
L'étude de Corpus par Analyse en Composantes Principales

Stéphanie Buisine* — Jean-Claude Martin**

* LCPI-ENSAM, 151 bd de l'Hôpital, 75013 Paris, France

stephanie.buisine@paris.ensam.fr

** LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France

martin@limsi.fr

RÉSUMÉ. L'objectif de cet article est de présenter une méthode d'analyse statistique : l'Analyse en Composantes Principales (ACP). Elle permet d'étudier les relations entre des variables multiples de manière exploratoire, c'est-à-dire lorsque l'exploitation des données n'est guidée par aucune hypothèse préalable. Par ailleurs, l'ACP permet de positionner les individus, ou groupes d'individus, vis-à-vis de la structure révélée des variables. Nous défendons l'intérêt de ce type d'analyse dans le cas de l'étude de corpus d'interaction. A titre d'illustration, nous présentons une expérience au cours de laquelle nous avons recueilli un corpus d'interaction multimodale d'utilisateurs avec des Agents Conversationnels Animés. L'ACP a permis d'extraire des profils comportementaux à partir des données multidimensionnelles, et de comparer deux groupes d'individus : adultes et enfants.

ABSTRACT. The aim of this paper is to present a statistical method: the Principal Component Analysis (PCA). It allows us to study the relations between multiple variables in an exploratory way, i.e. when no preexisting hypothesis guides the data analysis. Besides, PCA allows us to position the individuals, or groups of individuals, with regards to the subsequent structure of variables. We recommend the use of such methods for analyzing interaction corpora. To illustrate this statement, we present an experiment in which we collected a corpus of multimodal interaction of users with Embodied Conversational Agents. PCA enabled us to extract some behavioral patterns from multidimensional data, and to compare two groups of users: adults and children.

MOTS-CLÉS : Analyse de corpus ; Statistiques ; Etude comportementale ; Multimodalité.

KEYWORDS: Corpora analysis; Statistics; Behavior study; Multimodality.

1. L'annotation de corpus

L'annotation de corpus est devenue un domaine de recherche à part entière dans la communauté étudiant l'Interaction Homme-Machine. Devant l'éparpillement des données et des outils (souvent développés en interne au sein des équipes de recherche), le besoin de capitaliser et de standardiser ces ressources s'est fait ressentir et a donné lieu récemment à plusieurs projets européens, parmi lesquels nous pouvons citer (voir aussi Dybkjaer & Bernsen, 2002) : le projet MATE¹ (Multi-level Annotation Tools Engineering), le projet ISLE² (International Standards for Language Engineering), et le projet NITE³ (Natural Interactivity Tools Engineering). Plusieurs workshops dans des conférences internationales ont également été consacrés aux corpus multimodaux (Maybury & Martin, 2002 ; Martin et al., 2004 ; Devillers et al., 2006 ; Martin et al., 2006). L'étude de corpus a un grand intérêt pour les études comportementales et/ou pour fournir des données étiquetées à un système ayant des capacités d'apprentissage.

L'annotation des corpus aboutit cependant généralement à une quantité considérable de données difficiles à manipuler et à résumer, en raison notamment d'une grande quantité de variables. Dans cet article, nous présentons une méthode statistique, l'Analyse en Composante Principale, et son application à l'analyse de corpus. Nous présenterons rapidement les principes généraux de cette méthode, puis nous illustrerons le type de résultats qu'elle permet d'obtenir, avant de conclure sur son intérêt dans l'exploitation de corpus.

