Par où couper pour aller à la plage ?

Dominique Boutet,[1] *Karine Martel,*[2]*Marion Blondel*[1]
(1) SFL, CNRS-Paris8, 59 rue Pouchet, 75017 Paris
(2) Laboratoire PALM (EA4659) Université de Caen Basse-Normandie, Esplanade de la Paix, 14032 Caen cedex
dominique_jean.boutet@orange.fr, karine.martel@unicaen.fr, marion.blondel@sfl.cnrs.fr

RÉSUMÉ

La multimodalité représente un véritable défi pour le traitement du langage, particulièrement quand on s'intéresse à la question de la segmentation du discours dialogique généré simultanément à travers les canaux vocaux et gestuels. Cette étude porte sur les gestes de pointages manuels employés par un locuteur entendant lors d'une tâche d'explication d'un itinéraire. Les auteurs examinent successivement les questions suivantes : quels sont les critères formels pertinents pour segmenter et identifier le bornage relatif aux unités gestuelles ? Quels sont les critères les plus appropriés pour annoter le signal vocal ? Quel est le degré de granularité le plus pertinent pour rendre compte de l'interaction éventuelle entre les gestes manuels et vocaux ? Ils fournissent une description et une annotation de ces gestes de pointage à l'aide du logiciel ELAN et proposent une approche *bottom-up* pour segmenter et catégoriser les tronçons pertinents, alors que les précédentes études ont associé la plupart du temps des critères formels et fonctionnels dans un cadre *top-down*.

ABSTRACT

Where do you switch to get the band?

Multimodality is a challenge to natural language processing, especially if one is interested in finding out how one should segment a dialogic discourse generated through vocal and gestural channels simultaneously. This paper focuses on the manual pointing gestures a hearing speaker uses while performing a map task. We will successively address the following issues: which formal criteria are relevant in segmenting and identifying terminals for gestural units? Likewise, what criteria can we suggest as appropriate to the prosodic vocal flow? Which degree of granularity is the more relevant to account for the potential interaction between vocal and manual 'gestures'? We not only provide a description and an annotation of these pointing gestures with the ELAN tool; we also claim for a bottom-up design for segmenting and categorizing the relevant chunks, while previous studies have usually mixed formal and functional criteria in a top-down strategy.

MOTS-CLÉS : segmentation, multimodalité, alignement, modèle d'annotations
KEYWORDS : segmentation, multimodality, alignment, template for annotations

JEP-TALN-RECITAL 2012, Atelier DEGELS 2012: Défi GEste Langue des Signes, pages 23–39,
Grenoble, 4 au 8 juin 2012.

1 Introduction

A quel niveau de l'analyse doit-on situer la multimodalité que l'on veut étudier ? Autrement dit, un constat fait de manière unanime, à savoir la multimodalité des phénomènes de sens, peut-il infuser l'analyse et jusqu'où ? La segmentation des unités constitue la première phase d'extraction des données d'un corpus et ne devrait subir aucune influence des objectifs de la recherche qu'elle permettra, faute de quoi elle enclencherait un processus circulaire. Les segments obtenus, et empruntant des modalités variées, peuvent-ils être simplement mis en présence, et ce indépendamment de leur inclusion dans une modalité spécifique, au seul motif qu'ils apparaissent en même temps ? On gagnerait certainement à explorer au préalable le sens à attribuer à chaque unité au sein de sa modalité et de la sémiose qui l'accompagne. Au fond, pour étudier la multimodalité, il faut choisir entre la recherche sur plusieurs modalités des traits qui composent un sens *a priori* (démarche *top down*), ou bien la recherche du sens qui émerge d'un ensemble d'unités multimodales dans une construction sémiotique d'unités, elles-mêmes porteuses de sens (démarche *bottom up*).

Nous examinerons ici quelques modes de segmentation qui amènent des questionnements quant au statut des unités -- gestuelles en particulier, mais également prosodiques vocales -- et des rapports qu'elles entretiennent avec la modalité verbale au niveau segmental. Il s'agit moins ici de pointer telle ou telle dysharmonie que d'essayer d'en extraire une démarche réflexive.

2 On segmente en référence à quoi ?

Un segment a une valeur sémiologique, certes mais par rapport à quoi ? Autrement dit, il nous faut comprendre si le lien entre le segment et ce par rapport à quoi il existe, est proprement référentiel (mis dans cette modalité pour quelque chose venant d'ailleurs) ou bien s'il est plus différentiel (émergence d'unités par différence dans la même modalité). Nous souhaitons retenir la conception sémiologique différentialiste pour de simples raisons méthodologiques de stabilisation des unités à considérer : si des unités gestuelles et prosodiques existent en tant que telles et montrent une extension trans-langagière, alors leur stabilisation ne peut dépendre entièrement d'une conception référentielle le plus souvent verbale et co-occurrente.

2.1 Selon la modalité

Deux types de segmentation peuvent être envisagés, celui d'une segmentation faite en référence à la modalité vocale-verbale co-occurrente ou co-présente, ou bien celui d'une segmentation en rapport avec la même modalité gestuelle. Autrement dit, la question posée est la suivante : segmente-t-on la gestualité en fonction d'un fait externe ou d'un fait interne ? La réponse à cette question loin d'être univoque révèle des chemins tortueux. Cette question de la référence ultime qui dirige la segmentation oriente y compris les discours tenus autour des phénomènes gestuels.

