

Critères de segmentation de la gestualité co-verbale

Gaëlle Ferré

LLING, Chemin de la Censive du Tertre, BP 81227, 44312 Nantes, cedex 3

Gaëlle.Ferre@univ-nantes.fr

RÉSUMÉ

La gestualité suscite un intérêt croissant chez les linguistes qui s'intéressent au caractère multimodal de la structuration de l'information : modalité verbale, vocale et visuelle. Cependant, la prise en compte des informations visuelles passe nécessairement par une réflexion sur les unités gestuelles. Qu'est-ce qu'un geste ? Comment subdiviser le flux gestuel en unités discrètes ? Quel degré de finesse est nécessaire dans la segmentation des unités gestuelles pour permettre leur mise en relation avec les informations relevant d'autres modalités ? A partir du corpus DEGELS fourni par les organisatrices de cet atelier de réflexion, nous décrivons les critères adoptés pour l'annotation de la gestualité co-verbale ainsi que de la direction du regard des locuteurs.

ABSTRACT

Segmentation criteria for the annotation of co-speech gestures

Gesture has been arousing a growing interest in linguists who are attracted by the multimodality of information structuring, organized into the verbal, vocal and visual modes. Yet, in order for visual information to be taken into account, one has to consider gesture units. What is a gesture? How can the constant flood of movement be subdivided into discrete units? What degree of fineness is necessary in the segmentation of gesture units to put them into relationship with information in other modes? Drawing on the DEGELS corpus provided by the organizers of the workshop, we describe the criteria adopted in our practice for the annotation of co-speech gesture as well as gaze direction.

MOTS-CLÉS : Annotation, gestualité, segmentation, unités.

KEYWORDS : Annotation, gesture, segmentation, units.

1 Introduction

Si la multimodalité prend une part croissante dans les études linguistiques, les corpus annotés d'enregistrements vidéos, prenant en compte la gestualité co-verbale, restent encore assez peu nombreux à l'heure actuelle. Cela ne signifie pas cependant qu'une réflexion n'a pas été conduite sur l'annotation de ces phénomènes linguistiques, bien au contraire. L'annotation de la gestualité a fait l'objet d'une réflexion concertée ces dix dernières années, au cours de divers projets de recherche, et c'est le travail de réflexion mené au cours d'un de ces projets que je souhaiterais présenter ici, lui-même alimenté par une longue pratique de ce type d'annotation dans les recherches personnelles des divers participants au projet. Il s'agit du projet ANR OTIM (ANR-08-BLAN-0239), dont le but est la constitution d'un corpus de français spontané, annoté dans diverses modalités (verbale, avec des annotations discursives et syntaxiques ; vocale, avec des annotations prosodiques ; et visuelle, avec des annotations des gestes et des postures).

Nous nous intéresserons ici à l'annotation des gestes, et plus précisément à la segmentation des unités gestuelles qui constituent la base d'une annotation multimodale, et présenterons une annotation du corpus DEGELS construite à partir d'une adaptation du schéma d'encodage proposé dans OTIM pour le corpus CID (Bertrand et al., 2008). Cette adaptation a permis d'une part de répondre aux consignes de l'atelier dans le but de faciliter le partage d'informations, mais aussi de corriger certaines imperfections du schéma d'encodage initial.

Sur le corpus CID, comme sur le corpus DEGELS, les annotations ont été entièrement réalisées de manière manuelle sous ANVIL (Kipp, 2001), ce qui a un impact certain sur le type d'annotation réalisé : il est impossible de noter tous les micro-mouvements des locuteurs car cela prendrait un temps considérable, ni d'avoir la précision d'une annotation automatique. Le gain par rapport à l'annotation automatique est cependant de pouvoir faire des inférences et des mises en relation entre signifiant et signifié sur le plan linguistique, ce qu'aucune machine n'est en mesure de réaliser à présent. La moindre précision par rapport à l'annotation automatique ne signifie pas cependant une absence totale de critères formels pour la segmentation. Ce sont ces critères que je vais tenter de préciser ici, en abordant dans un premier temps les gestes manuels, puis les mouvements de tête, de la face et du buste, qui présentent leurs propres particularités, pour finir par la direction du regard.

