; item 4 = G ; item 5 = O ; item 6 = R ; item 7 = I ; item 8 = C ; item 9 = A ; interprétation combinée = « *Podgorica* »

La combinaison récursive de telles règles peut être représentée par un arbre, comme par exemple avec la fig. 3 représentant une des traductions observées de l'ALT « *jeudi soir* » présent dans la phrase (1) donnée précédemment.

L'imbrication empile, par construction, les formes associées à chacune de ces règles et détermine ainsi les formes à produire en LSF pour exprimer la combinaison sémantique ainsi représentée.

La partie notée à l'aide du symbole [...] correspond à un sous-arbre dans la phrase (1). Elle décrit l'événement se déroulant le « *jeudi soir* » mais son détail n'est pas donné ici.

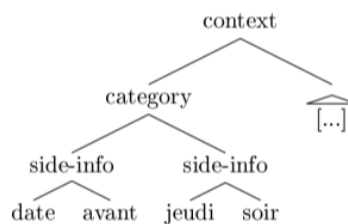


Figure 3 : Arbre représentant une des traductions observées de *jeudi soir* et son lien avec le reste de la phrase (1).

3.2 Description du corpus

Les données bilingues alignant français écrit et LSF sont très rares. Le corpus utilisé dans cette étude a été constitué par Filhol & Tannier (2013) et est le premier corpus français écrit-LSF créé dans un but de recherche académique. Il est composé de 40 brèves journalistiques en français qui ont ensuite été signées par trois traducteurs professionnels natifs en LSF, constituant ainsi un total de 120 vidéos (soit une heure de LSF journalistique). Chacune des vidéos propose une vue de face et de profil, permettant de saisir au mieux les aspects articulatoires, et donc linguistiques, de la LSF. Les traductions n'ont pas été réalisées de manière instantanée, et sont le résultat d'une préparation en amont de la part des traducteurs. Ces 40 brèves journalistiques ont été sélectionnées parmi un ensemble d'environ 980 phrases grâce à la prise en compte de quatre critères en lien avec l'analyse de la temporalité linguistique : i) la précédence chronologique entre des événements ; ii) la durée (d'un événement ou séparant deux événements distincts) ; iii) la relation de causalité entre des événements ; iv) et la notion de répétition pour un événement.

4 Formalisation et observations

En appliquant ces formalismes sur le corpus des 40 brèves, nous disposons au total de 95 ALT formalisés en français, dont 7 sont de nature calendaire (e.g. *le 1er février 2003*), 51 sont de nature déictique (e.g. *mercredi matin*), 14 sont de nature anaphorique (e.g. *la veille*), 21 sont relatifs à un événement (e.g. *Dix ans après l'évacuation musclée de l'église Saint-Bernard*).¹

Pour les données en LSF, en prenant en compte les traductions des trois locuteurs, nous avons pu identifier un peu plus de 1 200 nœuds participant aux traductions clairement identifiables de 83 ALT de la partie écrite (française) du corpus. Les 12 traductions d'ALT non repérées sont essentiellement les ALT constitués d'une simple préposition, telle que « *après* » ou « *lors* », pour lesquels les traductions sont difficilement repérables. En prenant en compte le fait que, dans ces 83 ALT traduits, il y a eu également 8 fois un oubli de la part d'un des traducteurs, cela nous fait un total de 241 traductions d'ALT observables et une moyenne d'environ 5 nœuds pour la traduction d'un ALT par locuteur.

À partir de ces formalisations, nous avons pu faire quelques observations, que nous pourrions illustrer avec des traductions des phrases ci-dessous issues du corpus :

- (2) *Le processus de sortie de coma d'Ariel Sharon se poursuivait progressivement mercredi, alors que son équipe soignante le dit hors de danger "immédiat" une semaine après son attaque cérébrale.* (phrase 1B)
- (3) *Les partis de gauche se retrouvent mercredi à la Mutualité pour leur premier sommet depuis 2002, où la préparation d'actions communes prendra le pas sur celle de l'alternance en 2007, au lendemain de la journée de mobilisation contre le Contrat première embauche (CPE).* (phrase 2J)
- (4) *Le Français Bernard Planche, 52 ans, retenu en otage en Irak pendant 35 jours, a retrouvé samedi la liberté près de Bagdad, et devrait très prochainement regagner la France.* (phrase 1R)
- (5) *Le chef des soldats rebelles du Timor oriental, Alfredo Reinado, a opposé une fin de non recevoir mercredi aux mesures présidentielles destinées à rétablir le calme dans le pays, rendant ainsi probable un dénouement rapide d'une crise qui perdurent depuis plusieurs semaines.* (phrase 1I)

