

## Types de gestes et utilisation de l'espace gestuel dans une description spatiale : méthodologie de l'annotation

Marion Tellier, Mathilde Guardiola, Brigitte Bigi

Laboratoire Parole et Langage, UMR 6057, CNRS & Université de Provence,  
5 avenue Pasteur - 13604 Aix en Provence  
marion.tellier@lpl-aix.fr , mathilde.guardiola@lpl-aix.fr , brigitte.bigi@lpl-aix.fr

**Résumé** Cet article présente la méthodologie utilisée par une équipe de 3 chercheuses du Laboratoire Parole et Langage pour annoter des corpus multimodaux. Nous décrirons à la fois le découpage automatique du signal audio en unités inter-pausales (IPU) pour segmenter la parole, les conventions de la Transcription Orthographique Enrichie (TOE) pour le verbal/sonore et enfin la typologie utilisée pour l'annotation du geste coverbal et l'espace gestuel. Ce travail nous permet de faire le bilan des techniques actuellement utilisées et développées dans notre laboratoire et au sein de l'ANR OTIM sur l'annotation multimodale de corpus. Il met également au jour un certain nombre de problèmes posés par l'annotation gestuelle.

**Abstract** This paper presents the methodology used by a team of 3 researchers of the Laboratory Parole et Langage to code multimodal data. We will first describe how we segmented the audio signal in Inter-Pausal Units (IPU), and then the conventions used to transcribe oral/verbal aspects with the Enriched Orthographic Transcription. We will also present the typologies used to code co-speech gestures and gesture space. This study enables us to introduce and discuss the techniques developed in our lab and in the OTIM project to code multimodal data. It also enables us to shed light on several current issues for gesture coding.

**Mots-clés :** multimodalité, geste, annotation, Transcription Orthographique Enrichie

**Keywords:** multimodality, gesture, coding, Enriched Orthographic Transcription

## 1 Objets de l'annotation

Nous avons transcrit et annoté le corpus en français. Préalablement à toute annotation, le signal de parole a été découpé automatiquement en unités inter-pausales (Interpausal Unit - désormais IPU -). Il s'agit de blocs de paroles bornés par des pauses silencieuses d'au moins 200 ms (durée variable selon les langues). Ensuite, la partie verbale a été transcrite avec les conventions de la Transcription Orthographique Enrichie (TOE) élaborées pour l'annotation du corpus CID (Corpus of Interactional Data) développées au LPL et approfondies dans le cadre du projet ANR OTIM (BLAN08-2\_349062 ; <http://www.lpl-aix.fr/~otim/>). L'intérêt de cette transcription est qu'elle permet de rendre compte de manière orthographique de phénomènes typiques de l'oral tels que les pauses remplies (euh, hum...), les amorces, les mots tronqués, les répétitions, les faux-départs, les rires, les prononciations atypiques... cette transcription rend possible l'alignement temporel des phonèmes nécessaire aux analyses phonétiques et permet également une automatiser des annotations morpho-syntaxiques et syntaxiques (Bertrand & al. 2008). Ceci a été fait pour les deux locuteurs.

Au niveau du geste, notre attention s'est portée sur tous les gestes coverbaux produits uniquement avec les mains. Nous avons annoté les types de gestes ainsi que l'espace gestuel dans lequel chaque geste est produit. Pour ce corpus spécifique, il nous a semblé pertinent d'attribuer des valeurs d'indications spatiales pour tous les gestes (déictiques ou non) qui servaient à donner des indications spatiales. L'annotation des gestes a été faite pour le locuteur masculin uniquement (locuteur B).

## 2 Description de la méthodologie de l'annotation

### 2.1 Aspects techniques

La première étape a donc été la segmentation en IPU. Chaque locuteur étant enregistré sur un canal audio distinct, ce découpage est bien adapté. En raison de sa nature formelle, objective et repérable automatiquement, l'IPU est souvent utilisée sur des corpus de taille importante. Ce découpage vise à faciliter la transcription mais aussi les étapes ultérieures de conversion graphème-phonème et d'alignement temporel avec le signal audio.

La segmentation en IPU a été faite automatiquement, avec un logiciel que nous avons développé et que nous diffusons sous licence GPL (<http://www.lpl-aix.fr/~bigi/logiciels.html>). Ce programme prend en entrée un fichier son au format .wav. Il recherche les zones de silences dans une fenêtre de largeur fixe. La valeur de rms (root-mean-square) des trames de la fenêtre est comparée à une valeur seuil fixée par l'utilisateur. Ce programme fournit un répertoire contenant autant de fichiers sons que de zones de parole trouvées, et un fichier TextGrid indiquant les intervalles parole/silence. Ces derniers sont indexés : gpd\_X pour les zones de parole, gpf\_X pour les zones de silences (où X est la valeur d'index). Le temps passé pour cette annotation est de 15 minutes car la segmentation obtenue a été vérifiée manuellement.