2 Les gestes manuels

La première question qui se pose lors de l'annotation des gestes manuels concerne le type d'unité que l'on va annoter. Prenons l'exemple du mouvement réalisé par l'expérimentatrice – et interlocutrice – dans DEGELS (entre 0.15 et 0.16 s dans le corpus) et reproduit en séquence dans la figure 1 ci-dessous. Il apparaît clairement que la main droite de l'interlocutrice est soulevée de son appui sur la cuisse, puis reposée dans une position légèrement modifiée, qu'elle va conserver jusqu'à la fin de l'extrait.

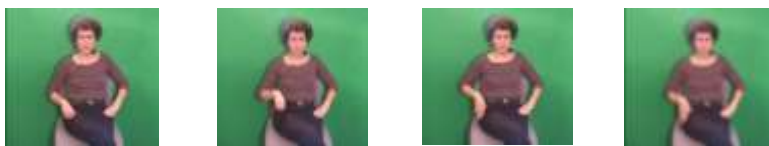


Figure 1 – Mouvement réalisé par l'expérimentatrice dans DEGELS.

Même si l'annotation s'est concentrée sur les gestes du locuteur et non sur ceux de l'interlocutrice, ce mouvement n'aurait de toute manière pas été annoté car il ne serait pas considéré comme un geste co-verbal, ni même comme un adaptateur (geste d'auto-contact¹), mais plutôt comme un changement de posture. Ainsi, pour être considéré comme geste co-verbal, un mouvement doit-il répondre aux critères suivants :

- être perceptible par l'annotateur,

¹ Les adaptateurs ne sont pas des gestes co-verbaux, mais ont été néanmoins annotés dans le projet OTIM dans la mesure où ils fournissent des indications sur la régulation des tours de parole. Aucun adaptateur n'a été annoté dans DEGELS, car le locuteur principal n'en produit pas sur cet extrait.

- opérer un contraste,
- participer d'une intention de communication.

Cela signifie que des mouvements de très petite amplitude, opérés qui plus est de manière très progressive ne seront pas perçus comme gestes. De même, certains gestes peuvent avoir une très petite amplitude (comme le fait de pianoter avec les doigts) et l'on peut imaginer que pour un locuteur atteint de tremblement, le geste devra se démarquer du mouvement généré par le tremblement afin d'être perçu comme geste, et donc opérer un contraste. De plus, pour que deux gestes enchaînés puissent être démarqués, il doit y avoir un contraste dans au moins une des caractéristiques entre les deux gestes (configuration de la main, direction du mouvement, type de mouvement, etc). Pour certains types de geste comme les mouvements de la tête vers la gauche ou la droite par exemple, le contraste opéré doit également être limité dans le temps. En effet, si un locuteur tourne la tête pendant une certaine durée, le mouvement sera interprété plutôt comme un changement de posture que comme un geste. A ma connaissance, aucune durée n'a été déterminée pour ce type de geste et l'annotateur hésite parfois entre geste et changement de posture, bien que le cas ne ce soit pas présenté dans DEGELS. Enfin, le mouvement doit participer d'une intention de communication pour être classé comme geste et c'est ce qui manque dans le celui qui illustré dans la figure 1. On pourrait imaginer que le fait de soulever la main (non nécessaire pour déplacer la main sur la cuisse) relève d'une intention de communication (un geste ébauché mais inachevé), mais cela n'est qu'une supposition et en l'absence de certitude, le mouvement ne sera pas considéré comme geste.

2.1 Segmentation des gestes manuels

Lorsque les gestes manuels sont produits en isolation, la/les mains est/sont d'abord en position de repos (muscles détendus, mains posées sur les cuisses par exemple), comme c'est le cas dans la figure 2 ci-dessous. Dans cette figure, les mains sont au repos, sans aucun mouvement dans les deux premières images. Les doigts commencent à s'écarter dans la troisième image pour ébaucher le geste. Le début du geste est donc noté entre la deuxième et la troisième image, car étant donné la granularité vidéo (25 images par seconde), le geste a commencé légèrement avant la troisième image. Ceci a été décrit également dans Ferré (2011). Pour la fin du geste, la frontière se situe juste avant l'image de retour à la position de repos (ainsi, dans DEGELS, les phases de rétraction et de rebond décrites dans la section suivante ne comptent pas dans la 'phrase gestuelle', contrairement à la segmentation qui a été adoptée pour OTIM).