Parmi les 241 traductions d'ALT obtenues, 200 ont été signés en tant qu'item 1 d'une règle *context* avec l'événement décrit en tant qu'item 2 ; 30 ont été signés en tant qu'item 2 d'une règle *side-info* avec l'événement décrit en tant qu'item 1 ; et 11 ont un placement autre par rapport à l'événement qu'ils décrivent. Ce choix de placement de l'élément temporel vis-à-vis de l'événement décrit ne diffère entre les signeurs pour la traduction d'un même ALT que pour 5 ALT déictiques, 1 ALT calendaire, et 3 ALT relatifs à un événement.

Enfin, parmi ces 83 ALT traduits, 20 ont été signés de la même manière par au moins deux signeurs. Les différences notables trouvées dans ses traductions sont, par exemple, l'ajout ou non d'un signe non obligatoire tel qu'un catégorisateur comme le signe DATE, la mobilité apparente de certains signes comme par exemple le signe U-N (Until-Now / du temps s'écoule jusqu'à ce moment), ou encore des oublis partiels de la part des traducteurs.

¹ Le corpus d'origine contenait également 2 adverbes de répétition (e.g. *pour la quatrième fois*) difficilement analysables avec le formalisme de Teissèdre (2012), et que nous avons ainsi décidé de ne pas prendre en compte pour le reste de l'analyse.

Nous pouvons observer tout d'abord que l'élément temporel se place très majoritairement en tant qu'item 1 d'une règle *context*, avec l'évènement décrit en tant qu'item 2. Un exemple de ce premier placement de l'élément temporel peut se trouver dans la figure 4 qui est un exemple de traduction de l'ALT « *une semaine après* » issu de la phrase (2).

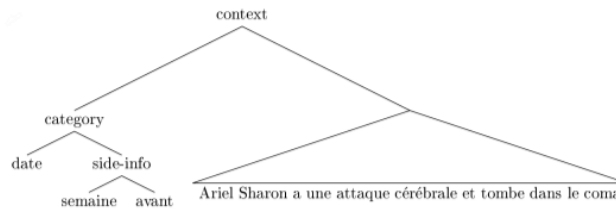


Figure 4 : Arbre représentant une des traductions observées de « *une semaine après* » et son lien avec le reste du contenu.

Le placement de l'élément temporel en tant qu'item 2 d'une règle *side-info*, avec l'évènement décrit en tant qu'item 1, a aussi pu être observé. Un exemple de ce deuxième placement de l'élément temporel peut se trouver dans la figure 5 qui est un exemple de traduction de l'ALT « *depuis 2002* » issu de la phrase (3).

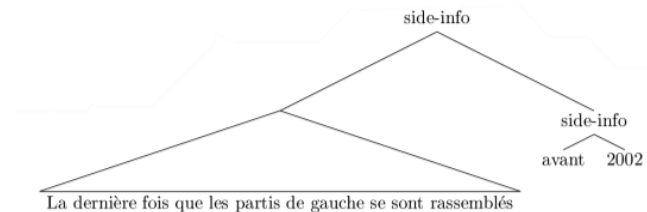


Figure 5 : Arbre représentant une des traductions observées de « *depuis 2002* » et son lien avec le reste du contenu.

Le placement des éléments temporels semble stable en LSF. En analysant plus en détail le placement de l'élément temporel en tant qu'item 2 d'une règle *side-info*, nous avons pu nous apercevoir que ce placement de l'élément temporel avait été utilisé pour les traductions en LSF des trois signeurs des 3 ALT du corpus à base relative à un évènement ayant pour forme en français « pendant + durée ». Ce propos peut être illustré grâce à une des traductions observées du groupe prépositionnel « *pendant 35 jours* » de la phrase (5) du corpus.

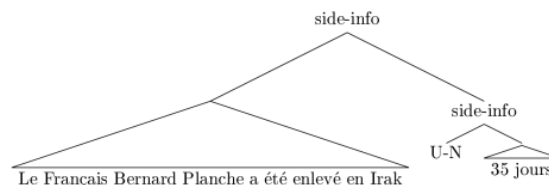


Figure 6 : Arbre représentant une des traductions observées de « *pendant 35 jours* » et son lien avec le reste du contenu.

