Vers une segmentation multimodale : l'utilisation des traductions et des gloses en français comme aide à l'identification des unités discursives en LSF.

Analyse basée sur le corpus *LS-Colin*

Herandi GARCÍA CONTRERAS

Sous la direction de Lydia Mai HO-DAC et Jérémie SEGOUAT .

2025-2026
DÉPARTEMENT DES SCIENCES DU LANGAGE
MASTER 1 – Linguistique, Informatique et Technologies du Langage

Remerciements

Je tiens tout d'abord à adresser mes sincères remerciements à mes deux encadrants de mémoire, Lydia Mai Ho-Dac et Jérémie SEGOUAT pour leur temps et leur accompagnement tout au long de ce projet.

Je tiens également à remercier chaleureusement ma mère, mon père et ma sœur, qui, depuis le Mexique, m'ont accompagnée et continuent de m'accompagner, et qui me soutiennent inconditionnellement tous les jours. Sans eux, je ne serais pas aujourd'hui en train de poursuivre mes rêves si loin de chez moi.

Mes remerciements vont également à mes trois annotatrices, Amira, Mathilde et Marie, pour leur temps, leur disponibilité et leur implication. Sans leur aide et leur participation, ce projet n'aurait pas pu être mené à bien.

Enfin, je souhaite remercier les trois enseignants du Master LITL, Lydia Mai Ho-Dac, Ludovic Tanguy et Cécile Fabre, qui ont contribué à faire naître en moi une grande passion pour ce domaine.

« Vers une segmentation multimodale : l'utilisation de traductions et de gloses en français comme aide à la segmentation des unités discursives en LSF. *Analyse basée sur le corpus LS-Colin* »

1. Introduction.

Les langues des signes (LS) sont des langues naturelles qui utilisent une modalité visuo-gestuelle permettant de transmettre du sens grâce à la combinaison d'articulateurs manuels et d'éléments non manuels, tels que le visage et le corps. Malgré leur complexité, les technologies développées pour les LS sont en retard significatif par rapport aux langues orales, ce qui limite l'accès à l'information et renforce les inégalités existantes.

La segmentation en unités minimales de discours (EDU) est une tâche fondamentale pour inverser cette situation, car elle constitue la base d'applications telles que la traduction automatique, l'alignement des sous-titres, entre autres. Cependant, la nature simultanée des LS et l'absence d'une forme écrite standard rendent ce processus difficile, empêchant l'application des techniques traditionnelles de traitement du langage parlé.

Cette recherche utilise le corpus LS-Colin, une ressource de référence pour la langue des signes française (LSF), afin de proposer une méthodologie de segmentation discursive s'appuyant sur les traductions et les gloses en français. L'objectif est d'évaluer dans quelle mesure ces ressources textuelles peuvent constituer un support fiable pour l'identification des unités discursives de base (BDU) en LSF, aussi bien pour des annotateurs humains que pour des systèmes de segmentation automatique.

Pour répondre à cette question, trois axes complémentaires ont été développés. Dans un premier temps, une exploration manuelle de la segmentation discursive a été réalisée à partir du modèle BDU proposé par Gabarró-López et Meurant (2016). Dans un deuxième temps, une expérimentation inter-annotateurs a permis d'évaluer l'applicabilité d'un guide d'annotation élaboré pour cette recherche auprès de trois participants aux profils distincts. Enfin, une expérimentation avec le segmentateur automatique DisCut, développé par l'équipe MELODI de l'IRIT (Ezzabady et al., 2021), a conduit au développement d'un annotateur automatique et d'un pipeline complet d'annotation et de segmentation dans ELAN, permettant d'automatiser l'ensemble du processus depuis l'extraction des traductions jusqu'à la création d'une nouvelle couche d'annotation.

CHAPITRE I

2. TAL et la langue de signes.

2.1 La langue des signes.

Les langues des signes sont les principaux moyens de communication des personnes sourdes et malentendantes. Elles constituent le cœur des communautés sourdes. (Moryossef & Goldberg, 2021) Ce sont des langues qui utilisent la modalité visuelle-gestuelle pour transmettre du sens à travers des articulations manuelles combinées à des éléments non manuels tels que le visage et le corps. Tout comme les langues parlées, les langues des signes sont des langues naturelles régies par leurs propres règles linguistiques (Sandler et Lillo-Martin, 2006). De même, la complexité et la richesse des langues des signes sont les mêmes que celles des langues orales. (Bull et al., 2020)

Selon différentes sources, on estime qu'il existe environ 300 langues des signes dans le monde. Cependant, il n'existe pas de chiffre exact, car chaque pays possède sa propre langue des signes, avec des variations internes selon les régions. Certains linguistes affirment même qu'il pourrait exister une langue des signes différente pour chaque communauté sourde.

Malgré cette diversité, ces langues ont en commun une forte présence de l'iconicité. Les formes peuvent être représentées naturellement par des gestes, ce qui crée un lien fort entre la forme et le sens dans les langues des signes, lien qui est moins présent dans les langues orales (Hadjadj et al., 2018).

Le traitement automatique des langues (TAL) est un domaine de recherche pluridisciplinaire à l'intersection de la linguistique et de l'informatique, visant à développer des systèmes informatiques capables d'analyser et de traiter le langage humain. Ce domaine de recherche est apparu au début des années 1950, en même temps que la mise au point des premiers ordinateurs. Il s'est depuis largement diversifié et ces applications se sont imposées dans la vie courante, avec les moteurs de recherche, les outils de veilles, d'opinion et de traduction automatique, les correcteurs orthographiques, ou la commande à la voix des téléphones portables. (Poibeau, 2019).

Historiquement, les progrès du TAL se sont concentrés presque exclusivement sur les langues orales, qu'elles soient écrites ou parlées. Par conséquent, la grande majorité des modèles actuels nécessitent une entrée linguistique basée sur la voix ou le texte, ce qui exclut environ 300 langues des signes et, par conséquent, près de 70 millions de personnes sourdes de l'accès aux technologies linguistiques modernes (Yin et al., 2021).

Face à cette réalité, le Traitement automatique des langues de signes (TALS) a commencé à s'imposer comme un domaine émergent qui cherche à appliquer les principes du TAL aux langues visogestuelles. Cependant, jusqu'à présent, la recherche sur le TALS a été dominée par la communauté de la vision par ordinateur, avec peu de participation directe du TAL (Yin et al., 2021). Néanmoins, les langues des signes sont des systèmes linguistiques complets qui présentent toutes les propriétés définissant les langues naturelles. L'intégration des théories et des techniques du TAL est donc fondamentale pour leur traitement informatique (Yin et al., 2021).

2.1.1 Les tâches fondamentales du TALS

Dans l'article de Yin, « *Including Signed Languages in Natural Language Processing* » (2021), il est mentionné qu'actuellement, le domaine du TALS (Traitement automatique des langues de signes) est organisé en différentes tâches techniques. L'une d'entre elles est la détection, dont l'objectif est d'identifier si un fragment de vidéo contient l'utilisation de signes. Bien que les modèles récents aient atteint des niveaux de performance élevés, il existe encore un manque de données correctement annotées permettant leur évaluation rigoureuse (Borg & Camilleri, 2019 ; Moryossef et al., 2020, cité par Yin et al., 2021). À cela

s’ajoute un manque de standardisation, ainsi que l’absence d’un système de codage stabilisé et officiel permettant de représenter numériquement les signes.

Contrairement aux langues vocales et écrites, les langues des signes ne disposent pas d’un système d’écriture ou de représentation universellement stabilisé. Il n’existe notamment pas d’équivalent direct à l’Alphabet phonétique international (API) pour la représentation phonétique, ni à un alphabet ou à un système logographique pour les langues écrites. Il n’existe donc pas, à ce jour, de moyen de transcription standardisé permettant de représenter les données signées dans un format directement adapté au traitement automatique des langues (TAL).

Une autre tâche importante concerne l’identification de la langue des signes, qui consiste à déterminer automatiquement quelle langue des signes est utilisée dans une séquence donnée (Gebre et al., 2013).

La segmentation constitue un autre axe central du domaine, dont l'objectif est d'identifier les limites entre les signes ou les phrases dans la vidéo. La plus part des approches existantes n'exploitent pas les indicateurs linguistiques fiables, tels que les traits prosodiques de la langue des signes : pauses, durée, expressions faciales ou ouverture oculaire (Sandler, 2010 ; Ormel & Crasborn, 2012, cité par Yin et al., 2021).

La traduction automatique de la langue des signes vers les langues orales a progressé grâce à des systèmes basés sur des annotations ou des estimations de la posture corporelle, mais elle reste confrontée à des défis importants, notamment la représentation des référents spatiaux et des relations discursives, qui sont essentiels pour la signification dans les langues visogestuelles (Yin et al., 2021). Enfin, la production automatique de signes à partir du langage oral en est à un stade de développement initial et s'appuie généralement sur des représentations intermédiaires basées sur les postures et l'animation numérique. (Yin et al., 2021).

2.2. Enjeux et importance du TALS

Dans un contexte de plus en plus numérisé, la recherche en traitement du langage naturel devrait contribuer à la création de technologies linguistiques accessibles à tous, y compris aux personnes sourdes, en garantissant leur droit de communiquer dans des langues qui reflètent leur expérience de vie et leur identité linguistique (Moryossef & Goldberg, 2021). L'exclusion des langues des signes des technologies modernes restreint non seulement l'accès à l'information, mais renforce également les inégalités sociales et communicatives qui existent depuis toujours (Padden & Humphries, 1988 ; Glickman & Hall, 2018). Il est donc essentiel de promouvoir le développement de technologies qui rendent les langues des signes accessibles et qui répondent de manière adéquate aux besoins réels des communautés qui les utilisent.

L’étude pour le traitement automatique des langues de signes peuvent donner lieu à de nombreuses applications bénéfiques, notamment une meilleure documentation des langues des signes en voie de disparition, des outils éducatifs pour les étudiants en langue des signes, des systèmes de consultation et de récupération d'informations dans des vidéos signées, des assistants personnels capables de répondre aux signes et des systèmes d'interprétation automatique en temps réel, entre autres possibilités (Yin et al., 2021).

De plus, si l'on considère le TALS d'un point de vue scientifique, travailler avec les langues des signes permet d'obtenir une compréhension plus complete des langues naturelles, en offrant une vision approfondie de la manière dont le sens est construit et transmis par la modalité visuelle et de la manière dont le langage s'appuie sur des concepts visuo-spatiaux (Yin et al., 2021). En ce sens, le traitement automatique de la langage des signes constitue non seulement un domaine intellectuellement stimulant, mais aussi un champ de recherche à fort potentiel d'impact social, capable de bénéficier directement aux communautés qui utilisent ces langues (Yin et al., 2021).

3. Principaux défis du TALS

3.1. Pénurie de données et de ressources pour le TALS

Les données sont essentielles au développement de tout outil de base de traitement du langage naturel, et les efforts actuels en matière de traitement des langues de signes sont souvent limités pour réaliser des tâches telles que celles décrites ci-dessus en raison du manque de données adéquates. Pour que les modèles de TALS puissent être mis en œuvre efficacement, ils doivent être développés à partir de données qui représentent fidèlement l'utilisation de la langue dans des contextes réels. Yin et Read (2020b), par exemple, soulignent qu'un ensemble de données idéal en langue des signes devrait inclure : un domaine étendu ; des données et un vocabulaire suffisants ; des conditions réelles ; une production naturelle des signes ; une démographie diversifiée des signants ; la participation de signants natifs ; et, le cas échéant, des annotations denses.

Cependant, la collecte et l'annotation de données en langue des signes qui se rapprochent de cet idéal nécessitent beaucoup de ressources, car il faut parfois jusqu'à 600 minutes de travail pour chaque minute de vidéo annotée, ce qui représente un investissement considérablement plus important que pour les données vocales ou textuelles (Hanke et al., 2020). En outre, le processus d'annotation nécessite souvent des connaissances spécialisées, ce qui rend difficile le recrutement et la formation d'annotateurs qualifiés. À cela s'ajoute la rareté des données en langue des signes librement accessibles, en particulier celles provenant de signants natifs et qui ne correspondent pas à des interprétations de la parole. Par conséquent, la collecte de données implique souvent un effort logistique et un coût économique élevés, car elle nécessite des processus d'enregistrement sur place.

Cette pénurie de données est également liée à des difficultés plus fondamentales concernant leur représentation et leur annotation numériques. En effet, l'absence d'un système de codage stabilisé et largement accepté pour représenter les langues des signes constitue un obstacle supplémentaire à la constitution de ressources numériques. Contrairement aux langues vocales, pour lesquelles il existe des conventions relativement stabilisées permettant de représenter certains aspects de la langue, les langues des signes ne disposent pas d'un système standardisé permettant de représenter de manière univoque l'ensemble des caractéristiques pertinentes du signe. Cette situation est également liée à un manque de consensus sur la définition des unités de sens à représenter, ainsi que sur les aspects du signe qui doivent être pris en compte et sur la manière de les formaliser. Il peut ainsi être difficile de déterminer quelles unités doivent faire l'objet d'une annotation et selon quelles conventions elles doivent être représentées.

Pour remédier à ce problème, Yin et al. (2021) proposent que pour accroître la disponibilité des données et faciliter le développement de modèles TALS applicables, il faudrait créer des outils qui simplifient ou automatisent partiellement les processus de collecte et d'annotation.

À cet égard, Yin et al. (2021) soutiennent que des outils capables d'effectuer des analyses automatiques, de détecter les limites des images, d'extraire des caractéristiques articulatoires, de suggérer des annotations lexicales, entre autres fonctions, permettraient d'externaliser une partie du processus d'annotation à des personnes non expertes et d'aider les linguistes à définir des unités de sens pertinentes. Ces technologies auraient donc un grand potentiel pour faciliter et accélérer la disponibilité de données de qualité.

3.2. Spécificités linguistiques des langues des signes

Les langues des signes posent de nouveaux défis pour le traitement du langage naturel en raison de leur modalité visuo-gestuelle, de leur caractère simultané, de leur organisation spatiale et de l'absence d'une forme écrite standardisée.

L'absence d'écriture rend incompatibles avec les langues des signes les procédures classiques du traitement du langage parlé, qui commencent généralement par la transcription du signal audio avant l'application des

techniques de TAL. En effet, les langues des signes ne disposent pas d'un format textuel pouvant servir d'entrée à ces outils, ce qui oblige à travailler directement à partir du signal vidéo brut (Moryossef & Goldberg, 2021).

De plus, si nous cherchions à traduire les langues des signes afin que les techniques de TAL puissent être appliquées, les signes totalement lexicaux ne représentent qu'une partie du discours et beaucoup d'entre eux n'ont pas d'équivalent direct dans la langue écrite (Bull et al., 2020). D'autres, appelés signes partiellement lexicaux, découlent de l'iconicité propre aux langues des signes et englobent des éléments tels que les référents spatiaux, les mouvements et la représentation de la taille ou de la forme des objets (Braffort & Filhol, 2014). Ces signes dépendent dans une large mesure du contexte communicatif. En outre, les signes peuvent être produits simultanément : le discours en langue des signes n'est pas une séquence strictement linéaire d'unités, mais peut impliquer plusieurs canaux articulatoires actifs en même temps (Filhol et al., 2014).

Un exemple illustratif proposé par Bull et al., 2020 est l'expression « la grande moustache » en LSF. Bien qu'il existe un signe pour « grand », ce signe isolé n'apparaît pas dans cette phrase. Au lieu de cela, la notion de taille est exprimée simultanément et contextualisée en modifiant le signe « moustache » lui-même, intégrant ainsi l'information de grandeur dans le geste principal

Ainsi, si la forme la plus canonique de diviser un texte en langue parlée est une séquence linéaire de mots, en raison de la simultanéité du langage des signes, le concept de « mot » dans le langage des signes est mal défini et ce langage ne peut être modélisé de manière totalement linéaire (Moryossef & Goldberg, 2021).

4. Segmentation du discours en LSF

4.1.Pertinence de l'étude

Tout traitement automatique du langage nécessite la définition préalable d'unités d'analyse. Cependant, dans le cas des langues des signes, l'unité minimale n'a pas encore été définie de façon opérationnelle, il n'existe pas d'unité équivalente au *token* ou à la phrase. En linguistique informatique, un token est une unité élémentaire de texte utilisée par les systèmes de traitement automatique du langage naturel. Il peut correspondre à un mot, une sous-partie de mot ou un caractère, selon la méthode de tokenisation utilisée (Stanford Natural Language Processing Group, s. f.). Cette absence d'unités de base clairement délimitées constitue un défi fondamental pour l'analyse linguistique et pour le développement d'outils automatiques appliqués aux langues des signes.

Face à cette difficulté, il est pertinent d'explorer la possibilité de segmenter le discours en langue des signes en segments possédant une certaine autonomie discursive, afin de pouvoir définir des unités d'analyse opérationnelles.

Yin et al. dans son article *Including Signed Languages in Natural Language Processing* (2021), soulignent que progresser vers une inclusion réelle des langues des signes en TAL requiert des modèles intégrant des spécificités linguistiques propres à ces langues. Dans ce cadre, cette étude porte sur la segmentation discursive ; une tâche clé dont dépendent de nombreuses applications avancées du traitement automatique des langues (Li, 2019).

Si la segmentation discursive a été largement étudiée dans les langues orales et écrites, son application aux langues des signes reste un domaine encore peu exploré. Cette limitation est renforcée à la fois par la rareté des corpus annotés et par la complexité méthodologique qu'implique le travail direct avec des données visuelles multimodales.

Dans ce contexte, le présent travail propose d'explorer la possibilité de segmenter la langue des signes en unités discursives et d'observer quels éléments permettent d'identifier les frontières entre ces unités .

L'analyse se concentre à la fois sur les éléments visuels présents dans la vidéo en langue des signes et sur les annotations disponibles, en particulier les gloses et les traductions.

Le choix de travailler à partir de traductions et de gloses répond à un double besoin, méthodologique et pratique. D'une part, comme le souligne Yin (2021), il est essentiel de développer des outils qui facilitent l'analyse et l'annotation des données en langue des signes, même pour les chercheurs qui ne maîtrisent pas pleinement cette langue. D'autre part, les systèmes actuels de traitement du langage naturel fonctionnent principalement sur des entrées textuelles, ce qui fait des traductions et des gloses en français un pont important entre les données signées et les outils informatiques existants.

N'étant pas locutrice de la langue des signes française (LSF), ce travail adopte une approche exploratoire. L'objectif est d'évaluer si les méthodes de segmentation développées pour les langues orales peuvent être adaptées aux langues des signes, et si les transcriptions et annotations peuvent servir de guide valable pour la segmentation, complétées par les indices visuels observables dans le matériel vidéo.

4.2. Approches existantes.

À ce jour, peu de travaux traitent directement sur la segmentation la langue des signes en unités discursives ce qui confirme qu'il s'agit d'un domaine de recherche encore émergent.

L'un des travaux les plus pertinents dans ce domaine est celui de Gabarró-López et Meurant (2016), qui analysent explicitement la problématique de la segmentation du discours signé et proposent une théorie adapté à cette modalité. Dans leur article *Slicing Your SL Data into Basic Discourse Units (BDUs): Adapting the BDU Model (Syntax + Prosody) to Signed Discourse*, les auteurs adaptent la théorie des Basic Discourse Units (BDU), initialement conçu pour les langues orales, aux langues des signes à l'aide d'une procédure en trois étapes : segmentation syntaxique basée sur la grammaire de dépendance, segmentation prosodique utilisant des signaux visuels équivalents aux signaux acoustiques (pauses, changements de rythme, expressions non manuelles), et identification des points de convergence entre les deux segmentations afin d'établir les unités discursives finales. Ce travail constitue l'une des rares tentatives de segmentation du discours dans les langues des signes ,et sert de référence pour les critères de délimitation des unités qui seront utilisés dans cette recherche.

Dans le domaine de la segmentation discursive automatique, Ezzabady et al. (2021) présentent MELODI, un système développé dans le cadre de la tâche partagée DISRPT 2021 pour la segmentation discursive multilingue et l'identification des connecteurs. Leur approche permet d'identifier automatiquement les frontières des unités discursives, à la fois au niveau de la phrase et à l'intérieur de celle-ci, en exploitant des données provenant de plusieurs langues. Ce type d'approche montre les possibilités de l'automatisation de la segmentation discursive et constitue un point de comparaison pertinent pour le présent travail

Pour leur part, Tofiloski, Brooke et Taboada (2009), dans *A Syntactic and Lexical-Based Discourse Segmenter*, présentent SLSeg, un segmentateur automatique d'EDU dans les textes écrits basé sur des règles linguistiques conçues manuellement à partir d'informations syntaxiques et d'indices lexicaux. Plus récemment, Li (2021) propose SEGBOT, un système automatique basé sur des réseaux neuronaux pour diviser les textes en EDU. Le modèle ne nécessite pas de règles manuelles ni de caractéristiques conçues à la main et surpasse les méthodes traditionnelles, démontrant que la segmentation automatique peut être effectuée de manière efficace et généralisable. Cependant, à ma connaissance, il n'existe à ce jour aucun segmentateur automatique de langue des signes.

Enfin, le Reference Manual for the Analysis and Annotation of Rhetorical Structure (Reese, Hunter, Asher, Denis et Baldridge, 2007), développé dans le cadre du projet DISCOR, détaille les critères opérationnels permettant de définir les unités discursives élémentaires (EDU) et de les relier entre elles à l'aide d'un ensemble limité de relations discursives suivant la SDRT. Pour le présent travail, ce manuel servira de guide méthodologique pour diviser les transcriptions du corpus.

5. Cadre théorique et méthodologie

La segmentation du discours ne constitue pas un domaine homogène d'un point de vue théorique ou méthodologique. Comme le soulignent divers auteurs, il existe de multiples façons de définir et de délimiter une unité discursive. Simon et Degand (2011) expliquent que tout fragment de discours se compose d'unités plus petites s'articulant de manière cohérente ; toutefois, la définition de ces unités constitutives varie considérablement d'un modèle à l'autre.

En ce sens, différents courants théoriques ont proposé diverses appellations pour désigner des unités comparables, telles que : les « énoncés » (Reboul & Moeschler, 1998), les *Elementary Discourse Units* (Mann & Thompson, 1988), les unités élémentaires du discours (Asher & Lascarides, 2003), les actes discursifs (Roulet et al., 2001), les actes discursifs de base (Steen, 2005), les unités minimales du discours (Degand & Simon, 2005, 2008 ; Hannay & Kroon, 2005) ou encore les unités discursives de base (Degand & Simon, 2009).

Dans le domaine du traitement automatique des langues (TAL), l'un des cadres les plus utilisés pour identifier et lier ces unités dans les textes écrits est la *Rhetorical Structure Theory* (RST), proposée par Mann et Thompson (1988). La RST modélise le texte comme une structure hiérarchique en arbre, dont les nœuds terminaux correspondent aux unités discursives élémentaires (EDU). D'autres méthodologies sont également employées pour la segmentation de productions écrites, à l'instar du modèle de Bâle (Ferrari, 2005 ; Ferrari et al., 2008).

Cependant, les données en langue des signes (LS) relevant d'une modalité gestuelle-visuelle et non de l'écrit, ces méthodologies ne sont adaptées que pour la segmentation des traductions textuelles. Pour le discours oral, Gabarró-López et Meurant (2016) identifient au moins six modèles principaux : le modèle de Genève (Roulet et al., 1985), le modèle Val.Es.Co. (Briz Gómez et al., 2003), le modèle de Fribourg (Groupe de Fribourg, 2012), le modèle de co-énonciation (Morel & Danon-Boileau, 1998), le modèle de démarcation par proéminence (Lombardi Vallauri, 2009) et le modèle des unités discursives de base (BDU) (Degand & Simon, 2005, 2009).

Ces modèles diffèrent par leur conception de l'unité et par leurs critères de délimitation. Certains privilégient une approche prosodique (pauses, unités tonales), comme le modèle de co-énonciation. D'autres s'appuient sur des critères pragmatiques (force illocutive, organisation interactive), comme les modèles de Genève ou Val.Es.Co. Enfin, le modèle BDU se distingue par l'intégration combinée de critères syntaxiques et prosodiques, ce qui lui confère une plus grande flexibilité descriptive (Gabarró-López & Meurant, 2016).

En ce qui concerne la segmentation de la langue des signes, la question s'avère encore plus complexe, faute de consensus établi. Bien que certaines études aient tenté de segmenter les discours signés en « phrases » à partir de repères visuels (Börstell et al., 2014 ; Fenlon et al., 2007 ; Jantunen, 2007), cette notion s'avère insuffisante pour rendre compte de la complexité structurelle du discours spontané.

Face à ce constat, Gabarró-López et Meurant (2016) proposent d'adapter le modèle BDU à l'analyse des LS. Son principal avantage réside dans sa polyvalence, tout en offrant un cadre plus systématique et moins dépendant d'interprétations subjectives. Dans ce cadre, les BDU sont conçues comme des « segments minimaux d'interprétation discursive » (Degand & Simon, 2009a).

5.1. Modèle théorique pour segmenter la langue des signes en unités discursives

Comme indiqué dans les travaux précédents, cette étude se base sur les propositions de Gabarró-López et Meurant (2016) dans leur article *Slicing Your SL Data into Basic Discourse Units (BDUs): Adapting the BDU Model (Syntax + Prosody) to Signed Discourse*. Dans cet article, les auteurs adaptent le modèle des *Basic*

Discourse Units (BDU), initialement conçu pour les langues orales, aux langues des signes, en suivant trois étapes principales : (1) Segmentation syntaxique, basée sur la Grammaire de Dépendances, (2) Segmentation prosodique, utilisant des signaux visuels équivalents aux signaux acoustiques (pauses, changements de rythme, expressions non manuelles). (3) Identification des points de convergence entre les deux segmentations, afin de déterminer les unités discursives finales.

5.1.1Segmentation syntaxique :

Le modèle BDU délimite les unités syntaxiques, appelées clauses, en s'appuyant sur le cadre de la Grammaire de Dépendance (GD), tel qu'il a été conçu pour le français parlé par Blanche-Benveniste et al. (1984).

Dans la perspective de la GD, le verbe constitue le noyau de la clause. Il convient toutefois de noter que ce rôle de noyau n'est pas exclusivement réservé au verbe ; il peut également être assuré par d'autres éléments tels que des pronoms, des noms ou des adjectifs. En suivant le modèle BDU, on distingue généralement cinq types de clauses selon leur structure de dépendance. (Gabarró-López & Meurant, 2016).

Il est important de préciser une distinction méthodologique : alors que dans le guide de Gabarró-López et Meurant (2016), la segmentation s'effectue directement sur le discours en langue des signes (LS), le présent travail s'appuie sur les gloses et les traductions pour réaliser la segmentation syntaxique.

En langue des signes, les gloses constituent un système de notation conventionnel permettant de représenter les signes et certains aspects de la structure du discours signé à l'aide de mots issus d'une langue vocale, généralement écrits en majuscules. Elles ne correspondent pas à une forme écrite de la langue des signes, mais servent principalement d'outil de transcription et d'analyse pour les chercheurs. Chaque glose renvoie généralement à un signe ou à une unité de sens, tandis que différentes conventions typographiques peuvent être utilisées pour rendre compte de certaines propriétés du discours signé. L'ordre des gloses vise notamment à respecter l'organisation syntaxique de la langue des signes, qui peut différer de celle de la langue vocale utilisée pour les noter.

La vidéo utilisée pour la réalisation de ce travail est disponible à l'adresse suivante : https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82.

1) Clauses à dépendance verbale

Comme l'indique leur nom, ce type de clause présente un verbe comme noyau.