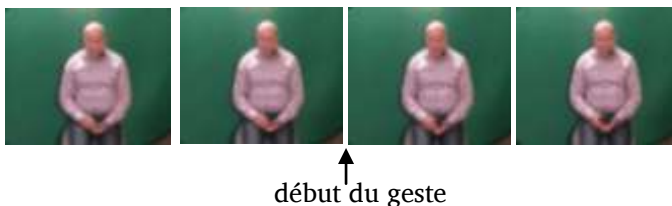


Figure 2 – Segmentation d'un geste depuis la position de repos.

La segmentation est différente lorsque deux gestes sont produits à la suite l'un de l'autre. Deux cas de figure se présentent alors : (a) le premier geste se termine par une

tenue des mains et dans ce cas, la frontière entre les deux gestes est posée après la fin de la tenue du premier geste (voir la section suivante pour une définition de la tenue du geste), et (b) la fin du premier geste est une phase dynamique et la frontière est posée immédiatement avant l'image qui figure soit un changement de direction du mouvement, soit un changement de la configuration de la main.

Dans le fichier d'annotation que je propose pour DEGELS, cette segmentation correspond aux 'phrases gestuelles' proposées par Kendon (2004 : 111) en termes de segmentation, mais à la typologie de McNeill (1992, 2005) pour ce qui concerne le lien entre le type de geste et la parole pour chaque étiquette. La typologie distingue entre les gestes iconiques, métaphoriques, déictiques, les battements purs (voir la section suivante pour d'autres types de battement), les emblèmes, à laquelle nous avons ajouté les adaptateurs dans le projet OTIM. Cette relation du geste avec le verbal possède une certaine influence sur la segmentation pour certains types de gestes bi-manuels asymétriques. Nous avons déjà noté dans Ferré (2011 : 39) que deux gestes, réalisés avec les deux mains, peuvent être produits en chevauchement, et alors que l'une des deux mains place un référent du discours dans l'espace, la seconde main, qui figure un deuxième référent, effectue un déplacement par rapport à la première main. On comptera donc ici deux unités distinctes.

2.2 Les phases gestuelles

Toujours en suivant la segmentation de Kendon (2004 : 112), chaque geste est ensuite segmenté en différentes 'phases gestuelles'. Dans le fichier d'annotation, cela correspond nécessairement aux pistes primaires exigées dans les consignes car les pistes primaires doivent constituer les plus petites unités dans Anvil. Les phases qui ont été retenues sont : la préparation ('preparation', mise en place des articulateurs), la réalisation ('stroke' : partie dynamique du geste), la tenue ('hold', les mains restent en tension mais ne bougent pas), la rétraction totale ou partielle (rétraction totale : retour des mains à une position de repos ; rétraction partielle : les mains n'atteignent pas la position de repos), le rebond ('recoil', les mains peuvent avoir un léger rebond lorsqu'elles sont posées sur la cuisse par exemple). A ces phases déterminées par Kendon, nous avons ajouté dans OTIM le battement lorsqu'un battement est réalisé au cours d'un autre geste (le plus souvent pendant la tenue).

2.2.1 Préparation

Dans le cas où le geste est réalisé de manière isolée (où les articulateurs partent d'une position de repos), il peut compter une phase de préparation (mais elle n'est pas obligatoire) : parfois, pour un mouvement de la main vers le bas par exemple, il est nécessaire dans un premier temps de lever la main. C'est ce qui se produit dans DEGELS à plusieurs reprises, ainsi que l'illustre la figure 3.

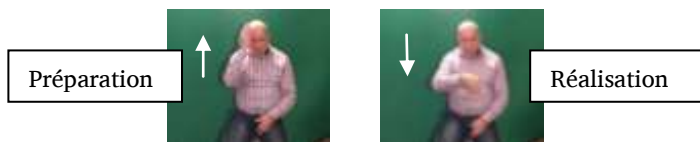


Figure 3 – Phases de préparation [image 1034] et de réalisation [image 1045] du geste

(la flèche indique la direction du mouvement).

Ici, l'on voit que le locuteur commence par lever la main droite pour ensuite la descendre pour indiquer une direction située devant lui, légèrement à gauche. La frontière entre les deux phases se situe juste après l'image montrant l'extension maximale de la préparation. La distinction entre la phase de réalisation et la préparation suppose donc un changement de direction du mouvement pour ce type de geste mais cela n'est pas nécessairement le cas. Ainsi, par exemple, dans le geste produit entre les images 1962 et 2008 sur « la route qui monte vers Sormiou », la phase de préparation consiste à changer la configuration manuelle par rapport à celle du geste précédent et ce n'est que lorsque cette nouvelle configuration est adoptée que le locuteur commence à lever la main en un mouvement qui évoque une route sinueuse, ainsi que le montre la figure 4.