Pour la traduction d'un petit nombre d'ALT, nous avons pu néanmoins observer le placement de l'élément temporel en tant qu'item 1 d'une règle *context* ou en tant qu'item 2 d'une règle *side-info*

indifféremment selon les signeurs, comme par exemple pour les traductions de l'ALT « *depuis plusieurs semaines* » de la phrase (5) du corpus illustrées dans les figures 7 et 8.

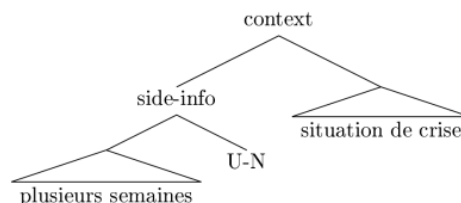


Figure 7 : Arbre représentant une première traduction observée de « *depuis plusieurs semaines* » et son lien avec le reste du contenu (placement de l'élément temporel en tant qu'item 1 d'une règle *context*).

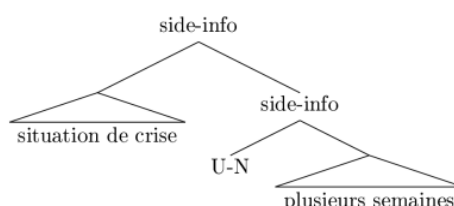


Figure 8 : Arbre représentant une seconde traduction observée de « *depuis plusieurs semaines* » et son lien avec le reste du contenu (placement de l'élément temporel en tant qu'item 2 d'une règle *side-info*).

5 Résultats

Ces observations nous ont conduits à proposer 3 types de généralisations (fig.9).

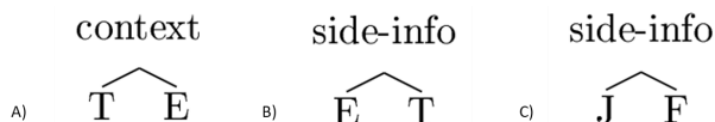


Figure 9 : Schémas des généralisations.

D'après nos résultats de formalisation, les éléments temporels en LSF - et cela peu importe leur base ou la syntaxe d'origine en français - se placent très majoritairement de deux manières distinctes vis-à-vis de l'événement qu'ils décrivent. Nous pouvons ainsi proposer deux premiers types de généralisations :

Généralisation A) : l'élément temporel, que nous noterons T, se place en tant qu'item 1 d'une règle *context* avec l'événement E qu'il décrit en tant qu'item 2. Nous avons pu voir cette généralisation avec les figures 4 et 7.

Généralisation B) : l'élément temporel T se place en tant qu'item 2 d'une règle *side-info* et l'événement E en tant qu'item 1. Nous avons pu voir cette généralisation avec les figures 5, 6 et 8.

Cette généralisation est utilisée pour traduire les ALT français à base relative à un événement ayant pour forme « pendant + durée » (fig. 6).

Nous avons également remarqué que pour les 5 ALT du corpus en français composé d'un jour de la semaine J ainsi que d'une focalisation concernant une partie du jour F, cela a été traduit 12 fois par l'élément J placé en tant qu'item 1 d'une règle side-info, et l'élément F en tant qu'item 2. Nous avons pu observer ceci d'ailleurs dans la figure 1. Cela correspond à la troisième généralisation que nous proposons, la généralisation C). Dans les cas où cette règle n'est pas observable, c'est qu'il y a eu un oubli partiel de la focalisation de la part d'un des traducteurs.

Les résultats sous la forme de seulement 3 types de généralisations que nous avons pu obtenir sont au moins en partie en lien avec la taille réduite et relativement peu diversifié du corpus du point de vue des types d'ALT présents. Certaines formes d'ALT ne sont ainsi pas présentes dans un nombre d'occurrences satisfaisant dans le corpus pour pouvoir faire des généralisations. Par exemple, nous ne trouvons dans le corpus que trois occurrences d'ALT possédant un opérateur de régionalisation et un opérateur de focalisation ayant une valeur différente de ID. L'impact que ces deux opérateurs peuvent avoir l'un sur l'autre ne peut donc pas être analysé de manière très approfondie.