2. L'Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales (ACP), appelée aussi « analyse géométrique des données » ou « analyse des corrélations » (Wolff, 2003 ; Wolff & Visser, 2005), permet d'analyser des données multivariées et de les visualiser sous forme de nuages de points dans des espaces géométriques. Par opposition aux analyses confirmatoires (par exemple l'Analyse de Variance), qui visent à tester des hypothèses fondées sur un raisonnement théorique, l'ACP est dite exploratoire, et permet de découvrir des relations entre variables sans avoir d'hypothèses préalables. L'intérêt majeur de l'ACP est d'offrir la meilleure visualisation possible des données multivariées, en identifiant le ou les plans dans lesquels la dispersion est maximale – mettant ainsi en évidence avec le maximum de précision les relations de proximité et d'éloignement entre les variables. Les « composantes principales » désignent les axes orthogonaux qui structurent les données et qui rendent compte de cette variance maximale.

¹ <http://mate.nis.sdu.dk/>

² <http://isle.nis.sdu.dk/>

³ <http://nite.nis.sdu.dk/>

L'ACP permet, dans un premier temps, de visualiser les relations entre les variables. Puis le chercheur, ou l'expert, doit réaliser un travail d'interprétation afin de renommer les axes (composantes principales) et ainsi de résumer les données. L'interprétation des axes est fondée sur l'analyse des contributions de chaque variable à chaque axe. Enfin, une fois l'espace des variables interprété, l'ACP permet de positionner les individus, ou les groupes d'individus, dans le nouvel espace et de visualiser leurs relations avec les axes d'une part, et leurs relations avec les autres groupes ou les autres individus d'autre part.

Pour mieux comprendre l'intérêt de l'ACP pour l'analyse de corpus, nous proposons de dérouler un exemple d'analyse appliqué à un corpus de comportements multimodaux.

3. Exemple d'application de l'ACP

L'objectif de l'expérience décrite ci-dessous était d'étudier le comportement multimodal d'utilisateurs face à des Agents Conversationnels dans le cadre d'un jeu. Pour cela, nous avons conçu un dispositif de simulation en Magicien d'Oz : les utilisateurs avaient la possibilité d'interagir avec les Agents par la parole, par le geste (utilisation d'un stylo sur une tablette-écran Wacom Cintiq 15X) ou par des commandes multimodales (combinaison de paroles et de gestes). Le scénario du jeu consistait à rendre service aux Agents en allant trouver l'objet manquant à chacun d'entre eux et en le leur apportant (Figure 1). Une caméra numérique, installée au-dessus de la tablette-écran, a permis un enregistrement audio-visuel du comportement des utilisateurs.



Figure 1. La situation expérimentale.

3.1. Méthode

Dix-sept utilisateurs ont participé à cette expérience. Ils se répartissent en deux groupes : 7 utilisateurs adultes (3 hommes, 4 femmes, âge moyen 25 ans) et 10 enfants (7 garçons 3 filles, âge moyen 11 ans).

Les vidéos recueillies ont été annotées à l'aide des outils Praat (<http://www.fon.hum.uva.nl/praat/>) et Anvil (Kipp, 2001). Nous avons ainsi pu analyser les constructions multimodales, et notamment l'intégration temporelle et sémantique des modalités :

- Intégration temporelle : chaque construction multimodale peut témoigner soit d'une intégration simultanée des modalités, lorsqu'il y a chevauchement entre parole et geste, soit d'une intégration séquentielle des modalités, lorsqu'il n'y a pas de chevauchement.
- Intégration sémantique : dans ce corpus, nous avons distingué les cinq types d'intégration sémantique suivants (Martin et al., 2001 ; Buisine & Martin, 2005) :

- *La Redondance* (ex : « Je voudrais aller dans la pièce rose s'il te plaît. » + [pointage sur la porte rose]),
- *La Complémentarité* (ex : « Tiens prends-le. » + [pointage sur un livre]),
- *La Concurrence* (ex : « Bonjour » + [pointage sur un gâteau] ou « La cafetière électrique » + [exploration au niveau du bureau]).
- *La Complémentarité dialoguée* : il s'agit de cas où les utilisateurs annoncent ou demandent l'autorisation verbalement avant de réaliser une action par le geste (ex : « Je peux prendre du gâteau ? » Réponse de l'Agent : « Sers-toi. » [l'utilisateur pointe sur une part de gâteau]).
- *La Répétition multimodale* : face à une erreur ou une absence de réaction du système, l'utilisateur peut être amené à répéter une commande. Dans 86% des cas, la répétition a été faite dans la même modalité que la commande initiale, mais dans 14% des cas, l'utilisateur a changé de modalité lors de la répétition. Nous avons comptabilisé ces cas dans les répétitions multimodales.