Figure 4 – (a) Fin du geste précédent [image 1962] ; (b) fin de la phase de préparation [image 1968] ; (c) fin de la phase de réalisation [image 1987].

2.2.2 Tenue et battement

La phase de tenue correspond à une séquence comprenant au moins deux images sans mouvement de la part du locuteur, mais où les mains sont toujours dans la configuration adoptée pour le geste. Cette phase peut intervenir avant et/ou après la phase de réalisation. Contrairement à Kendon, qui cite plusieurs auteurs, nous ne distinguons pas la tenue produite avant la réalisation ('prestroke hold') de celle produite après ('poststroke hold'), dans la mesure où cette phase peut se substituer à la réalisation dans les gestes 'statiques'. DEGELS n'offre aucun exemple de ce type de geste, mais au cours des annotations produites sur le CID, nous avons remarqué que certains gestes impliquent un mouvement des articulateurs dans leur phase pertinente de réalisation, alors que d'autres sont statiques. Ainsi, juste avant l'exemple donné en figure 5, les mains de la locutrice étaient posées sur ses cuisses, puis en disant « j'ai lu le résumé », la locutrice les soulève pour les placer dans la configuration illustrée (phase de préparation) et les laisse sans faire de mouvement avant de les reposer. On a donc dans ce geste une phase de préparation et une tenue mais pas de phase de réalisation.



Figure 5 – Geste statique (tenu) sans phase de réalisation.

Enfin, l'on remarque dans les corpus de conversation spontanée, que la tenue du geste,

lorsqu'elle est présente, n'est pas toujours tout à fait parfaite. Le locuteur peut avoir à certains moments de la tenue un très léger mouvement de l'ordre d'un centimètre. Il importe que ce mouvement ne soit pas rapide pour ne pas opérer un contraste et ne pas être compris comme un battement. Et précisément, lorsqu'un mouvement rapide vers l'avant, les côtés ou vers le bas est réalisé alors que les mains sont placées dans la configuration pour la phase de réalisation et que le type de geste n'implique pas ce type de mouvement dans son sémantisme², alors, le locuteur réalise un battement. Ce battement compte comme une phase du geste dans la mesure où il n'est pas réalisé pour lui-même mais au cœur d'un autre geste. En ce qui concerne la segmentation, la tenue commence juste avant l'arrêt du mouvement et s'achève juste après. C'est l'inverse pour le battement.

2.2.3 Réalisation

La phase de réalisation du geste est une phase dynamique qui apporte son sémantisme au mouvement. Elle n'est pas nécessairement précédée d'une phase de préparation, comme le montre la figure 6.



Figure 6 – Deux gestes produits en séquence.

La figure 6 illustre deux gestes produits en séquence par le locuteur. 6(a) correspond au geste iconique qui accompagne « le rond-point du pouce » [image 1101, fin de la phase de réalisation] et 6(b) est la fin de la réalisation du geste métaphorique qui accompagne « avec le musée d'art euh » [image 1115]. Sitôt la configuration atteinte en 6(a), le locuteur effectue une rotation du poignet et une ouverture de la main pour atteindre la configuration en 6(b). Plusieurs interprétations sont possibles pour ce deuxième geste. On peut penser que l'aspect signifiant du geste est la tenue de la main paume ouverte orientée vers le haut (c'est le point de vue adopté par Kendon qui nomme ce type de geste 'Open Hand Supine' ou 'Palm up', op. cit. : 264, bien qu'il ne soit pas exactement certain qu'il s'agisse du même type de geste). Dans ce cas, le geste comporte une préparation et une tenue. Mais on peut aussi penser à l'instar de Kipp (2004 : 267, 'hand flip') que l'aspect signifiant du geste est précisément cette rotation du poignet et dans ce cas, la rotation constitue la phase de réalisation du geste. C'est le point de vue qui a été adopté ici et dans OTIM. On voit alors que le deuxième geste enchaîne directement sur le premier avec une phase de réalisation et la frontière se situe donc ici à l'image 1101.