Le format de fichier TextGrid a été choisi en raison de sa simplicité et de la portabilité qui en découle. C'est le format natif du logiciel Praat (Boersma, 2011), mais il peut être lu par d'autres logiciels d'annotation (notamment Elan), et par tous les éditeurs de texte. C'est un fichier ASCII qui contient une entête, suivie d'une liste d'intervalles avec : le temps de début, le temps de fin et le label de chaque intervalle. La Figure 1 présente le début du fichier obtenu après la segmentation en IPU du locuteur B.

Le fichier .wav du locuteur A a été segmenté en 15 IPU ; le fichier son du locuteur B en 14 IPU.

Les fichiers TextGrid des annotations de la parole ont ensuite été importés dans le logiciel Elan, logiciel gratuit développé au Max Planck Institute de Nijmegen. L'ensemble des annotations gestuelles a été réalisé avec ce logiciel par une troisième annotatrice. Ce travail a pris environ 5 heures.

7 pistes ont été définies : TOE locuteur A, TOE locuteur B, Tokens (= Mots du locuteurs B), Type de geste principal locuteur B, Type de geste secondaire locuteur B, Valeur indication spatiale et Espace gestuel. L'utilisation de deux pistes pour annoter les gestes permet de rendre compte des superpositions de deux types de gestes. Par exemple, lorsque un battement est superposé sur un métaphorique, la piste *geste principal* étiquette le métaphorique et la piste *geste secondaire* montre la durée pendant laquelle

apparaissent les battements superposés. Cette deuxième piste sert aussi à indiquer lorsqu'un geste a deux dimensions (une dimension iconique et une dimension déictique, par exemple). Pour faciliter la lecture, des couleurs pastel ont été attribuées aux pistes et des couleurs vives pour les types de gestes (voir Figure 1).

Les pistes « Type de geste principal » et « Type de geste secondaire » sont des annotations qui portent sur le sens du geste. La piste « valeur indication spatiale » donne une indication fonctionnelle et la piste « espace gestuel » apporte des informations relatives aux coordonnées spatiales du geste.

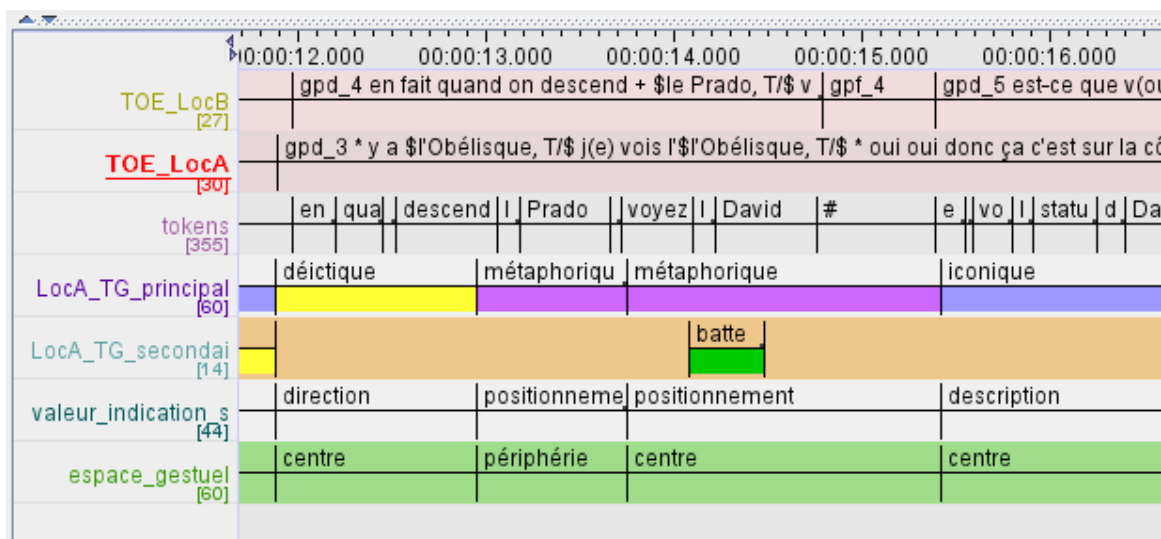


Figure 1 : Pistes définies pour l'annotation du geste

## 2.2 Aspects théoriques

### Transcription des aspects verbaux et sonores

Pour la transcription de la parole, nous avons opté pour la Transcription Orthographique Enrichie (TOE) qui a été développée pour transcrire le corpus CID : Corpus of International Data (Bertrand & al, 2008). Elle permet de rendre compte de manière orthographique de la plupart des phénomènes typiques de l'oral ce qui facilite ensuite la détection de phonèmes par les aligneurs automatiques.