(**1) Exemple:**

TEMPORALITÉ : 00:19.04 – 00:19.05
GLOSES : [ENVIE SAUTER BARRIÈRE]
TRADUCTION :« Le cheval a envie de sauter »

2) Clauses à dépendance non verbale

Bien que le verbe soit généralement considéré comme le noyau prototypique de la clause, il existe des productions où d'autres éléments constituent eux-mêmes une clause à dépendance non verbale. Cela se produit, par exemple, lorsqu'un locuteur répond simplement "oui" à une question. Dans les langues des signes, des signes à fonction pronominale, nominale ou adjectivale peuvent également jouer le rôle de noyau.

(**2) Exemples :**

TEMPORALITÉ : 00:00:04 – 00:00:07
GLOSES : [HISTOIRE CHEVAL]
TRADUCTION : « C'est l'histoire… d'un cheval. »

TEMPORALITÉ : 00:05:51 – 00:05:53
GLOSES : [CHOU ROUGE]
TRADUCTION : « C'est le chou rouge. »

3) Clauses à dépendance elliptique

Aucun exemple de ce type de clause n'a été trouvé dans le corpus. Toutefois, le guide mentionne qu'une clause est considérée elliptique lorsqu'elle présente une structure incomplète mais peut être interprétée comme une clause à dépendance verbale grâce au contexte discursif (Tanguy et al., 2012). Les réponses brèves et les interventions concises en interaction appartiennent généralement à cette catégorie.

4)Clauses à dépendance interrompue

Cette catégorie n'est pas représentée dans le corpus, mais inclut toutes les clauses verbales, non verbales ou elliptiques qui ont été interrompues par divers facteurs, tels que l'introduction d'une nouvelle idée, l'intervention de l'interlocuteur ou tout autre élément contextuel.

5) Clauses contenant un élément non dépendant

Ce type de clause comprend les cas où certains éléments, appelés adjoints, restent en dehors de la structure de dépendance verbale. Parmi eux se trouvent les marqueurs discursifs ou connecteurs logiques, qui peuvent constituer à eux seuls une clause indépendante.

(**3) Exemples :**

TEMPORALITÉ: 00:06:18 – 00:06:18
GLOSES : **[**APRES (esq)**] [** ouvrir la cocotte **]**
TRADUCTION : « Ensuite on ouvre la cocotte »

TEMPORALITÉ : 00:06:58 – 00:06:58
GLOSES : **[**ALORS**] [** enlever le couvercle**]**
TRADUCTION : « Puis ouvrir la cocotte, »

5.1.2 Segmentation prosodique :

Dans le modèle BDU, la délimitation prosodique des unités discursives s'appuie sur trois signaux acoustiques principaux : les pauses silencieuses, l'allongement syllabique (lorsqu'une syllabe est trois fois plus longue que les syllabes environnantes) et une augmentation brutale de la fréquence fondamentale (f0), supérieure à dix demi-tons au sein de la syllabe. (Gabarró-López & Meurant, 2016)

1) Pauses

Le premier indicateur prosodique adapté à la modalité signée correspondrait aux pauses, définies comme des périodes d'absence totale de production manuelle dans le flux du discours (Fenlon, 2010). Ces pauses comprennent à la fois les interruptions pendant lesquelles les mains restent croisées, détendues le long du corps, et celles pendant lesquelles elles se trouvent dans l'espace neutre (Notarrigo & Meurant, 2014).

2) Signes allongés

Le deuxième indicateur prosodique considéré correspond aux sign holds ou signes allongés, équivalents fonctionnels de l'allongement syllabique dans la langue parlée. Un *sign hold* se produit lorsque la configuration manuelle d'un signe est maintenue pendant une période plus longue que prévu, tandis qu'un signe allongé peut se manifester par un ralentissement, une répétition ou une exagération du mouvement. Bien que ce type d'allongement puisse apparaître au début, au milieu ou à la fin du signe (Notarrigo & Meurant, 2014), seuls les cas qui se produisent en position finale sont pris en compte à des fins de segmentation (Gabarró-López & Meurant, 2016).

3) Clignement des yeux

Enfin, le troisième indicateur prosodique visuel pris en compte est le clignement des yeux (*eye blinks*), intégré comme signal complémentaire car il contribue à segmenter le discours en unités rythmiques de manière comparable au rôle joué par les ascensions de f0 dans la langue parlée (Pfau & Quer, 2010 ; Herrmann, 2012). Cependant, tous les clignements des yeux ne sont pas pertinents sur le plan prosodique, car ils peuvent remplir d'autres fonctions linguistiques ou répondre à des facteurs physiologiques (Wilbur, 1994 ; Sze, 2008 ; Herrmann, 2012). C'est pourquoi l'identification des clignements prosodiques est limitée à ceux qui coïncident avec un autre marqueur prosodique non manuel (mouvement des sourcils, ouverture des yeux, regard, mouvement de la tête, mouvement du corps, gestes de la bouche et expressions faciales) car cette combinaison constitue l'un des indicateurs les plus fréquents de la frontière d'une unité discursive, après les pauses et les *sign holds* (Gabarró-López & Meurant, 2016).

5.1.3 Segmentation BDU :

La segmentation BDU est effectuée lorsque les limites des segmentations prosodiques et syntaxiques convergent.

5.2. Segmentation des traductions en français de la LSF

La section précédente, *5.1. Modèle théorique pour segmenter la langue des signes en unités discursives* , constitue une adaptation de la théorie proposée par Gabarro López et Meurant pour la segmentation de la langue des signes en BDU. La principale différence entre leur travail et celui présenté ici réside dans la démarche de segmentation : dans leur étude, la segmentation syntaxique était réalisée directement sur la vidéo signée afin de permettre ensuite l'identification des BDU. Dans le présent travail, nous avons utilisé les gloses pour réaliser la segmentation syntaxique.

Cependant, afin d'identifier le plus grand nombre possible d'éléments et d'indices susceptibles d'indiquer une frontière entre unités discursives, il m'a semblé pertinent de réaliser également une segmentation directement sur les traductions en français. Cette segmentation permettrait ainsi de disposer d'un maximum d'indices et a faciliter, dans la suite de l'analyse, l'identification des frontières syntaxiques et des BDU dans les données en LSF.

Ainsi, pour la segmentation discursive des traductions en français, le *Reference Manual for the Analysis and Annotation of Rhetorical Structure* (Reese et al., 2007) a été adopté comme cadre méthodologique. Ce manuel propose d'identifier des *Elementary Discourse Units* (EDU), comprises comme des unités minimales du discours remplissant une fonction communicative reconnaissable.

La segmentation a été effectuée manuellement et marquée par le symbole <***>, conformément aux indications du *Reference Manual for the Analysis and Annotation of Rhetorical Structure* (Reese et al., 2007).

Les exemples suivants sont des extraits des traductions de « Histoire du cheval » (00:00:04–00:00:54) et de « Recette de cuisine » (00:05:45–00:07:06), disponibles à ce lien : https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82.

1. Segmentation de base.

Nous avons d'abord commencé par la segmentation de base. Le manuel indique qu'une limite de segment est placée après les signes de ponctuation qui marquent la fin d'une phrase (point, point d'interrogation, point d'exclamation).

(4) Exemple :

> ***Extrait de la traduction de « Histoire du cheval » (00:00:04–00:00:54) :*** *« C'est l'histoire… d'un cheval.<***> Il galope, et soudain, il freine, surpris par la barrière.<***> Il aperçoit une vache qui rumine, ils se regardent.<***> »*

Chaque phrase constitue une unité discursive indépendante, car elle introduit un nouvel événement dans la progression narrative.

2. Segmentation interne de la phrase

Comme le proposent Reese et al. (2007), les clauses qui remplissent une fonction discursive discernable, en particulier lorsqu'elles introduisent de nouveaux événements, ont été segmentées au sein d'une même phrase.

(5) Exemple :

> ***Extrait de la traduction de « Histoire du cheval » (00:00:04–00:00:54):*** *« Il recule, prend son élan et saute,<***> mais il heurte la barrière et tombe,<***> et se retrouve les pattes étendues .<***> »*

Ici, les actions successives du cheval ont été segmentées car chacune correspond à un événement narratif distinct et contribue à la progression temporelle du récit.

3. Coordination et succession d'événements

Dans les textes narratifs et procéduraux, la coordination des verbes introduit généralement plusieurs actions consécutives. Conformément au manuel, les verbes coordonnés ont été segmentés lorsque chacun apporte une action pertinente.

(6) Exemple :

> ***Extrait de la traduction de Traduction « Recette de cuisine » (00:05:45–00:07:06):*** *« Ensuite on ouvre la cocotte,<***> et on enlève le couvercle qu'on pose à côté,<***> c 'est à peu près à moitié cuit, <***> »*

Bien qu'elles fassent partie d'une même séquence procédurale, chaque action correspond à une étape distincte de la procédure et a donc été considérée comme une EDU indépendante.

4. Clauses subordonnées ayant une valeur discursive.

Les subordonnées ont été segmentées lorsqu'elles introduisent des informations pertinentes pour la progression du discours, en particulier lorsqu'elles expriment une cause, une conséquence ou une explication.

(7) Exemple :

> ***Extrait de la traduction de « Histoire du cheval » (00:00:04–00:00:54) :*** *« Il a mal, il souffre, <***> la vache regarde derrière elle et voit un oiseau qui est perché sur la barrière,<***> »*

La proposition relative *qui est perché sur la barrière* apporte de nouvelles informations nécessaires à la compréhension de la scène suivante, elle est donc conservée dans le même segment, sans créer de discontinuité.

5. Particularités du texte procédural (recette)

Dans le cas de la recette, le discours présente une structure hautement séquentielle et directive. C'est pourquoi la plupart des verbes à l'impératif ou des constructions instructives ont été traités comme des EDU indépendants.

(8) Exemple :

> ***Extrait de la traduction de Traduction « Recette de cuisine » (00:05:45–00:07:06) :*** *« Mettre le tout dans une cocotte minute.<***> mettre le couvercle et fermer la cocotte minute.<***> Mettre le sifflet (ce qui tourne et fait de la vapeur)<***> Faire cuire jusqu'à ce que le sifflet émette la vapeur et tourne.<***> »*

Ce type de segmentation permet d'observer clairement la correspondance entre chaque instruction verbale et les unités exprimées par la suite en langue des signes.

6. Marqueurs discursifs et pauses

Certains marqueurs discursifs présents dans le corpus oral, tels que *Alors*, *Ensuite*, *Ah !*, *Voilà*, ainsi que les pauses et les reformulations, ont également été considérés comme des indices de segmentation.

(9) Exemple :

> ***Extrait de la traduction de « Recette de cuisine » (00:05:45–00:07:06) :*** *Ah ! <***>Je coupe le feu,<***> le feu s'éteint,<***> et j'enlève le sifflet.<***>*

Ces marqueurs remplissent une fonction d'organisation du discours et reflètent des transitions cognitives pertinentes pour l'analyse multimodale ultérieure.

6. Le corpus LS-Colin

6.1. Contenu

Pour l'analyse de la segmentation discursive en langue des signes, cette étude s'appuie sur le corpus LS-Colin, constitué dans le cadre du projet de recherche intitulé *Langues des signes : analyseurs privilégiés de la faculté de langage ; apports croisés d'études linguistiques, cognitives et informatiques (traitement et analyse d'image) autour de l'iconicité et de l'utilisation de l'espace* (projet n° LACO 39), (Cuxac, 2002). Ce corpus constitue aujourd'hui une référence pour l'étude de la langue des signes française (LSF).

Le projet, d'une durée de 24 mois, était placé sous la responsabilité scientifique de **Christian Cuxac**, professeur en Sciences du langage à l'Université Paris 8 (UMR 7023). Il a été mené par une équipe interdisciplinaire composée de chercheurs en sciences du langage, en sciences informatiques, ainsi que d'enseignants et d'étudiants issus de ces deux domaines. Plusieurs équipes partenaires ont contribué à sa réalisation, notamment le laboratoire LaLIC (Paris IV), le LIMSI/CNRS et l'IRIT (Université Paul Sabatier) (Cuxac, 2002).

Selon le rapport final (Cuxac, 2002), le projet visait à : Mettre en évidence l'iconicité propre à la LSF, en particulier sa grammaire spatiale. Fournir des données de haute qualité aux chercheurs en linguistique. Offrir du matériel utile aux spécialistes du traitement de l'image et des technologies du langage, compte tenu de l'absence de forme écrite standardisée pour la LSF.

Afin de refléter la richesse de la langue, le corpus intègre différents genres discursifs. Parmi les types de productions, on trouve :

- Des récits basés sur des images (tels que les histoires du *Cheval* et de l'*Oiseau*),
- Des discours argumentatifs,
- Des interactions spontanées,
- Des explications thématiques ou expositives.

Les documents ont été produits par 13 signants aux profils variés (âge, sexe, profession et origine régionale), ce qui garantit la diversité linguistique. Au total, le corpus rassemble environ 6 heures d'enregistrement, filmées avec 3 caméras simultanées.

Afin de privilégier le caractère naturel du discours, la dynamique suivante a été adoptée : chaque signant préparait son intervention pendant environ 20 minutes, puis était filmé en interaction avec un interlocuteur. Les tâches qui leur ont été demandées étaient les suivantes :

- Racontez l'histoire en images du *Cheval*.
- Racontez l'histoire en images de l'*Oiseau*.
- Exprimez-vous sur l'un des thèmes suivants :
    - - le passage à l'euro,
    - - ou les événements du 11 septembre 2001.
- Expliquer leur recette de cuisine préférée.
- Pour ceux qui avaient étudié la linguistique à Paris 8, expliquer un thème du programme comme s'il s'agissait d'un cours.
- Racontez à nouveau l'histoire du *Cheval*.

6.2. Annotation, segmentation et traduction

Les transcriptions, annotations et traductions du corpus LS-Colin ont été réalisées par des participants disposant d'une expérience préalable dans la transcription et la traduction de langues des signes. Afin d'établir des critères de segmentation cohérents, les auteurs ont comparé les transcriptions produites par plusieurs annotateurs pour une même séquence et ont pris en compte les choix méthodologiques de chacun (Cuxac, 2002).

2.4 Grille d'analyse (valable pour un monologue)

Niveaux			Définitions	Indices de segmentation
Discours			interaction/thème	changement de thème
Enoncés			cohérence sémantique et syntaxique	changement d'actant, de plan, suppression d'actant, clignement de paupières, cht direction regard
Mimique			point de vue du signeur/discours	croisement regard/ « oui » x3 dictum (le fait) / modus (comment c'est dit)
unités signifiantes	Signe	**Syntagme**	cohérence syntaxique d'un groupe de morphèmes	clignement de paupières (pas toujours) pic de tension musculaire sur noyau prédicatif
		Morphème	pas de cohérence syntaxique : signes standards	pauses
	paramètres	**Morphèmes liés**	niveaux des paramètres. ex: emplacement tête = activité cérébrale	
unités non signifiantes	**Phonétique**		description articulatoire des paramètres	modification de la configuration, changement d'emplacement
	Prosodie		rythme: long, bref, répétitions, tension	

Tableau 1: Guide d'analyse avec les différents indices de segmentation qui ont été utilisés.

6.3. Données analysées dans ce travail Annotation.

La vidéo utilisée pour la réalisation de ce travail est disponible à l'adresse suivante : https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82.

Deux extraits distincts ont été analysés :

l'histoire du *Cheval* : du second 00:00:04 au second 00:00:54.

Recette de cuisine préférée : de la minute 00:05:45 à la minute 00:07:06.

6.4. Disponibilité réelle du corpus et ses limites

Dans la pratique, lorsque l'on recherche le corpus en ligne, celui-ci est disponible sur la plateforme COCOON Huma (https://cocoon.huma-num.fr) dont l'organisation n'est pas tout à fait claire et peut rendre difficile la localisation de l'ensemble des vidéos disponibles. Une fois les ressources identifiées, on constate que les fichiers peuvent être téléchargés aussi bien au format vidéo qu'au format d'annotation. Les annotations sont disponibles dans des formats tels que XML/Pangloss, avec des options de conversion vers XML/TEI, ainsi que dans le format de conservation XML/ELAN. Quant aux vidéos, elles sont au format MP4, tant pour la conservation que pour la diffusion.

Cependant, en examinant le contenu des annotations, on constate que toutes les rubriques décrites dans le rapport du corpus (*Langues des signes : analyseurs privilégiés de la faculté de langage ; apports croisés d'études linguistiques, cognitives et informatiques (traitement et analyse d'image) autour de l'iconicité et de l'utilisation de l'espace (projet n° LACO 39), (Cuxac, 2002).*) ne sont pas effectivement présentes. En pratique, seules les gloses et la traduction sont disponibles, et celles-ci n'apparaissent pas de manière systématique dans tous les discours : elles prédominent principalement dans les discours narratifs et les recettes, tandis que d'autres types de ressources ne bénéficient pas de ce niveau d'annotation. C'est pourquoi nous avons choisi de travailler avec deux discours différents : un récit et une recette de cuisine du même locuteur. Bien qu'il aurait été tout aussi pertinent d'analyser d'autres niveaux d'annotation, ceux qui sont disponibles sont suffisants pour les objectifs de cette recherche. En outre, il est important de souligner que, même s'il ne s'agit pas d'un corpus entièrement annoté, la rareté des corpus accessibles comprenant une traduction oblige dans de nombreux cas à s'adapter aux limites du matériel disponible, sans pour autant invalider la pertinence de l'analyse effectuée.

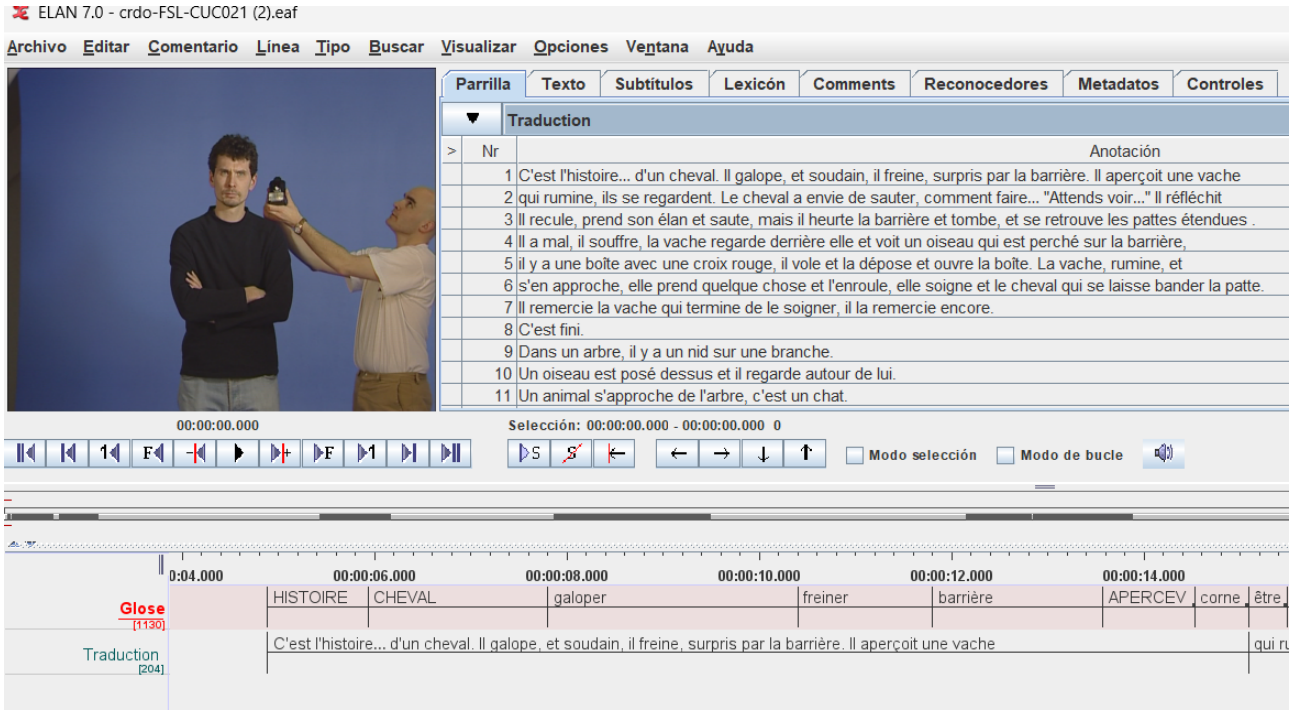


Figure 1: Aperçu de l'interface ELAN montrant les couches d'annotation disponibles

7. Une première application pratique sur le corpus LS-Colin

7.1. Segmentation de la LSF dans le corpus LS-Colin

Comme indiqué dans la partie *5. Cadre théorique et méthodologie*, la segmentation a été réalisée en s'appuyant sur les travaux de Gabarró-López et Meurant (2016), présentés dans leur article « Slicing Your SL

Data into Basic Discourse Units (BDUs): Adapting the BDU Model (Syntax + Prosody) to Signed Discourse» . Toutes les annotations ont été réalisées à l'aide du logiciel ELAN.

Avant de présenter les résultats, il est nécessaire de rappeler que la segmentation proposée doit être considérée comme expérimentale et ne peut être prise comme une référence pour une segmentation en BDU définitive ou pleinement validée. Étant donné que je ne suis pas locutrice de cette langue, ni experte en BDU et que la segmentation en BDU reste très subjective, alors les décisions adoptées, bien que guidées par un cadre méthodologique, ne garantissent pas l'identification entièrement adéquate de tous les indices de nature prosodique et, encore moins, ceux de caractère syntaxique.

7.1.1 Segmentation pragmatique.

Le processus d'annotation a commencé par la segmentation pragmatique, car, étant donné mon manque de compétence en langue des signes française (LSF), ce niveau était méthodologiquement plus accessible. Pour ce faire, quatre lignes d'annotation ont été ajoutées dans le logiciel ELAN :

1. une première ligne destinée à la segmentation des pauses ;
2. une deuxième ligne pour identifier les signes allongés (*signes allongés*) ;
3. une troisième ligne dédiée à l'annotation des phénomènes non manuels pertinents pour la structuration discursive, tels que les clignements des yeux accompagnés d'un mouvement des sourcils, l'ouverture des yeux, la direction du regard, les mouvements de la tête, les mouvements corporels, les gestes buccaux et les expressions faciales ;
4. et une quatrième ligne destinée à l'harmonisation des trois précédentes, afin de proposer une segmentation pragmatique unifiée.

Les pauses ont été marquées en premier lieu, car elles étaient relativement faciles à identifier. Dans le discours procédural (recette de cuisine), cette tâche a été encore plus simple, car les annotations comprenaient souvent la glose « position neutre », ce qui permettait de confirmer que ces segments correspondaient à des pauses. La segmentation a été effectuée systématiquement avant chaque pause identifiée.

Comme dans tout travail d'annotation linguistique, le processus de segmentation prosodique a soulevé diverses questions méthodologiques, notamment en ce qui concerne l'identification des signes allongés. Des questions ont été soulevées, telles que : qu'implique exactement l'exagération d'un signe ? Comment définir un signe allongé ? À partir de quel seuil temporel peut-on considérer qu'un signe est effectivement allongé ?

Bien que ces phénomènes soient décrits comme suit dans la bibliographie :

> « Un allongement peut être observé lorsque la configuration manuelle d'un signe est maintenue pendant une période plus longue que prévu, ou lorsque le signe présente un ralentissement, une répétition ou une exagération du mouvement. Ce type de phénomène peut apparaître au début, au milieu ou à la fin du signe (Notarrigo & Meurant, 2014) ; cependant, à des fins de segmentation discursive, seuls les allongements en position finale sont pris en compte » (Gabarró-López & Meurant, 2016).

Néanmoins, dans la pratique, il s'est avéré difficile de déterminer dans quels cas il s'agissait d'une véritable exagération et dans quels cas la durée du signe répondait plutôt à des besoins expressifs du contenu.

C'est pourquoi, dans ce travail, ont été considérés comme des cas d'exagération les signes qui apparaissaient nettement plus allongés dans les gloses par rapport à d'autres signes, ainsi que ceux dans lesquels le locuteur semblait renforcer le signe par une légère répétition du mouvement accompagnée d'une légère accentuation de la tête, comme s'il cherchait à souligner discursivement le contenu exprimé.

En ce qui concerne les phénomènes non manuels (clignements des yeux accompagnés de mouvements des sourcils, ouverture des yeux, direction du regard, mouvements de la tête, mouvements corporels, gestes buccaux et expressions faciales), leur segmentation a suscité moins de doutes méthodologiques, car dans la plupart des cas, ces éléments étaient clairement identifiables.

Toutefois, certaines difficultés sont apparues lorsque l'orateur faisait des signes avec le visage incliné vers le bas, ce qui rendait difficile, voire pratiquement impossible, d'observer si certains traits non manuels étaient

présents. Cette limitation montre que, si les phénomènes non manuels constituent des indicateurs pertinents pour la segmentation, il est nécessaire de disposer d'enregistrements dans lesquels le visage et le corps du signant sont visibles en continu afin de pouvoir les analyser de manière fiable.

7.1.2 Segmentation syntaxique.

Dans un second temps, une segmentation syntaxique a été réalisée. Pour ce niveau d'analyse, une ligne d'annotation supplémentaire a été ajoutée dans ELAN. La segmentation s'est appuyée principalement sur les gloses fournies dans le corpus et, dans certains cas, sur la traduction en français comme outil complémentaire afin de confirmer ou d'affiner certaines décisions analytiques.

Par exemple, face à la séquence suivante : 00:00:10 – 00:00:16

(10) Exemple:

Extrait des gloses de « Histoire du cheval » 00:00:10 – 00:00:16 :

« freiner barrière APERCEVOIR cornes de VACHE être sur ses sabots ruminer »

Dans un premier temps, il avait été interprété que *APERCEVOIR* dépendait de *freiner* et de *barrière*, c'est-à-dire que le cheval avait aperçu la barrière et freinait pour cette raison. Cependant, la traduction a permis de comprendre que le cheval apercevait en réalité la vache :

Extrait de traductions de « Histoire du cheval » 00:00:10 – 00:00:16 :

« Il aperçoit une vache qui rumine. »

En conséquence, la segmentation syntaxique finale retenue est la suivante :

[freiner barrière] [APERCEVOIR cornes de VACHE] [être sur ses sabots ruminer]

La procédure a commencé par une lecture attentive des gloses, accompagnée de l'observation de la vidéo, dans le but de délimiter les unités en suivant les critères proposés dans le manuel. Cependant, il est rapidement apparu que, dans le discours procédural (recette de cuisine), de nombreux verbes apparaissaient isolés, contrairement aux exemples décrits dans le manuel. Pour cette raison, chaque action ou verbe a été segmenté comme une unité indépendante.

Les gloses ont ensuite été extraites et les limites de segmentation ont été matérialisées à l'aide de crochets « [ ] », afin de visualiser plus clairement les unités syntaxiques identifiées. Ces segmentations syntaxiques sont présentées en annexe.

7.1.3 Segmentation des BDU.

Comme indiqué dans le manuel, la segmentation en BDU doit être établie lorsque les limites identifiées dans les segmentations prosodique et syntaxique coïncident.

Ainsi, une nouvelle ligne d'annotation intitulée « BDU » a été ajoutée dans ELAN, dans laquelle les limites ont été marquées uniquement lorsque les deux segmentations coïncidaient. Dans certains cas, les limites n'étaient pas parfaitement alignées ; cependant, lorsque l'écart entre elles était minime, il a été choisi de positionner la limite de la BDU à un point intermédiaire entre les deux marques.

Ce processus s'est révélé relativement plus rapide, la principale difficulté du travail résidant dans la réalisation des segmentations prosodique et syntaxique préalables. Par ailleurs, il a été décidé de représenter ces segmentations directement sur les gloses, également présentées en annexe, afin de faciliter l'observation et la comparaison entre la segmentation syntaxique et la segmentation finale en BDU.