La difficulté de mettre en correspondance un alignement temporel entre le vocal-verbal et les gestes co-occurrents ne laisse que très peu d'espoir d'établir une relation biunivoque entre ce que véhiculent le canal gestuel et le contenu vocal-verbal. On doit

donc s'en remettre aux détours d'un alignement strict. Le découpage proposé par Kendon (Kendon 1972), largement repris (Kendon 1980; McNeill 1992 ; Guidetti 2002 ; Colletta et al. 2009 ; Ferré et al. 2007), dont certains se sont inspirés (Allwood, J. et al. 2004), que d'autres ont complété (Bressem 1998) ou ont précisé (Kita, van Gijn & van der Hulst 1998) est basé sur des caractéristiques proprement gestuelles. Il correspond bien au deuxième type de segmentation. L'emboîtement des entités gestuelles en *Phrases*, puis *Phases* et *Unités* augure bien d'un découpage en référence à la seule modalité gestuelle. Pourtant, malgré ce découpage, l'alignement entre des unités constituées pour les modalités gestuelle et verbale constitue un point aveugle des recherches sur la gestualité (McNeill & Duncan 2000 ; Quek et al. 2002). Les objectifs des études mettant en œuvre l'annotation des gestes tournent très majoritairement autour des rapports entretenus entre les deux canaux du point de vue de la sémiose (Kendon 1988), de l'interaction (Mondada 2009), de l'origine du langage (Kendon 1991 ; Corballis 2002) ou de la cognition (McNeill 1992 ; Rizzolatti & Arbib 1998). Nous différencions bien les objectifs d'une annotation mettant en présence le vocal-verbal et la gestualité co-expressive, de la référence par rapport à laquelle on segmente la gestualité. Cette dernière segmentation gagne même à être indépendante de ce avec quoi elle sera mise en relation. Les conditions d'une étude convenable de la réalité des interactions entre modalités exigent même que les segmentations d'unités soient effectuées pour elles-mêmes, chacune dans leur modalité ; l'alignement ou le quasi-alignement temporel servant de lieu d'interactions. On l'aura compris, le risque sinon est de créer les conditions d'une analyse asymétrique assujettissant la gestualité à la verbalité. Tant qu'il s'agit de la même modalité (par exemple pour les langues des signes, désormais LS), il n'y a finalement que la recherche d'assignation de formes à une fonction ou à un sens. Légitime au sein d'une même modalité, cette démarche pose un questionnement méthodologique majeur dès lors qu'on souhaite aborder les situations/les événements sur un plan transmodal. On accrocherait artificiellement une segmentation d'un canal à une catégorisation d'un autre canal dont on ne connaît pas le domaine d'extension et l'empan que couvre chaque segment. On l'a vu, les modalités de segmentation largement en vigueur dans la gestualité évite cet écueil. Pour autant, la segmentation de la gestualité est-elle exempte de toute contamination du canal vocal-verbal ? Dans les faits, la partie signifiante de l'unité gestuelle, le *stroke,* est définie par McNeill comme le pic de l'effort du geste. C'est pendant cette phase que la signification du geste est exprimée (McNeill 1992:83 et 375–376). Il ajoute que cette phase de *stroke* est synchronisée avec les segments langagiers avec lesquels il est co-expressif. Ainsi deux traits définitoires — kinésique et sémantique — délimitent cette phase des autres (la préparation ou la rétraction). Kita *et al* commentent d'ailleurs ce point en disant que le *stroke* se définit autant formellement que fonctionnellement (Kita, van Gijn & van der Hulst 1998 : 27). Ainsi, la part sémantique de la segmentation gestuelle repose-t-elle sur une connaissance langagière verbale *a priori*. Quand bien même des précautions seraient prises à l'instar de ce qui se passe pour les LS (la lemmatisation, Johnston 2008), on n'a pas actuellement, pour la gestualité, de lexique établi sur la base duquel on pourrait procéder à une segmentation. Le passage entre phases gestuelles (préparation, stroke, rétraction), autrement dit la segmentation de la gestualité même, est donc en partie tributaire de traits sémantiques de la modalité verbale.

Nous n'aborderons pas ici l'autre manière de segmenter très en rapport avec un

étiquetage verbal, parce que ce procédé est désormais très marginal.

Un raisonnement implicite sous-tend les remarques précédentes : la segmentation ne pouvant être faite qu'en fonction de critères, dans l'absolu il faut que ceux-ci relèvent du même support que celui qu'on segmente. La mise en parallèle de phénomènes relevant de modalités différentes ressort plutôt d'un niveau d'analyse que de celui de la segmentation. Mettre en place cette référence transmodale dès la segmentation revient à forcer l'analyse dans le sens d'une asymétrie : la gestualité ne peut alors qu'être dépendante du canal vocal-verbal

2.2 Selon le type de catégories

La segmentation peut dépendre de plusieurs types de catégories : fonctionnelle, formelle (discussion sur quelques définitions formelles du pointage, Wilkins 2003), sémantique (discussion par Povinelli, Bering & Giambrone 2003 sur la forme du pointage et sa signification chez le chimpanzé, mais aussi Calbris 1990).

Un découpage en rapport avec une distinction fonctionnelle tel que les déictiques/anaphoriques verbaux peut montrer une différenciation sur la gestuelle (Kendon & Versante 2003 : 134). Les déictiques pourraient répondre à des gestes de pointage dans l'espace (situés), tandis que les pointages à valeur anaphorique répondraient à des déplacements ou des directions axiales (antériorité, postériorité sur une « ligne discursive »). C'est en tout cas l'hypothèse que nous avons formulée (Boutet et al. 2011 : 18). On peut aussi postuler que la gestualité présente cette distinction entre des déictiques renvoyant à un espace situé et à des pointages plus anaphoriques (pointages abstraits pour McNeill 1992 : 173) en rapport avec ce qui a été déposé gestuellement dans l'espace discursif. Cette distinction linguistique serait ainsi transférable sur des phénomènes gestuels. C'est ce type de transfert que nous essayons de valider ici sur la gestualité. 61% des anaphores verbales du corpus sont précédées par un geste de pointage relevant d'un mouvement secondaire (mouvement non aligné avec le vecteur général du pointage principal). 72% de ces mouvements sont faits en aller-retour et parallèlement, 68% des mouvements en aller-retour sont associées à des anaphores verbales. On a bien ici une utilisation préférentielle de mouvements en aller-retour non alignés avec le vecteur principal du pointage, mouvements qui déterminent ainsi un pointage secondaire marquant des anaphores verbales.