Les conventions de la TOE sont basées sur celles développées à l'origine par le GARS : Groupe Aixois de Recherche en Syntaxe (Blanche-Benveniste & Jeanjean 1987). Il n'y a ni ponctuation, ni correction de la syntaxe (les phénomènes du français parlé sont reproduits fidèlement). Le groupe OTIM (Blache & al, 2008 ; Blache & al, 2010) y a ajouté des précisions sur des spécificités de prononciation avec si besoin des transcriptions codées en pseudo-orthographe. Les caractéristiques principales de la TOE utilisées pour la transcription de ce corpus sont résumées ci-après (Tableau 1).

```
File type = "short ooTextFile"
Object class = "TextGrid"

0
87.92

<exists>
1

"IntervalPiste"
"TOE"
0
87.92
30
0
4.6012
"dummy"
4.6012
```

```

5.5458
"gpd_1 "
5.5458
6.8052
"gp_f_1 "
6.8052
8.0647
"gpd_2 "
8.0647
11.8957
"gp_f_2 "

```

Figure 2 : Extrait du textgrid obtenu après segmentation en IPU

Phénomène	Convention	Exemples
Elision	<i>On met entre parenthèses ce qui aurait dû être prononcé.</i>	<i>i(l)s sont, p(u)is, t(u) as □</i>
Prononciation particulière	<i>On utilise la forme [orthographe, prononciation].</i>	<i>[je suis, chui], [stats, stateu]</i>
Amorce	<i>C'est une disfluece, phénomène propre à l'oral, notées par un tiret '-' à la fin. □</i>	<i>co- co- comment tu veux que je fasse, i- ils</i>
Noms propres	<i>entre dollars suivi du type de nom propre □</i>	<i>\$la statue d(u) David, T/\$</i>
Séquences incompréhensibles	<i>par une astérisque</i>	<i>et le * il est parti</i>

Tableau 1 : Exemples de la convention pour la TOE

Il va de soi que certains de ces phénomènes peuvent être imbriqués. Ci-après, deux exemples d'IPU transcrites en TOE du locuteur B :

```
gpd_1 ben [di(s)ons, dzon] si on si on est à le à $l'Escale Borély, T/$
```

```
gpd_8 et là à [c(e), ss] rond-point sur la droite c'est plein d(e) restaurants euh ça s'appelle $l'Escale Borély T/$
```

La transcription de la parole a été faite manuellement selon les conventions de la TOE par une seconde annotatrice avec le logiciel Praat. Cette annotation a pris 3 heures environ. La transcription orthographique enrichie permet de générer, par une conversion graphème-phonème, une suite de phonèmes la plus proche possible de ce qui a réellement été prononcé par le locuteur. Ce sont les outils de phonétisation et d'alignement automatique qui déterminent l'annotation phonétique. La première étape consiste en une conversion graphème-phonème. Le convertisseur graphème-phonème (phonétiseur) permet de retranscrire des phrases de leur forme orthographique sous leur forme phonétique. Cette étape de conversion est destinée à produire la séquence de phonèmes nécessaire à l'aligneur. À partir de la TOE, la transcription orthographique truquée seule est fournie au convertisseur qui s'appuie sur un dictionnaire de formes fléchies et un ensemble de règles modifiables (exception, liaison, etc.). En sortie, le convertisseur fournit une suite de tokens phonétisés en SAMPA. Ensuite, l'aligneur employé, développé au LORIA par D. Fohr et Y. Laprie, est basé sur des modèles de Markov cachés (Hidden Markov Models). L'alignement est effectué IPU par IPU. À partir de la séquence de phonèmes et du signal audio, l'aligneur fournit en sortie la localisation temporelle de chaque phonème sur le signal. Il fournit également l'alignement temporel des mots. Ce dernier est particulièrement intéressant pour l'annotation du geste car il permet de voir aisément la synchronie entre geste et mot et il facilite le travail d'annotation gestuelle.