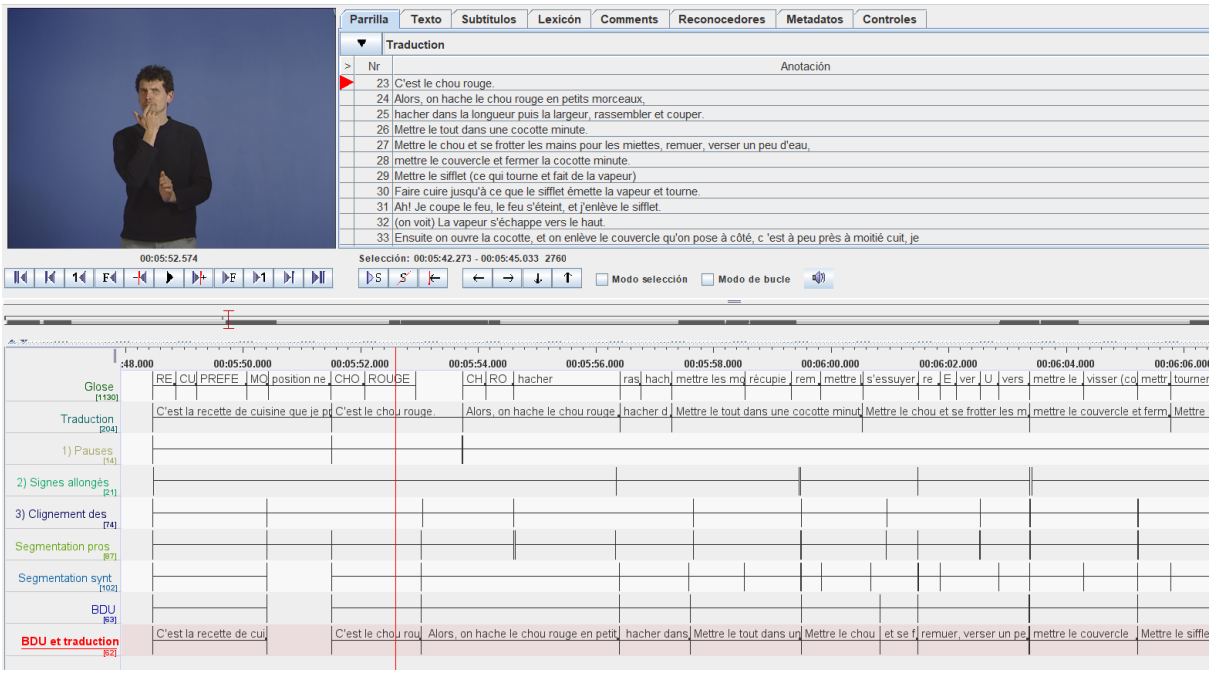


Figure 2: Exemple de fichier ELAN intégrant l'ensemble des niveaux d'annotation : les trois lignes de segmentation prosodique, la ligne de segmentation syntaxique, la ligne de segmentation en BDU et la ligne des traductions par BDU.

7.2 Résultats de la première application pratique sur le corpus LS-Colin

Une fois les deux premières phases d'annotation réalisées, en suivant les lignes directrices proposées par Gabarró-López et Meurant (2016) dans leur article *Slicing Your SL Data into Basic Discourse Units (BDUs)*, je considère qu'il a été possible de mener à bien la segmentation proposée, en m'appuyant sur les gloses et les traductions pour la segmentation syntaxique.
De plus, en tenant compte du fait que je ne suis ni utilisatrice ni locutrice de cette langue, je considère être parvenue à identifier de manière pertinente plusieurs indices prosodiques utiles à la segmentation.

Néanmoins, il serait nécessaire de valider ces annotations avec une personne locutrice de la langue, capable de réaliser sa propre segmentation en BDU, afin de pouvoir comparer ses résultats avec ceux de ce travail. Ce n'est qu'à travers cette comparaison qu'il serait possible de confirmer de manière plus rigoureuse si la segmentation obtenue peut être considérée comme satisfaisante.

À ce stade, et dans la limite des connaissances mobilisées dans cette étude, l'annotatrice responsable des gloses et des traductions du corpus (Marie-Anne Sallandre) a su restituer de nombreux indices visuels et prosodiques dans ses annotations. En particulier, des phénomènes tels que les pauses et les clignements des yeux semblent partiellement reflétés dans l'usage de la ponctuation (points, virgules) dans les traductions.

Par ailleurs, ce travail montre que la collaboration avec des personnes locutrices de la langue des signes en l'occurrence de la langue des signes française, est indispensable pour valider les annotations et permettre d'envisager les étapes suivantes de l'analyse.

Chapitre II

8. Objectifs et problématique expérimentale

Après la première application pratique sur le corpus LS-Colin, il a été jugé pertinent d'évaluer dans quelle mesure les traductions et les gloses en français peuvent constituer un support fiable pour la segmentation discursive en LSF (langue des signes française), en cherchant à reproduire la démarche de la première exploration, mais cette fois avec d'autres annotateurs.

L'objectif est de déterminer si ces traductions peuvent faciliter à la fois le travail des annotateurs humains ne disposant pas de connaissances approfondies en langue des signes et l'utilisation de segmentateurs automatiques appliqués aux traductions textuelles.

L'expérimentation se déroulera en plusieurs étapes. Dans un premier temps, trois personnes présentant des profils et des niveaux d'expérience différents en LSF et en annotation linguistique segmenteront les mêmes extraits en suivant un guide d'annotation commun. Le degré d'accord inter-annotateurs sera ensuite analysé afin d'évaluer la cohérence des segmentations produites.

Dans un deuxième temps, un segmentateur automatique sera testé sur les traductions des mêmes discours, préalablement segmentés par les annotateurs humains. Les résultats obtenus seront comparés afin d'observer le degré de convergence ou de divergence entre les segmentations réalisées par les trois annotateurs et celles produites par l'outil automatique.

À partir de cette comparaison, nous chercherons à déterminer dans quelle mesure la segmentation automatique des traductions pourrait servir de base à la segmentation ultérieure des vidéos en LSF. Cette analyse permettra d'identifier les écarts entre les segmentations humaines et automatiques, d'évaluer l'apport potentiel de ces outils dans le processus d'annotation et de mettre en évidence les critères susceptibles d'en améliorer la qualité. Elle contribuera également à préciser les conditions dans lesquelles une telle méthodologie pourrait être mise en œuvre de manière efficace.

9. Élaboration du guide d'annotation

9.1 Critères pour la conception du guide d'annotation

Nous avons utilisé la même méthode d'annotation que celle présentée dans la section 5.2 « Segmentation de la langue des signes », à savoir le modèle des Basic Discourse Units (BDU) proposé par Gabarró-López et Meurant (2016), qui adaptent la notion de BDU à la modalité signée afin de proposer une méthode de segmentation du discours en langue des signes. Toutefois, nous avons adapté cette proposition et élaboré un guide d'annotation spécifique.

En effet, la proposition de ces autrices ne constitue pas à proprement parler un guide d'annotation entièrement structuré. L'article ne décrit pas de manière détaillée la procédure concrète de segmentation dans ELAN et ne fournit pas suffisamment d'exemples opérationnels permettant une application uniforme par des annotateurs aux profils variés. Pour cette raison, cet article a été utilisé comme base théorique afin de construire un guide d'annotation plus détaillé, plus explicite et directement exploitable dans le cadre de cette recherche.

Le guide a été placé en annexe. Ce guide présente plusieurs différences importantes par rapport à l'étude originale. Tout d'abord, les autrices réalisent toute le processus de segmentation directement à partir de la vidéo en langue des signes, tandis que, dans ce travail, la segmentation syntaxique est effectuée à partir des gloses et des traductions françaises présentes dans les annotations ELAN du corpus LS-COLIN.

Par ailleurs, l'article indique que seuls les signes allongés situés en fin de signe ou en fin d'unité discursive doivent être pris en compte:

> *« Les maintiens de signe (sign holds) et les signes allongés (lengthened signs) constituent l'équivalent des syllabes allongées [...] seuls les maintiens apparaissant en fin de signe sont pris en compte pour la segmentation »* (Gabarró-López & Meurant, 2016, p. 86, notre traduction).

Dans cette recherche, il a toutefois été décidé de considérer également les signes dont la durée était au moins trois fois supérieure à celle des signes environnants, même lorsqu'ils n'apparaissaient pas en position finale. Cette décision a été prise afin de ne pas exclure d'éventuels indices prosodiques pertinents pour la segmentation et de faciliter leur identification par des annotateurs présentant différents niveaux d'expérience.

Le guide a également été enrichi par des explications plus détaillées et par des exemples supplémentaires destinés à faciliter la compréhension des critères d'annotation. Des exemples complets issus d'un récit et d'une recette de cuisine ont notamment été intégrés afin d'illustrer l'application des règles de segmentation dans des contextes discursifs plus larges et plus naturels.

L'un des principaux objectifs de ce guide était de permettre son utilisation non seulement par des spécialistes de la langue des signes ou de l'annotation linguistique, mais également par des étudiants issus d'autres disciplines ou par des personnes ne disposant d'aucune expérience préalable en annotation ou en LSF. L'ambition était de leur fournir les outils nécessaires pour identifier les indices discursifs décrits dans le protocole et appliquer les critères de segmentation de manière cohérente.

9.2 Difficultés rencontrées pendant la conception du guide

L'élaboration du guide d'annotation a soulevé plusieurs difficultés méthodologiques et pratiques. Parmi les principales figurent l'ambiguïté de certains phénomènes discursifs et prosodiques, le manque de consensus dans la littérature concernant la segmentation des langues des signes, les différences structurelles entre la LSF et le français, ainsi que l'absence de modèles de référence suffisamment détaillés pour servir de base directe à cette recherche.

L'une des difficultés les plus importantes concernait la plateforme d'annotation ELAN. Tout d'abord, la fenêtre de visualisation de la vidéo est relativement réduite, ce qui rend l'observation de certains marqueurs non manuels subtils, tels que les clignements des yeux ou certaines expressions faciales brèves, particulièrement difficile. Or, ces éléments peuvent constituer des indices pertinents pour la segmentation prosodique.

Par ailleurs, il s'est avéré complexe d'expliquer clairement la manière dont les segments doivent être sélectionnés dans ELAN. Dans le cas de la segmentation prosodique, par exemple, l'annotation n'est pas créée au moment précis où apparaît un marqueur tel qu'un clignement des yeux. Au contraire, le segment est interrompu à cet endroit, de sorte que les limites visuelles des annotations indiquent directement la présence de ce type d'indice. Cette logique de fonctionnement facilite l'interprétation ultérieure de la segmentation, mais elle n'est pas nécessairement intuitive pour des personnes peu familières avec les outils d'annotation.

Une autre difficulté importante concernait la définition des signes allongés. Même pour un observateur ne maîtrisant pas la langue des signes, il n'est pas toujours évident de déterminer à partir de quel moment un signe peut être considéré comme suffisamment allongé pour constituer un indice prosodique pertinent. Certaines situations demeurent donc ambiguës et ne présentent pas nécessairement une unique interprétation correcte. Cette incertitude a conduit à l'adoption de critères relativement souples au sein du guide.

Enfin, l'absence d'annotations comparables et de gold standard a rendu difficile la validation de certaines décisions méthodologiques. De nombreux critères ont dû être élaborés progressivement à partir de l'observation du corpus et de l'interprétation des phénomènes décrits dans la littérature. Les différences structurelles entre la LSF et le français ont également constitué un défi important : certains marqueurs

discursifs présents dans la langue signée ne trouvent pas toujours d'équivalent direct dans les traductions françaises, ce qui peut entraîner des divergences entre l'organisation du discours dans les deux modalités.

10. Protocole expérimental

L'expérimentation a été mise en place afin d'évaluer l'applicabilité du guide d'annotation proposé auprès d'annotateurs présentant différents niveaux d'expérience en annotation linguistique et en langue des signes française (LSF). Elle visait également à identifier les difficultés rencontrées au cours du processus d'annotation et à mettre en évidence les aspects du guide nécessitant des améliorations.

10.1 Participants

Trois annotateurs présentant des profils distincts ont été sélectionnés afin de tester le guide d'annotation dans des conditions variées et d'évaluer dans quelle mesure cette tâche peut être réalisée par des personnes possédant différents niveaux de connaissances en linguistique et en LSF.

Le premier annotateur ne possédait aucune expérience préalable en annotation linguistique ni connaissance de la LSF. Il était étudiant dans le domaine de l'économie du bâtiment.

Le deuxième annotateur disposait d'une expérience en annotation linguistique, mais ne possédait aucune connaissance de la LSF. Il s'agissait d'une étudiante en Master Linguistique, Informatique et Technologies du Langage.

Enfin, le troisième annotateur était sourd, locuteur de LSF et possédait une expérience préalable en annotation. Il était étudiant en troisième année de Licence Langues, Littératures et Civilisations Étrangères et Régionales (LLCER), parcours Traduction-Médiation Langue des Signes Française – Langue Étrangère et Seconde.

10.2 Données annotées

Les participants ont été invités à segmenter deux discours extraits d'un enregistrement du corpus LS-COLIN.

Les deux extraits sélectionnés correspondaient à un récit, intitulé L'histoire de l'oiseau, situé entre 00:27:09 et 00:28:08, ainsi qu'à une recette de cuisine située entre 00:31:04 et 00:34:04. Le premier extrait a une durée approximative d'une minute, tandis que le second dure environ trois minutes.

Les discours sont disponibles dans ici: https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82.

Initialement, les deux discours devaient être annotés par l'ensemble des participants. Toutefois, en raison de contraintes de temps, seule l'annotation complète de la recette de cuisine réalisée par l'annotateur 2 a pu être obtenue. Les annotateurs 1 et 3 n'ont pas été en mesure de terminer l'annotation de ce second discours dans le délai prévu pour l'expérimentation.

10.3 Procédure d'annotation

L'ensemble des participants a reçu un dossier compressé contenant trois éléments : le guide d'annotation, la vidéo à segmenter et un modèle d'annotation ELAN.

Ce modèle comprenait deux discours différents préalablement annotés, à savoir un récit et une recette de cuisine. Ces exemples avaient pour objectif de servir de référence méthodologique et d'aider les participants à comprendre l'application concrète des critères décrits dans le guide.

Les conditions dans lesquelles chacun des annotateurs a réalisé l'annotation sont présentées ci-dessous.

Annotateur 1 — aucune expérience en annotation linguistique ni en LSF

L'annotation a été réalisée en présentiel dans un environnement calme et sans distraction. Un ordinateur a été mis à la disposition du participant et une présentation du fonctionnement de base d'ELAN a été effectuée. Cette présentation portait notamment sur l'ouverture des fichiers, le chargement d'un projet, la création, la modification et la suppression de segments, ainsi que sur l'utilisation des repères temporels.

Le participant a ensuite été invité à lire le guide d'annotation et à poser toute question susceptible de se présenter au cours du processus.

Une fois la tâche commencée, l'annotatrice a travaillé de manière autonome. Lorsque des interrogations concernant la méthodologie ou l'interprétation des consignes apparaissaient, des clarifications étaient fournies. L'ensemble des questions posées a été consigné afin d'identifier les parties du guide nécessitant des explications supplémentaires.

En revanche, aucune question portant sur des décisions concrètes de segmentation n'a reçu de réponse. Il n'a notamment jamais été indiqué si un indice prosodique ou syntaxique particulier devait être considéré comme présent ou absent. Cette précaution visait à éviter toute influence sur les résultats obtenus.

L'annotation du discours narratif a nécessité environ 1 h 40.

Annotateur 2— expérience en annotation linguistique mais aucune connaissance de la LSF

L'annotation a été réalisée à distance depuis le domicile du participant.

Avant le début de la tâche, une explication en présentiel portant sur le fonctionnement général d'ELAN et sur les objectifs de la segmentation a été fournie. Le même dossier que celui remis à l'annotateur 1 lui a ensuite été transmis.

Au cours de l'annotation, le participant pouvait poser ses questions par message. Comme pour l'annotateur 1, seules les questions relatives à l'interprétation du guide ou aux aspects techniques de l'annotation recevaient une réponse. Aucune indication concernant les solutions attendues ou les décisions de segmentation à adopter n'a été fournie.

L'annotation du discours narratif a nécessité environ 1 h 30.

Annotateur 3 — locuteur de LSF, sourd et expérience préalable en annotation

L'annotation a été réalisée en présentiel dans un environnement calme et sans distraction. Nous avons répondu aux questions portant sur le guide d'annotation et sur la tâche demandée.

Contrairement aux deux autres participants, cette annotatrice s'est principalement appuyée sur la vidéo signée pour réaliser la segmentation syntaxique vu qu'elle est locutrice de la LSF et possédait déjà une expérience de l'annotation. Les gloses et les traductions françaises ont été utilisées essentiellement comme support permettant de localiser les segments dans ELAN et d'aligner les annotations avec la structure déjà présente dans le corpus.

L'annotation du discours narratif a nécessité environ deux heures.

10.4 Difficultés rencontrées par les annotateurs

L'analyse des retours des participants a permis d'identifier plusieurs difficultés récurrentes liées à la fois au guide d'annotation, à l'utilisation du logiciel ELAN et à la nature même de la tâche de segmentation.

Annotateur 1 — aucune expérience en annotation linguistique ni en LSF

La principale difficulté rencontrée par l'annotatrice concernait la compréhension des unités syntaxiques. Ne disposant d'aucune formation en linguistique, certains concepts utilisés dans le guide lui étaient peu familiers et ont nécessité des explications complémentaires. Cette observation suggère que certaines notions théoriques gagneraient à être davantage explicitées ou illustrées par des exemples supplémentaires.

L'annotatrice a également signalé plusieurs difficultés liées à l'utilisation d'ELAN. Certaines manipulations lui ont paru peu intuitives. Des problèmes techniques ou des erreurs de manipulation l'ont parfois contrainte à répéter certaines opérations, ce qui a ralenti le processus d'annotation.

Annotateur 2— expérience en annotation linguistique mais aucune connaissance de la LSF

L'annotateur a principalement exprimé des doutes concernant la segmentation syntaxique.

Ici un exemple des questions qu'elle s'est posée:

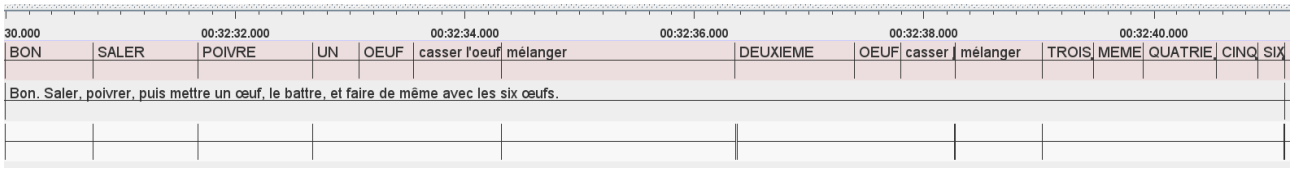


Figure 3 : Capture d'écran d'ELAN illustrant l'exemple fourni par l'annotatrice 2.

> *« Au départ, j'ai annoté ["saler"], ["poivrer"], ["un œuf casser l'œuf et "mélanger"], car on dirait des actions distinctes. Mais ensuite, j'ai été tentée de regrouper ["deuxième œuf casser mélanger"], car cela me semblait davantage former une seule séquence d'actions. Finalement, j'ai choisi de les séparer parce que le guide indiquait qu'il fallait le faire, mais la décision ne me paraissait pas toujours évidente. »*

Il a aussi indiqué qu'au fur et à mesure de l'avancement de la tâche, sa compréhension de la structure du discours évoluait, ce qui l'amenait parfois à modifier ses critères de segmentation. Cette évolution progressive de ses décisions lui a fait craindre un manque de cohérence dans l'ensemble de son annotation.

Cette observation met en évidence la difficulté de définir des frontières syntaxiques stables à partir des seules gloses et traductions, en particulier pour des annotateurs ne connaissant pas la LSF.

L'identification des clignements des yeux et des autres marqueurs non manuels a également représenté une difficulté importante. La qualité de la vidéo ainsi que la taille réduite de la fenêtre de visualisation dans ELAN rendaient leur observation particulièrement complexe pour une personne non signante. Cette difficulté était accentuée par les mouvements fréquents de la tête du locuteur, qui rendaient parfois difficile la distinction entre des yeux ouverts et fermés.

Les signes allongés ont également suscité plusieurs interrogations. Dans certaines situations, il était difficile de déterminer si un phénomène observé correspondait à un véritable allongement prosodique ou simplement à une répétition du signe.

Enfin, l'annotateur a mentionné que, malgré l'utilisation d'un ordinateur performant, ELAN pouvait parfois présenter des ralentissements. Il a toutefois souligné que l'explication préalable du fonctionnement du logiciel lui avait été très utile pour mener à bien la tâche.

Annotateur 3 — locuteur de LSF, sourd et expérience préalable en annotation

L'annotatrice a indiqué que certaines gloses et traductions présentes dans le corpus ne reflétaient pas toujours fidèlement le contenu du discours signé.

Elle a également expliqué que l'objectif exact de la tâche ne lui avait pas semblé totalement clair au début de l'expérimentation et qu'un certain temps d'adaptation avait été nécessaire avant de comprendre précisément ce qui était attendu. Cette remarque suggère que le guide pourrait être complété par des explications plus détaillées ou par des supports pédagogiques supplémentaires, tels qu'un tutoriel vidéo présentant des exemples complets de segmentation.

La segmentation syntaxique a constitué la principale difficulté rencontrée par cette participante. Selon elle, s'appuyer principalement sur les gloses ne correspond pas à une démarche naturelle pour une personne utilisatrice de la LSF, en raison des différences structurelles importantes entre les langues signées et les langues vocales. Pour cette raison, elle a préféré effectuer une partie de son analyse sur papier et fonder ses décisions de segmentation avant tout sur l'observation directe de la vidéo.

Son procédé consistait à observer d'abord le discours signé afin d'identifier les frontières de segmentation à partir des informations visuelles, puis à consulter les gloses pour faire correspondre ces décisions à la structure d'annotation présente dans ELAN. Cette observation suggère que les gloses et le guide d'annotation se sont révélés moins utiles pour une utilisatrice experte de la LSF que pour les participants ne possédant pas de connaissances dans cette langue.

Dans l'ensemble, les retours des trois participants montrent que les principales difficultés concernent la compréhension des critères syntaxiques, l'identification des indices prosodiques a cause de la qualité et taille de la video et l'utilisation de l'interface ELAN.

11. Expérimentation de segmentation automatique avec l'outil DisCut

Afin d'évaluer la faisabilité d'une segmentation automatique fondée sur les traductions françaises du corpus, il a été décidé d'utiliser l'outil DisCut, développé par l'équipe MELODI de l'IRIT (Institut de Recherche en Informatique de Toulouse), composée notamment de Morteza Ezzabady, Philippe Muller et Chloé Braud.

L'objectif de cette expérimentation est de comparer les segmentations produites automatiquement par DisCut avec celles réalisées par les annotateurs humains décrits dans la section précédente. Cette comparaison doit permettre d'évaluer dans quelle mesure les résultats de la segmentation automatique appliquée aux traductions françaises pourraient être utilisés comme base pour la segmentation des vidéos en LSF.

11.1 Origine de DisCut : le challenge DISRPT

DisCut a été développé dans le cadre de DISRPT (Discourse Relation Parsing and Treebanking), une série de campagnes internationales d'évaluation (shared tasks) consacrées à l'analyse automatique du discours. Ces campagnes réunissent des chercheurs issus de différentes institutions afin de développer et de comparer des méthodes automatiques destinées à identifier les structures discursives dans diverses langues et selon différents cadres théoriques.

Parmi les tâches proposées par DISRPT, la première concerne la segmentation des unités discursives (Discourse Unit Segmentation across Formalisms). Son objectif est d'identifier automatiquement les frontières entre unités discursives dans des textes écrits. Étant donné que les différents corpus et cadres d'annotation reposent sur des conventions de segmentation parfois très différentes, cette tâche vise à encourager le développement de méthodes suffisamment flexibles pour s'adapter à une grande variété d'approches théoriques.

L'outil utilisé dans ce travail correspond à la participation de DisCut à l'édition DISRPT 2021. Lors de cette campagne, le système a obtenu des résultats compétitifs et a montré que la combinaison de corpus issus de différentes langues et de différents cadres d'annotation pouvait contribuer à améliorer les performances de la segmentation discursive automatique.

Le choix de DisCut pour cette expérimentation repose sur plusieurs raisons. Tout d'abord, il s'agit d'un système spécifiquement conçu pour la segmentation discursive, qui constitue précisément l'objet de cette recherche. De plus, l'outil a déjà été évalué dans le cadre d'une campagne internationale, ce qui fournit une référence méthodologique solide pour comparer les résultats obtenus par les annotateurs humains.

11.2 Fonctionnement du segmenteur DisCut

11.2.1 Cadre théorique

Contrairement à de nombreux systèmes conçus pour un seul modèle discursif, DisCut a été développé afin de s'adapter aux principaux cadres théoriques représentés dans les corpus de DISRPT 2021.

Le premier est la Rhetorical Structure Theory (RST) proposée par Mann et Thompson (1988). Dans cette approche, un texte est découpé en petites unités discursives appelées Elementary Discourse Units (EDUs). Ces unités sont ensuite reliées entre elles par différentes relations discursives, telles que la cause, l'explication ou le contraste, afin de former une structure hiérarchique représentant l'organisation globale du discours. Il s'agit du cadre le plus représenté dans les corpus de la campagne.

Le deuxième est la Segmented Discourse Representation Theory (SDRT) développée par Asher et Lascarides (2003). Comme la RST, ce modèle cherche à décrire l'organisation du discours à partir de relations entre segments. Toutefois, il autorise des structures plus complexes, notamment des segments imbriqués les uns dans les autres. Pour les besoins de la campagne DISRPT, ces structures ont été simplifiées et transformées en séquences linéaires afin de rendre la tâche de segmentation comparable entre les différents corpus.

Enfin, certains corpus utilisent le cadre du Penn Discourse Treebank (PDTB) (Prasad et al., 2008). Contrairement aux deux approches précédentes, ce modèle ne cherche pas d'abord à découper le texte en unités discursives. Il s'intéresse principalement aux connecteurs discursifs, tels que mais, parce que ou donc, ainsi qu'aux segments qu'ils relient. La segmentation y est donc abordée de manière plus indirecte.

Malgré leurs différences, ces trois approches ont un objectif commun : identifier les différentes unités qui composent un discours et décrire les relations qui les unissent. Dans le cadre de DISRPT, cette tâche est formulée comme un problème d'étiquetage de séquences. Chaque mot du texte reçoit une étiquette indiquant s'il correspond au début d'une unité discursive, à l'intérieur d'un connecteur discursif ou à aucun de ces cas, selon la convention BIO (Begin, Inside, Outside)

11.2.2 Données utilisées

Le système a été entraîné et évalué sur les corpus fournis dans le cadre de DISRPT 2021, soit seize corpus répartis dans onze langues différentes. Ces langues comprennent le français, l'espagnol, le portugais, l'anglais, l'allemand, le néerlandais, le russe, le turc, le basque, le mandarin et le persan.

Les corpus présentent une forte diversité en termes de taille, de domaine textuel et de cadre d'annotation. Cette diversité constitue un défi important pour les systèmes de segmentation automatique, et c'est précisément pour cette raison que ce système représente un candidat intéressant pour la segmentation de traductions en français de la langue des signes française.

11.2.3 Architecture et méthodologie de DisCut

Dans leur travail, Ezzabady et al. (2021) présentent l'architecture et la méthodologie du système DisCut développé pour la campagne DISRPT 2021. DisCut s'appuie sur une modification de l'architecture de ToNy, un système ayant obtenu les meilleurs résultats lors de la campagne DISRPT 2019.

Son fonctionnement repose sur plusieurs mécanismes d'apprentissage automatique permettant d'analyser le contexte des mots et d'identifier les frontières discursives dans un texte. Le système utilise notamment mBERT (*Multilingual Bidirectional Encoder Representations from Transformers*), un modèle de langue

multilingue capable de prendre en compte le contexte dans lequel apparaît chaque mot. Cette information est complétée par d'autres modules permettant d'améliorer la représentation des tokens et la qualité des prédictions (Ezzabady et al., 2021).