Un autre type de découpage permet de segmenter en fonction d'éléments formels, qu'ils soient ou non du même canal. La difficulté essentielle provient alors de la quasi absence de système de transcription adapté aux LS (à l'exception d'Hamnosys, Hanke 2004), ou à la gestualité. Nous n'insisterons pas sur ce point (Boutet & Garcia 2006 : 33) qui constitue pourtant un obstacle majeur à une annotation lemmatisable pour les LS par exemple (pour une exception voir Johnston 2008). Selon une approche formelle dans un cadre transmodal, une des grandes difficultés de la segmentation consiste à trouver un tempo commun entre les unités gestuelles et les unités verbales et vocales (Quek et al. 2002). Ces battements par minute ou par seconde (tempo) doivent être communs au canal voco-verbal et au canal gestuel, en l'absence de connaissances sur l'organisation exacte du sens pour la gestualité. En effet, ne connaissant pas le contenu sémantique de ce que l'on mesure exactement, il faut se donner la même jauge pour pouvoir le faire. Les

phénomènes formels voco-verbaux sont d'une durée très brève, de l'ordre de 200 millièmes de seconde, or il n'y a pas actuellement de moyen simple et direct de procéder à un découpage labellisé de cet ordre de durée pour la gestualité (*Cf. supra* absence de système de transcription). Pour ce corpus, nous avons procédé à une segmentation progressive (voir *infra*) d'une piste « Phase Geste » reprenant les catégories de Kendon, en passant par une piste segmentant les « directions du pointage », pour segmenter encore les mouvements de chaque main lors des pointages. Pour ces mouvements, nous arrivons à une durée moyenne équivalente à celle des mots du corpus (M : 0,22s), comme le montre le tableau (1).

Entrée de la piste MvtMD (nbre)	Durée moyenne des annotations en s.
FLEXION (26)	0,218
A-R FLEXION (13)	0,292
EXTENSION (26)	0,229
A-R EXTENSION (5)	0,261
ABDUCTION (27)	0,209
A-R ABDUCTION (11)	0,420
ADDUCTION (25)	0,217
A-R ADDUCTION (6)	0,290

TABLEAU 1 – Durée moyenne par mouvement (*A-R* pour *aller-retour*)

Les seules catégories formelles relativement 'étiquetables' relèvent *i/* de l'espace, *ii/*de critères physiques (vitesse, accélération, segments physiologiques), *iii/* de critères s'apparentant à du verbal et provenant des LS (paramètres tels que la configuration, l'emplacement, l'orientation, le mouvement). Nous avons choisi de recourir à des critères relevant de l'espace (*i/*) et des aspects physique et physiologique (*ii/*). Ainsi le mouvement et la direction constituent un critère formel retenu pour la main. Ce codage se fait sur une matrice physiologique selon deux degrés de liberté (Flexion/Extension et Abduction/Adduction pour plus de détails voir Boutet 2008). En outre, afin de préciser la forme du pointage, nous avons eu recours à une analyse débouchant sur une définition formelle du pointage. Pour cela nous avons découplé l'alignement, le mouvement et la directionnalité dans les pointages (Boutet et al. 2011:16 et 17).

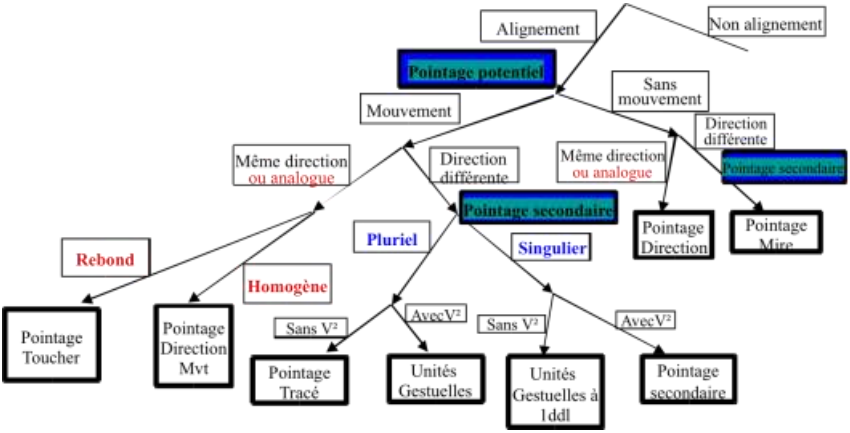

Figure 1-Arbre de décision des paramètres du pointage

Cette opération nous a permis, d'une part, d'inclure bon nombre de pointages gestuels qui n'ont pas une forme canonique (index étendu, les autres doigts étant repliés) tels que des conformations de doigts tous étendus et collés ou bien un pouce étendu les autres doigts repliés et, d'autre part, de prendre en compte, en tant qu'instance de pointage, des directions de mouvements qui ne sont pourtant pas parallèles à l'alignement des segments ('direction différente' dans le schéma de la figure 1).

Le pointage est alors défini comme l'alignement d'au moins trois segments adjacents distaux (première disjonction dans le schéma). La direction des mouvements associés aux pointages marquent la différence entre des pointages que l'on a considérés comme principaux — même direction que l'alignement — et d'autres secondaires qui interviennent en sus de pointages principaux ou d'Unités Gestuelles lorsque la direction ne correspond pas à l'alignement. Ces distinctions se retrouvent dans la typologie formelle mise en place pour le pointage, illustrées ci-dessous :

				
Figure-2 Pointage Direction 00:49.285	Figure-3 Pointage Mire 01:06.000	Figure-4 Pointage Tracé 00:31.559	Figure-5 Pointage Dir.Mouv. 00:21.508	Figure-6 Pointage Toucher 00:18.000

- Pointage Direction : alignement d'au moins 3 segments adjacents distaux sans mouvement (exemple Figure-2, Time code 00 49 285)
- Pointage Mire : alignements segmentaux sans mouvement dont la direction est donnée par le vecteur du regard, *marqué* par l'extrémité du pointage (exemple Figure-3, Time code 0106 000)
- Pointage Tracé : alignements segmentaux avec au moins un mouvement de direction différente de l'alignement, ici le traçage d'un cercle fait sur un plan horizontal (exemple Figure-4, Time code 00 31 559).
- Pointage Direction/Mouvement : alignements segmentaux, avec un mouvement de même direction que l'alignement (exemple Figure-5, Time code 00 21 508).
- PointageToucher : alignements segmentaux avec un mouvement en aller-retour de même direction que l'alignement et dont le point de rebroussement présente un rebond (exemple Figure-6, Time code 00 18 000).