2.2.4 Rétraction totale et partielle

La rétraction totale consiste à un retour à une position de repos (main pendante, posée

² On pense par exemple à un geste accompagnant un verbe de type « frapper » qui serait alors considéré comme un geste iconique, mimant l'action de frapper.

sur un accoudoir, sur les cuisses, etc.) c'est-à-dire une position qui n'implique aucune tension musculaire comme c'est le cas par exemple de la main gauche dans la figure 6(a), par rapport à 6(b) qui montre une légère ouverture de la main. Si la phase de réalisation consiste en un mouvement vers le bas et qu'il n'y a aucun changement de configuration manuelle entre la réalisation et la position de repos, alors il n'y a pas de phase de rétraction, celle-ci étant englobée dans la réalisation. Parfois, le locuteur initie un mouvement vers la position de repos, ou bien referme la main par exemple après l'avoir ouverte lors de la réalisation mais sans atteindre la position de repos et enchaîne ensuite immédiatement sur un autre geste, on compte alors une rétraction partielle.

2.2.5 Rebond

Le cas ne se présente pas dans DEGELS, mais il arrive que la main du locuteur ait un léger rebond lorsque la rétraction est rapide sur les cuisses du locuteur ou sur un accoudoir, par exemple. C'est une phase purement physiologique dans laquelle la main se soulève légèrement avant de retomber. La segmentation suit le principe adopté pour les autres phases : la frontière de début de l'unité est placée juste avant que la main se soulève et la frontière de fin est placée juste après que la main soit posée.

2.3 Les unités gestuelles

A l'instar de Kendon (2004 : 111-124), nous distinguons une unité supérieure dans la hiérarchie des unités manuelles : les unités gestuelles ('gesture units'). Ceci constitue une adaptation par rapport au schéma d'encodage utilisé dans OTIM. Selon Kendon, (op. cit., p. 111), les unités gestuelles constituent des excursions de la / des main(s) du locuteur, basées en partie sur la segmentation des phrases gestuelles. Les étiquettes permettent de noter si les unités gestuelles sont réalisées avec une seule main, avec les deux mains, de manière symétrique ou asymétrique.



Figure 7 – Deux unités gestuelles distinctes mais de même nature.

Ainsi que le montre la figure 7, à deux reprises, le locuteur se touche le pouce en évoquant la « statue du pouce ». Les deux gestes impliquent les deux mains et constituent deux unités gestuelles distinctes dans la mesure où ils sont séparés par la rétraction partielle dont nous avons vu plus haut qu'elle ne faisait pas partie du geste.

Dans la figure 8 ci-dessous apparaissent également deux unités gestuelles distinctes, mais celles-ci sont cette fois de nature différente. Dans un premier temps, le locuteur enchaîne deux gestes bimanuels (déictique et métaphorique). Ces deux gestes constituent une seule unité dans la mesure où la main ne retourne pas à une position de repos entre les deux. Mais ensuite, le locuteur rétracte la main gauche (bien que la phase de rétraction ne fasse pas partie du geste) et continue la séquence avec un geste réalisé de la main droite. J'ai décidé que dans la mesure où le locuteur change de mode de gestualisation, alors il entame une nouvelle unité gestuelle.



Figure 8 – Deux unités gestuelles distinctes de nature différente.

Un cas légèrement différent se présente dans la figure 9 ci-dessous. Le locuteur produit un geste iconique avec la main droite, qui compte comme une unité gestuelle sur la première tire. Puis il enchaîne directement avec un geste métaphorique réalisé également avec la main droite (dernière tire sur la figure). On pourrait donc penser que ce métaphorique fait partie de la même unité gestuelle que le geste iconique. Cependant, pendant la production de ce second geste, un léger métaphorique est produit avec la main gauche du locuteur (piste 5), en chevauchement avec le métaphorique produit par la main droite. Ces légers écartements de la main gauche ne me semblent pas faire partie des gestes produits avec la main droite et passent d'ailleurs relativement inaperçus puisqu'ils n'avaient pas été notés lors de la première édition de l'atelier. Ici, dans la mesure où les deux mains réalisent chacune des gestes en chevauchement, on peut compter une unité gestuelle dans laquelle les deux mains fonctionnent de manière asymétrique.



Figure 9 – Deux mains asymétriques.