L'architecture combine ainsi plusieurs composants de traitement permettant au système de déterminer, pour chaque token, s'il correspond ou non au début d'une nouvelle unité discursive. Les auteurs décrivent notamment l'utilisation d'un CNN au niveau des caractères et d'une couche BiLSTM pour traiter les représentations produites par le modèle de langue avant la prédiction finale (Ezzabady et al., 2021).

L'une des principales caractéristiques de DisCut réside toutefois dans sa stratégie d'apprentissage multilingue et multi-corpus. Plutôt que d'entraîner un modèle entièrement indépendant pour chaque langue, les auteurs cherchent à exploiter les similarités entre des langues appartenant à une même famille linguistique. Ainsi, pour les langues romanes, les données espagnoles, portugaises et françaises sont progressivement combinées afin de construire des modèles bénéficiant d'un volume de données d'apprentissage plus important. Une stratégie comparable est appliquée aux langues germaniques, avec la combinaison des données anglaises, allemandes et néerlandaises (Ezzabady et al., 2021).

Le système est ensuite adapté aux corpus individuels à l'aide d'une étape de *fine-tuning*. Cette stratégie permet de bénéficier à la fois des caractéristiques communes apprises à partir de plusieurs langues et des spécificités propres à chaque corpus. Les auteurs évaluent ainsi différentes configurations et sélectionnent, pour chaque corpus, le modèle donnant les meilleurs résultats sur les données de développement correspondantes (Ezzabady et al., 2021).

11.2.4 Résultats obtenus dans DISRPT 2021

Lors de l'évaluation finale de DISRPT 2021, DisCut s'est classé troisième parmi les quatre systèmes participants.

Les performances les plus faibles ont été observées dans la tâche d'identification des connecteurs discursifs, notamment sur le corpus mandarin.

Malgré cela, la stratégie d'apprentissage multilingue a montré son efficacité dans plusieurs cas. DisCut a notamment obtenu les meilleurs résultats de la campagne pour la segmentation sur les corpus allemand et espagnol, deux langues ayant bénéficié de l'entraînement conjoint avec des langues apparentées.

11.3 Démarche d'adaptation au corpus LS-COLIN

Comme indiqué précédemment, DisCut est un outil conçu pour la segmentation discursive de textes écrits. Afin de l'appliquer au corpus LS-COLIN, plusieurs étapes d'adaptation des données ont été nécessaires.

Dans un premier temps, les gloses ont été extraites manuellement à partir des annotations ELAN et utilisées comme entrée du système. Cependant, cette première tentative n'a pas permis d'obtenir des résultats exploitables : le système semblait entrer dans une boucle d'exécution sans produire de segmentation, il est probable que les gloses s'éloignent fortement des données linguistiques sur lesquelles le système a été entraîné.

Face à cet échec, il a été décidé de recourir aux traductions françaises du corpus. Celles-ci ont été extraites puis soumises à DisCut, qui a cette fois permis de générer une segmentation automatique des discours.

L'hypothèse de départ était que, même si DisCut ne peut pas analyser directement un discours signé, les traductions françaises pourraient contenir suffisamment d'indices discursifs pour permettre l'identification automatique de certaines frontières discursives. Une fois la segmentation obtenue, les résultats ont été reportés dans ELAN et alignés manuellement avec les données du corpus, en s'appuyant à la fois sur les

gloses et sur la vidéo afin d'identifier la correspondance entre les segments produits automatiquement et les segments du discours signé.

Ce travail de correspondance a été réalisé sur les mêmes discours que ceux utilisés lors des expérimentations avec les annotateurs humains présentées précédemment.

11.4 Difficultés techniques

Certains éléments identifiés comme des segments contenaient des expressions telles que « bon », « alors » ou « donc », voire des phrases complètes, qui n'étaient pas systématiquement représentées dans les gloses. Dans d'autres cas, les gloses avaient déjà été alignées sur une autre partie du discours, ce qui compliquait l'identification des correspondances entre les différents niveaux d'annotation. Par ailleurs, il arrivait que les traductions et les gloses présentent des divergences de structuration : la traduction pouvait introduire une idée dans un ordre différent de celui des gloses, ou contenir des reformulations qui inversaient partiellement l'organisation de l'information entre les deux niveaux.

Par conséquent, il n'était pas toujours possible de transférer directement les frontières de segmentation vers la vidéo en se basant uniquement sur les gloses. Une phase d'interprétation des données était nécessaire afin de déterminer les correspondances entre les différents niveaux d'annotation et de décider si certains éléments devaient être pris en compte comme indices de segmentation ou ignorés.

12. Méthodes d'évaluation

Une fois les annotations des trois participants obtenues et la segmentation basée sur les résultats de DisCut complétée, une analyse comparative des données a été réalisée afin d'évaluer le degré de correspondance entre les différents types de segmentation.

12.1 Extraction des données

Les fichiers d'annotation produits par ELAN sont au format .eaf, un format basé sur XML. Dans un premier temps, plusieurs fichiers ont été examinés manuellement à l'aide d'un éditeur de texte afin de comprendre leur structure interne, d'identifier l'organisation des différentes couches d'annotation (tiers) et de repérer les informations pertinentes pour l'analyse.

Sur la base de ces observations, plusieurs tests ont été effectués sur une annotation modèle préalablement réalisée par la chercheuse afin de mettre en place une procédure d'extraction automatique des données nécessaires à la comparaison.

Par la suite, plusieurs scripts en Python ont été développés, certains ayant été affinés à l'aide de Claude. Ces scripts s'appuient principalement sur la bibliothèque pympi-ling, spécialisée dans la lecture, la modification et l'analyse de fichiers ELAN (.eaf). Cette bibliothèque permet d'accéder aux différents tiers d'annotation, d'extraire les informations temporelles et de manipuler les données contenues dans les corpus annotés.

12.2 Comparaison des segmentations

Le premier programme (*programme_1.py*) développé avait pour objectif de comparer une même couche d'annotation (dans ce cas, la couche représentant les BDU) entre deux fichiers différents. Le programme recherchait les annotations coïncidant dans les fichiers 1 et 2, c'est-à-dire celles dont les débuts et les fins correspondaient avec une tolérance de 250 ms (la tolérance pouvant être modifiée dans le script).

Après l'exécution de ce programme, nous avons constaté que le nombre de segments coïncidant était le plus souvent nul. Pour cette raison, il a été jugé plus pertinent de développer un second programme (*programme_2.py*), qui compare également une même couche d'annotation (dans ce cas, la couche

représentant les BDU) entre deux fichiers différents, mais en se concentrant uniquement sur les frontières de segmentation.

Ce programme ne prend en compte que les temps de fin (end times) et non simultanément les temps de début et de fin. Ce choix méthodologique s'explique par le fait que chaque frontière de segmentation correspond, par définition, à la fin d'un segment et au début du suivant. Ainsi, si les deux valeurs avaient été prises en compte, chaque frontière aurait été comptabilisée deux fois, ce qui aurait artificiellement augmenté les résultats.

Le programme no. 2 permettait de définir :

- l'intervalle temporel correspondant au discours à analyser ;
- une marge de tolérance temporelle permettant de considérer que deux frontières étaient équivalentes.

Plusieurs tests ont été réalisés avec différentes marges de tolérance : 100 ms, 200 ms, 300 ms, 400 ms et 500 ms.

Les résultats ont montré qu'une tolérance supérieure à 500 ms pouvait entraîner des correspondances artificielles, dans la mesure où plusieurs unités (gloses ou segments complets) pouvaient se situer dans cet intervalle. Dans ce cas, deux annotateurs pouvaient avoir placé des frontières autour d'un même signe sans que celles-ci correspondent réellement, tout en étant considérées comme équivalentes par le programme.

À l'issue de ces tests, une tolérance de 250 millisecondes a été retenue comme offrant le meilleur compromis entre flexibilité et précision. Tous les résultats présentés dans cette étude reposent donc sur cette valeur.

Pour chaque comparaison, le programme fournit :

- le nombre total de frontières présentes dans chaque fichier ;
- le nombre de correspondances détectées ;
- le nombre de frontières non appariées ;
- la distance temporelle (en millisecondes) entre les frontières considérées comme équivalentes ;
- ainsi que la liste des frontières présentes dans un fichier mais absentes dans l'autre.

Ce dispositif a été appliqué à l'ensemble des combinaisons possibles entre les différents annotateurs et la segmentation produite par DisCut :

- Anotateur 1 – Anotateur 2
- Anotateur 1 – Anotateur 3
- Anotateur 2 – Anotateur 3
- DisCut – Anotateur 1
- DisCut – Anotateur 2
- DisCut – Anotateur 3

Cette procédure a permis de mesurer le degré d'accord entre annotateurs humains, puis de comparer ces résultats avec ceux obtenus par la segmentation automatique.

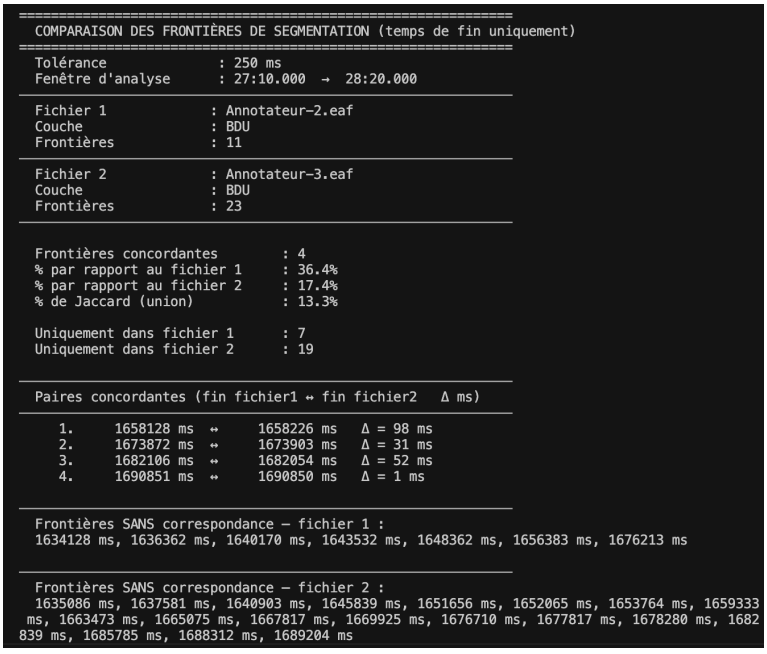


Figure 4 : Capture d'écran de la sortie du programme n° 2.

12.3 Analyse des frontières non concordantes

Une fois les correspondances et divergences entre les segmentations de BDU identifiées, un troisième programme (*programme_3.py*) a été développé afin d'analyser spécifiquement les frontières non appariées entre deux documents.

Ce programme prend en entrée une liste de temps correspondant à l'ensemble des frontières identifiées par un annotateur mais non appariées avec celles de l'autre annotateur. Ces dernières sont extraites des sections « frontières sans correspondance – fichier 1 » et « frontières sans correspondance – fichier 2 », issues de la sortie du programme 2, et sont directement intégrées dans le script. Nous fournissons également le fichier dans lequel la frontière n'a pas été détectée lors de la première analyse.

Pour chacun de ces temps, le programme vérifie automatiquement l'existence d'une annotation correspondante dans les couches de segmentation syntaxique ou prosodique du document analysé.

Comme précédemment, une tolérance de 250 ms a été appliquée.

Le programme fournit les informations suivantes :

- le nombre de frontières ayant trouvé une correspondance dans l'une des couches de segmentation syntaxique ou prosodique et la liste de temps de ces frontieres;
- le nombre de frontières sans correspondance et la liste de temps de ces frontières ;
- le type de correspondance identifiée (syntaxique, prosodique ou les deux).

Ce traitement a été appliqué à l'ensemble des divergences observées :

- Anotateur 1 – Anotateur 2
- Anotateur 2 – Anotateur 1
- Anotateur 2 – Anotateur 3
- Anotateur 3 – Anotateur 2
- Anotateur 1 – Anotateur 3
- Anotateur 3 – Anotateur 1
- DisCut – les trois annotateurs

L'analyse inverse (annotateurs → DisCut) n'a pas été réalisée, dans la mesure où la segmentation produite par DisCut ne dispose pas de couches équivalentes de segmentation syntaxique ou prosodique.

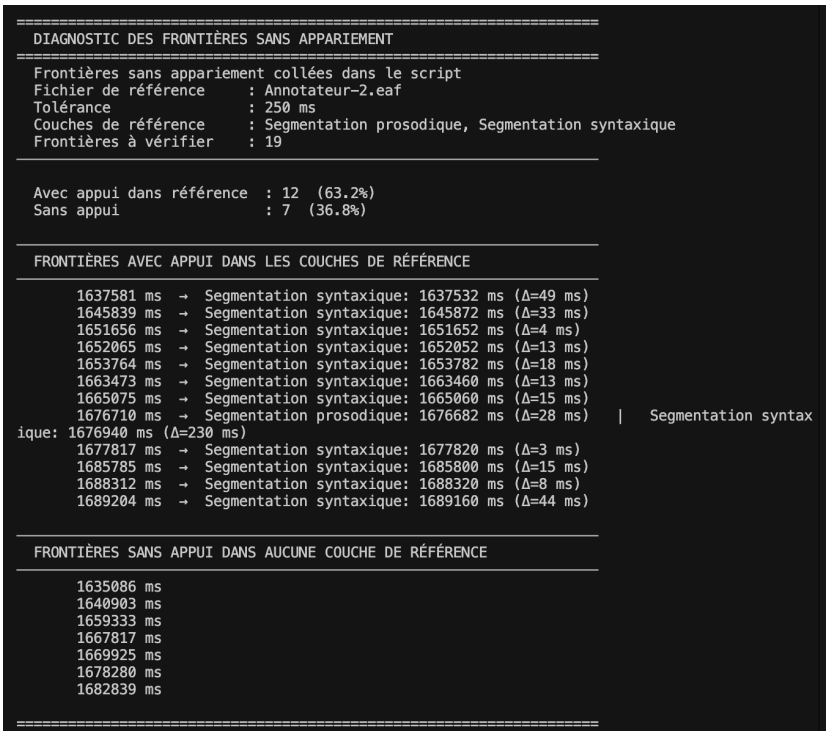


Figure 5 : Capture d'écran de la sortie du programme n° 3.

12.4 Analyse du consensus

Enfin, un quatrième programme *(programme_4.py)* a été développé afin d'analyser le niveau de consensus entre les annotateurs pour chaque couche d'annotation.

Ce programme prend en entrée l'ensemble des fichiers annotés ainsi que la couche à analyser. Il extrait automatiquement toutes les frontières identifiées par les différents annotateurs et génère une liste unique contenant l'ensemble des frontières distinctes observées. Une tolérance de 250 ms est appliquée afin de regrouper les frontières considérées comme équivalentes. En sortie, le programme génère un document au format texte (.txt) contenant la liste des frontières uniques.

Dans un second temps, un autre programme *(programme_5.py)* vérifie, pour chacune de ces frontières, quels annotateurs l'ont identifiée. Il produit alors un tableau dans lequel chaque frontière est représentée par une ligne et chaque annotateur par une colonne. La présence d'une annotation est indiquée par la valeur « 1 » dans la colonne correspondante.

Cette représentation permet d'analyser plus finement les convergences et les divergences entre annotateurs, ainsi que le niveau de consensus atteint pour chaque couche d'annotation.

Les tableaux produits ont ensuite été exportés vers des feuilles de calcul Excel afin de faciliter leur analyse, leur comparaison et la création des visualisations présentées dans la section des résultats.

```
[OK] Annot1 (Annotateur-1.eaf | 'Pauses') — 17 frontières dans la fenêtre
[OK] Annot2 (Annotateur-2.eaf | 'Pauses') — 34 frontières dans la fenêtre
[OK] Annot3 (Annotateur-3.eaf | 'Pauses') — 19 frontières dans la fenêtre

  TABLE DE CONCORDANCE DES FRONTIÈRES
  Tolérance : 250 ms    |    12 frontières    |    3 annotateurs

             ms  h:mm:ss.ms      | Annot1 | Annot2 | Annot3 |
        1631940  0:27:11.940     |        |    1   |        |
        1636460  0:27:16.460     |        |    1   |    1   |
        1636838  0:27:16.838     |    1   |        |        |
        1637210  0:27:17.210     |        |    1   |        |
        1651568  0:27:31.568     |    1   |        |    1   |
        1656773  0:27:36.773     |        |    1   |        |
        1657720  0:27:37.720     |        |    1   |        |
        1669827  0:27:49.827     |        |        |    1   |
        1688808  0:28:08.808     |        |        |    1   |
        1690500  0:28:10.500     |    1   |        |        |
        1690939  0:28:10.939     |        |    1   |    1   |
        1694128  0:28:14.128     |        |    1   |        |
          TOTAL                  |    3   |    7   |    5   |
              %                  |  25.0 %|  58.3 %|  41.7 %|
  Frontières identifiées par TOUS : 0  (0.0%)
```

Figure 6: Capture d'écran de la sortie du programme n° 5.

12.5 Organisation et exploitation des résultats

Les résultats produits par les deux programmes ont ensuite été exportés au format CSV afin de faciliter leur analyse. Le document regroupant l'ensemble des tableaux est présenté en annexe.

Les différents tableaux s'organisent de la façon suivante :

Tableau 1 : Comparaison des segmentations BDU
Cette table regroupe les résultats générés par le premier programme et contient les informations suivantes :

- la paire de fichiers comparés (fichier A / fichier B) ainsi que les couches d'annotation correspondantes ;
- le nombre total de frontières de segmentation identifiées dans chaque fichier (A et B) ;

- le nombre de frontières appariées (concordantes) ;
- le nombre de frontières non appariées pour chaque fichier (sans pair A / sans pair B) ;
- le coefficient de similarité de Jaccard ;
- le pourcentage de concordance A → B ;
- le pourcentage de concordance B → A ;

Tableau 2 : Analyse des frontières non concordantes des BDU
Ce tableau rassemble les résultats générés par le second programme et inclut les informations suivantes :

- la paire de documents analysés ;
- le fichier comparé ;
- le fichier de référence ;
- le nombre total de frontières non appariées (à analyser) ;
- le nombre de frontières ayant trouvé un appui (correspondance) ;
- le pourcentage de frontières avec appui ;
- le nombre de correspondances avec la couche syntaxique ;
- le nombre de correspondances avec la couche prosodique ;
- le nombre de correspondances simultanément présentes dans les deux couches ;
- le nombre de frontières sans appui ;
- le pourcentage de frontières sans appui.

La combinaison de ces deux analyses permet d'évaluer à la fois le niveau global d'accord entre annotateurs humains et les causes spécifiques des divergences observées. Elle permet également de situer la segmentation produite par DisCut par rapport aux segmentations humaines.

En complément, un tableau a été créé sur une page distincte pour le calcul du consensus, correspondant à la sortie du *programme 5*. Ce tableau comprend des feuilles de comparaison pour les BDU, la segmentation syntaxique, la segmentation prosodique, les pauses, les signes allongés et les clignements des yeux. Une dernière feuille analyse le consensus global, en regroupant les résultats des analyses précédentes.

Ainsi, le fichier Excel correspondant aux résultats des traitements du chapitre 2 comprend au total neuf feuilles de calcul et se trouve en annexe.

13.Resultats

13.1 Résultats des annotations réalisées par les annotateurs humains

Le tableau suivant présente le nombre de frontières identifiées pour chaque couche d'annotation et pour chaque annotateur.

Tier	Annotateur 1	Annotateur 2	Annotateur 3
Pauses	3	7	5
signes allongés	4	7	16
Clignement des yeux	22	14	4
Segmentation Prosodique	25	18	25
Segmentation syntaxique	43	39	56
BDU	16	11	22

Tableau 2 : nombre total de frontières identifiées par chaque annotateur

13.1.1 Resultats BDU

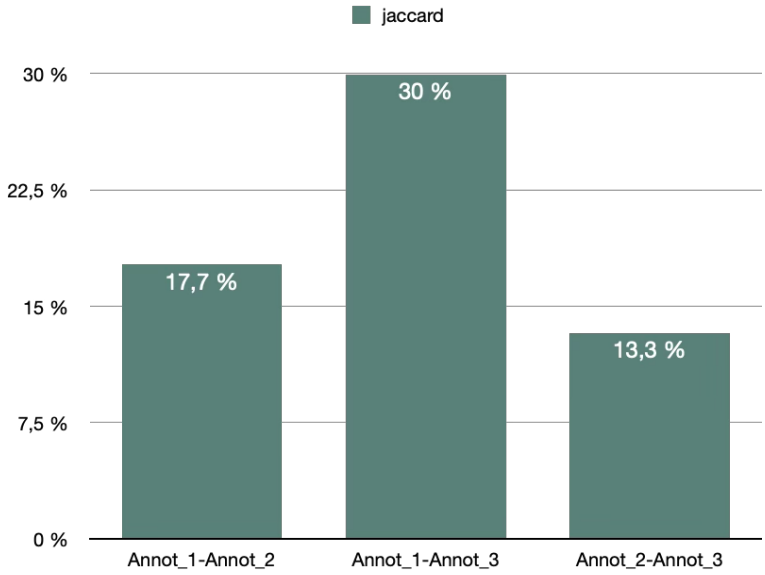
Dans le cadre de l'analyse, l'objectif principal étant la comparaison des Unités Discursives de Base (BDU), l'analyse a d'abord porté sur cette couche, en calculant le pourcentage de concordance entre annotateurs afin d'identifier les paires présentant le plus haut niveau d'accord.

Comme le montre le tableau no.2, l'annotateur 1 a identifié 16 frontières de BDU, l'annotateur 2 en a identifié 11 et l'annotateur 3 en a identifié 26.

Une première analyse a consisté à identifier le nombre de frontières ayant coïncidé à la fois au niveau du début et de la fin du segment. Ce premier calcul a montré qu'il n'existait que deux segmentations présentant une correspondance stricte de ce type, toutes deux entre l'annotateur 1 et l'annotateur 3. Pour toutes les autres comparaisons, aucune frontière de ce type n'a été observée.

En raison de ce faible niveau de concordance, il a été décidé de ne pas poursuivre l'analyse en se basant uniquement sur des correspondances strictes début–fin, et de se concentrer plutôt sur l'ensemble des frontières concordantes, indépendamment de l'alignement exact des bornes initiales et finales.

Après comparaison des annotations entre les différents annotateurs, les résultats suivants ont été obtenus :



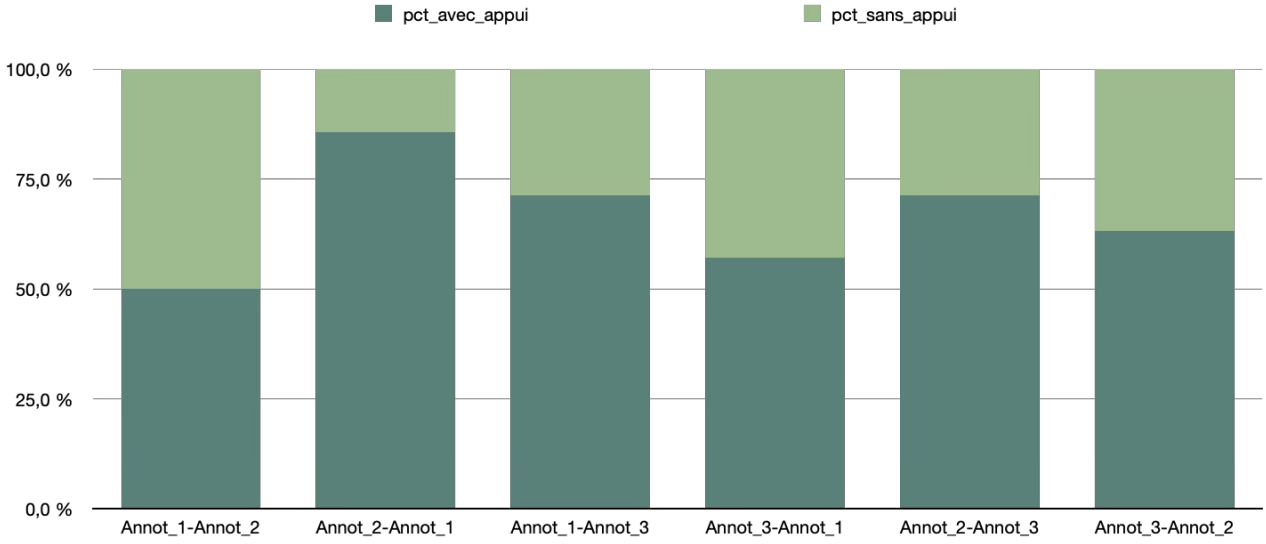
Graphique 1 : pourcentage d'accord entre annotateurs pour la couche « BDU »

Comme l'illustre la graphique 1 le taux de concordance le plus élevé est observé entre l'annotateur 1 (sans expérience préalable en annotation ni en LSF) et l'annotateur 3 (ayant une expérience dans les deux domaines). À l'inverse, le niveau d'accord le plus faible est constaté entre l'annotateur 2 et l'annotateur 3, bien que ces deux participants possèdent une expérience en annotation.

Cette divergence peut s'expliquer en partie par le fait que l'annotateur 2 a identifié un nombre total de BDU plus faible (11), ce qui réduit les probabilités de correspondance avec les frontières proposées par les autres annotateurs.

En termes absolus, 4 correspondances ont été observées entre l'annotateur 1 et l'annotateur 2, 9 entre l'annotateur 1 et l'annotateur 3, et 4 entre l'annotateur 2 et l'annotateur 3.

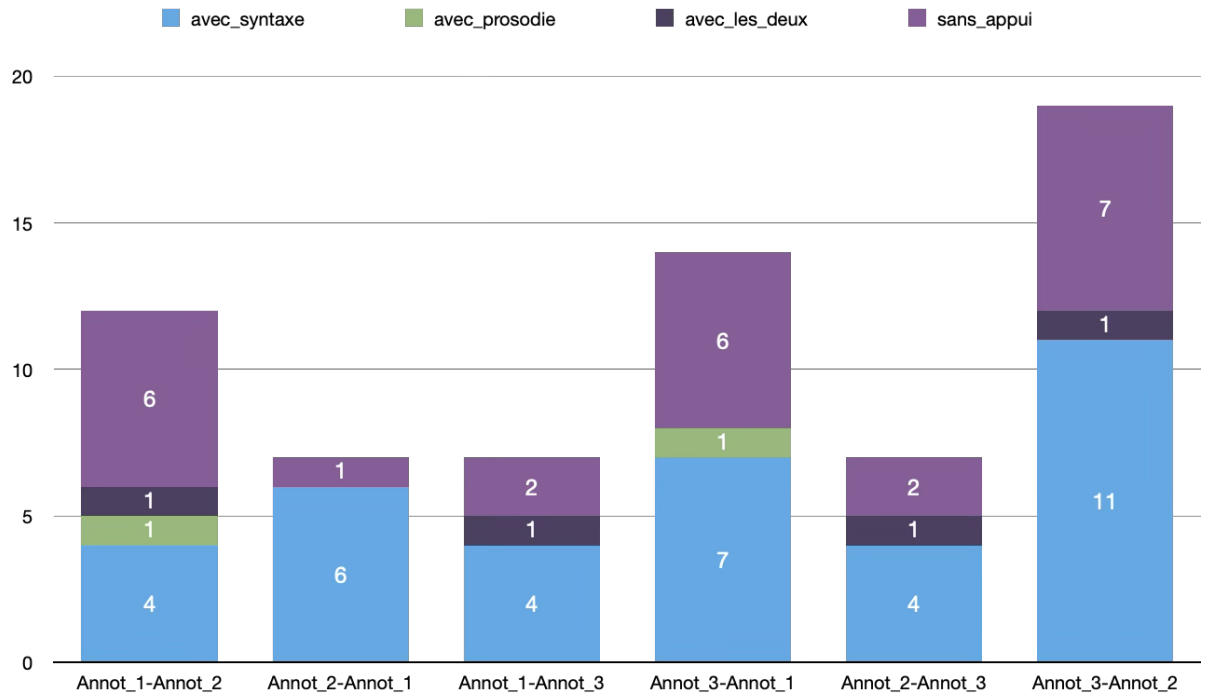
Par la suite, il a été analysé si les frontières non identifiées comme BDU coïncidaient néanmoins avec d'autres éléments présents dans le document de référence, tels qu'une piste prosodique ou syntaxique. À partir de cette analyse, les résultats suivants ont été obtenus :



Graphique 2 : représentation du pourcentage de frontières non concordantes ayant ou non un appui syntaxique ou prosodique préalablement identifié par l'autre annotateur.