Le bornage de ces catégories n'est pas toujours facile et le passage d'un pointage à l'autre s'exerce parfois sans que l'on puisse aisément les distinguer. Toutefois, avant même de considérer le bornage, les catégories de pointage présentées ici relèvent de distinctions formelles. On notera pourtant que la catégorisation formelle qui débouche sur une typologie de pointages a des répercussions fonctionnelles : référence ponctuelle vers un locus par le pointage toucher ; pointage mire associant le regard indiquant un lieu distant à atteindre marqué de manière homologue par la distance entre la main portant le pointage et les yeux ; délimitation d'une zone par le pointage Tracé ; valeur présentative/constative pour le pointage directionnel sans mouvement associant plusieurs doigts dans l'alignement. Enfin, on l'a vu les pointages secondaires sont

fortement représentés à proximité des anaphores verbales co-occurrentes.

Parallèlement à ce travail d'annotation des pointages, un codage prosodique a été initié. Il s'agit pour l'heure d'une première tentative qui s'appuie sur le découpage du signal brut et qui consiste à relever différents aspects d'un point de vue formel. Premièrement, on a noté la nature du contour intonatif (CI) des groupes de souffle (en l'occurrence des énoncés complets), à partir d'une typologie d'intonèmes limitée, inspirée par les travaux de Rossi (1999) et qui comprend des contours simples unidirectionnels tels que les contours Plat, Montant ou Descendant et des contours multidirectionnels enchainant 2 contours simples ou plus, comme Montant-Descendant, Descendant-Montant-Descendant, etc. Deuxièmement, on a noté la durée de ces groupes de souffle (en ms), ainsi que celle des syllabes pénultième (l'avant dernière syllabe du dernier mot quand celui-ci n'est pas monosyllabique) et finale (la syllabe finale étant marquée en français par un allongement et/ou un pic de la fréquence fondamentale de la voix qui porte la structure accentuelle). Troisièmement, on a noté la fréquence maximum au cours de chaque énoncé (en Hz). Quatrièmement, on a noté le nombre de syllabes par groupe de souffle.

Ce découpage est réalisé à l'aide du logiciel d'analyse et de transcription phonétique Praat (Boersma & Weeninck 2012). Les objectifs visés, sont d'une part, de décrire les phénomènes observés en termes de gestes vocaux -- les contours intonatifs marqués par des progressions plus ou moins amples dans des plages fréquentielles aigues ou graves, peuvent en effet être considérés comme des gestes – et d'autre part, de croiser *a posteriori* les deux types de segmentation : gestuel et vocal-verbal, en évitant la « contamination » évoquée plus haut.

3 On borne en vertu de quel(s) critère(s) ?

3.1 Doit-on définir une segmentation sur la base de contenus ou de bornages ?

La réponse est d'emblée une segmentation en fonction d'un contenu. Toutefois, certains critères de bornage glissent parfois vers un critère relevant du contenu. Ainsi pour la piste « Phase Geste », le mouvement est utilisé comme critère de bornage : tant que ça bouge on compte une même séquence (Kendon 2004 : 112 ; Mcneill 1992 : 376). Lorsque ça ne bouge plus ou de manière différente, on arrête et la séquence est bornée en *sortie* tandis qu'une nouvelle borne *d'entrée* lui est jointive. Ceci permet de différencier le passage entre « Stroke » et « Tenue », « Stroke » et « Rétraction » dans la piste « Phase Geste ». Mais ce bornage ne dit rien des raisons du passage entre « Préparation » et « Stroke ». Un autre critère permet de décider quel élément catégoriel va être utilisé après la borne, il s'agit de la présence/absence d'un geste ayant une signification. Ces deux critères, l'un fonctionnel portant sur le contenu du segment («*'expression' of the gesture* », Kendon 2004 : 112) et l'autre formel bornant par le mouvement général se complètent sans être de même nature. Que clôt réellement le bornage ? Un changement perceptible dans le mouvement. Certes, mais parfois le changement opère sur le bras avant qu'il n'affecte la main, qui elle est encore dans la partie signifiante du geste. La coarticulation est en jeu ici et elle donne lieu à des études plutôt par le versant informatique (Kipp, Neff & Albrecht 2007, Segouat 2009, Ojala, Salakoski & Aaltonen 2009). Les bornes sortantes et entrantes ne sont donc pas toujours jointives. Ce fait ne

relève pas d'un simple problème de temps à moyenner. Le membre supérieur a, d'une part, une densité qui lui permet d'exprimer encore une chose ici et déjà rien, voire autre chose, sur une autre partie, au même moment et, d'autre part, il possède plusieurs degrés de liberté par partie (doigts, main, avant-bras, bras) susceptibles d'être investis par diverses 'expressions' au même moment (pour l'expressivité du pointage voir Kendon & Versante 2003 : 134 *sq.* ; pour la LSF à un niveau formel, voir Boutet & Garcia 2007 : 106). Les pointages secondaires en sont une bonne illustration dans la mesure où des micromouvements sur les doigts sont sans impact sur la main ou l'avant-bras.