Étant donné le découpage en IPU, les transcritteurs n'ont pas dû noter les pauses silencieuses, exceptées celles internes aux IPU (donc inférieures à 200 ms mais perceptibles). Les chevauchements de parole, qui

peuvent être repérés automatiquement grâce aux IPU, n'ont pas été transcrits non plus. La transcription spécifie par ailleurs certaines réalisations phonétiques particulières. A priori ces réalisations phonétiques en parole conversationnelle apparaissent plus étendues, en forme et en type, que celles habituellement recensées sur des corpus de parole lue ou contrôlée. Pour ces derniers, une grande partie des phénomènes phonétiques particuliers (type élision ou réduction par exemple) est automatisable par implémentation de règles puis constitution d'un lexique de variantes systématiques testées lors de la phase d'alignement (« je suis » réalisé /SHi/ ; « je sais » réalisé /Se/).

En revanche, à côté de ces réalisations phonétiques particulières habituelles, on peut constater des élisions non standards, des substitutions et/ou ajouts de phonèmes, etc. (« expérience » réalisé /perja~s/ ; « demande » réalisé /ma~/), qui peuvent rendre la constitution d'un lexique inefficace en raison d'un trop grand nombre d'entrées et de variantes possibles par entrée. De plus, les schwas réalisés (en particulier en finale) ou les sons épenthétiques sont aussi spécifiés (« rappelle quand » réalisé /rape12ka~t2/).

Ces différentes contraintes ont conduit à opérer une désambiguïsation manuelle préalable au stade de la transcription, pour tenter de pallier les limites des outils actuellement disponibles et davantage adaptés à de la parole lue ou contrôlée sur du français standard. C'est en ce sens que nous parlons de transcription orthographique enrichie, à partir de laquelle deux transcriptions sont automatiquement dérivées :

- une transcription orthographique standard, dont sont dérivés les « tokens orthographiques », est destinée aux modules d'analyse textuelle.
- une transcription orthographique « truquée », dont sont dérivés les « tokens phonétiques », est destinée au convertisseur graphème-phonème.

L'usage de la TOE n'a pas d'influence directe sur l'annotation gestuelle et son analyse. En revanche, elle permet de générer automatiquement des annotations syntaxiques et phonétiques permettant une analyse multi-niveaux du discours. Cette analyse n'a pas été faite pour le présent corpus.

### Transcription des aspects gestuels

Les types de gestes ont été annotés selon la typologie de McNeill (1992, 2005) c'est-à-dire *déictique* (Figure 3), *iconique* (Figure 4), *métaphorique* et *battements*. Elle a été enrichie par 4 autres types de gestes : les *emblèmes*, les *Butterworth*, les *interactifs* et les *gestes avortés* (voir Tableau 2). Cette typologie a déjà été utilisée par Tellier et Stam (2010). Un vocabulaire contrôlé a été utilisé pour l'annotation des types de gestes principaux et secondaires, il est constitué des 8 types de gestes mentionnés dans le Tableau 2.

Déictique	<i>Geste de pointage</i>
Iconique	<i>Geste illustratif d'un concept concret</i>
Métaphorique	<i>Geste illustratif d'un concept abstrait</i>
Battement	<i>Geste rythmant la parole, sans contenu sémantique</i>
Emblème	<i>Geste culturel, conventionnel</i>
Butterworth	<i>Geste de recherche lexicale</i>
Interactif	<i>Geste adressé à l'interlocuteur pour la gestion de l'interaction</i>
Avorté	<i>Geste esquissé mais avorté</i>

Tableau 2 : Typologie des gestes utilisée pour l'annotation



Figure 1 : [vous allez vers les Goudes]



Figure 2 : [y a un rond point]

Nous avons également annoté l'utilisation de l'espace gestuel en utilisant les 4 zones principales définies par McNeill, 1992 (voir Figure 5) à savoir *centre centre*, *centre*, *périphérie* et *extrême périphérie*. L'annotation de l'espace gestuel rend possible l'évaluation de la taille des gestes (amples ou non) et permet de déterminer si un locuteur utilise particulièrement une zone plutôt qu'une autre. Afin de rendre compte fidèlement de la taille des gestes, nous avons choisi d'étiqueter l'espace gestuel d'un geste en prenant le point le plus ample du mouvement (quelque soit la phase). Ainsi si un geste démarre dans le *centre*, se déplace dans l'*extrême périphérie* et se termine dans le *centre centre*, l'espace gestuel codé sera *extrême périphérie* car c'est le point le plus ample. Cela correspond la plupart du temps au « stroke » du geste mais ce n'est pas systématique. Autre choix méthodologique, c'est la main qui sert de point de référence pour l'espace gestuel sauf dans les cas où les doigts sont signifiants et qu'ils sont dans une autre zone que le reste de la main. La Figure 6 exemplifie ce type de situation : la majeure partie de la main est dans la zone *centre centre* alors que le pouce est dans le *centre*, or dans ce cas le pouce est l'élément signifiant du geste, c'est pourquoi l'espace gestuel est codé *centre*. L'annotation de l'espace gestuel pose aujourd'hui beaucoup de problèmes car la grille de McNeill se révèle insuffisante principalement parce qu'elle est en 2 dimensions alors que le geste est produit dans un espace en 3 dimensions. Ainsi, un geste produit dans l'espace *centre centre* avec la main proche du ventre est codé comme un geste produit dans le même espace mais avec le bras tendu alors que ces deux gestes n'ont pas la même amplitude. Voilà pourquoi nous avons rajouté une catégorie *bras tendu devant* dans le vocabulaire contrôlé pour coder l'espace gestuel.