Comme on peut l'observer sur la graphique 2 , dans l'ensemble des cas analysés, au moins la moitié des frontières coïncident avec une piste syntaxique, prosodique ou avec une combinaison des deux.

Une analyse plus détaillée des données permet de mettre en évidence les différentes distributions. Les valeurs numériques indiquées à l'intérieur des barres correspondent au nombre exact d'annotations identifiées pour chaque catégorie de segment.



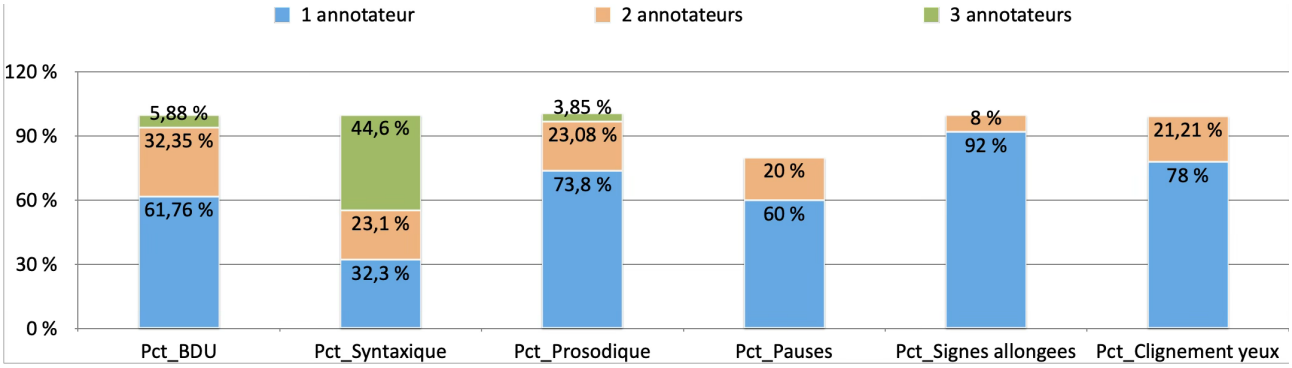
Graphique 3 : représentation de la distribution des frontières non concordantes selon la présence d'un appui syntaxique, prosodique, des deux, ou d'aucun appui. Les valeurs indiquées dans les barres correspondent au nombre de frontières par catégorie.

Comme on peut l'observer dans la graphique 3, dans la majorité des cas, c'est la piste syntaxique qui a été identifiée ; toutefois, en l'absence d'une piste prosodique, la BDU n'a pas été consolidée. Au total, seuls deux cas présentent une piste prosodique sans présence d'une piste syntaxique. De même, quatre cas ont

été identifiés dans lesquels les deux pistes coïncident, mais où l'annotateur a décidé (pour des raisons de critère méthodologique ou par omission involontaire) de ne pas marquer la frontière de BDU. Il convient néanmoins de souligner qu'un pourcentage non négligeable d'annotations ne présente aucune piste associée.

13.1.2 Consensus sur chaque couche d'annotation

Enfin, il a été décidé d'examiner le niveau de consensus des trois annotateur pour chaque couche d'annotation. Les résultats obtenus sont présentés ci-dessous.



Graphique 4 : représentation en pourcentage du consensus entre les trois annotateurs pour chaque couche d'annotation.

Dans la graphique 4, on observe qu'à l'exception de la segmentation syntaxique, toutes les autres couches présentent une majorité de frontières identifiées par un seul annotateur. Ce phénomène est particulièrement marqué pour les signes allongés, où seulement 8 % des cas (2 frontières) ont été validés par deux annotateurs.

De manière générale, aucune des couches (pauses, signes allongés, clignements des yeux) ne présente de cas dans lesquels les trois annotateurs convergent simultanément sur la même frontière. Enfin, le 3,85 % (2 frontières) observé dans la colonne *pct_prosodique* indique un consensus entre les trois annotateurs. Toutefois, après analyse de ces deux frontières, il apparaît qu'elles reposent sur des indices prosodiques différents.

13.2 Résultats des annotations avec Discut

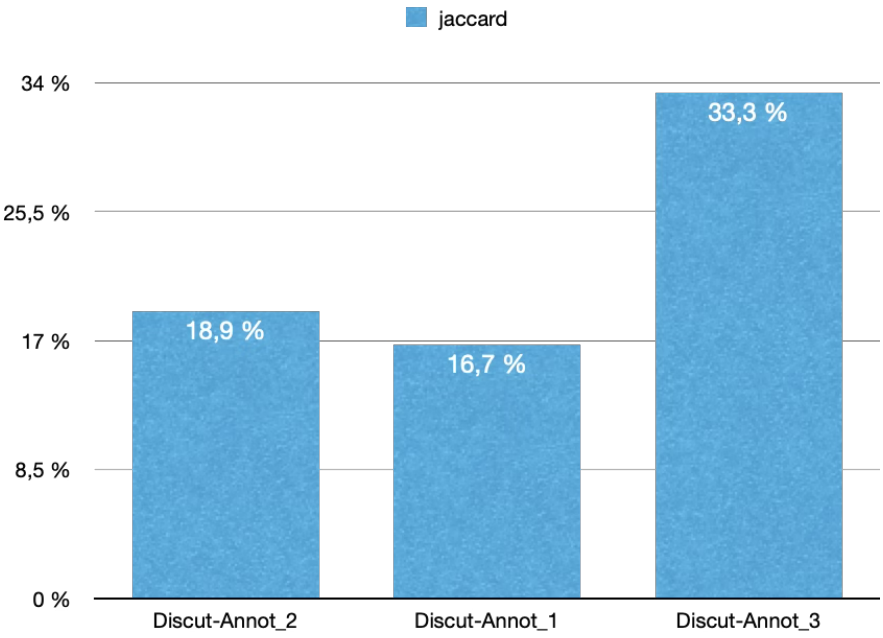
Le système DisCut a identifié 33 frontières de BDU, contre 16 pour l'annotateur 1, 11 pour l'annotateur 2 et 26 pour l'annotateur 3.

Ici, nous avons comparé les annotations des trois annotateurs humains avec celles obtenues en suivant les segmentations proposées par DisCut sur les traductions en français de la langue des signes française, lesquelles ont servi de guide pour l'annotation réalisée sous ELAN.

Dans un premier temps, comme pour la comparaison entre les annotateurs humains, Nous avons analysé si certains segments coïncidaient à la fois au niveau du début et de la fin. Dans ce cadre, il a été observé que DisCut et l'annotateur 1 ne présentaient aucune correspondance de ce type. Avec l'annotateur 2, une seule correspondance a été identifiée, tandis que l'annotateur 3 en présentait trois.

Même si le nombre de correspondances est légèrement plus élevé que celui observé entre les annotateurs humains, qui n'ont trouvé qu'une seule correspondance de ce type, le faible nombre de cas de ce genre a conduit, conformément à la démarche adoptée pour les annotateurs humains, à privilégier une analyse de

la concordance générale des frontières sans exiger une correspondance stricte des bornes initiales et



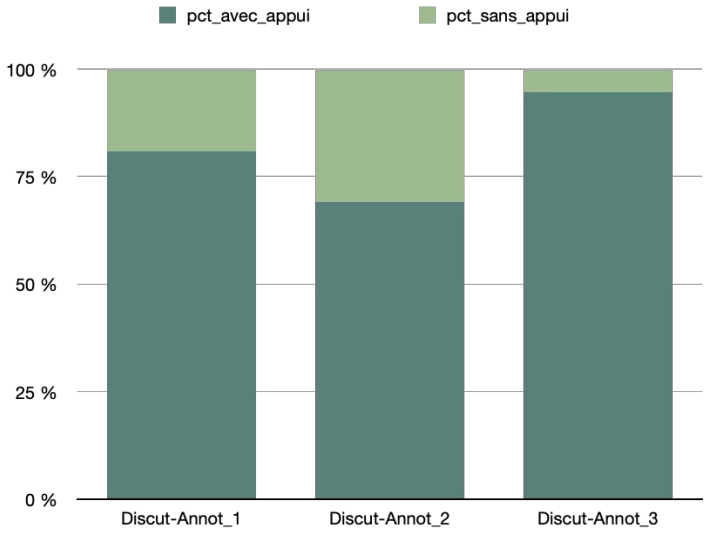
Graphique 5 : représentation en pourcentage des frontières concordantes entre DisCut et les trois annotateurs humains.

finales. Cette approche a permis d'obtenir les résultats suivants.

Comme on peut l'observer, le plus fort taux de concordance est obtenu avec l'annotateur 3, avec un total de 14 correspondances. Viennent ensuite l'annotateur 2 et l'annotateur 1, chacun avec 7 correspondances.

Cependant, bien que le nombre absolu de correspondances soit identique pour les annotateurs 1 et 2, le pourcentage final diffère, car l'annotateur 2 a identifié un nombre total de BDU plus faible, ce qui réduit mécaniquement la probabilité de correspondance avec les segments produits par le système.

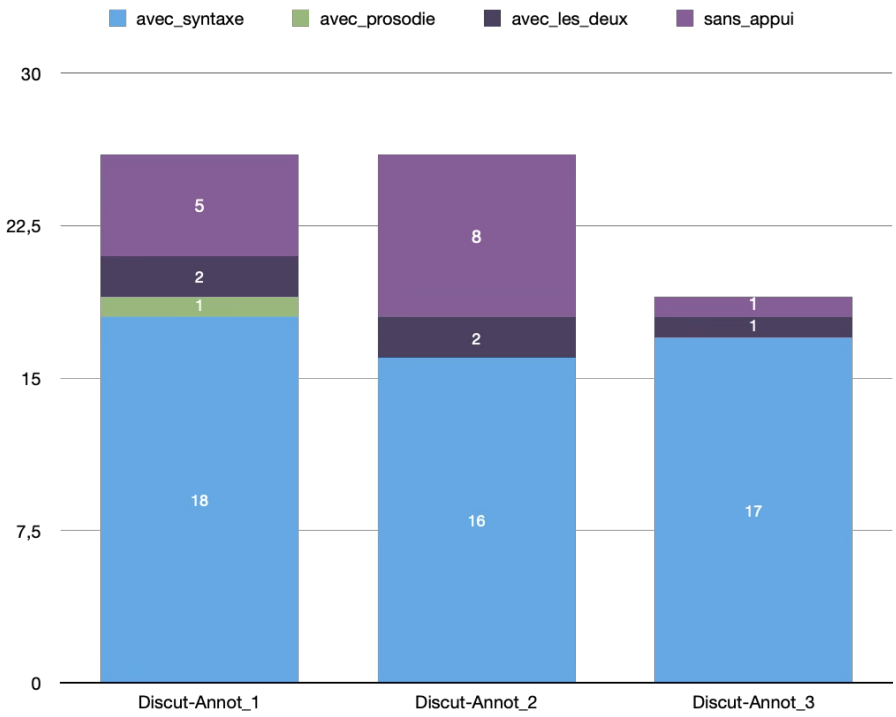
Par la suite, les segmentations de BDU produites par DisCut et n'ayant pas de correspondance avec celles des annotateurs humains ont été analysées. Dans ces cas, il a été vérifié si les évaluateurs avaient identifié au moins une piste prosodique ou syntaxique sur le même intervalle temporel, ce qui a conduit aux résultats suivants :



Graphique 6: représentation du pourcentage de frontières non concordantes ayant ou non un appui syntaxique ou prosodique préalablement identifié par l'autre annotateur.

Les résultats de la graphique 6, montrent que, dans au moins 69 % des cas, les trois annotateurs avaient identifié une piste prosodique ou syntaxique, correspondant à la segmentation produite par DisCut.

Une analyse plus détaillée des données permet de mettre en évidence les différentes distributions. Les valeurs numériques indiquées à l'intérieur des barres correspondent au nombre exact d'annotations identifiées pour chaque catégorie de segment.



Graphique 7 : représentation de la distribution des frontières non concordantes selon la présence d'un appui syntaxique, prosodique, des deux, ou d'aucun appui. Les valeurs indiquées dans les barres correspondent au nombre de frontières par catégorie.

Comme on peut l'observer dans la graphique 7 , de manière similaire à ce qui avait été constaté dans la comparaison entre annotateurs humains, la piste syntaxique est celle qui s'aligne le plus fréquemment avec les BDU non identifiées par les évaluateurs. À l'inverse, les occurrences de pistes prosodiques isolées sont très rares ; un seul cas de ce type a été relevé.

L'observation la plus notable concerne toutefois la présence de cinq cas dans lesquels les pistes syntaxique et prosodique coïncident simultanément, sans que l'annotateur n'ait jugé pertinent de marquer la frontière de BDU.

13.3 Analyse générale

L'analyse globale des résultats met en évidence plusieurs tendances structurantes concernant les niveaux de consensus entre annotateurs et entre les annotateurs et le système DisCut.

En premier lieu, la couche syntaxique apparaît comme celle présentant le niveau d'accord le plus élevé de manière générale. Ce résultat est particulièrement intéressant dans la mesure où la segmentation syntaxique constitue, en théorie, l'une des tâches les plus complexes et avait été identifiée comme la plus difficile à réaliser par les annotateurs.

À l'inverse, la couche prosodique apparaît comme la moins stable et la moins consensuelle. Elle présente les taux de correspondance les plus faibles et est également la couche la moins fréquemment associée à des

frontières lors de l'analyse des segments non appariés. Cette tendance peut s'expliquer par le fait qu'il est difficile d'observer avec clarté la vidéo lors de l'annotation, ce qui rend l'identification de certains marqueurs prosodiques incertaines.

Le nombre total de frontières identifiées par chaque annotateur dans les différentes couches d'annotation confirme que la segmentation syntaxique est nettement plus fournie (43, 39 et 56 frontières selon les annotateurs) que la segmentation prosodique (25, 18 et 25). De manière similaire, les BDU varient fortement selon les annotateurs (16, 11 et 22), ce qui contribue également aux différences de taux de concordance observées.

Un autre point intéressant concerne la couche prosodique. L'annotateur 3, locutrice de la LSF, identifie un nombre nettement plus élevé de signes allongés (16 contre 4 et 7 pour les autres annotateurs), ce qui suggère une sensibilité accrue à certains phénomènes gestuels. À l'inverse, les clignements des yeux sont principalement identifiés par les annotateurs non locuteurs (22 et 14 contre 4), ce qui peut indiquer des stratégies d'annotation différentes selon l'expérience linguistique et perceptive.

Il est également notable que l'annotateur 1, en dépit de l'absence d'expérience préalable en annotation, en linguistique ou en langue des signes française, a pu identifier un nombre significatif de pistes, ce qui suggère que son apport a été pertinent pour l'étude. Par ailleurs, la paire Annotateur 1–Annotateur 3 présente le plus haut niveau de concordance entre annotateurs humains (30 %), ce qui est particulièrement intéressant dans la mesure où l'annotateur 3 possède une expérience à la fois en langue des signes et en annotation. Ces résultats semblent indiquer que l'inclusion de profils variés, y compris non experts, constitue une approche pertinente pour ce type de tâche.

Dans ce cadre, il est également intéressant de noter que DisCut présente globalement un niveau d'accord avec les annotateurs humains supérieur à celui observé entre certaines paires d'annotateurs humains. Toutefois, ce résultat doit être nuancé par le fait que DisCut produit un nombre plus élevé de frontières de BDU (33 contre 16, 11 et 22 pour les annotateurs humains), ce qui augmente mécaniquement les possibilités de recouvrement.

Par ailleurs, contrairement à ce qui avait été initialement supposé, l'annotateur 3, pourtant locutrice de la langue des signes française, n'est pas celui qui diverge le plus de DisCut. Au contraire, il présente le niveau d'accord le plus élevé avec le système (33,3 %). Cela est particulièrement intéressant dans la mesure où son annotation syntaxique repose principalement sur l'observation directe du signal signé, alors que les gloses ont été utilisées comme support secondaire pour guider ses décisions. À l'inverse, l'annotateur 2, qui ne possède pas de compétence en langue des signes mais une expérience en annotation et en linguistique, et qui a réalisé la segmentation syntaxique uniquement à partir des gloses et traductions (approche plus proche de celle de DisCut), obtient un taux de concordance plus faible (16,7 %). Cela suggère que la segmentation produite par DisCut peut converger davantage avec une segmentation fondée sur l'observation du signal signé.

Enfin, le niveau d'accord global sur les BDU reste relativement faible dans l'ensemble des comparaisons, avec un maximum de 33,3 % avec DisCut et 30 % entre annotateurs humains. Cette faible stabilité confirme que cette annotation dépend fortement de choix interprétatifs individuels et souligne la nécessité d'une standardisation plus rigoureuse ainsi que d'une révision et d'un ajustement de la grille d'annotation utilisée pour les annotateurs humains.

CHAPITRE III

14. Automatisation de DisCut

14.1 Justification

À la suite de l'analyse comparative entre les annotations des annotateurs humains et celles produites par DisCut, il a été observé que, dans certains cas, le taux d'accord entre les annotateurs humains et le système automatique était supérieur au taux d'accord entre les annotateurs humains eux-mêmes. Pour cette raison, il a semblé pertinent d'explorer la possibilité d'une automatisation complète de la procédure de segmentation.

Cette approche permettrait de produire une première segmentation en BDU de manière largement automatisée. Une telle segmentation ne nécessiterait ensuite qu'une phase de vérification et de correction manuelle afin de s'assurer de la cohérence et de la qualité des résultats obtenus.

Le gain de temps potentiel est considérable. En effet, la segmentation manuelle basée sur les résultats fournis par DisCut reste une tâche longue à réaliser. De plus, lorsque l'annotation est effectuée de manière séquentielle (segmentation prosodique, puis syntaxique, puis BDU), le temps de traitement augmente significativement.

Les annotateurs ayant participé à cette étude, mentionnés dans le chapitre 2, ont consacré entre une heure et demie et deux heures à la segmentation d'environ une minute de discours signé. À l'inverse, une procédure automatisée permettrait de générer une première proposition de segmentation en quelques secondes seulement. L'intervention humaine pourrait alors se limiter à une tâche de validation et d'ajustement, réalisable en quelques minutes plutôt qu'en plusieurs heures.

Dans cette perspective, l'automatisation ne viserait pas à remplacer entièrement l'annotateur humain, mais plutôt à lui fournir une base de travail permettant d'accélérer considérablement le processus d'annotation tout en conservant un contrôle humain sur les résultats finaux.

14.2 Méthodologie

14.2.1 Processus actuel de segmentation des BDU à partir des résultats de DisCut

Actuellement, le processus de segmentation des BDU à partir des résultats de DisCut repose sur plusieurs étapes manuelles. Il est d'abord nécessaire d'extraire les traductions françaises depuis ELAN, puis de les soumettre à DisCut afin d'obtenir une segmentation automatique. Les résultats produits doivent ensuite être réintégrés dans le processus d'annotation et utilisés comme support pour la relecture des gloses et des traductions, afin de guider la segmentation dans ELAN.

Afin d'évaluer dans quelle mesure cette tâche pourrait être automatisée, nous avons cherché à identifier les différentes étapes et les décisions réalisées par l'annotateur lorsqu'il exploite les segmentations produites par DisCut pour les reporter dans ELAN. L'objectif est de déterminer quelles opérations reposent sur des règles explicites susceptibles d'être reproduites par un programme et lesquelles nécessitent au contraire une interprétation humaine.

À titre d'exemple, l'extrait suivant permet d'illustrer la démarche actuellement utilisée pour transférer manuellement les segmentations de DisCut dans ELAN. L'analyse de ce cas concret permet également d'évaluer si ce processus pourrait être reproduit, totalement ou partiellement, par un programme informatique.

L'exemple suivant est extrait de l'« histoire du cheval » du corpus LS-COLIN. Le passage analysé, disponible à l'adresse suivante : https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82 , correspond à l'intervalle compris entre 00:00:04 et 00:00:54.

(11) Exemple

Extrait de Traduction de « histoire du cheval »:

« C'est l'histoire… d'un cheval. Il galope, et soudain, il freine, surpris par la barrière. Il aperçoit une vache qui rumine, ils se regardent. Le cheval a envie de sauter, comment faire… "Attends voir…" Il réfléchit. Il recule, prend son élan et saute, mais il heurte la barrière et tombe, et se retrouve les pattes étendues… »

Extrait des Gloses correspondantes de « histoire du cheval » :

HISTOIRE CHEVAL galoper freiner barrière APERCEVOIR vache être sur ses sabots ruminer regarder (vers le cheval) bon ! ENVIE sauter barrière COMMENT sauter attendre voir reculer galoper sauter heurter tomber gésir au sol AVOIR MAL

Si l'on considère uniquement les segments en bleu, on obtient la représentation suivante de l'organisation des couches de gloses et de traductions dans ELAN :

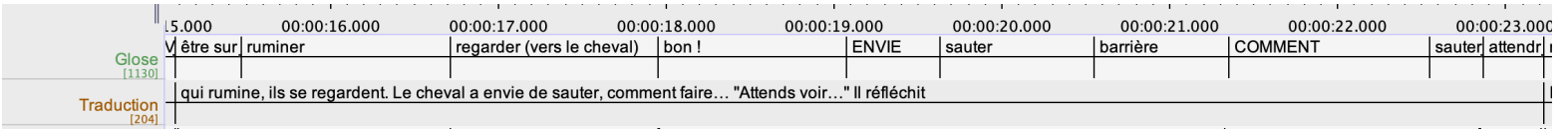


Figure 2 : représentation des traductions et des gloses dans ELAN.

On observe dans la figure 2 que le segment de traduction s'étend de la glosse « être sur ses sabots » jusqu'à la glosse « Attends voir… ». Toutefois, la segmentation produite par DisCut sur la traduction est la suivante :

[C'est l'histoire… d'un cheval.] [Il galope,] [et soudain,] [il freine,] [surpris par la barrière.] [Il aperçoit une vache] [qui rumine,] [ils se regardent.] [Le cheval a envie de sauter,] [comment faire…] ["Attends voir…"] [Il réfléchit] [Il recule, prend son élan] [et saute,] [mais il heurte la barrière] [et tombe,] [et se retrouve les pattes étendues.]

Si l'on tente de suivre les gloses afin de retrouver leur équivalent dans la segmentation produite par DisCut sur la traduction, on obtient une représentation de ce type dans ELAN :

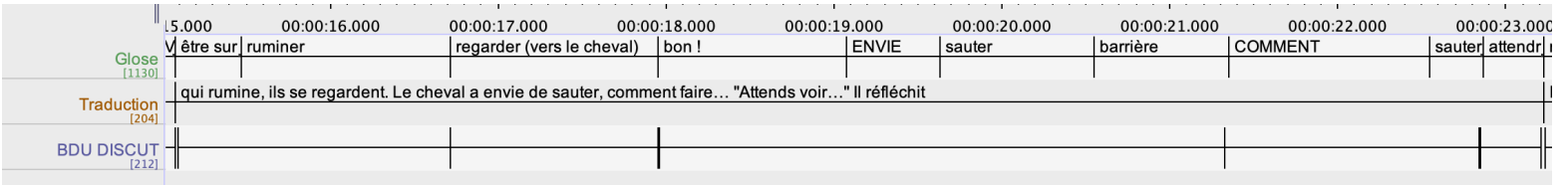


Figure 3 : représentation des traductions, des gloses et de la segmentation réalisées sous ELAN en suivant les segmentations proposées par DisCut.

On observe que la traduction doit être segmentée en unités plus fines pour pouvoir être alignée avec les sorties de DisCut. En suivant cette segmentation, les gloses peuvent être restructurées de la manière suivante :

[être sur ses sabots ruminer] [regarder (vers le cheval)] [bon ! ENVIE sauter barrière] [COMMENT sauter] [attendre voir]

Cependant, bien que cette tâche puisse sembler intuitive pour un annotateur humain, elle pourrait être plus complexe à formaliser pour un système automatique. Certaines unités présentes dans les gloses,

comme *« bon ! »* ou *« barrière »*, ne sont pas explicitement présentes dans la traduction de notre exemple . À l'inverse, des expressions comme *« comment faire »*apparaissent dans la traduction alors que les gloses contiennent *« comment sauter »*. De plus, certains éléments, comme *« il réfléchit »*, ne sont pas directement réalisés dans les gloses mais sont implicites dans des segments tels que *« attendre voir »*.

Pour cette raison, même sur un extrait court et relativement simple, la conception d'un système capable de gérer automatiquement ces correspondances et ces nuances s'avère complexe. Néanmoins, une première expérimentation est proposée dans cette section afin d'évaluer la faisabilité d'une telle approche.

L'objectif final serait de construire un pipeline complet comprenant :

1. l'extraction automatique des traductions du corpus ELAN ;
2. l'envoi des traductions à DisCut pour segmentation automatique ;
3. la récupération des résultats de segmentation ;
4. la création d'une nouvelle couche d'annotation dans ELAN basée sur ces résultats ;
5. l'alignement avec les gloses et les traductions afin de guider la segmentation;
6. Création d'un nouveaux document .eaf contenant la segmentation automatique.

Dans un premier temps, seules les étapes 4 et 5 seront explorées, afin de tester la faisabilité de l'approche avant d'envisager l'implémentation complète du pipeline.

14.2.2 Conception et réalisation d'un annotateur automatique basé sur la segmentation de DisCut

Pour cette partie du processus, nous nous sommes appuyés sur Claude afin de développer le programme, dans la mesure où, comme expliqué précédemment, il s'agissait d'une tâche relativement complexe que nos compétences actuelles ne permettaient pas de réaliser entièrement de manière autonome.

Chaque étape du développement a été explicitée de manière progressive, en formulant des requêtes détaillées afin d'obtenir un code correspondant aux besoins spécifiques de l'analyse. Plusieurs versions du programme ont ainsi été générées, puis évaluées successivement. Certaines fonctionnalités ont été modifiées, ajustées ou supprimées en fonction de leur pertinence pour l'objectif de l'étude.

Le programme *(annotateur*_automatique.*py)* prend en entrée un fichier texte (.txt) contenant les résultats de la segmentation produite par DisCut, ainsi qu'un fichier ELAN (.eaf) dans lequel se trouvent les couches de gloses et de traductions. Il nécessite également la définition de l'intervalle temporel (exprimé en minutes ou en secondes) correspondant au passage du discours à traiter, afin de limiter l'analyse à la portion pertinente du corpus et de permettre une segmentation ciblée de ce segment spécifique.