Même pour des critères de bornage identiques et *a priori* jointifs (la borne sortante d'une catégorie correspond à la borne entrante dans une autre), la densité et les possibilités de mouvement du membre supérieur rendent plus qu'improbable la pertinence d'une seule piste pour coder les phases des gestes (plus de 32 degrés de liberté par membre supérieur). Il en faudrait plusieurs pour rendre compte de l'entrecroisement des Unités Gestuelles et sortir de la limitation qu'impose de fait la linéarité vocale à la gestualité coverbale. Nous avons choisi ici d'annoter les mouvements *au niveau de* la main et de ne coder que certains degrés de liberté (flexion/extension et abduction/adduction dans la piste « Mvt Main »). La pronosupination n'a pas été codée, mais, par contre, des mouvements redevables de l'avant-bras ou du bras et ayant une répercussion directionnelle sur la main au titre de l'un des deux degrés de liberté signalés ont bien été pris en compte. Un mouvement proximal/distal par rapport à l'alignement ont également fait l'objet d'annotation. A l'assujettissement du canal de la gestualité au canal vocal/verbal pour des raisons de contenus, comme nous l'avons vu au-dessus, s'ajoute un écrasement des multiples possibilités simultanées gestuelles au profit de la prise en considération d'un seul phénomène gestuel à la fois.

Le critère de mise en mouvement non significative qui permet de caractériser la phase de « préparation » exclut pourtant à cet endroit tous les mouvements qui sont de fait jugés comme non pertinents dans l'unité gestuelle. Ainsi les mouvements que nous avons qualifiés de secondaires et qui pour les pointages sont perpendiculaires au vecteur principal que constitue l'alignement (index, alignement doigts-paume voire avant-bras) ne sont pas pris en compte dans la segmentation de McNeill. Ceci est dû à une segmentation mixte entre contenu et bornage. Ainsi, parmi ces mouvements perpendiculaires à l'alignement, seuls les Pointages Tracés qui dessinent les ronds-points successifs dans le corpus sont retenus comme gestes iconiques matinés de pointage, ils ne sont pas retenus en vertu de la présence de mouvements successifs perpendiculaires, mais à cause du contenu qu'ils délimitent. Ce type d'approche sélectionne des contenus gestuels mais ne s'interroge guère sur les conditions formelles d'émergence de ces contenus.

3.2 Granularité de la segmentation : repérage et segmentation

Nous avons vu que ne considérer qu'une seule piste pour rendre compte de phénomènes gestuels revenait à considérer que la main portait avant tout des caractéristiques fonctionnelles ou sémantiques holistiques (continuum 4 "*Global &Synthetic*" McNeill 2005:10–11). Si la solution réside *a priori* dans la multiplication des pistes, une question subsiste : peut-on se contenter de segmenter de façon parallèle (avec des pistes totalement indépendantes) lorsqu'on segmente très finement ? A l'instar de ce que

préconisent Kendon et McNeill, il semble qu'on ait besoin d'un repérage préalable, avec des unités segmentées pouvant servir de repères, autrement dit certaines pistes peuvent aussi servir de cadre de référence à d'autres. Il s'agit d'un jeu entre bornes et segment. Une première segmentation (par exemple sur la forme des pointages) permet de repérer des unités qui peuvent être segmentées elles-mêmes en unités de mouvements associés. C'est le cas ici de la piste « Forme Pointage » qui sert de repère à une autre piste « Mvt » qui permet d'affiner la segmentation. C'est aussi le cas entre les pistes « Forme Pointage » et « Type Pointage ». Ainsi pour la piste « Forme Pointage » entre 01 :10 : 200 et 01 : 12 : 200, on a un seul segment qualifié de 'phalangedospaumeavant-bras' décomposé en trois segments — 'direction', 'direction/mouvement' et 'direction' — sur la piste enfant « Type Pointage ». Cette décomposition n'est pas très productive dans le corpus, puisque sur 99 segments de la piste « Forme Pointage », seuls 5 d'entre eux sont décomposés dans la piste « Type pointage ». On peut dire ici que l'approche formelle est très productive et correspond environ à 95% à un découpage fonctionnel.

On trouve ce même type de segmentations en cascade pour d'autres pistes. Toujours à partir de la piste « Forme Pointage MD », et pour les mêmes 99 segments dans le corpus, on trouve 196 segments pour la piste « Mvt MD ». 40 % des segments de la piste parent « Forme Pointage » sont subdivisés entre 2 et 9 segments. En outre, la piste « Direction Mvt MD », qui reprend et précise de manière égocentrée[1] les directions des mouvements de la piste allocentrée[2] « Mvt MD », présente 240 segments. Elle subdivise 56% des segments de la piste « Forme Pointage » entre 2 et 14 segments.

Il eut été difficile de segmenter à ce point ce corpus sans s'appuyer sur une pré-segmentation. Celle-ci commence d'ailleurs dès la piste « Phase Geste ».

La granularité (ou le tempo) à laquelle on arrive pour la piste « Mvt Md » (Moyenne des segments 0,28 s) est équivalente à celle de la piste « *Word* » (Moyenne 0,22 s), comme cela a été dit plus haut. Toutefois, pour la modalité gestuelle, il a fallu 3 pistes avec des segmentations successives tandis que la segmentation vocale-verbale est automatique à partir d'une piste de transcription orthographique pour la modalité vocale (utilisation d'EasyAlign).

En continuité avec ce qui a été dit, dans la section concernant le bornage, à propos de la difficulté de situer précisément les limites de la pertinence d'un geste (articulation Préparation-Stroke-Enchaînement-Stroke-Rétraction), 25% des « Enchaînements » donnent lieu à un pointage. Dans la même idée, 18% des 11 segments « Préparation » sensés ne contenir aucun geste signifiant contiennent pourtant une forme de pointage. On voit ici, d'une part, la difficulté à considérer ce qui est significatif dans un geste, et, d'autre part, la nécessité de multiplier le nombre de pistes à même de croiser l'information.