TYPES DE GESTES ET UTILISATION DE L'ESPACE GESTUEL DANS UNE DESCRIPTION SPATIALE :  
MÉTHODOLOGIE DE L'ANNOTATION

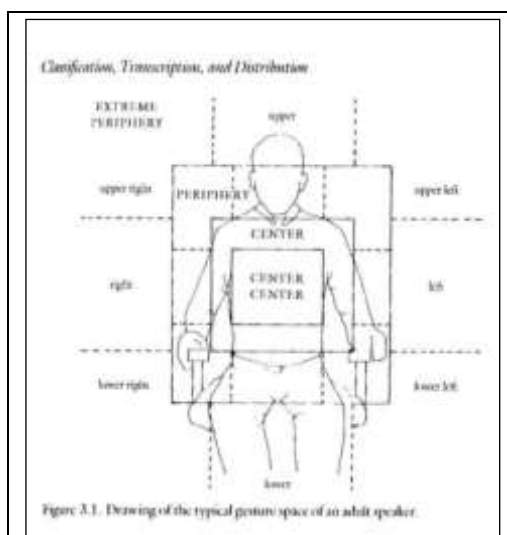


Figure 5 : L'espace gestuel (McNeill, 1992 : 89)

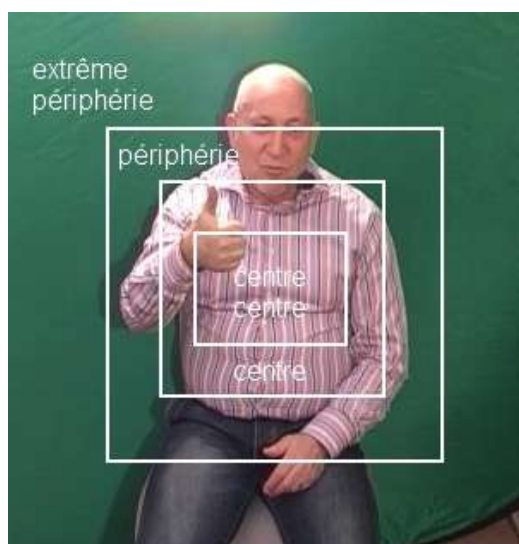


Figure 6 : [vous avez la statue du pouce là]

Pour les besoins de ce corpus, il nous a semblé intéressant d'attribuer une valeur d'indication spatiale aux gestes qui étaient utilisés pour donner des indications géographiques. Nous avons créé un vocabulaire contrôlé avec les quatre valeurs suivantes : *anaphorique* (lorsqu'un geste fait référence à un geste précédemment produit), *description* (lorsque le geste sert à décrire un point de repère, dans sa forme par exemple), *direction* (lorsque le geste donne un sens ou une direction) et *positionnement* (lorsque le geste place les éléments les uns par rapport aux autres). Cette typologie ne préexistait pas à l'annotation mais a été élaborée pendant le travail sur ce corpus.

Le Tableau 3 synthétise les vocabulaires contrôlés utilisés dans le schéma d'encodage pour l'annotation du geste coverbal.

Type de geste	<i>avorté</i> <i>battement</i> <i>Butterworth</i> <i>déictique</i> <i>emblème</i> <i>iconique</i> <i>métaphorique</i>
Espace gestuel	<i>centre centre</i> <i>centre</i> <i>périphérie</i> <i>extrême périphérie</i> <i>bras tendu devant</i>
Valeur indication spatiale	<i>anaphorique</i> <i>déscription</i> <i>direction</i> <i>positionnement</i>

Tableau 3 : Synthèse du schéma d'encodage pour le geste

### 3 Résultats

#### 3.1 Au niveau du verbal sonore

La TOE du corpus pour les locuteurs A et B durant la minute transcrite rend compte d'un certain nombre de phénomènes. Ils sont résumés dans le tableau 4.