Le programme final utilise les bibliothèques suivantes :

xml.etree.ElementTree (bibliothèque standard de Python) : elle permet de lire et d'écrire des fichiers XML. Les fichiers .eaf d'ELAN étant structurés en XML, cette bibliothèque est utilisée pour analyser l'ensemble de la structure du fichier, notamment les timeslots, les tiers et les annotations.
unicodedata (bibliothèque standard) : elle est utilisée pour la normalisation des caractères Unicode. Dans ce travail, elle a notamment permis de résoudre certains problèmes liés à des ligatures typographiques, identifiés lors des tests sur le corpus (en particulier dans l'exemple de la recette de cuisine), où certaines configurations de caractères provoquaient des erreurs d'exécution.
re (bibliothèque standard) : elle fournit des outils d'expressions régulières. Elle a été utilisée pour extraire les segments encadrés par des crochets [ ] dans la sortie de DisCut, ainsi que pour nettoyer les espaces et la ponctuation.
rapidfuzz (bibliothèque externe, installée pour ce projet) : elle permet de comparer des chaînes de caractères en tenant compte de petites variations. Elle implémente des méthodes de *fuzzy matching* (méthode permettant de comparer deux textes et de mesurer leur similarité même lorsqu'ils ne sont pas identiques) , ce qui rend possible l'identification de correspondances entre les segments produits par DisCut

et le contenu du tier « Traduction » du fichier .eaf, même en présence de différences orthographiques ou de tokenisation.

Le programme effectue les étapes suivantes :

Dans un premier temps, il lit le fichier ELAN (.eaf) et construit un dictionnaire des timeslots. Cette étape permet d'accéder à l'ensemble des tiers du fichier et de récupérer leurs annotations associées avec leurs temps exacts de début et de fin.

Ensuite, le programme construit un texte de référence en concaténant toutes les annotations du tier « Traduction » situées dans la fenêtre temporelle définie par l'utilisateur. Le résultat est un texte continu représentant la traduction française du discours en LSF. Ce texte est ensuite normalisé (mise en minuscules, suppression de la ponctuation et normalisation Unicode en NFKC) afin de faciliter la comparaison avec la sortie de DisCut.

Dans une étape suivante, le programme extrait les segments produits par DisCut. Ceux-ci sont délimités par des crochets [ ] dans la sortie du système. Une expression régulière permet d'extraire le contenu de chaque segment, produisant ainsi une liste ordonnée de segments textuels (par exemple : « Il galope , », « surpris par la barrière . », etc.).

L'étape la plus complexe correspond à l'alignement entre texte et temps. Pour chaque segment de DisCut, le programme cherche à déterminer sa position dans la vidéo, c'est-à-dire les timestamps du fichier ELAN auxquels il correspond. Étant donné que le tier « Traduction » n'est pas aligné mot à mot avec les gloses mais constitue une paraphrase globale structurée en blocs, l'alignement direct n'est pas possible.

Le programme procède donc d'abord à une recherche exacte du segment dans le texte global de la traduction, puis, en cas d'échec, applique une méthode de *fuzzy matching*. Une fois le segment localisé, sa position relative dans le texte global est calculée (sous forme de ratio compris entre 0 et 1). Ce ratio est ensuite projeté sur le nombre total de gloses du discours afin d'estimer la glosse correspondant au début du segment. Les timestamps finaux sont récupérés directement à partir de cette glosse dans le tier correspondant du fichier ELAN.

Un curseur séquentiel garantit que les segments sont traités dans l'ordre du discours : le programme ne revient jamais en arrière dans la liste des gloses, ce qui permet d'éviter des alignements incohérents.

Les segments dont le score de similarité est inférieur à un seuil de 65 % sont ignorés. Ce choix correspond à la stratégie adoptée lors de l'annotation manuelle, consistant à écarter les segments courts et peu informatifs produits par DisCut (par exemple « bon » ou « alors »), qui ne trouvent pas de correspondance claire dans les gloses.

Enfin, le programme crée un nouveau tier intitulé « BDU DISCUT AUTOMATIQUE » dans la hiérarchie du fichier ELAN. Pour chaque segment aligné, il génère une annotation de type ALIGNABLE_ANNOTATION sur ELAN en utilisant les timestamps de la glosse correspondante, et le texte du segment DisCut comme contenu. Les identifiants des annotations et des timeslots sont incrémentés à partir des valeurs maximales présentes dans le fichier afin d'éviter tout conflit. Le fichier original n'est jamais modifié : une copie est enregistrée avec le suffixe _DISCUT_AUTO.eaf.

14.2.3 Limitations importantes

Le programme produit une approximation et non une segmentation et annotation parfaitement fiable. La principale limite réside dans le fait que le tier « Traduction » constitue une paraphrase libre des gloses, sans correspondance mot à mot entre les deux niveaux d'annotation. Cette absence d'alignement direct rend l'appariement entre les segments de DisCut et les données du corpus intrinsèquement incertain.

Pour cette raison le résultat du programme est ainsi conçu comme une aide à l'annotation, destinée à être vérifiée et corrigée manuellement dans ELAN, et non comme un substitut à l'annotation humaine.

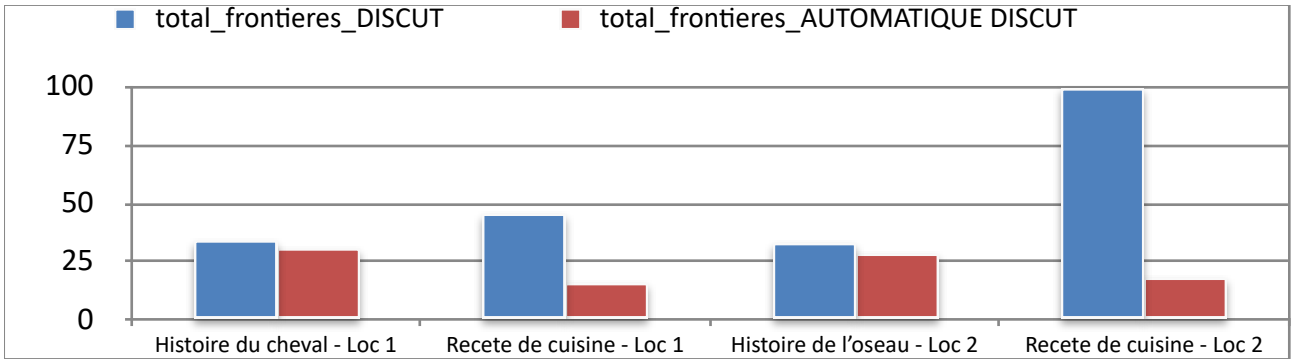
14.3 Résultats de l'annotateur automatique sur ELAN

Le programme a été testé sur quatre discours différents, produits par deux locuteurs distincts : (https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82)

1. « Histoire du cheval » (00:00:04 – 00:00:54), locuteur 1
2. « Recette de cuisine » (00:05:45 – 00:07:06), locuteur 1
3. « Histoire de l'oiseau » (00:27:09 – 00:28:08), locuteur 2
4. « Recette de cuisine » (00:31:04 – 00:34:04), locuteur 2

Après avoir réalisé la segmentation automatique de ces quatre discours, chaque résultat a été comparé à l'aide du "*programme_2.py*" même programme que celui utilisé pour les évaluations présentées dans la section 12.2 (Comparaison des segmentations).

Les segmentations manuelles, construites à partir de l'exploitation directe des résultats de DisCut, ont ainsi été comparées aux segmentations produites par le système automatique d'alignement et de segmentation. Cette comparaison a permis d'évaluer les différences entre les deux approches en termes de nombre total de frontières identifiées par chacun des systèmes. Les résultats obtenus sont présentés dans la graphique suivant :



Graphique 8 : représentation du nombre total de frontières identifiées lors de la segmentation manuelle basée sur les résultats de DisCut et de la segmentation automatique à partir des mêmes résultats ; les colonnes correspondent aux différents discours.

Comme on peut le constater sur la graphique 8, dans les cas de « Histoire du cheval » et de « Histoire de l'oiseau », le nombre de segments produits est pratiquement similaire entre les deux approches, avec seulement une différence de 3 et 5 frontières de moins identifiées par le système automatique DisCut. En revanche, les écarts les plus importants apparaissent dans les deux « recettes de cuisine », en particulier celle du locuteur 2, où la différence atteint 85 frontières. Cette variation suggère une plus grande complexité de segmentation pour ce type de discours.

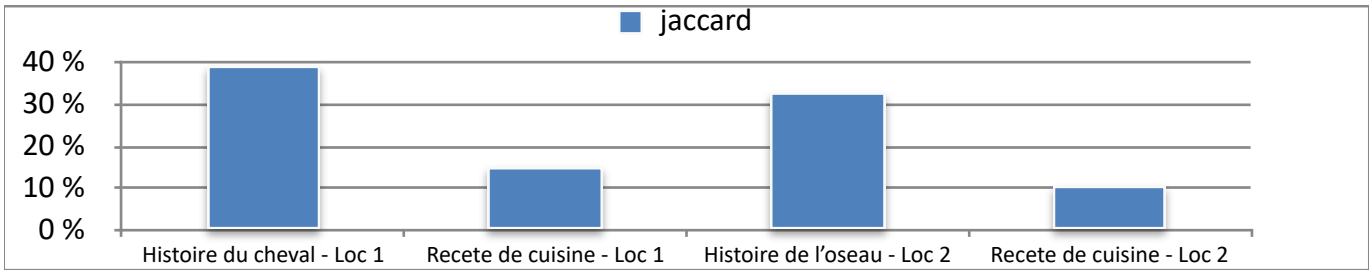
À partir de ces premiers résultats, une première analyse a consisté à identifier le nombre de frontières ayant été détectées par les deux annotations, ainsi que celles pour lesquelles une correspondance stricte a été observée à la fois au niveau du début et de la fin du segment. Les résultats sont présentés dans le tableau suivant.

Discours	Nombre total de frontières concordantes	Concordance des frontières (début et fin)
Histoire du cheval - Loc 1	18	12
Recete de cuisine - Loc 1	8	3
Histoire de l'oseau - Loc 2	15	5
Recete de cuisine - Loc 2	11	6

Tableau 3 : nombre total de frontières concordantes et nombre total de segments présentant une concordance stricte entre le début et la fin.

Comme on peut l'observer sur le tableau 3, le nombre de concordances obtenues lorsque l'on exige une correspondance simultanée des bornes initiales et finales reste relativement faible. Pour cette raison, et conformément aux analyses précédentes, les analyses suivantes se concentreront principalement sur les frontières concordantes, indépendamment du fait qu'elles soient alignées strictement au début et à la fin.

Sur cette base, les pourcentages de concordance ont ensuite été calculés afin de mieux visualiser les résultats. La graphique suivante en présente la distribution:



Graphique 9 : représentation du pourcentage de concordance des frontières entre DisCut annoté manuellement et automatiquement ; les colonnes représentent les différents discours.

On observe ici que le pourcentage d'accord reste inférieur à 40 % dans tous les cas. Le niveau de concordance le plus élevé est obtenu pour l'histoire du cheval du locuteur 1, avec 39,1 %, suivi de l'histoire de l'oiseau du locuteur 2, avec 32,6 %.

En revanche, les résultats sont nettement plus faibles pour les recettes de cuisine, avec 15,1 % d'accord pour le locuteur 1 et 10,3 % pour le locuteur 2.

Enfin, le score F1 du système d'annotation automatique a été calculé pour chacun des discours. Les résultats obtenus sont présentés ci-dessous :

Discours	Précision	Rappel	F1-score
Histoire du cheval	60,0 %	52,9 %	56,3 %
Recette de cuisine (Loc 1)	50,0 %	17,8 %	26,2 %
Histoire de l'oiseau	53,6 %	45,5 %	49,1 %
Recette de cuisine (Loc 2)	61,1 %	11,0 %	18,6 %

Tableau 4 : représentation de la précision, du rappel et du score F1 du système d'annotation automatique basé sur les résultats de DisCut pour chaque discours.

Les résultats obtenus montrent des différences nettes selon le type de discours. Dans les discours narratifs (*Histoire du cheval* et *Histoire de l'oiseau*), le système automatique atteint des valeurs de F1-score plus élevées, ce qui indique un meilleur équilibre entre précision et rappel. Dans ces cas, le rappel, défini comme la proportion de frontières ou de gloses réelles correctement récupérées par le système par rapport à l'ensemble des éléments de référence, est relativement plus élevé, ce qui reflète une meilleure capacité du système à couvrir la structure complète du discours.

Ce comportement peut s'expliquer par la nature des discours narratifs, qui présentent une organisation séquentielle, linéaire et cohérente des événements. Cette structure facilite l'alignement entre la traduction et les gloses, car il existe une correspondance plus stable entre l'ordre du discours segmenté par DisCut et la progression des gloses dans le fichier ELAN. Par conséquent, le modèle basé sur une propagation

progressive (curseur textuel et projection vers la séquence de gloses) s'adapte mieux à ce type de données, réduisant les erreurs d'omission.

En revanche, dans les discours correspondant à des recettes de cuisine, les résultats montrent une performance significativement plus faible, en particulier en termes de rappel. Dans ces cas, le système ne parvient pas à récupérer une partie importante des gloses présentes dans la segmentation de référence, ce qui entraîne un nombre plus élevé de faux négatifs.

Cette dégradation des performances s'explique par la structure discursive des recettes, caractérisée par une forte densité d'instructions, des répétitions lexicales et des variations dans la granularité des actions. Contrairement aux discours narratifs, la correspondance entre la traduction segmentée et les gloses n'est ni linéaire ni uniforme, ce qui affecte directement l'hypothèse du système basée sur la proportionnalité entre la position textuelle et la séquence de gloses.

Dans l'ensemble, ces résultats montrent que le système automatique est plus robuste dans des contextes narratifs structurés, tandis qu'il présente des limites importantes dans des discours plus fragmentés et non linéaires, comme les recettes de cuisine.

Un modèle de langage aurait pu être intégré au système afin d'améliorer la correspondance sémantique entre les segmentations produites par DisCut et les gloses, notamment en permettant une meilleure prise en compte des nuances entre les traductions et les gloses. Cependant, dans le cadre de ce travail et dans une première phase d'exploration, une approche plus simple a été choisie. L'objectif était de garantir la reproductibilité des résultats, en évitant des variations possibles à chaque exécution du programme, et d'assurer une meilleure traçabilité des annotations produites.

15. Mise en place d'un pipeline pour l'automatisation de l'annotation et de la segmentation

15.1 Méthodologie du pipeline d'annotation et de segmentation

Après une première série de tests de l'annotateur automatique, le pipeline envisagé au début de ce chapitre a été développé afin d'automatiser l'ensemble du processus de segmentation et d'annotation dans ELAN.

Le pipeline s'exécute dans un environnement contenant toutes les bibliothèques ainsi que les versions de Python nécessaires au fonctionnement de DisCut. Il prend en entrée un fichier d'annotation au format ELAN (.eaf), qui doit contenir un tier intitulé « Traduction ».

Le traitement est entièrement automatisé et se déroule en plusieurs étapes successives. Dans un premier temps, le système extrait les données du fichier ELAN fourni en entrée en lisant les annotations présentes dans le tier *Traduction*. Ces données constituent le matériau de base utilisé pour la segmentation.

Dans un second temps, une étape de détection automatique des discours est appliquée. Les annotations sont regroupées en unités discursives en fonction des pauses temporelles observées. Un seuil de 2000 ms est utilisé comme critère de segmentation : toute interruption supérieure à cette durée est interprétée comme une frontière entre deux discours distincts.

Une fois les discours identifiés, une étape de préparation des données est réalisée afin de les adapter au format attendu par DisCut. Chaque discours est exporté sous la forme d'un fichier texte brut (.txt), organisé dans la structure requise par le système, à savoir : DISCUT/data/NOM/NOM.txt

Ces fichiers texte sont ensuite utilisés comme entrée du segmentateur automatique DisCut. Le script principal de DisCut est exécuté pour chaque discours et produit une segmentation structurée du texte en unités discursives.

Les résultats de segmentation sont ensuite extraits des fichiers _brac_meta.txt générés par DisCut. Ces fichiers constituent l'une des sorties du programme et contiennent la segmentation produite automatiquement par le modèle. Les unités discursives y sont représentées sous forme de segments délimités par des crochets (« [ ] »), ce qui explique la présence du terme *brac* dans le nom du fichier.

Dans une étape suivante, les segments extraits des fichiers _brac_meta.txt sont alignés avec le corpus ELAN existant. Cet alignement est effectué à l'aide du programme annotateur_automatique.py, qui met en relation les segments produits par DisCut avec les annotations de traduction et les gloses présentes dans le corpus. Cette étape permet d'attribuer à chaque segment une position temporelle cohérente avec la structure des annotations existantes.

Enfin, les segments validés sont intégrés dans un nouveau tier ELAN intitulé « BDU DISCUT AUTOMATIQUE ». Chaque segment est ajouté avec ses bornes temporelles correspondantes, créant ainsi une nouvelle couche d'annotation au sein du fichier.

Le système produit en sortie un nouveau fichier ELAN nommé nom_du_fichier_DISCUT_AUTO.eaf. Ce fichier conserve l'ensemble des tiers originaux du corpus et inclut le nouveau tier de segmentation automatique. Il contient ainsi les segments discursifs générés par DisCut, alignés temporellement avec les annotations existantes, permettant une analyse conjointe de la segmentation automatique et de la structure annotée du corpus.

15.2 Résultats du pipeline d'annotation et de segmentation.

Dans un premier temps, le pipeline a été appliqué à l'ensemble d'un fichier ELAN contenant huit discours différents. Ce fichier a été choisi parce qu'il incluait les quatre discours qui ont servi de base aux analyses présentées tout au long des trois chapitres précédents, ce qui permettait de comparer directement les performances obtenues avec celles déjà observées :

- « Histoire du cheval » (00:00:04 – 00:00:54), locuteur 1
- « Recette de cuisine » (00:05:45 – 00:07:06), locuteur 1
- « Histoire de l'oiseau » (00:27:09 – 00:28:08), locuteur 2
- « Recette de cuisine » (00:31:04 – 00:34:04), locuteur 2

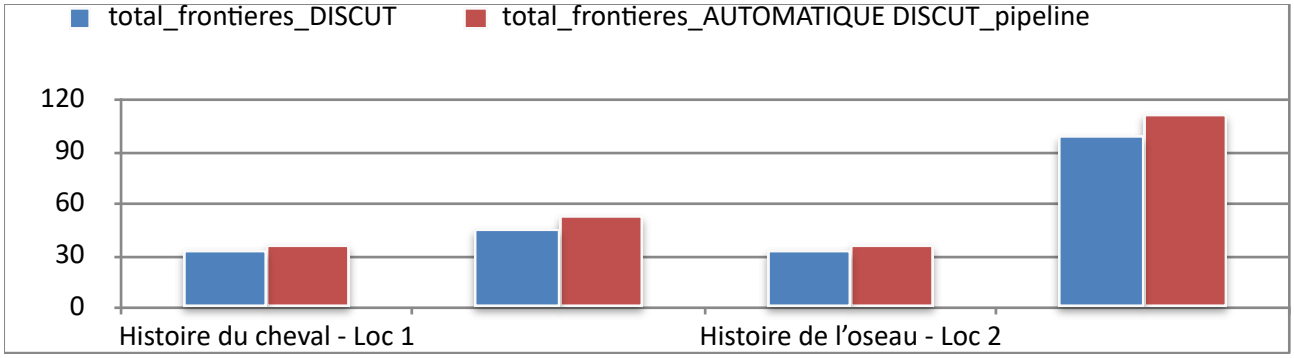
L'ensemble des résultats de la segmentation automatique sur les huit discours est présenté en annexe.

Comme indiqué précédemment, une analyse des résultats obtenus dans ELAN après l'exécution du pipeline a ensuite été réalisée. L'objectif était de vérifier le bon fonctionnement de l'ensemble de la chaîne de traitement, composée de plusieurs étapes successives, et d'évaluer si cette complexification du processus avait eu un impact sur les performances de l'annotateur automatique.

Cette analyse a été menée en comparant les résultats produits par le pipeline (*pipeline_discut_elan.py*) aux segmentations réalisées manuellement à partir des résultats de DisCut appliqués aux traductions. En effet, l'objectif principal de ce pipeline était de reproduire de manière automatique, et avec le plus haut degré de fidélité possible, les annotations obtenues manuellement à partir des segmentations générées par DisCut.

Une première analyse a consisté à comparer ces résultats à l'aide du même protocole que celui utilisé dans la section 12.2 (Comparaison des segmentations) avec e programme_2.py.

Nous avons commencé par comparer le nombre total de frontières identifiées par chacune des deux annotations dans les différents discours. À partir de ces résultats, nous avons généré le graphique suivant.



Graphique 10 : représentation du nombre total de frontières identifiées lors de la segmentation manuelle basée sur les résultats du pipeline pour l'annotation automatique à partir des mêmes résultats ; les colonnes correspondent aux différents discours.

Les résultats obtenus montrent une réduction notable des écarts entre les deux annotations. En comparaison avec la Figure 8 présentée dans la section 14.3 (Résultats de l'annotateur automatique sur ELAN), on observe que l'utilisation du pipeline permet de rapprocher davantage les frontières détectées par le système automatique de celles identifiées manuellement. Cette amélioration est particulièrement visible dans le cas des discours de type « recette de cuisine », qui constituaient auparavant les segments les plus problématiques lors du traitement discours par discours. Dans la configuration en pipeline, la différence entre les segmentations a été réduite à 12 frontières, contre 84 lors des premiers tests effectués en mode individuel.

On observe également que le nombre de frontières détectées par le pipeline est plus élevé et plus proche de celui de l'annotation manuelle issue de DisCut sur l'ensemble des discours. À l'inverse, la configuration discours par discours produisait systématiquement un nombre de frontières inférieur à celui de l'annotation de référence, ce qui suggère une sous-segmentation du corpus dans cette configuration.

À partir de ces premiers résultats, une seconde analyse a été menée afin d'identifier le nombre de frontières détectées simultanément par les deux annotations, ainsi que les cas de correspondance stricte au niveau des débuts et des fins de segments. Les résultats détaillés de cette analyse sont présentés dans le tableau suivant.

Discours	Nombre total de frontières concordantes	Concordance des frontières (début et fin)
Histoire du cheval - Loc 1	25	14
Recete de cuisine - Loc 1	18	10
Histoire de l'oseau - Loc 2	16	2
Recete de cuisine - Loc 2	42	12

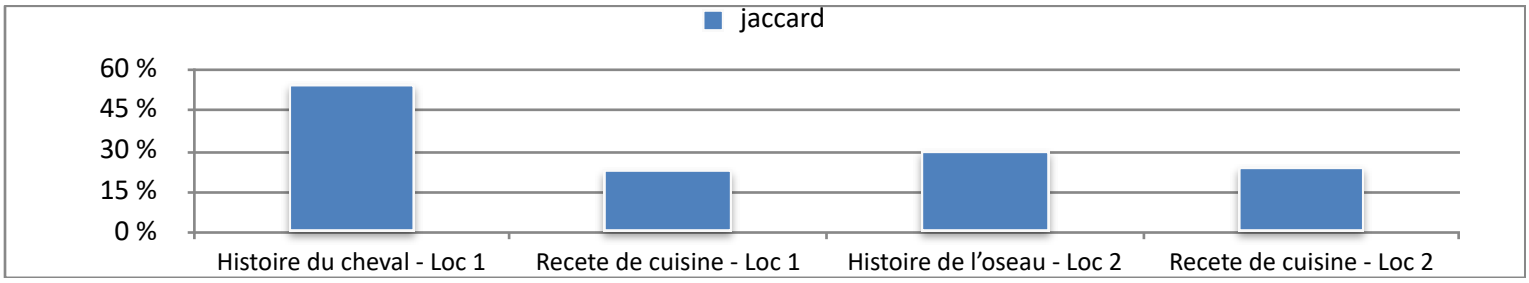
Tableau 5 : nombre total de frontières concordantes et nombre total de segments présentant une concordance stricte entre le début et la fin.

Comme on peut l'observer dans le tableau 5, le nombre de concordances obtenues lorsque l'on exige une correspondance simultanée des bornes initiales et finales reste relativement faible. Toutefois, ces résultats sont globalement plus élevés que ceux présentés dans le tableau 3. On observe en effet une augmentation significative des concordances de frontières début–fin dans plusieurs discours : elles passent de 12 à 14 pour l'histoire du cheval, de 3 à 10 pour la recette de cuisine (locuteur 1), et de 6 à 12 pour la recette de cuisine (locuteur 2). Seule exception, l'histoire de l'oiseau (locuteur 2), pour lequel le nombre de concordances diminue légèrement, passant de 5 lors du traitement discours par discours à 2 dans la configuration en pipeline.

Malgré cette amélioration globale, les concordances strictement alignées sur les bornes initiales et finales restent relativement faibles lorsqu'on les compare au nombre total de frontières concordantes. Pour cette raison, et conformément aux analyses précédentes, les analyses suivantes se concentrent principalement

sur les frontières concordantes, indépendamment de la stricte correspondance des débuts et des fins de segments.

Sur cette base, les pourcentages de concordance ont ensuite été calculés afin de mieux visualiser les résultats. La graphique suivante en présente la distribution.



Graphique 11 : représentation du pourcentage de concordance des frontières entre DisCut annoté manuellement et le pipeline avec la segmentation automatiquement ; les colonnes représentent les différents discours.

On observe une augmentation générale des taux de concordance dans la configuration en pipeline par rapport au traitement discours par discours. Ainsi, le pourcentage de concordance passe de 39,1 % à 54,3 % pour l'histoire du cheval, de 15,1 % à 23,1 % pour la recette de cuisine (locuteur 1), et de 10,3 % à 24,7 % pour la recette de cuisine (locuteur 2). En revanche, une légère diminution est observée pour l'histoire de l'oiseau, dont le taux de concordance passe de 32,6 % en mode discours par discours à 30,2 % dans la configuration en pipeline.

Enfin, le score F1 du système a été calculé pour chacun des discours. Les résultats obtenus sont présentés ci-dessous.

Discours	Précision	Rappel	F1-score
Histoire du cheval - Loc 1	75,8 %	73,5 %	74,6 %
Recette de cuisine - Loc 1	40,0 %	40,0 %	40,0 %
Histoire de l'oiseau - Loc 2	50,0 %	48,5 %	49,2 %
Recette de cuisine - Loc 2	43,8 %	42,0 %	42,9 %

Tableau 6 : représentation de la précision, du rappel et du score F1 du système d'annotation automatique utilise dans le pipeline.

Les résultats globaux montrent des différences importantes entre la configuration discours par discours et le pipeline global appliqué à l'ensemble du fichier ELAN. Le pipeline global (Tableau 6) obtient un F1-score micro de 48,4 %, avec une précision de 49,0 % et un rappel de 47,9 %. À l'inverse, la configuration discours par discours (Tableau 2) présente des performances plus faibles, avec un F1-score de 34,2 %, une précision légèrement plus élevée (56,5 %) mais un rappel très bas (24,5 %).

Ces résultats montrent que la configuration discours par discours produit des segmentations plus restrictives : elle favorise la précision, mais elle détecte moins de frontières discursives. Le pipeline global, en revanche, permet un meilleur équilibre entre précision et rappel, ce qui conduit à de meilleurs résultats globaux.

Une observation importante et quelque peu inattendue concerne l'amélioration globale des performances du système d'annotation automatique (annotateur_automatique.py) lorsqu'il est intégré dans le pipeline complet. Les résultats montrent que cette configuration est plus stable et plus efficace que l'approche

consistant à traiter chaque discours séparément. On observe notamment moins de segments non détectés et un meilleur alignement entre les segments automatiques et les annotations de référence.

Au niveau des performances par discours, le pipeline global présente des résultats relativement stables et homogènes. Les scores de F1 varient entre 40 % et 75 %, avec une performance particulièrement élevée pour le discours « Histoire du cheval – Loc 1 », qui atteint 74,6 %. Les autres discours présentent des résultats un peu plus faibles, mais globalement meilleurs que ceux obtenus avec la configuration discours par discours. Cela montre une amélioration générale lorsque le système est appliqué à tout le document.

Ces différences peuvent s'expliquer par l'importance du contexte global dans la segmentation automatique. Dans la configuration en pipeline, le traitement de l'ensemble du fichier ELAN permet de mieux identifier les transitions entre les discours et les structures du corpus. À l'inverse, le traitement discours par discours réduit ce contexte, ce qui entraîne des pertes d'information et plus d'erreurs de détection.