[1] L'espace et en particulier ici les directions des mouvements sont appréciés en fonction d'un cadre de référence dans lequel le corps du locuteur constitue l'élément organisateur. Les directions *avant*, *arrière*, *haut*, *bas*, *gauche* et *droite* dépendent bien de la position du corps.

[2] Ce type de cadre de référence dépend ici des possibilités articulatoires de chaque segment. Le nombre de directions dépend donc du nombre et de la géométrie que les degrés de liberté imposent pour chaque segment.

4 Schéma d'annotation

En fonction des réflexions présentées ci-dessus, nous avons établi un schéma d'annotation qui respecte le principe, sinon la lettre, des problématiques référentielles de la segmentation ainsi que de celles des catégories visées par cette segmentation. A ce titre, le choix délibéré d'une segmentation formelle sous-tend l'ensemble des pistes. En vue du partage de la segmentation, les catégories classiques de segmentation de la gestualité coverbale proposés par Kendon ont été mises en œuvre en même temps qu'en question (de façon réduite ici à la piste « Phase geste »). Toutes les pistes qui ont pu être l'objet d'un vocabulaire contrôlé en ont été systématiquement dotées. Le nombre de grands articulateurs de la gestualité a été traité selon trois instances : *main droite*, *main gauche* et *deux mains*. Les pistes afférentes ont toutes été démultipliées en fonction. Enfin la part gestuelle et la part vocale ont été annotées en plus d'une transcription orthographique et d'une décomposition phonologique.

4.1 Détails des pistes

En dehors de la piste « Phase geste » dont on a parlé précédemment, trois pistes *filles* lui ont été associées : « Forme pointage », « Type pointage » et « Mvt ». Tout d'abord, la piste « Forme pointage » a donné lieu à une description générique des configurations manuelles redevables d'un pointage. Quatre items s'en dégagent : premièrement l'item *canonique*, pour lequel l'index est totalement étendu les autres doigts étant repliés dans le poing ; deuxièmement, l'item *phalangesalignées*, pour lequel on remarque au moins un alignement des trois phalanges digitales, quel que soit le doigt (selon la définition que nous avons proposée pour le pointage en section 2.2) ; troisièmement, l'item *phalangesdospaume*, pour lequel le dos de la paume s'ajoute à l'alignement distal; enfin l'item *phalangesdospaumeavant-bras*, qui prolonge l'alignement. La raison essentielle à ces distinctions réside dans les différences d'implications possibles du locuteur dans le pointage. La deuxième piste fille est « Type pointage » pour laquelle cinq items ont été présentés et développés dans la dernière partie de la section 2.2. Pour la troisième et dernière piste fille appelé « Mvt », il s'agit de donner les directions des mouvements associées au geste candidat pour être des pointages. Ces directions sont allocentrées sur la main et seront d'ailleurs exprimées par les degrés de liberté de la main (à l'exception de la prono-supination) et par deux directions d'un vecteur reprenant l'alignement du pointage [proximal *vs* distal]. Nous avons distingué pour chacun de ces items deux instances possibles, soit le *mouvement simple* et le *mouvement en aller-retour*.

Cette dernière piste « Mvt » reprend les mouvements, y compris les plus furtifs, au niveau de la main. Comme nous l'avons dit plus haut, ces mouvements sont notés selon un cadre de référence allocentré, c'est-à-dire, d'une part, qu'on ne situe pas la position de la main selon un repérage égocentré (avant, arrière, gauche, droite, haut ou bas) et, d'autre part, qu'on resitue les mouvements en fonction des possibilités de mouvement de la main et au niveau de celle-ci. L'hypothèse ici est que des mouvements peuvent prendre leur sens sur le substrat qui les génère. Ceci répond à un principe évoqué plus haut de non assujettissement d'une modalité par une autre et de proximité référentielle de la segmentation. On segmente des mouvements de la main au plus près des possibilités de la main et non par rapport à un espace dans lequel elle se meut. La situation des

mouvements dans un repère cartésien, tout comme l'équation qui préside à une courbe dans un plan ou un tracé dans l'espace, constitue un dessin d'une fonction ou d'une signification qu'il reste à mettre en équation. Si le dessin présente un intérêt évident, si un repérage permet de l'orienter et de le situer par rapport à un autre, on ne peut tout de même pas confier la segmentation au seul repérage dans un espace qui constitue avant tout une étendue en l'absence de laquelle il n'y aurait simplement pas de tracé possible. L'espace égocentré est un support dont les axes hiérarchisent certaines informations. Il est important d'en relever l'organisation et de l'extraire. Pour autant, on doit également segmenter en fonction d'autres cadres de référence plus proches des segments qui bougent. Ceci n'empêche pas que l'égocentration constitue un véritable lieu d'organisation, notamment référentiel.

A ce titre, deux pistes filles de la piste « Mvt » situent l'alignement et le mouvement dans un cadre égocentré : la piste « Direction alignement » et la piste « Direction mouvement ». Ces deux pistes partagent le même vocabulaire contrôlé d'orientation composé de 27 items qui reprennent des directions simples (les six faces d'un cube), et des directions composées précisant pour chaque face vers quelle diagonale l'alignement pointe ou la paume s'oriente. On a également adjoint un item '*sans*' (direction particulière). Ainsi, en plus d'une segmentation du mouvement au niveau de la main selon un cadre centré sur la main (allocentré), on a noté les changements d'orientation de la paume. Ces deux informations ne sont pas homologues, puisqu'on peut avoir une orientation vers l'avant de la paume avec un mouvement d'adduction sans que l'orientation vers l'avant ne s'en trouve modifiée. Ces deux types d'informations complémentaires permettent de vérifier notamment si les pointages secondaires relèvent d'une organisation égocentrée ou non.