Phénomène	Locuteur A	Locuteur B
Elision	5	12
Prononciation particulière	0	2
Amorces	0	1
Pauses courtes	2	5
Noms propres (toponymes)	3	20
Inaudible	2	3

Tableau 4 : TOE du corpus pour les locuteurs A et B

On remarque que le locuteur B utilise un grand nombre de noms propres, tous des toponymes, ce qui est fortement lié à la tâche de maptask qui élicite ce genre d'items lexicaux. Il y a également 12 élisions, phénomène propre à l'oral. La qualité de l'enregistrement permet d'avoir très peu de séquences inaudibles ou inintelligibles (notées par « \* » dans la transcription, afin de signaler à l'aligneur de ne pas aligner ces séquences sur le signal).

#### 3.2 Au niveau du gestuel

Soixante gestes principaux ont été annotés et 14 gestes secondaires. Parmi ces derniers, on trouve 10 séries de battements superposés ce qui donne un total de 70 gestes. On remarque donc que la très grande majorité des battements sont superposés à d'autres gestes, principalement à des déictiques, iconiques et



TYPES DE GESTES ET UTILISATION DE L'ESPACE GESTUEL DANS UNE DESCRIPTION SPATIALE :  
MÉTHODOLOGIE DE L'ANNOTATION

métaphoriques. Vingt-huit gestes du locuteur masculin sont des déictiques, ce qui est bien sûr lié à la maptask. On remarque également dans le Tableau 5 une quantité importante d'iconiques utilisés principalement pour décrire des lieux.

Types de geste	Nombre gestes principaux	Nombre gestes secondaires
Avorté	2	
Battement	1	10
Butterworth	3	
Déictique	28	2
Emblème	1	
Iconique	16	2
Métaphorique	11	

Tableau 5: Nombre de gestes par type

Au niveau de l'espace gestuel, sur les 70 gestes (60 principaux + 10 battements superposés), 23 sont produits dans le centre et 25 dans la périphérie (voir Tableau 6).

Espace gestuel	Nombre de gestes
Bras tendu devant	0
Centre centre	8
Centre	23
Périphérie	25
Extrême périphérie	14

Tableau 6 : Répartition des gestes dans l'espace gestuel

En regardant la répartition *type de geste/espace gestuel*, on constate que le *centre centre* est assez peu utilisé et que n'y sont produits que des Butterworth et des iconiques. L'ensemble des iconiques est principalement réparti sur les 3 plus petits espaces. Les déictiques, quant à eux, sont répartis entre les 3 plus grands espaces. Il est intéressant de noter que la majorité des gestes produits dans l'extrême périphérie (9 sur 12) sont des déictiques (voir Figure 7). Une fois encore cela peut être dû à la tâche.