En conclusion, les résultats montrent que le pipeline global est la configuration la plus efficace. Il permet un meilleur équilibre entre précision et rappel, et améliore la détection des frontières discursives. Le traitement discours par discours est plus précis localement, mais il perd en couverture et donne des résultats globaux moins bons.

15. Discussion

15.1 Apports

Ce travail a permis de dégager plusieurs apports méthodologiques et empiriques dans le domaine de la segmentation discursive en langue des signes française.

En premier lieu, un guide d'annotation a été élaboré afin de permettre la segmentation des unités discursives de base (BDU) dans des vidéos de LSF à partir du corpus LS-Colin. Ce guide adapte le modèle de Gabarró-López et Meurant (2016) à un contexte opérationnel concret, en rendant les critères de segmentation accessibles à des annotateurs présentant des profils variés. Il constitue ainsi une ressource méthodologique directement exploitable pour de futures recherches dans ce domaine.

En deuxième lieu, le protocole expérimental mis en place a permis d'évaluer l'applicabilité de ce guide auprès de trois annotateurs aux niveaux d'expérience distincts, en annotation linguistique et en langue des signes française. Les résultats obtenus confirment que des annotateurs non experts sont en mesure d'identifier des indices discursifs pertinents, et que l'inclusion de profils variés constitue une approche pertinente pour ce type de tâche. Cette observation ouvre des perspectives pour la constitution de corpus annotés à plus grande échelle, sans nécessiter une expertise linguistique spécialisée en LSF.

En troisième lieu, ce travail propose une analyse exploratoire du niveau d'accord inter-annotateurs sur plusieurs couches d'annotation: segmentation syntaxique, prosodique et BDU, ce qui permet d'identifier les sources de divergence les plus fréquentes et de préciser les aspects du guide nécessitant une standardisation plus rigoureuse.

En quatrième lieu, une comparaison systématique entre les segmentations humaines et celles produites par le segmentateur automatique DisCut a été réalisée. Cette comparaison a permis d'évaluer dans quelle mesure les traductions françaises du corpus peuvent servir de support à une segmentation automatique des vidéos en LSF. Les résultats montrent que, dans certains cas, le niveau d'accord entre DisCut et les annotateurs humains est comparable, voire un peu supérieur, au niveau d'accord observé entre les annotateurs humains eux-mêmes.

Enfin, un annotateur automatique et un pipeline complet d'annotation et de segmentation ont été développés, permettant d'automatiser l'ensemble du processus, depuis l'extraction des traductions dans ELAN jusqu'à la création d'un nouveau tier de segmentation dans le fichier .eaf. Ce pipeline constitue un

prototype fonctionnel dont les résultats, bien que perfectibles, ouvrent la voie à une automatisation partielle du processus d'annotation, susceptible de réduire considérablement le temps de travail nécessaire à la constitution de corpus annotés en langue des signes.

15.2 Limites

Malgré les apports décrits précédemment, ce travail présente plusieurs limites importantes qui doivent être prises en compte dans l'interprétation des résultats.

La première limite concerne le nombre restreint d'annotateurs ayant participé à l'expérimentation. Avec seulement trois participants, il n'est pas possible de généraliser les conclusions relatives au niveau d'accord inter-annotateurs ni à l'applicabilité du guide à un public plus large. Une expérimentation impliquant un plus grand nombre de participants, aux profils plus diversifiés, serait nécessaire pour valider les tendances observées.

En lien avec ce premier point, la subjectivité inhérente à la tâche d'annotation constitue une limite structurelle importante. La segmentation repose sur des décisions interprétatives qui varient selon les annotateurs, leur expérience et leur compréhension des critères proposés. Malgré les efforts de standardisation du guide, certaines frontières discursives restent ambiguës et ne font pas l'objet d'un consensus clair, ce qui se reflète dans les faibles taux d'accord inter-annotateurs observés, en particulier pour la segmentation prosodique.

Une troisième limite tient aux contraintes de temps qui ont pesé sur l'ensemble du processus. En raison de ces contraintes, seul un discours a pu être annoté dans sa totalité par l'ensemble des trois participants, la recette de cuisine n'ayant été complétée que par l'annotateur 2. Cela réduit la comparabilité des résultats et limite la portée des analyses réalisées.

Par ailleurs, l'outil DisCut présente des limitations importantes dans le cadre de cette application. Celui-ci a été entraîné sur des corpus de textes écrits dans plusieurs langues, mais n'a jamais été conçu ni évalué pour traiter des traductions de discours signés. La distance entre les données d'entraînement de DisCut et les traductions françaises du corpus LS-Colin, produites à partir d'une langue visuo-gestuelle et présentant une structure discursive spécifique, constitue une source d'erreur non négligeable, en particulier pour les discours de type procédural, comme les recettes de cuisine.

Enfin, l'approche adoptée dans ce travail repose sur l'utilisation des traductions et des gloses comme substitut à l'analyse directe du signal signé. Si cette stratégie permet de contourner la nécessité d'une maîtrise approfondie de la LSF, elle introduit inévitablement des pertes d'information. Certains marqueurs discursifs propres à la langue des signes, notamment les éléments simultanés, les référents spatiaux ou certains phénomènes prosodiques non manuels, ne trouvent pas toujours d'équivalent dans les traductions françaises, ce qui limite la complétude et la fidélité de la segmentation obtenue.

15.3 Perspectives futures

Les résultats obtenus dans ce travail ouvrent plusieurs pistes de recherche et de développement pour la suite.

En premier lieu, il serait pertinent d’élargir le protocole expérimental à un plus grand nombre d’annotateurs, en incluant notamment davantage de locuteurs natifs de la LSF. Une telle extension permettrait de valider de manière plus rigoureuse le guide d’annotation proposé et d’affiner les critères de segmentation les plus ambigus, en particulier ceux liés à la couche prosodique. Par ailleurs, l’organisation de sessions de formation préalables à l’annotation, ainsi que la mise en place de cycles de révision et d’ajustement collectif du guide, pourraient contribuer à améliorer le niveau d’accord entre les annotateurs.

En deuxième lieu, les outils développés dans ce travail : le guide d'annotation, l'annotateur automatique et le pipeline complet, pourraient être appliqués à d'autres corpus de langue des signes disponibles, afin d'évaluer leur transférabilité à d'autres langues signées ou à d'autres types de discours. Une telle généralisation permettrait de mieux comprendre dans quelles conditions cette approche est applicable et d'identifier les ajustements nécessaires selon les caractéristiques des corpus traités.

En troisième lieu, il serait intéressant d'améliorer les performances de l'annotateur automatique, en testant d'autres approches comme l'utilisation des modèles de langage ou d'autres méthodes de segmentation existentes. Une autre piste importante concerne l'adaptation des traductions dans les corpus. On pourrait chercher à rapprocher autant que possible les traductions du texte présent dans les gloses, ce qui faciliterait la segmentation et annotation automatique.

Enfin, à plus long terme, lorsque les méthodes de segmentation seront plus précises, il serait possible de segmenter plusieurs corpus et d'analyser les correspondances observées à chaque frontière. On pourrait par exemple réaliser des analyses de transferts ou étudier des phénomènes propres à la langue des signes française. Cela permettrait de vérifier dans quelle mesure les segmentations basées sur les gloses et les traductions correspondent au contenu réellement présent dans les vidéos signées, et d'envisager ainsi une nouvelle manière de valider ou d'améliorer la segmentation.

Les résultats de ce travail suggèrent qu'une approche combinant annotation humaine et segmentation automatique pourrait constituer une stratégie efficace pour accélérer la constitution de corpus annotés en langue des signes, notamment pour la segmentation des BDU (unités de base du discours). Dans cette perspective, le rôle de l'annotateur humain ne serait pas de réaliser toute la segmentation, mais de valider, corriger et enrichir une première proposition générée automatiquement. Ce type de collaboration représente une voie prometteuse pour répondre au problème de la rareté des ressources annotées dans le domaine du traitement automatique des langues des signes.

15. Conclusion

Ce travail s'est inscrit dans le domaine encore peu exploré du traitement automatique des langues des signes, en proposant une approche méthodologique originale pour la segmentation discursive en langue des signes française à partir du corpus LS-Colin. Face au défi que représente l'absence d'unités d'analyse clairement définies pour les langues des signes, l'objectif principal était d'évaluer dans quelle mesure les traductions et les gloses en français pouvaient constituer un support fiable pour l'identification des unités discursives de base (BDU).

Pour répondre à cette question, trois axes complémentaires ont été développés tout au long de ce mémoire. Dans un premier temps, une exploration manuelle de la segmentation discursive en LSF a été réalisée à partir du modèle BDU proposé par Gabarró-López et Meurant (2016), en s'appuyant sur les gloses et les traductions du corpus comme substitut à l'analyse directe du signal signé pour la segmentation syntaxique. Cette première application a permis de confirmer la faisabilité de l'approche et d'identifier les principaux défis méthodologiques liés à ce type de tâche.

Dans un deuxième temps, une expérimentation inter-annotateurs a été mise en place afin d'évaluer l'applicabilité du guide d'annotation élaboré pour cette recherche. Les résultats ont montré que des annotateurs présentant des profils très différents , y compris sans expérience préalable en annotation ni en langue des signes, sont en mesure d'identifier des indices discursifs pertinents à partir des gloses, des traductions et de l'observation de la vidéo. Si les taux d'accord inter-annotateurs restent relativement faibles, en particulier pour la couche prosodique, ils reflètent la complexité intrinsèque de la tâche et soulignent la nécessité d'une standardisation plus rigoureuse des critères de segmentation.

Dans un troisième temps, une expérimentation avec le segmentateur automatique DisCut a permis d'évaluer la faisabilité d'une segmentation automatique fondée sur les traductions françaises du corpus. Les résultats ont montré que, dans certains cas, le niveau d'accord entre DisCut et les annotateurs humains est comparable à celui observé entre les annotateurs humains eux-mêmes. Sur cette base, un annotateur

automatique et un pipeline complet d'annotation et de segmentation ont été développés, permettant d'automatiser l'ensemble du processus depuis l'extraction des traductions jusqu'à la création d'un nouveau tier de segmentation dans ELAN. Les performances obtenues, bien que perfectibles, constituent un résultat encourageant pour une première exploration dans ce domaine.

De manière générale, ce travail confirme que les traductions et les gloses en français peuvent effectivement servir de pont entre les données signées et les outils de traitement automatique du langage. Si cette approche ne permet pas de capturer l'ensemble des marqueurs discursifs propres à la langue des signes, elle offre une base de travail accessible et reproductible, susceptible d'être utilisée par des chercheurs ne maîtrisant pas la LSF.En ce sens, elle répond à l'un des défis fondamentaux du domaine, notamment la rareté des ressources annotées et la difficulté de recruter des annotateurs compétents en langue des signes.

Ce mémoire s'inscrit ainsi dans une démarche exploratoire et contributive. Les résultats, les outils et les réflexions méthodologiques présentés constituent des points de départ pour de futures recherches visant à améliorer l'intégration des langues des signes dans le traitement automatique des langues.

16. Bibliographie

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press. https://hal.science/hal-04829706

Blanche-Benveniste, C., Deulofeu, J., Stéfanini, J., & Van den Eynde, K. (1984). *Pronom et syntaxe : L'approche pronominale et son application au français*. SELAF.

Benamara, F., & Taboada, M. (2015). Mapping Different Rhetorical Relation Annotations : A Proposal. *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, 147-152. https://doi.org/10.18653/v1/S15-1016

Borg, M., & Camilleri, K. P. (2019b). Sign Language Detection "in the Wild" with Recurrent Neural Networks. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1637-1641. https://doi.org/10.1109/ICASSP.2019.8683257

Börstell, C., Mesch, J., & Wallin, L. (2014). Segmenting the Swedish Sign Language corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Kristoffersen, & J. Mesch (Eds.), *Proceedings of the LREC2014 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel* (pp. 7–10). European Language Resources Association (ELRA). https://www.sign-lang.uni-hamburg.de/lrec/pub/14023.html

Braffort, A., Choisier, A., Collet, C., Dalle, P., Gianni, F., Lenseigne, F., & Segouat, J. (s. d.). *Toward an Annotation Software for Video of Sign Language, Including Image Processing Tools and Signing Space Modelling*.

Briz Gómez, A., Hidalgo, A., Padilla, X., Pons, S., Ruiz Gurillo, L., Sanmartín, J., Albelda, M., & Fernández, M. (2003). *Un sistema de unidades para el estudio del lenguaje coloquial.* Oralia, 6, 7–61.

Bull, H., Braffort, A., & Gouiffès, M. (2020). MEDIAPI-SKEL - A 2D-Skeleton Video Database of French Sign Language With Aligned French Subtitles. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Éds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (p. 6063-6068). European Language Resources Association. https://aclanthology.org/2020.lrec-1.743/

Cuxac, C., Braffort, A., Choisier, A., Choisier, C., Collet, C., Dalle, P., Fusellier-Souza, I., Gherbi, R., Jausions, G., Jirou, G., Lejeune, F., Lenseigne, B., Monteilliard, N., Risler, A., & Sallandre, M.-A. (2002). *Projet LS-Colin : rapport de fin de recherche*. Programme Cognitique « Langage et cognition », Projet LACO 39.

Degand, L., & Simon, A.-C. (2005). Minimal Discourse Units: Can we define them, and why should we? In Aurnague, M., Bras, M., Le Draoulec, A., & Vieu, L. (eds) (ed.), Proceedings of SEM-05. Connectors, discourse framing and discourse structure: from corpus-based and experimental analyses to discourse theories (pp. 65-74). https://hdl.handle.net/2078.5/241862

Degand, L., & Simon, A.-C. (2009). On identifying basic discourse units in speech: Theoretical and empirical issues. *Discours, 4*. https://doi.org/10.4000/discours.5852

Degand, L., & Simon, A.-C. (2009). On identifying basic discourse units in speech: Theoretical and empirical issues. *Discours, 4*. https://doi.org/10.4000/discours.5852

Ezzabady, M. K., Muller, P., & Braud, C. (2021). Multi-lingual discourse segmentation and connective identification: MELODI at DISRPT2021. *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, 22–32. https://doi.org/10.18653/v1/2021.disrpt-1.3

Fauré, L. (1999). Mary-Annick Morel, Laurent Danon-Boileau, Grammaire de l'intonation. L'exemple du français. *Cahiers de praxématique*, (32), 238-242. https://doi.org/10.4000/praxematique.2877

Fenlon, J., Denmark, T., Campbell, R., & Woll, B. (2007). Seeing sentence boundaries. *Sign Language and Linguistics*, *10*(2), 177-200. https://doi.org/10.1075/sll.10.2.06fen

Fenlon, J. J. (2010). Seeing sentence boundaries: The production and perception of visual markers signalling boundaries in signed languages [Doctoral dissertation, University College London]. UCL Discovery.

Ferrari, A. (Ed.). (2005). Rilievi. Le gerarchie semantico-pragmatiche di alcuni tipi di testo. Cesati.

Ferrari, A., Cignetti, L., De Cesare, A.-M., Lala, L., Mandelli, M., Ricci, C., & Roggia, C. E. (2008). *L'interfaccia lingua-testo. Natura e funzioni dell'articolazione informativa dell'enunciato*. Edizioni dell'Orso.

Filhol, M. (s. d.). *Grammaire récursive non linéaire pour les langues des signes*.

Gabarro-Lopez, S., & Meurant, L. (2016). *Slicing Your SL Data into Basic Discourse Units (BDUs) Adapting the BDU Model (Syntax + Prosody) to Signed Discourse*.

Gabarró-López, S., & Meurant, L. (2014). *When nonmanuals meet semantics and syntax : A practical guide for the segmentation of sign language discourse*.

Gebre, B. G., Wittenburg, P., & Heskes, T. (2013). Automatic sign language identification. In *2013 20th IEEE International Conference on Image Processing (ICIP)* (pp. 2626–2630). IEEE. https://doi.org/10.1109/ICIP.2013.6738541

Glickman, N. S., & Hall, W. C. (Éds.). (2018). *Language Deprivation and Deaf Mental Health*. Routledge. https://doi.org/10.4324/9781315166728

Hadjadj, M. N., Filhol, M., & Braffort, A. (2018). Modeling French Sign Language : A proposal for a semantically compositional system. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Éds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). https://aclanthology.org/L18-1671/

Hanke, T., Schulder, M., Konrad, R., & Jahn, E. (2020). Extending the Public DGS Corpus in size and depth. In Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives (pp. 75–82). European Language Resources Association (ELRA).

Hannay, M., & Kroon, C. (2005). Acts and the relationship between discourse and grammar. *Functions of Language*, *12*, 87-124. https://doi.org/10.1075/fol.12.1.05han

Herrmann, A. (2012). Prosody in German Sign Language. In P. Prieto & G. Elordieta (Eds.), *Prosody and meaning* (pp. 349–380). De Gruyter Mouton.

Jantunen, T. (2007). The equative sentence in Finnish Sign Language. *Sign Language & Linguistics*, *10*(2), 113-143. https://doi.org/10.1075/sll.10.2.04jan

Li, J. (2019). Neural discourse segmentation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)* (pp. 6539–6541). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2019/949

Lombardi Vallauri, E. (2009). *La struttura informativa: Forma e funzione negli enunciati linguistici*. Carocci.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory : Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, *8*(3). https://doi.org/10.1515/text.1.1988.8.3.243

Monteiro, C. D. D., Mathew, C. M., Gutierrez-Osuna, R., & Shipman, F. (2016). Detecting and Identifying Sign Languages through Visual Features. *2016 IEEE International Symposium on Multimedia (ISM)*, 287-290. https://doi.org/10.1109/ISM.2016.0063

Moryossef, A., Yin, K., Neubig, G., & Goldberg, Y. (2021). Data Augmentation for Sign Language Gloss Translation. In D. Shterionov (Éd.), *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)* (p. 1-11). Association for Machine Translation in the Americas. https://aclanthology.org/2021.mtsummit-at4ssl.1/

Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S., & Narayanan, S. (2020). Real-time sign language detection using human pose estimation. In A. Bartoli & A. Fusiello (Eds.), *Computer Vision – ECCV 2020 Workshops* (pp. 237–248). Springer. https://doi.org/10.1007/978-3-030-66096-3_17

Notarrigo, I., & Meurant, L. (2014). *Nonmanuals and markers of (dis)fluency in French Belgian Sign Language (LSFB)*. https://www.semanticscholar.org/paper/Nonmanuals-and-markers-of-(dis)fluency-in-French-Notarrigo-Meurant/002260a9949444d2100ea2cf89b853b89fec49c6

Ormel, E., & Crasborn, O. (2012). Prosodic correlates of sentences in signed languages: A literature review and suggestions for new types of studies. *Sign Language Studies, 12*(2), 279–315. https://doi.org/10.1353/sls.2011.0019

Padden, C., & Humphries, T. (1988). Deaf in America: Voices from a culture. Harvard University Press.

Pfau, R., & Quer, J. (2010). Nonmanuals: Their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign languages* (pp. 381–402). Cambridge University Press. https://doi.org/10.1017/CBO9780511712203.018

Poibeau, T. (2019). Le traitement automatique des langues : tendances et enjeux. Lalies, 39, 7–65.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*(pp. 2961–2968). European Language Resources Association (ELRA). https://aclanthology.org/L08-1093/

Reboul, A., & Moeschler, J. (1998). Pragmatique du discours : De l'interprétation de l'énoncé à l'interprétation du discours. Armand Colin.

Reese, B., Hunter, J., Asher, N., Denis, P., & Baldridge, J. (s. d.). *Reference Manual for the Analysis and Annotation of Rhetorical Structure (Version 1.0)*.

Roulet, E. (s. d.). *Vers une approche modulaire de l'analyse du discours'*.

Sandler, W. (2010). The Uniformity and Diversity of Language : Evidence from Sign Language. *Lingua. International Review of General Linguistics. Revue Internationale De Linguistique Generale*, *120*(12), 2727-2732. *https://doi.org/10.1016/j.lingua.2010.03.015*

Sandler, W., & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals* (1re éd.). Cambridge University Press. https://doi.org/10.1017/CBO9781139163910

Simon, A.-C. et Degand, L. (2011). L'analyse en unités discursives de base : pourquoi et comment ? Langue française, 170(2), 45-59. https://doi.org/10.3917/lf.170.0045.

Stanford Natural Language Processing Group. (s. f.). *Stanford tokenizer*. Stanford University. https://nlp.stanford.edu/software/tokenizer.html

Sze, F. Y. B. (2008). Blinks and intonational phrasing in Hong Kong Sign Language. In J. Quer (Ed.), *Signs of the time: Selected papers from TISLR 2004* (pp. 83–107). Signum.

Tofiloski, M., Brooke, J., & Taboada, M. (2009). A syntactic and lexical-based discourse segmenter. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers on - ACL-IJCNLP '09*, 77. https://doi.org/10.3115/1667583.1667609

Wilbur, R. B. (1994). Eyeblinks & Asl Phrase Structure. *Sign Language Studies*, (84), 221-240.

Yin, K., Moryossef, A., Hochgesang, J., Goldberg, Y., & Alikhani, M. (2021a). Including Signed Languages in Natural Language Processing. In C. Zong, F. Xia, W. Li, & R. Navigli (Éds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)* (p. 7347-7360). Association for Computational Linguistics. *https://doi.org/10.18653/v1/2021.acl-long.570*

17. ANNEXE

17.1 Résultats de la première exploration sur le corpus LS-Colin – chapitre I

Dans cette section, sont présentés les résultats de la première exploration du corpus LS-Colin.
Deux discours différents ont été utilisés : « Histoire du cheval » (00:00:04 – 00:00:54) et « Recette de cuisine » (00:05:45 – 00:07:06). Ces données sont accessibles à l'adresse suivante : https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-d098bd34-f696-3c40-82eb-b1e5a2bc2e82.

17.1.1 Représentation de la segmentation syntaxique des gloses

La segmentation est représentée à l'aide des crochets [].

Gloses de l'Histoire du cheval :

[HISTOIRE CHEVAL] [galoper] [freiner barrière][APERCEVOIR cornes de VACHE][être sur ses sabots ruminer] [regarder (vers le cheval)] [bon !] [ENVIE sauter barrière][COMMENT sauter] [attendre voir] [reculer] [galoper] [sauter][sauter] [heurter][tomber] [gésir au sol][AVOIR MAL] [cornes de VACHE observer la situation] [tourner la tête][LA OISEAU être là][boite croix boite] [tenir la boite dans ses serres][voler][déposer la boite][ouvrir la boite][corne de VACHE ruminer] [avancer][prendre qqch avec la bouche][enrouler la bande][SOIGNER][enrouler la bande][CHEVAL posture patte droite tendue][se faire bander la patte][MERCI SOIGNER]posture au repos [MERCI] posture au repos [FINI]

Gloses de la Recette de cuisine :

[RECETTE CUISINE PREFERE MOI] position neutre [CHOU ROUGE] [CHOU ROUGE hacher][rassembler (esq) hacher][mettre les morceux dans un récipient] [récupient cylindrique (cocotte)] [remuer] [mettre les morceaux dans le récipient] [s'essuyer les mains] [remuer] [EAU verser UN PEU] [verser] [mettre le couvercle] [visser (cocotte)] [mettre le sifflet] [tourner (sifflet)] [LÀ (sur cocotte) ? mettre le sifflet][JUSQU'À (esq)][CUIRE JUSQU'A LA (sifflet) vapeur (sifflet) tourner][ah !] [MOI COUPER feu qui s'éteint] [enlever le sifflet] [vapeur qui s'échappe VAPEUR] position neutre [APRES (esq)] [ouvrir la cocotte] [enlever le couvercle] [CUIRE MOITIE ENVIRON] [MOI se frotter les mains] [POMME couper en morceaux] [mettre (début) DEUX TROIS POMME] [mettre (x3)] position neutre [PLUS][MIEL UN CUILLERE] [prendre une cuillère de miel][a mettre dans la cocotte] [PLUS] [VINAIGRE V.I.N.A.I.G.R.E] [mettre UNE DEUX CUILLERE] [verser le vinaigre dans la cuillère] [la mettre dans la cocotte] [verser le vinaigre dans la cuillère] [a mettre dans la cocotte] position neutre[PLUS] [SEL] [PLUS] [SUCRE ROUX] [mettre DEUX CUILLERE] [mettre dans la cocotte] [SENTIR JUSTE ENSEMBLE (esq)] [mettre le couvercle de la cocotte] [visser le couvercle] [CUIRE JUSQUA (esq)] [sifflet qui tourne] [JUSQU'À siffler sui tourne] [COUPER (feu)] [LAISSER CINQ MINUTES] [ALORS] [enlever le couvercle] [et sentir] [remuer avec une cuillère] [ODEUR PRESTIGIEUX] [AVEC] [MANGER AVEC VIANDE CÔTE PORC] [VOILA]

17.1.2 Représentation des gloses après la segmentation en BDU

Après avoir réalisé la segmentation syntaxique et prosodique, nous obtenons les groupes de gloses suivants, qui correspondent à des BDU.

Segmentation des gloses de l'Histoire du cheval après segmentation en BDU:

[HISTOIRE CHEVAL**] [**galoper**] [**freiner barrière**]**[**APERCEVOIR cornes de VACHE être sur ses sabots ruminer regarder (vers le cheval)**] [**bon ! ENVIE sauter barrière **]**[**COMMENT sauter**] [**attendre voir reculer **]**[**galoper**]**[**sauter **]**[**sauter heurter tomber gésir au sol**]**[AVOIR MAL**]**[cornes de VACHE observer la situation**]**[**tourner la tête LA OISEAU être là **]**[**boite croix boite tenir la boite dans ses serres**]**[voler**]**[**déposer la boite ouvrir la boite **]**[**corne de VACHE ruminer**]**[**avancer prendre qqch avec la bouche **]**[**enrouler la bande SOIGNER enrouler la bande**]**[CHEVAL posture patte droite tendue se faire bander la patte MERCI **]**[**SOIGNER **]**posture au repos **[**MERCI**]** posture au repos **[**FINI **]**

Segmentation des gloses de la Recette de cuisine après segmentation en BDU:

[RECETTE CUISINE PREFERE MOI**]** position neutre **[**CHOU ROUGE**] [**CHOU ROUGE hacher**] [** rassembler (esq) hacher **]**[mettre les morceaux dans un récipient récupient cylindrique (cocotte)**] [**remuer mettre les morceaux dans le récipient**] [** s'essuyer les mains**] [** remuer EAU verser UN PEU verser**] [**mettre le couvercle visser (cocotte)**] [**mettre le sifflet tourner (sifflet) **] [**LÀ (sur cocotte) ? mettre le sifflet **] [**JUSQU'À (esq) CUIRE JUSQU'A LA (sifflet) vapeur (sifflet) tourner**] [**ah ! MOI COUPER feu qui s'éteint **] [**enlever le sifflet**] [** vapeur qui s'échappe VAPEUR**]** position neutre **[**APRES (esq) ouvrir la cocotte enlever le couvercle**] [** CUIRE MOITIE ENVIRON**] [** MOI se frotter les mains **] [**POMME couper en morceaux**] [**mettre (début) DEUX TROIS POMME **] [**mettre (x3)**]** position neutre **[**PLUS MIEL UN CUILLERE**] [**prendre une cuillère de miel la mettre dans la cocotte **] [**PLUS VINAIGRE V.I.N.A.I.G.R.E**] [**mettre UNE DEUX CUILLERE**] [** verser le vinaigre dans la cuillère**] [**la mettre dans la cocotte verser le vinaigre dans la cuillère**] [**a mettre dans la cocotte**]** position neutre**[** PLUS SEL **] [**PLUS SUCRE ROUX **] [**mettre DEUX CUILLERE mettre dans la cocotte**] [** SENTIR JUSTE ENSEMBLE (esq)**] [**mettre le couvercle de la cocotte visser le couvercle**] [**CUIRE JUSQUA (esq) sifflet qui tourne**] [**JUSQU'À siffler sui tourne COUPER (feu)**] [** LAISSER CINQ MINUTES**] [**ALORS enlever le couvercle et sentir**] [**remuer avec une cuillère**] [**ODEUR PRESTIGIEUX**] [**AVEC MANGER AVEC VIANDE CÔTE PORC VOILA**]**

7.1.3 Segmentation des traductions du corpus LS-Colin en français selon le modèle EDU et BDU .

La segmentation est représentée à l'aide de <***>.