En résumé, ce qui compte pour la segmentation des unités gestuelles est : (a) la fin d'une phrase gestuelle qui n'enchaîne pas directement sur une autre phrase gestuelle, et (b) un changement de mode de gestualisation (passage d'un geste à une seule main à un geste bimanuel et vice versa, ou passage d'un geste bimanuel symétrique à un geste bimanuel asymétrique et vice versa).

3 Les gestes non manuels

Les gestes non manuels sont beaucoup plus simples à annoter dans le schéma d'encodage proposé ici, dans la mesure où il ne possède qu'un seul niveau hiérarchique. En effet, alors que les gestes manuels sont annotés en termes d'unités gestuelles, elles-mêmes décomposées en phrases gestuelles, à leur tour décomposées en phases gestuelles, le tout formant trois pistes d'annotation (multipliées pour les phrases et les phases par le fait que le geste est bimanuel, réalisé avec la main gauche ou réalisé avec la main droite), les gestes non manuels ne sont décrits que sur une seule piste.

3.1 La tête

En ce qui concerne la tête, il faut distinguer les gestes des changements de posture. L'annotation repose sur le principe que la posture de repos par défaut est lorsque la tête

est orientée dans l'alignement du corps du locuteur, sans tension au niveau du cou. Tout mouvement par rapport à cette position de repos peut être interprété comme un changement de posture ou comme un geste. Un changement de posture implique qu'une fois la nouvelle posture adoptée, la tête ne bouge plus. En revanche, aucune durée n'a été déterminée dans OTIM pour distinguer entre geste et changement de posture. Parmi les gestes, on compte les 'nods' (acquiescements) et les 'shakes' (gestes de négation). Mais aussi, les 'tilts' (inclinaisons de la tête sur le côté), les 'pointings' (pointages du menton), les 'jerks' (rejets de la tête vers l'arrière) et les 'beats' (mouvements du menton vers le bas sans valeur d'acquiescement). Le principe de segmentation adopté est le même que pour les gestes manuels : le geste commence juste avant l'image où la tête quitte la position de repos, et s'arrête juste après le retour à la position de repos. La précision est cependant moins grande que pour les gestes manuels dans la mesure où (a) certains gestes de la tête ont une très petite amplitude et, tout en étant perceptibles lorsque la vidéo défile à vitesse réelle, deviennent difficiles à distinguer de la position de repos dans l'annotation image par image, et (b) la position de repos n'est pas elle-même définie au pixel près.

3.2 Les sourcils

Ekman et al. (2002) ont répertorié dans leur guide d'annotation des mouvements de la face un grand nombre de mouvements des sourcils. Dans OTIM et DEGELS, cependant, seules deux positions ont été retenues 'raising' (sourcils haussés) et 'frowning' (sourcils froncés). Ceci s'explique par le type de corpus utilisés : il s'agit de corpus de type conversationnel à faible charge émotionnelle, dans lesquels le visage des locuteurs n'est pas filmé en gros plan. Cette absence de gros plan rend difficile l'annotation fine de mouvements de faible amplitude, mais les différents mouvements relevés par Ekman et al. ne sont pas non plus fréquents sur ce type de corpus où la charge émotionnelle des locuteurs est faible. Une étude des émotions exigerait de travailler sur un tout autre type de corpus. En d'autres termes, pour un travail sur le lien entre le verbal et le non-verbal hors émotions, les deux valeurs sont amplement suffisantes. La segmentation des mouvements des sourcils repose sur les principes adoptés pour la tête sans la distinction posture / geste. La position par défaut (non annotée) correspond à celle de la figure 10(a). En 10(b), le locuteur hausse les sourcils et en 10(c), c'est l'interlocutrice qui fronce les sourcils.

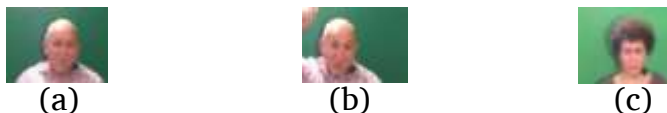


Figure 10 – Mouvements des sourcils. (a) Position de repos ; (b) sourcils haussés ; (c) sourcils froncés.

La segmentation dans Anvil se fait comme pour la tête juste avant le début du mouvement pour l'onset et juste après la fin du mouvement pour l'offset. Les mouvements des sourcils présentent cependant une particularité que ne présentent pas les mouvements de tête : à l'instar des sourires (annotés en partie dans OTIM, mais pas dans DEGELS), le début des mouvements des sourcils est facile à identifier car les sourcils passent très rapidement de la position de repos à la position haute ou basse. En

revanche, le retour à la position de repos se fait le plus souvent de manière très progressive et la fin de la segmentation est donc plus difficile à repérer et donc moins précise.