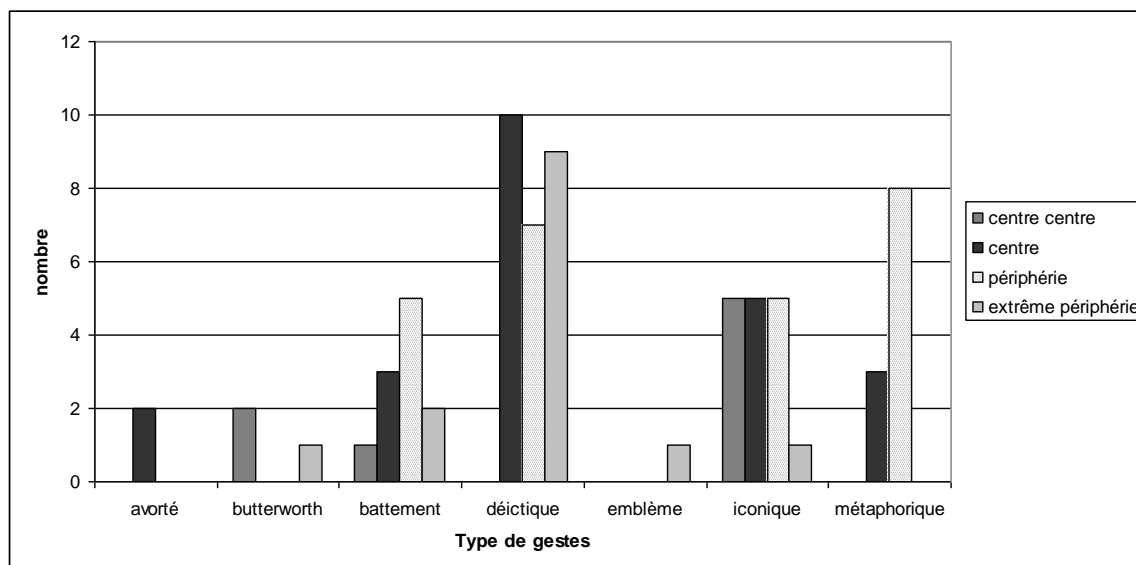


Figure 7 : Répartition des types de gestes dans l'espace gestuel

Il est difficile de tirer des conclusions sur un corpus si réduit. Cependant, nous pouvons faire quelques comparaisons avec la répartition des types de gestes dans l'espace gestuel étudiée par McNeill (1992 : 88). Ce dernier remarque en effet que les iconiques sont surtout produits dans les régions *centre centre* et *centre* et que les déictiques s'étendent dans la *périphérie* et l'*extrême périphérie*. Il note également que les battements sont répartis dans l'espace. En revanche, la plupart des métaphoriques qu'il a analysés sont produits dans le *centre* alors que dans le présent corpus, ils sont plus nombreux en *périphérie*. Cela dit, il convient de ne pas pousser la comparaison trop loin dans la mesure où la tâche proposée par McNeill à ses sujets est une tâche de narration (qui élicite des gestes et une utilisation de l'espace différents de la *maptask*). De plus, la variation interindividuelle dans la production gestuelle est reconnue pour être très importante. En somme la production gestuelle est tâche dépendante et sujet dépendante.

Enfin, nous avons attribué des fonctions à 44 gestes utilisés pour indiquer des directions. Le graphique de la Figure 8 montre la répartition statistique de l'attribution de ces fonctions selon les types de geste. La fonction *anaphorique* permet de faire référence à un élément déjà mentionné dans le discours et concerne 7 gestes (5 déictiques et 2 iconiques), la *description* sert à décrire un lieu et concerne également 7 gestes (6 iconiques et 1 métaphorique). La fonction de *direction* permet d'indiquer des sens de déplacements et concerne 15 gestes (dont 11 déictiques) et enfin la fonction de *positionnement* qui permet au locuteur de localiser des éléments dans l'espace gestuel est représentée par 15 gestes (dont 10 déictiques). Le lien entre type de gestes et fonction apparaît très cohérent. On s'attend en effet à ce que les déictiques apparaissent dans la formulation de directions et des positionnements. Ils servent également à faire des références anaphoriques. Enfin, on s'attend à ce que des iconiques soient privilégiés dans la description d'items.

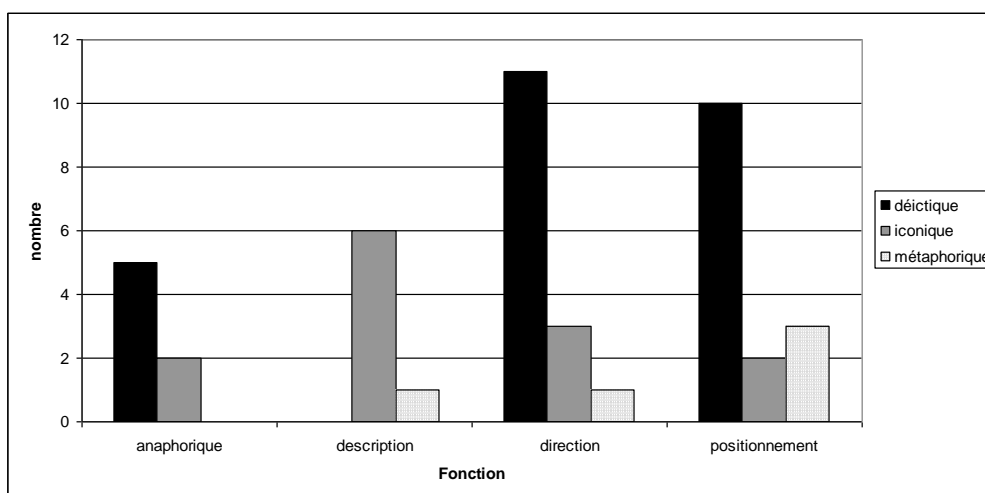


Figure 8 : Attribution des fonctions en fonction des types de geste

### 3.3 Interaction verbale/gestuelle

L'annotation en token (mots) facilite comme nous l'avons déjà évoqué l'annotation gestuelle, notamment pour la synchronisation geste/mot. Elle facilite également le calcul du taux gestuel qui est souvent nécessaire dans les études sur les co-verbaux. Il se calcule en divisant le nombre total de gestes par le nombre total de mots (donné par Elan dans les statistiques). Dans ce corpus le taux est de 0,2 ce qui est assez élevé. Bien sûr, sur ce corpus, calculer le taux gestuel ne présente pas vraiment d'intérêt mais dans le cas où l'on voudrait comparer le taux gestuel de différents locuteurs engagés dans une même tâche langagière, ce serait une donnée pertinente.