Traductions extraites de l'Histoire du cheval depuis le corpus LS-Colin :

"C'est l'histoire… d'un cheval. Il galope, et soudain, il freine, surpris par la barrière. Il aperçoit une vache
qui rumine, ils se regardent. Le cheval a envie de sauter, comment faire… "Attends voir…" Il réfléchit
Il recule, prend son élan et saute, mais il heurte la barrière et tombe, et se retrouve les pattes étendues .
Il a mal, il souffre, la vache regarde derrière elle et voit un oiseau qui est perché sur la barrière,
il y a une boîte avec une croix rouge, il vole et la dépose et ouvre la boîte. La vache, rumine, et s'en approche, elle prend quelque chose et l'enroule, elle soigne et le cheval qui se laisse bander la patte.
Il remercie la vache qui termine de le soigner, il la remercie encore.
C'est fini."

Segmentation des traductions de l'Histoire du cheval en EDU :

C'est l'histoire… d'un cheval.<***> Il galope,<***> et soudain,<***> il freine, surpris par la barrière.<***> Il aperçoit une vache qui rumine,<***> ils se regardent.<***> Le cheval a envie de sauter,<***> comment faire…<***>"Attends voir…"<***> Il réfléchit<***> Il recule, prend son élan et saute,<***> mais il heurte la barrière et tombe,<***> et se retrouve les pattes étendues .<***> Il a mal, il souffre,<***> la vache regarde derrière elle et voit un oiseau qui est perché sur la barrière,<***> il y a une boîte avec une croix rouge,<***> il vole et la dépose et ouvre la boîte.<***> La vache, rumine, et s'en approche,<***> elle prend quelque chose et l'enroule,<***> elle soigne et le cheval qui se laisse bander la patte.<***> Il remercie la vache qui termine de le soigner,<***> il la remercie encore.<***> C'est fini.<***>

Segmentation des traductions de l'Histoire du cheval avec les résultats finaux des BDU.

C'est l'histoire… d'un cheval. <***>Il galope, <***>et soudain, il freine, surpris par la barrière. <***> Il aperçoit une vache qui rumine, ils se regardent.<***> Le cheval a envie de sauter, <***>comment faire… <***>"Attends voir…" Il réfléchit Il recule, <***>prend son élan<***> et saute,<***> mais il heurte la barrière et tombe, et se retrouve les pattes étendues . <***> Il a mal, il souffre,<***> la vache regarde <***>derrière elle et voit un oiseau qui est perché sur la barrière, <***> il y a une boîte avec une croix rouge, <***> il vole <***> et la dépose et ouvre la boîte. <***>La vache, rumine, <***> et s'en approche, elle prend quelque chose et l'enroule, elle soigne <***> et le cheval qui se laisse bander la patte. Il remercie la vache <***> qui termine de le soigner, <***> il la remercie encore. <***> C'est fini. <***>

Traductions extraites de la Recette de cuisine depuis le corpus LS-Colin :

"C'est la recette de cuisine que je préfère.
C'est le chou rouge.
Alors, on hache le chou rouge en petits morceaux
hacher dans la longueur puis la largeur, rassembler et couper.
Mettre le tout dans une cocotte minute.
mettre le couvercle et fermer la cocotte minute.
Mettre le sifflet (ce qui tourne et fait de la vapeur)
Faire cuire jusqu'à ce que le sifflet émette la vapeur et tourne.
Ah! Je coupe le feu, le feu s'éteint, et j'enlève le sifflet.
(on voit) La vapeur s'échappe vers le haut.
Ensuite on ouvre la cocotte, et on enlève le couvercle qu'on pose à côté, c 'est à peu près à moitié cuit, je
Alors…prends deux ou trois pommes et les coupe en morceaux, tout mettre
Ajouter du miel.
Prendre une cuillère de miel et la mettre dans la cocotte.
Ajouter du vinaigre,
en mettre une ou deux cuillères (verser d'abord le vinaigre dans la cuillère et la mettre dans la cocotte).
Ajouter du sel et du sucre roux, en mettre deux cuillères.
Euh, je pense que c'est bon, c'est complet
Fermer la cocotte et la visser puis cuire jusqu'à ce que le sifflet tourne.
Attendre jusqu'à ce que le sifflet tourne, puis couper le feu et laisser 5 minutes.

Puis ouvrir la cocotte, et remuer avec la cuillère, ça sent la cuisine de chef !
On peut manger avec de la viande de porc, les deux vont très bien ensemble…
Voilà”

Segmentation des traductions de la Recette de cuisine en EDU :

C'est la recette de cuisine que je préfère.<***> C'est le chou rouge.<***> Alors, on hache le chou rouge en petits morceaux<***> hacher dans la longueur puis la largeur, rassembler et couper.<***> Mettre le tout dans une cocotte minute.<***> Mettre le chou et se frotter les mains pour les miettes,<***> remuer, verser un peu d'eau, Mettre le sifflet (ce qui tourne et fait de la vapeur)<***> Faire cuire jusqu'à ce que le sifflet émette la vapeur et tourne.<***> Ah!<***> Je coupe le feu,<***> le feu s'éteint,<***> et j'enlève le sifflet.<***> (on voit)<***> La vapeur s'échappe vers le haut.<***> Ensuite on ouvre la cocotte,<***> et on enlève le couvercle qu'on pose à côté,<***> c 'est à peu près à moitié cuit, je<***> Alors…prends deux ou trois pommes et les coupe en morceaux,<***>
tout mettre<***> Ajouter du miel.<***> Prendre une cuillère de miel et la mettre dans la cocotte.<***> Ajouter du vinaigre,<***> en mettre une ou deux cuillères (verser d'abord le vinaigre dans la cuillère et la mettre dans la cocotte).<***> Ajouter du sel et du sucre roux, en mettre deux cuillères.<***>Euh, je pense que c'est bon, c'est complet<***> Fermer la cocotte et la visser puis cuire jusqu'à ce que le sifflet tourne.<***> Attendre jusqu'à ce que le sifflet tourne, puis couper le feu et laisser 5 minutes.<***> Puis ouvrir la cocotte, et remuer avec la cuillère, ça sent la cuisine de chef !<***> On peut manger avec de la viande de porc, les deux vont très bien ensemble…<***> Voilà<***>

Segmentation des traductions de la Recette de cuisine jusqu’aux résultats finaux des BDU.

“C'est la recette de cuisine que je préfère. <***> C'est le chou rouge. <***> Alors, on hache le chou rouge en petits morceaux <***> hacher dans la longueur puis la largeur, rassembler et couper. <***>
Mettre le tout dans une cocotte minute. <***> mettre le couvercle et fermer la cocotte minute.<***> Mettre le chou <***> et se frotter les mains pour les miettes,<***> remuer, verser un peu d'eau, <***> Mettre le sifflet (ce qui tourne et fait de la vapeur)<***> Faire cuire jusqu'à ce que le sifflet émette la vapeur et tourne. <***> Ah! Je coupe le feu, le feu s'éteint, <***> et j'enlève le sifflet. <***>
(on voit) La vapeur s'échappe vers le haut. <***> Ensuite on ouvre la cocotte, et on enlève le couvercle qu'on pose à côté, <***> c 'est à peu près à moitié cuit, <***> je <***> Alors…prends deux ou trois pommes et les coupe en morceaux, <***> tout mettre <***> Ajouter du miel. <***> Prendre une cuillère de miel et la mettre dans la cocotte. <***> Ajouter du vinaigre, <***> en mettre une ou deux cuillères <***> (verser d'abord le vinaigre dans la cuillère <***> et la mettre dans la cocotte). <***>
Ajouter du sel <***> et du sucre roux, <***> en mettre deux cuillères. <***> Euh, je pense que c'est bon, c'est complet. <***> Fermer la cocotte et la visser <***> puis cuire jusqu'à ce que le sifflet tourne. <***> Attendre jusqu'à ce que le sifflet tourne, puis couper le feu <***> et laisser 5 minutes. <***> Puis ouvrir la cocotte, <***> et remuer avec la cuillère, <***>ça sent la cuisine de chef ! <***>
On peut manger avec de la viande de porc, <***> les deux vont très bien ensemble…
Voilà <***>

Guide d'annotation

Segmentation des unités discursives (Modèle BDU)

Introduction

Ce guide s'appuie sur les propositions de Gabarró-López et Meurant (2016) dans leur article *Slicing Your SL Data into Basic Discourse Units (BDUs): Adapting the BDU Model (Syntax + Prosody) to Signed Discourse*. Les auteurs y adaptent le modèle des Basic Discourse Units (BDU), initialement conçu pour les langues orales, aux langues des signes, en suivant trois étapes principales :

1. **Segmentation syntaxique**
2. **Segmentation prosodique**
3. **Segmentation de BDU. (**Identification des points de convergence entre la segmentation prosodique et la segmentation syntaxique)

La segmentation est réalisée dans le logiciel **ELAN**. Seuls les discours disposant à la fois d'une **traduction** et de **gloses** sont segmentés. Les couches contenant les gloses et la traduction ne sont **jamais modifiées**.

Partie 1 — Segmentation syntaxique

1.1 Qu'est-ce qu'une unité syntaxique ?

Une unité syntaxique est un groupe de signes qui forment ensemble une idée ou une action cohérente. Pour identifier ces groupes, on s'appuie sur deux notions clés : le noyau et les dépendants.

Noyau : Signe principal autour duquel s'organisent les autres. Le plus souvent un verbe (action, état, événement), mais, dans certains cas, peut aussi être un nom, un pronom ou un adjectif.

Dépendants : Signes qui complètent le noyau. On les repère en posant des questions simples : Qui fait l'action ? Sur qui / quoi ? Quand ? Où ? Comment ?

Ce modèle s'appuie sur la Grammaire de Dépendance développée pour le français parlé (Blanche-Benveniste et al., 1984 ; 1990).

1.2 Procédure d'annotation

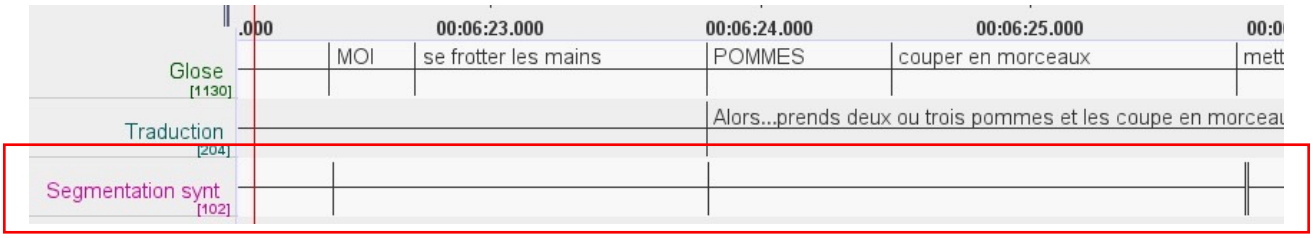
1. Lire attentivement les gloses du passage à annoter.
2. Identifier les noyaux.
3. Consulter la traduction française pour confirmer l'interprétation et pouvoir segmenter le noyau avec ses dépendants, notamment si vous n'êtes pas locuteur de la langue des signes.
4. En cas de doute, visionner la vidéo et vérifier la traduction avant de prendre une décision.
5. Délimiter les unités dans ELAN en sélectionnant à la souris l'étendue de chaque unité.

1.3 Critère principal de segmentation

Situation	Décision d'annotation	Exemple
Apparition d'un nouveau noyau verbal	Créer une nouvelle unité syntaxique contenant le noyau verbal et ses dépendants.	`[freiner barrière] [APERCEVOIR cornes de VACHE]`
Le noyau n'est pas un verbe	Une unité syntaxique peut aussi être construite autour d'un nom, d'un adjectif, d'un pronom ou d'une expression descriptive lorsqu'il n'y a pas de verbe	[HISTOIRE CHEVAL] [MIEL UN CUILLERE] [ODEUR PRESTIGIEUX]
Deux verbes successifs	Segmenter chaque verbe en une unité distincte, sauf si les deux forment clairement une seule action inséparable.	`[couper] [mélanger]`
Élément discursif autonome	Segmenter comme unité indépendante les marqueurs discursifs, commentaires, réactions ou organisateurs du discours.	`[après] [d'abord] [bon !] [MERCI]`

1.4 Méthode de segmentation

Pour faire la segmentation, nous effectuons la sélection du début jusqu'à la fin de la section.



1.5 Rôle de la traduction française

La traduction française est utilisée comme outil complémentaire pour confirmer ou affiner certaines décisions de segmentation, en particulier lorsque les gloses seules sont ambiguës.

Exemple — Séquence 00:00:10 – 00:00:16 :

Gloses : `freiner barrière APERCEVOIR cornes de VACHE être sur ses sabots ruminer`

Une première lecture pourrait suggérer qu'APERCEVOIR dépend de *freiner* et de *barrière*, c'est-à-dire que le cheval aurait aperçu la barrière et freiné pour cette raison. Cependant, la traduction française indique :

« Il aperçoit une vache qui rumine. »

La segmentation finale, les crochets [ ] étant utilisés comme symboles de délimitation des segments, est donc la suivante :

`[freiner barrière] [APERCEVOIR cornes de VACHE] [être sur ses sabots ruminer]`

1.5 Exemples du corpus LS COLIN

Les crochets [ ] servent de marqueurs des frontières de segmentation syntaxique.

Récit (00:00:04 - 00:00:54)

[HISTOIRE CHEVAL] [galoper] [freiner barrière] [APERCEVOIR cornes de VACHE]
[être sur ses sabots ruminer] [regarder (vers le cheval)] [bon !]
[ENVIE sauter barrière] [COMMENT sauter] [attendre voir] [reculer]
[galoper] [sauter] [sauter] [heurter] [tomber] [gésir au sol] [AVOIR MAL]
[cornes de VACHE observer la situation] [tourner la tête]
[LA OISEAU être là] [boite croix boite] [tenir la boite dans ses serres]
[voler] [déposer la boite] [ouvrir la boite] [corne de VACHE ruminer]
[avancer] [prendre qqch avec la bouche] [enrouler la bande] [SOIGNER]
[enrouler la bande] [CHEVAL posture patte droite tendue]
[se faire bander la patte] [MERCI SOIGNER] [MERCI] [FINI]

Discours procédural. (00:05:48 - 00:07:06)

[RECETTE CUISINE PREFERE MOI] [CHOU ROUGE] [CHOU ROUGE hacher]
[rassembler (esq) hacher] [mettre les morceaux dans un récipient]
[récipient cylindrique (cocotte)] [remuer] [EAU verser UN PEU]
[verser] [mettre le couvercle] [visser (cocotte)] [mettre le sifflet]
[CUIRE JUSQU'A LA (sifflet) vapeur (sifflet) tourner] [MOI COUPER feu qui s'éteint]
[enlever le sifflet] [vapeur qui s'échappe VAPEUR]
[APRES (esq)] [ouvrir la cocotte] [enlever le couvercle]
[CUIRE MOITIE ENVIRON] [POMME couper en morceaux]
[mettre (début) DEUX TROIS POMME] [MIEL UN CUILLERE]
[prendre une cuillère de miel] [mettre dans la cocotte]
[VINAIGRE V.I.N.A.I.G.R.E] [mettre UNE DEUX CUILLERE]
[verser le vinaigre dans la cuillère] [mettre dans la cocotte]
[SEL] [SUCRE ROUX] [mettre DEUX CUILLERE] [mettre dans la cocotte]
[SENTIR JUSTE ENSEMBLE (esq)] [mettre le couvercle de la cocotte]
[CUIRE JUSQU'A (esq)] [COUPER (feu)] [LAISSER CINQ MINUTES]
[enlever le couvercle] [remuer avec une cuillère] [ODEUR PRESTIGIEUX]
[AVEC MANGER AVEC VIANDE CÔTE PORC] [VOILA]

Partie 2 — Segmentation prosodique

2.1 Principe général

Dans le modèle BDU, la segmentation prosodique consiste à délimiter le discours en unités prosodiques. Cette étape intervient **après** la segmentation syntaxique. Les deux types de segmentation étant indépendants, il est recommandé de masquer la couche contenant les unités syntaxiques avant de commencer l'annotation prosodique, afin d'éviter des résultats biaisés.

La segmentation repose sur l'identification de trois indices visuels principaux, qui correspondent aux indices acoustiques utilisés pour les langues parlées dans le modèle BDU.

2.2 Couches d'annotation concernées

Cette partie mobilise **quatre couches** dans ELAN, dans l'ordre suivant :

1. **Pauses**
2. **Signes allongés**
3. **Clignement des yeux**
4. **Segmentation prosodique**

Les trois premières couches servent à consigner les indices observés. La quatrième couche (Segmentation prosodique) synthétise ces observations en marquant les frontières d'unités.

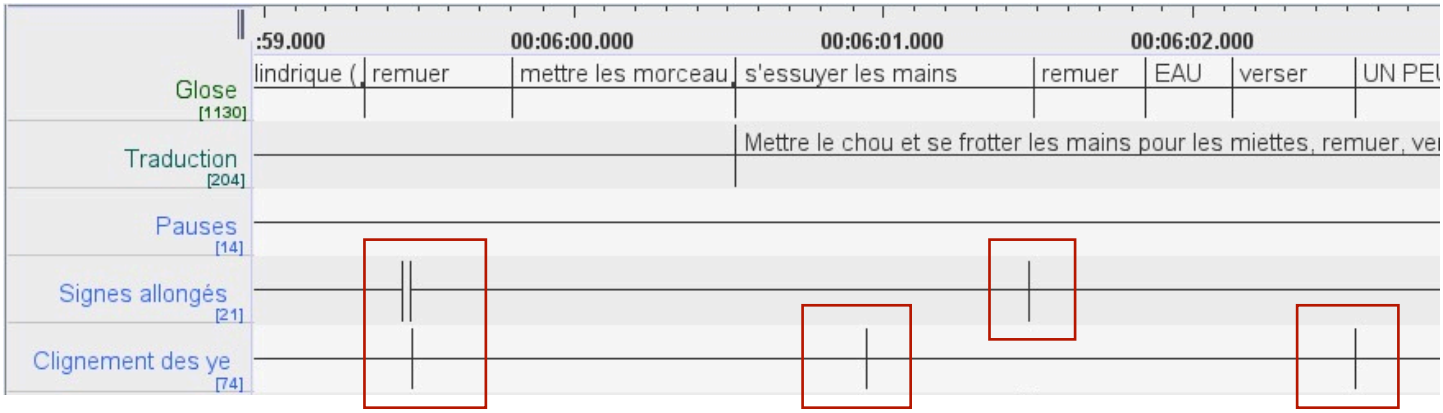
Les indices peuvent apparaître isolément ou combinés.

2.3 Indices prosodiques utilisés pour la segmentation

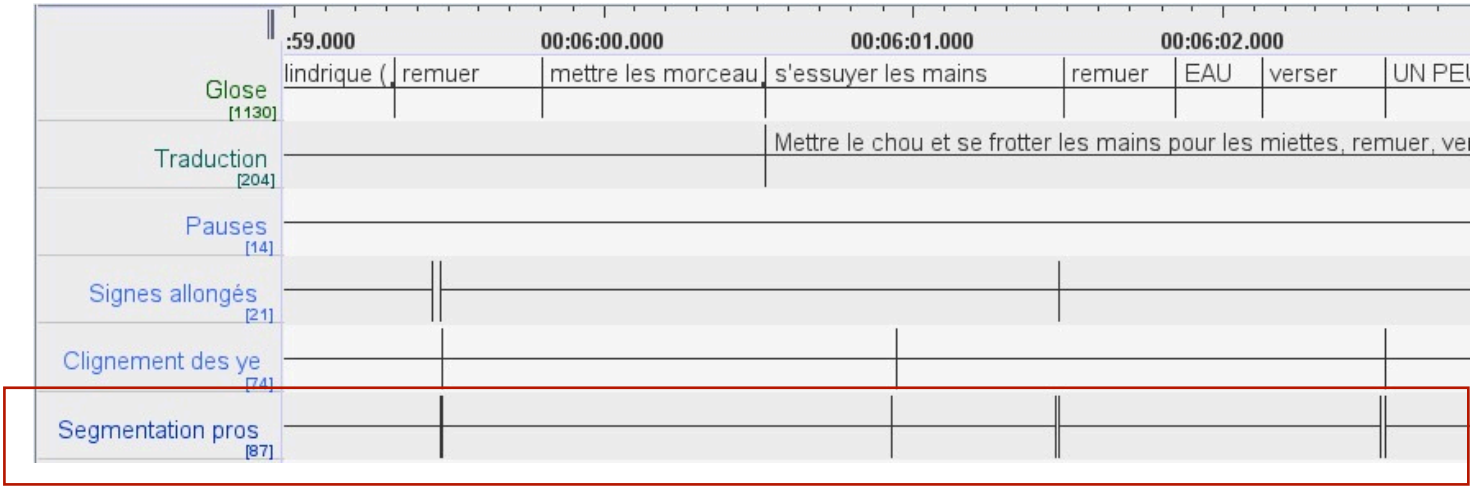
Couche ELAN	Comment l'identifier ?
Pauses	Les mains arrêtent de signer : elles peuvent rester immobiles, être relâchées le long du corps, croisées ou placées dans l'espace neutre..
Signes allongés	Le signe paraît nettement plus long que les signes voisins (environ trois fois plus long). par maintien ou ralentissement du mouvement.
Clignements des yeux	Le clignement apparaît en même temps qu'au moins un autre indice non manuel : **(-mouvement des sourcils, -ouverture des yeux,** **-direction du regard,** - **mouvement de la tête ou du corps,** - **gestes de la bouche,** - **expressions faciales).**

2.4 Méthode de segmentation

On fait glisser la souris sur la partie que l'on souhaite sélectionner, puis on arrête la sélection lorsqu'il y a un clignement des yeux, une pause ou un allongement, surtout en fin d'énoncé. La ligne sert alors de point de repère.

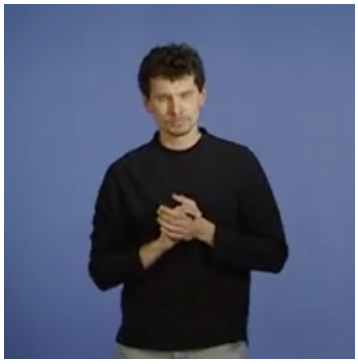


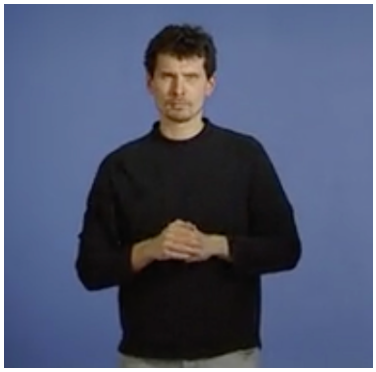
Ensuite, dans la couche de **segmentation prosodique,** on synthétise tous les indices prosodiques observés. Pour cela, on fait glisser la souris sur la section de segmentation prosodique, puis on arrête la sélection lorsqu'un indice prosodique déjà identifié auparavant apparaît.



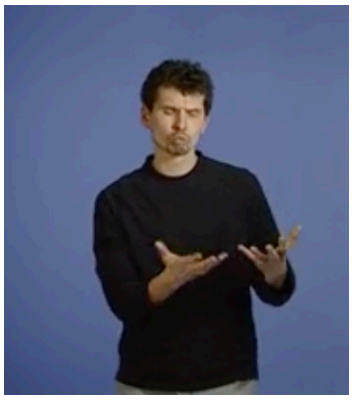
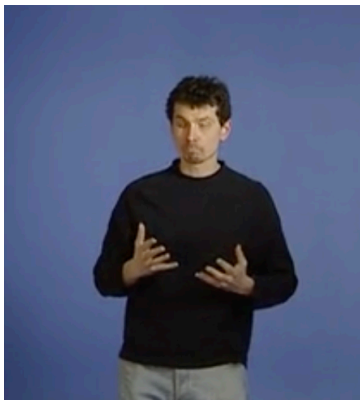
2.5 Exemples du corpus LS COLIN

Pauses:





Clignement des yeux :

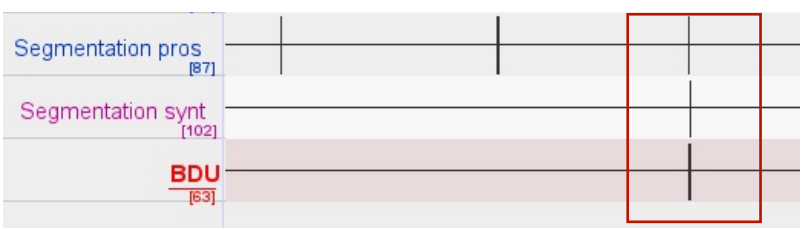
Partie 3 — Segmentation BDU

3.1 Principe

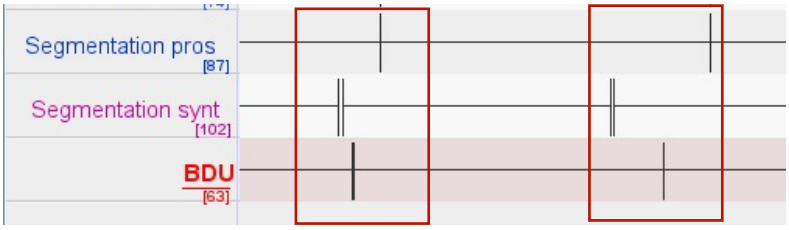
La segmentation BDU est effectuée dans une dernière étape, **après** les segmentations syntaxique et prosodique. Une unité BDU est délimitée aux endroits où les frontières des deux segmentations **convergent**.

3.2 Méthode de segmentation

La segmentation se fait sur la couche d'annotation **BDU** en faisant glisser la souris jusqu'à ce que la ligne de délimitation de la segmentation prosodique et celle de la segmentation syntaxique convergent. S'il n'y a que la segmentation prosodique, la segmentation n'est pas effectuée



Cependant, les lignes ne sont pas toujours parfaitement alignées ; elles peuvent néanmoins être prises en compte si la distance entre elles est faible.



Récapitulatif des couches d'annotation

Couche	Partie	Action de l'annotateur	Modifiable
Gloses	—	Aucune	Non
Traduction	—	Aucune	Non
Pauses	Prosodique	Marquer chaque pause observée	Oui
Signes allongés	Prosodique	Marquer chaque singe allongé	Oui
Clignements des yeux	Prosodique	Marquer les clignements accompagnés d'un autre indice non manuel	Oui
Segmentation prosodique	Prosodique	Marquer une frontière à chaque indice prosodique	Oui
Segmentation syntaxique	Syntaxique	Marquer une frontière à chaque nouvelle unité syntaxique	Oui
BDU	BDU	Marquer une frontière là où les couches 6 et 7 convergent	Oui

Déclaration sur l'honneur de non-plagiat

(à joindre au mémoire à la fin du document)

Je soussigné.e,

Nom, Prénom : Herandi GARCIA CONTRERAS
Régulièrement inscrit.e à l'Université de Toulouse II Jean Jaurès
N° étudiant : 22508938

Année universitaire : 2025 - 2026

certifie que le document joint à la présente déclaration est un travail original, que je n'ai ni recopié ni utilisé des idées ou des formulations tirées d'un ouvrage, article ou mémoire, en version imprimée ou électronique, sans mentionner précisément leur origine et que les citations intégrales sont signalées entre guillemets.

Fait à : Toulouse

Le : 18 septembre 2026

Signature :