6 Conclusion

L'objectif de cette étude était d'effectuer un premier pas vers la traduction automatique des ALT du français vers la LSF. La formalisation d'ALT issues d'un corpus en français grâce aux travaux de Teissèdre (2012), ainsi que la formalisation de leurs traductions en LSF grâce aux travaux de Filhol & Hadjadj (2016), nous ont permis d'obtenir des données formalisées que nous avons pu comparer afin de pouvoir ensuite proposer trois premières règles générales de passage d'un formalisme à l'autre permettant un début de traduction automatique d'ALT du français vers la LSF.

Ces règles de traduction proposées sont un premier apport à la question de la traduction automatique du français vers la LSF car il n'existe pas de règles déjà existantes de traductions automatiques pour la LSF. Il n'existe également pas de telles règles pour d'autres langues des signes nous permettant de vérifier si ces règles se retrouvent dans différentes langues des signes.

Ces règles restent cependant très générales et mériteraient dans le futur d'être améliorées, par exemple avec l'aide de nouvelles analyses sur de nouveaux corpus.

Il reste que cela serait grâce à ce genre de règles que nous pourrions par la suite passer à des animations à l'aide d'avatar. Ces animations pourront alors être proposées à des locuteurs natifs en LSF, ce qui permettra alors de modifier ou d'affiner d'avantage ces premières règles de traduction.

Remerciements

Ce travail s'insère dans le cadre du projet DGLFLF « *temporalité linguistique en LSF et en français écrit : création d'un outil d'aide à l'acquisition de marqueurs temporels chez l'enfant sourd signeur* » (2016-2018).

Références

BATTISTELLI D., COUTO J., MINEL J.-L., SCHWER S. (2008). Représentation algébrique des expressions calendaires et vue calendaire d'un texte. Actes de *TALN'08*, Avignon.

- BATTISTELLI D. (2011). *Linguistique et recherche d'information : la problématique du temps*. Lavoisier.
- EHRMANN M., HAGEGE C. (2009). Proposition de caractérisation et de typage des expressions temporelles en contexte. Actes de *TALN'09*, Senlis.
- FILHOL M., HADJADJ M.N., TESTU B. (2013). A rule triggering system for automatic Text-to-Sign translation. *Universal access in the information society* 15, 487-498.
- FILHOL M., HADJADJ M. N., CHOISIER A. (2014). Non-manual features: the right to indifference. In *6th Workshop on the Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation*, Reykjavik, Iceland.
- FILHOL M., TANNIER X. (2014). Construction of a French–LSF corpus. In *Language Resources and Evaluation Conference, Building and Using Comparable Corpora*.
- FILHOL M., HADJADJ M. N. (2016). Juxtaposition as a form feature - syntax captured and explained rather than assumed and modelled. In *Language resources and evaluation conference (LREC), Representation and processing of Sign Languages*.
- FUSELLIER-SOUZA I., LEIX J. (2005). L'expression de la temporalité en Langue des Signes Française (LSF). Actes du colloque *Conceptualisation et Surdit *, 207-230.
- GOFFIC P. L. (1994). *Grammaire de la phrase franaise*. Hachette.
- KERVAJAN L. (2011). *Contribution   la traduction automatique Franais/Langue des Signes Franaise (LSF) au moyen de personnages virtuels*. PhD Linguistique, Universit  de Provence - Aix-Marseille.
- MAEDER C. (1994). *Espace, temps et relations temporo-logiques chez le sujet sourd :  tude comparative de sujets sourds et entendants dans le maniement des marqueurs spatio-temporels en LSF et en franais*. PhD thesis. Universit  de Nancy II - U.F.R des Sciences du Langage, France.
- MORRISSEY S. (2008). *Data-driven Machine Translation for Sign Languages*. PhD thesis, Dublin City University, Dublin, Ireland.
- PUSTEJOVSKY P., CASTAÑO J., INGRIA R., SAURI R., GAIZAUSKAS R., SETZER A., KATZ G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg.
- SINTE A. (2015). *Le temps en langue des signes*. Presses universitaires de Rennes.
- TEISSEDRE C. (2012). *Analyse s mantique automatique des adverbiaux de localisation temporelle : application   la recherche d'information et   l'acquisition de connaissances*. PhD thesis. Universit  Paris X – Nanterre, France.
- WILMET M. (2003). *Grammaire critique du franais*. Duculot.
- WOLFE R., ETHIMIOUS E., GLAUERT J., HANKE T., MCDONALD J., SCHNEPP J. (2016). Special issue on recent advances in sign language translation and avatar technology. *Universal Access in the Information Society* 15, 485–486.