3.3 Le buste

Beaucoup plus complexe dans OTIM, l'annotation des mouvements du buste a été limitée dans DEGELS à deux valeurs 'forwards' (vers l'avant) et 'backwards' (vers l'arrière), le locuteur changeant relativement peu de posture dans ce court enregistrement. Le principe de segmentation est une fois encore identique à celui des autres niveaux : le début de chaque étiquette est fixé juste avant le début du mouvement et la fin des étiquettes juste après le retour à la position de repos. La position de repos est lorsque le locuteur se tient droit, le dos appuyé sur le dossier du siège, sans pression perceptible. Il se penche vers l'avant lorsque son dos décolle du siège et vers l'arrière lorsqu'une pression sur le dossier et un recul des épaules est perceptible. DEGELS présente un avantage pour l'annotation des mouvements du buste par rapport au CID, le corpus utilisé dans OTIM : la vue de profil permet de mieux appréhender les mouvements vers l'avant ou vers l'arrière du buste.

4 Le regard

L'annotation de la direction du regard comporte deux types d'information. Certaines informations sont d'ordre interactionnel : regard vers l'interlocuteur, regard vers une partie du corps du locuteur (pointage du regard : nous avons vu l'an dernier dans l'atelier DEGELS 2011 que le regard peut instancier un geste manuel et constituer un pointage). D'autres informations ne sont pas d'ordre interactionnel : sur le côté (à droite, à gauche), en haut / en bas (à droite, à gauche). Les changements d'orientation du regard sont le plus souvent accompagnés d'une fermeture des paupières. La segmentation se fait donc juste avant la réouverture des paupières pour le début de chaque étiquette et juste après la première image de fermeture des paupières pour la fin de chaque étiquette. Les fermetures des paupières n'ont pas été annotées dans DEGELS mais est prévue dans le schéma d'encodage d'OTIM. Dans DEGELS, lorsque le regard est orienté vers le bas, il n'est pas toujours possible de savoir si les paupières sont ouvertes ou fermées (c'est le cas notamment sur tout le début de l'enregistrement). Dans ce cas, l'annotation de la direction du regard n'a pas été réalisée. Il en a été de même lorsque les paupières du locuteur sont presque closes et que la direction du regard est alors incertaine. Lorsque le regard change de direction sans fermeture des paupières, nous avons segmenté le début du changement avant la première image sur laquelle il s'affiche. Il est certain que lors d'un changement de direction du regard, celui-ci effectue un changement progressif avant d'atteindre la nouvelle direction. Cependant, nous avons choisi de ne pas annoter cette mise en place du regard en raison de la pénibilité que présente cette annotation par rapport à l'annotation des autres types de mouvement. Pour finir, il est important de noter que ce type d'annotation ne permet de noter que la direction du regard et non ce que perçoit effectivement le participant ou ce sur quoi il se concentre.

L'ensemble des critères de segmentation des gestes est repris dans la Table 1 ci-dessous.

Critères formels	Perception		Segmentation
<p>Toute unité gestuelle</p>	<p>Mouvement perceptible par l'annotateur</p> <p>Contraste avec ce qui précède et ce qui suit</p> <p>Intention de communication</p>		<p>Début : image précédant le début du mouvement</p> <p>Fin : image précédant le retour au repos ou le changement de direction ou de configuration de la main</p>
Gestes manuels	Segmentation		
Phases	Préparation	Mise en place des articulateurs	
	Tenue	Absence de mouvement de la/des main(s) dans la configuration utilisée lors de la réalisation	
	Battement	Mouvement rapide vers l'avant, les côtés ou vers le bas est réalisé alors que les mains sont placées dans la configuration pour la phase de réalisation	
	Réalisation	Phase dynamique qui apporte son sémantisme au mouvement	
	Rétraction	Totale	Retour à la position de repos
		Partielle	Mouvement vers la position de repos sans que celle-ci soit atteinte
	Rebond	Après le retour au repos, la main se soulève légèrement en rebondissant sur un objet ou une partie du corps	
Phrases	Unités sémantiques qui commence à la préparation et se termine après la réalisation ou la tenue en excluant la rétraction		
Unités	Excursions de la / des main(s) du locuteur sans retour à la position de repos ou sans changement de mode (1 main vs. 2 mains par exemple)		

Gestes non manuels	Tête	Sourcils	Buste	Regard
Une seule unité (pas de subdivision) répondant aux critères formels				

TABLE 1 – Critères de segmentation des gestes manuels et non manuels du corpus DEGELS.