Au niveau vocal, le codage initié représente une première ébauche de travail. Celui-ci consiste en un travail de segmentation du flux sonore qui s'appuie sur des unités intonatives potentiellement applicables à des énoncés qui n'ont pas le même degré de complexité (dont la structure syntaxique peut varier) et d'achèvement (qui peuvent être des phrases interrompues). Un premier objectif est de mettre les temps forts du codage vocal en regard avec les configurations et mouvements manuels privilégiés dans ce travail (flexion/extension et abduction/adduction) et de découvrir quels sont les liens fonctionnels entre gestes et paroles. Un second est de raffiner ces points d'articulations et la qualité des indices de codage. Il ne s'agit pas encore d'objectiver la difficulté à caractériser les principes de bornage des unités prosodiques. Nous nous focalisons d'abord sur la nature du contour intonatif des énoncés (forme globale, pic de fréquence fondamentale, et pattern final correspondant au dernier segment du contour i.e. *similaire au contour global* dans le cas des contours simples et *descendant* ou *autre* dans le cas des contours complexes ou multidirectionnels). Le contour intonatif est considéré comme un objet prosodique saillant (et par conséquent la première piste explorée) nécessairement rattaché à un versant fonctionnel, qui peut être appréhendé selon une première alternative. Soit il apparaît que les unités intonatives désignées plus haut et les *Formes*, *Types de pointages* et *Mouvements* convergent, au sein d'un processus de segmentation du flux verbal, en unités discursives à l'origine de l'organisation de l'interaction et de la mise en œuvre de la structure informative. Soit il apparaît qu'une telle convergence est

d'abord la trace d'une synchronisation rythmique sur le plan moteur en réponse à un contexte informationnel. Au total 37 contours on été répertoriés dont 17 contours simples et 20 contours multidirectionnels.

4.2 Tableau synthétique des critères de segmentation ventilés par piste

Critères / Pistes	Type seg.	Bornage	Alignement	Déport	Complétude	Dépendance
Phase Geste	Mixte : formel/ fonctionnel	Exclusif	Sur data manuel	Temps (*stroke*)	Complet	Aucune sauf *stroke* vers le verbal
Forme pointage	Formel	Exclusif	Sur data manuel	Aucun	Partiel (pointages)	Aucune
Type pointage	Formel	Exclusif	Sur data manuel	Aucun	Partiel (pointages)	Aucune
Mvt	Multiple : formel (allocentré)	Non Exclusif (en fait exclusif)	Sur data manuel	Lieu (sur la main)	Partiel (pointages)	Aucune
Direction alignement	Multiple : formel (allocentré, Mvt)	Exclusif	Sur annotations, Mvt, manuel	Lieu (sur la main)	Partiel (pointages)	Avec Mvt bornage externe segments
Direction mouvement	Multiple : formel (allocentré, Mvt)	Exclusif	Sur annotations Mvt, manuel	Lieu (sur la main)	Partiel (pointages)	Avec Mvt bornage externe segments
Ortho	Formel	Non exclusif en interne	Non aligné	Aucun	Complet	Aucune
Words	Formel	Exclusif	Sur data et annotations (Ortho)	Aucun	Complet	Avec Ortho

			automatique			
Phono	Formel	Exclusif	Sur data et annotations (Words) automatique	Aucun	Complet	Avec Words
Phones	Formel	Exclusif	Sur data et annotations (Phono) automatique	Aucun	Complet	Avec Phono
CI	Formel	Exclusif…	Sur data	-	-	-
Pic F0	avec CI	-	Sur data	-	-	-

5 Conclusion

En conclusion, il apparaît que l'étude de la multimodalité nous confronte à un nombre de principes/questions méthodologiques important, même lorsque l'on revient à un niveau de segmentation épuré. *Multimodalité* signifie regroupement de canaux, de formes, de fonctions d'expression qui entrent en interaction, en synergie pour aboutir à la fabrication du sens, mais qui s'amalgament de telle manière, que l'on n'a pas encore réussi à mettre au point un véritable outil de visualisation de ce fonctionnement ou système.

Dans cette étude, en partant des questions essentielles telles que celle de l'origine de la segmentation, de la délimitation des critères de segmentation et de la granularité de la segmentation, nous proposons une première 'batterie' ou grille d'indices propres à la modalité gestuelle dont les prolongements sont orientés vers la compréhension du déroulement de l'interaction (avec les planification et réalisation éventuelles de configuration de gestes ?) et l'articulation des aspects vocaux.

ALLWOOD, J., CERRATO, L., DYBKJAER, L., JOKINEN, K., NAVARRETTA, C. et PAGGIO, P., éditeurs (2004). The MUMIN multimodal codingscheme. Proc. Workshop on Multimodal Corpora and Annotation.

BOERSMA, P. et WEENINK, D. (2012). Praat: doing phonetics by computer (Version 5.1). www.praat.org, 2012.

BOUTET, D. (2008). Une morphologie de la gestualité : structuration articulaire. *Cahiers de linguistique analogique, n°*5. (Abell), pages 80–115.

BOUTET, D.,BLONDEL, M.,CAËT, S., BEAUPOIL, P. et MORGENSTERN. A. (2011). Tu pointes ou tu tires ?! Annotation sous ELAN des pointages d'un 'entendant vocalo-gestualisant'. Actes du premier Défi Geste Langue des Signes, 15–27. Montpellier: TALN.

BOUTET, D.et GARCIA, B. (2006). Finalités et enjeux linguistiques d'une formalisation graphique de la langue des signes française (LSF). *Glottopol*(7), pages 32–52.

BOUTET, D.et GARCIA, B.. (2007). Compositionnalité morphophonétique de la langue des signes française (LSF) et exploration des relations structurales entre paramètres. *TAL* 48-3. Modélisation et traitement des langues des signes, pages 93–114.

BRESSEM, J. (1998). *Notatinggestures–Proposal for a formbased notation system of coverbalgestures*. Manuskript.

CALBRIS, G.(1990). *The Semiotics of French Gestures*. Advances in semiotics. Bloomington: Indiana UniversityPress.