Le fait d'avoir une piste présentant les différents mots sur la ligne temporelle rend plus précise la recherche de certaines co-occurrences gestes/mots. Par exemple, si l'on regarde dans les statistiques produites par Elan, on voit que le nom « David » apparaît 4 fois. On peut ensuite rechercher s'il est accompagné de gestes et, le cas échéant, de quels types il s'agit. La première mention du « David » (à 14.22) est accompagnée d'un métaphorique car le locuteur suppose que son interlocutrice connaît cette référence. Comme ce n'est pas le cas, il accompagne sa seconde occurrence du nom « David » en la remplaçant dans un contexte explicité « la statue du David » (15.91) et en produisant un geste iconique avec cette expression qui illustre le référent de manière plus concrète. Ceci fait, les deux mentions suivantes de la statue (21.6 et 27.1) sont accompagnées d'un geste déictique car il ne s'agit plus de décrire un élément du décor mais de s'en servir pour localiser le parcours.

On peut aussi imaginer qu'une étude sur la synchronisation entre gestes et parole serait facilitée par ce type de transcription en token-mots.

## 4 Conclusion

Ce travail sur corpus nous a permis de mettre à l'épreuve différentes techniques utilisées au sein du Laboratoire Parole et Langage et au sein de l'ANR OTIM. Des techniques et outils tels que la segmentation automatique en IPU et la TOE facilitent l'annotation semi-automatique de corpus et permettent un traitement plus rapide de certaines modalités notamment vocales (non abordées ici). Il reste que l'annotation du geste demeure un processus long et instable car les différentes grilles et typologies ne font pas toujours consensus et demandent une large part d'interprétation du codeur. Même si la typologie de McNeill (1992) est encore à l'heure actuelle la plus utilisée par la communauté, elle laisse de côté certains gestes et les frontières entre certains types demeurent encore poreuses. Cela dit, le fait qu'une grande part de la communauté scientifique des gestualistes utilise cette typologie permet de faire des comparaisons entre différentes études et différents résultats. Une base commune est donc indispensable.

Le travail sur l'espace gestuel demeure encore très peu satisfaisant et présente trop d'imprécisions. L'utilisation de capteurs de mouvements pour enregistrer des coordonnées spatiales du geste devrait pouvoir

apporter plus de précision sur ce type de données mais on peut se demander si cette technique n'est pas trop invasive et n'affecte pas la production gestuelle.

## Références

BERTRAND R., BLACHE P., ESPESSE R., FERRÉ G., MEUNIER C., PRIEGO-VALVERDE B., RAUZY S. (2008). Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement automatique des langues (TAL)* 49 (3), 105-134.

BLACHE P., BERTRAND R., BIGI B., BRUNO E., CELA E., ESPESSE R., FERRÉ G., GUARDIOLA M., HIRST D., MAGRO E.-P., MARTIN J.-C., MEUNIER C., MOREL M.-A., MURISASCO E., NESTERENKO I., NOCERA P., PALLAUD B., PRÉVOT L., PRIEGO-VALVERDE B., SEINTURIER J., TAN N., TELLIER M., RAUZY S. (2010). Multimodal Annotation of Conversational Data. *Proceedings of Linguistic Annotation Workshop* (2010 juillet 15-16 : Uppsala, SWEDEN), 186-191.

BLACHE P., BERTRAND R., FERRÉ G. (2008). Creating and exploiting multimodal annotated corpora. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (Marrakech, MOROCCO). Marrakech: ELDA.

BOERSMA P., WEENINK D. (2011). *Praat: doing phonetics by computer*. <http://www.praat.org>

BLANCHE-BENVENISTE C., JEANJEAN C. (1987). *Le français parlé, Transcription et édition*. Paris :Didier-Erudition/ InaLF, 2e éd.

MCNEILL D. (1992). *Hand and Mind : What gestures reveal about thought*. Chicago : The University of Chicago Press.

MCNEILL D. (2005). *Gesture & thought*. Chicago: The University of Chicago Press.

SLOETJES H., WITTENBURG, P. (2008). Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (Marrakech, MOROCCO). Marrakech: ELDA.

TELLIER M., STAM G. (2010). Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. In Actes du Colloque *Spécificités et diversité des interactions didactiques : disciplines, finalités, contextes*. INRP, 24-26 juin 2010, Lyon.