5 Conclusion

Les critères de segmentation de la gestualité co-verbale que nous avons présentés ici – et qui ont pour objet l’annotation du corpus DEGELS – s’inscrivent dans ce que Boutet (2008 : 82) nomme un ‘repérage axial’ (haut/bas, gauche/droite) et présentent donc une précision moindre par rapport au ‘repérage polaire’ décrit dans le même article. Ce type de segmentation correspond cependant mieux aux besoins de l’analyse de l’interaction conversationnelle définis dans le projet OTIM dont il s’inspire. Une segmentation de type morphologique ou même des mesures très précises réalisées avec des systèmes de capture de mouvement apportent une information si dense qu’il peut être difficile parfois de la mettre en relation avec des actes discursifs.

L’annotation présente une segmentation hiérarchique des gestes manuels, organisée autour de trois types d’étiquette, basés sur les travaux de Kendon (2004) et McNeill (1992, 2005) : les ‘unités gestuelles’ sont formées par les déplacements des mains ou des doigts sans retour à la position de repos. Nous distinguons également des gestes produits avec une seule main de ceux produits avec deux mains, de manière symétrique (les deux mains réalisent une seule unité gestuelle) ou asymétrique (chaque main réalise une unité gestuelle différente de manière plus ou moins simultanée). Ces unités gestuelles se décomposent en une ou plusieurs ‘phrases gestuelles’ dont le sémantisme se distingue des autres, ainsi que leurs caractéristiques formelles telles que la direction du mouvement, la configuration de la main, etc. Enfin, les phrases gestuelles sont elles-mêmes décomposées en ‘phases gestuelles’ où l’on distingue la partie pertinente du geste (qui lui donne son sémantisme) des phases de mise en place et de rétraction des articulateurs ou encore de leur tenue.

Les gestes non manuels tels que les mouvements de tête, des sourcils et du buste ont également été présentés ici. Leur segmentation est moins complexe – en termes de hiérarchie – que celle des gestes manuels puisque seul l’équivalent des ‘phrases gestuelles’ est annoté. Il est important cependant de savoir que le principe de segmentation (à quelle image débute et se termine l’unité) adopté pour les gestes non manuels est strictement identique à celui qui a été adopté pour les gestes manuels.

Enfin, la segmentation de la direction du regard des locuteurs a également été présentée, avec ses difficultés particulières et notamment le fait qu’elle est dépendante de l’ouverture des paupières contrairement aux autres annotations sur ce type de corpus.

Pour finir cet article, il me semble que la démarche de mise en commun des pratiques

d'annotation et de segmentation proposée par l'atelier DEGELS est très important pour la communauté des gestualistes en France car elle n'a pas été effectuée jusque-là et il est extrêmement difficile d'établir un dialogue entre les chercheurs d'une communauté sans un minimum de partage des principes de base de l'annotation qui pourront plus tard être transmis aux jeunes chercheurs.

BERTRAND, R., et al. (2008). Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *TAL* 49, pages 105-133.

BOUTET, D. (2008). Une Morphologie De La Gestualité : Structuration Articulaire, *Cahiers De Linguistique Analogique* 5, pages 80-115.

EKMAN, P., FRIESEN, W. V et HAGER, J. C. (2002). The Facial Action Coding System (2nd ed.). <http://www.face-and-emotion.com/dataface/facs/manual/TitlePage.html>.

FERRÉ, G. (2011). Annotation multimodale du français parlé. Le cas des pointages. In *Proceedings of TALN - Atelier Degels*, Montpellier, 1er juillet 2011, pages 29-43.

KENDON, A. (2004). *Gesture. Visible Action as Utterance*. CUP, Cambridge.

KIPP, M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark, pages 1367-1370.

KIPP, M. (2004) *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Boca Raton, Florida.

MCNEILL, D. (1992). *Hand and Mind : What Gestures Reveal about Thought*. The University of Chicago Press, Chicago and London.

MCNEILL, D. (2005). *Gesture & Thought*. The University of Chicago Press, Chicago and London.