COLLETTA, J-M., KUNENE, R.,VENOUIL, A.,KAUFMANN, V.etSIMON, J-P. (2009). Multimodal Corpora, Multi-track Annotation of Child Language andGestures. vol. 5509, pages54–72. (Lecture Notes in Computer Science). Springer Berlin / Heidelberg. http://www.springerlink.com.gate3.inist.fr/content/h02381708g7804k4/abstract/ (26 mars, 2012).

CORBALLIS, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton: Princeton UniversityPress.

FERRÉ, G., BERTRAND,R.,BLACHE, P.,ESPESSER, R.,et RAUZY, S.(2007). Intensive Gestures in French and their Multimodal Correlates. *Interspeech, Antwerp*, pages 690–693. Antwerp, Belgium. http://hal.archives-ouvertes.fr/index.php?halsid=1iq62kdrg2ngdrnfcgs27vdli2&view_this_doc=hal-00173729&version=1 (21 janvier, 2012).

GUIDETTI, M. (2002). The emergence of pragmatics: forms and functions of conventionalgestures in young French children. *First Language* 22(3), pages 265 –285. doi:10.1177/014272370202206603 (18 janvier, 2012).

HANKE, T. (2004). HamNoSys—Representingsignlanguage data in languageresources and languageprocessingcontexts. *Workshop on the Representation and Processing of SignLanguages on the occasion of the Fourth International Conference on LanguageResources and Evaluation*, 1–6. Lisbon: ELDA.

JOHNSTON, T. (2008). Corpus linguistics and signedlanguages: no lemmata, no corpus. *3rd Workshop on the Representation and Processing of SignLanguages*, pages 82–88. OnnoCrasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd, Inge Zwitserlood.

KENDON, A. (1972). Somerelationshipsbetween body motion and speech. *Studies in dyadic communication*, pages177–210. PergamonPress. New York: Siegman, A., Pope, B.

KENDON, A. (1980). Gesticulation and Speech: Two Aspects of the Process of Utterance. *The Relation Between Verbal and Nonverbal Communication*, pages 207–227. Mouton. The Hague: Key, M. R.

KENDON, A. (1988). How Gestures Can BecomelikeWords. *Cross-Cultural Perspective in Nonverbal Communication*, pages 131–141. C.J. Hogrefe. Toronto□; Lewiston, N.Y.: Fernando Poyatos.

KENDON, A.(1991). SomeConsiderations for a Theory of LanguageOrigins. *Man* 26(2). (New Series), pages 199–221. (5 mars, 2010).

KENDON, A.(2004). *Gesture: Visible Action asUtterance*. Cambridge: Cambridge UniversityPress.

KENDON, A.etVERSANTE, L. (2003). Pointing by hand in « Neapolitan ». *PointingWhereLanguage, Culture, and Cognition Meet*, pages 109–138. Lawrence Erlbaum Associates Publishers. Mahwah, London: Kita, Sotaro.

KIPP, M., NEFF, M. et ALBRECHT, I.(2007). An annotation scheme for conversationalgestures: how to economically capture timing and form. *LanguageResources and Evaluation* 41(3), pages325–339. doi:10.1007/s10579-007-9053-5 (26 mars, 2012).

KITA, S., GIJN, I.van,etHULST, H.G. van der. (1998). Movement Phases en Signs and Co-speech Gestures, and Their Transcription by HumanCoders. vol. 1371, pages23–35. (Lecture Notes in Computer Science). Springer Berlin / Heidelberg. http://www.springerlink.com/content/4w673515335h6703/abstract/ (26 mars, 2012).

MCNEILL, D.(1992). *Hand and mind □: whatgesturesreveal about thought*. Chicago□; London: University of Chicago press.

MCNEILL, D.(2005). *Gesture and thought*. Chicago (Ill.)□; London: University of Chicago Press.

MCNEILL, D.etDUNCAN, S. (2000). Growth Points in Thinking-for-Speaking. *Language and gesture*, 141–161. Cambridge UniversityPress. (Language, Culture & Cognition). Cambridge: David McNeill.

MONDADA, L. (2009). Emergent focused interactions in public places: A systematicanalysis of the multimodal achievement of a commoninteractionalspace. *Journal of Pragmatics* 41(10), pages1977–1997. doi:10.1016/j.pragma.2008.09.019 (29 mars, 2012).

OJALA, S., SALAKOSKI, T.etAALTONEN, O. (2009). Coarticulation in sign and speech. *NEALT PROCEEDINGS SERIES*, vol. 6, pages21–24. Odense Denmark: Costanza Navarretta Patrizia Paggio Jens Allwood Elisabeth AlsénYasuhiroKatagiri.

POVINELLI, D., BERINGET, J.M.,GIAMBRONE. (2003). Cimpanzees' « Pointing »:AnotherError of the Argument by analogy? *PointingWhereLanguage, Culture, and Cognition Meet*, pages 35–68. Lawrence ErlbaumAssiociates. Mahwah, London: Sotaro Kita.

QUEK, F., MCNEILL, D.,BRYLL, R.,DUNCAN,S., MA, X-F., KIRBAS, C., MCCULLOUGHK.E. et ANSARI, R. (2002). Multimodal humandiscourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.* 9(3), pages171–193. doi:10.1145/568513.568514 (26 mars, 2012).

RIZZOLATTI, G. etARBIB. M.A.(1998). Languagewithinourgrasp. *Trends in Neurosciences* 21(5), pages188–194. doi:10.1016/S0166-2236(98)01260-0.

ROSSI, M. (1999). *L'intonation_: le système du français*. Paris_: Ophrys.

SEGOUAT, J. (2009). A Study of SignLanguage Coarticulation. *SIGACCESS Newsletter*Accessibility and Computing(Issue 93), pages 31–38.

WILKINS, D. (2003). WhyPointingWith the Index Finger Is Not a Universal. *Pointing Where Language, Culture, and Cognition Meet*, pages 171–216. Lawrence Erlbaum Associates. Mahwah, London: Kita, Sotaro.