3.2. Résultats

Le corpus a donné lieu à l'analyse des variables suivantes : comportement verbal (ex : syntaxe), prise d'initiative, comportements sociaux, comportements gestuels (ex : forme des gestes), et enfin comportements multimodaux. Dans cet article, nous ne développerons que l'analyse des comportements multimodaux.

Globalement, la proportion d'utilisation des modalités se répartit comme suit : 47,53% des commandes ont été réalisées avec le stylo uniquement, 31,28% ont été réalisées avec la parole uniquement, et 21,19% ont été réalisées de façon

multimodale en faisant un usage combiné de la parole et du stylo, ce qui représente un effectif brut de 117 constructions multimodales.

Dans le but d'identifier des profils de comportement multimodal pour les deux groupes d'utilisateurs (adultes et enfants), nous avons réalisé une ACP à l'aide du logiciel SPSS (www.spss.com). Nous y avons inclus les 11 variables suivantes :

- | | |
|-------------------------------|--|
| 1. Nombre TOTAL de commandes, | 7. % CONCURRENCE, |
| 2. % PAROLE, | 8. % complémen. dialoguée COMP-D, |
| 3. % GESTE, | 9. % REPETITION multimodale, |
| 4. % MULTIMODALITE, | 10. % Constructions simultanées SIM, |
| 5. % REDONDANCE, | 11. % Constructions séquentielles SEQ. |
| 6. % COMPLEMENTARITE, | |

La matrice de corrélation nous indique que les corrélations positives les plus fortes sont :

MULTIMODALITE et SIM	r = 0,82
REDONDANCE et SIM	r = 0,69
MULTIMODALITE et REDONDANCE	r = 0,59

Cela signifie que les quelques sujets ayant beaucoup utilisé la multimodalité ont eu tendance à intégrer les modalités de façon simultanée et redondante. Les autres patterns temporels et sémantiques que nous avons observés sont donc sans doute attribuables à des sujets ayant moins utilisé la multimodalité.

Les corrélations négatives les plus fortes sont :

PAROLE et GESTE	r = -0,78
GESTE et SIM	r = -0,66
GESTE et MULTIMODALITE	r = -0,64

Les sujets ayant beaucoup utilisé le geste ont donc eu tendance à utiliser peu la parole, la multimodalité et la simultanéité.

A l'issue de l'ACP, nous avons retenu 4 axes expliquant 78,3% de la variance totale du nuage de points (un total de 70% de variance expliquée est généralement considéré comme acceptable). Les coordonnées factorielles des variables (coefficients de corrélation avec les axes), élevées au carré, indiquent la qualité de la représentation de chaque variable (le maximum étant de 1). Dans notre cas, les variables les mieux représentées sont le GESTE (qualité de la représentation = 0,91), la REPETITION (0,84), la MULTIMODALITE (0,83), la COMPLEMENTARITE (0,83) et la COMP-D (0,83). La variable la moins bien représentée est SEQ (0,55) : en dessous de 0,6 la qualité de la représentation est en effet considérée comme faible. Cela signifie que l'interprétation de cette variable devra s'accompagner de précautions importantes.

Les Figures 2 et 3 représentent le nuage des variables dans les plans 1-2 et 3-4 respectivement. Pour l'interprétation de chaque axe, nous avons retenu les variables dont la contribution à la variance de l'axe était supérieure à la contribution moyenne. Ainsi, l'axe 1 oppose MULTIMODALITE – REDONDANCE – SIM

d'un côté et GESTE – REPETITION de l'autre côté. Cet axe représente donc la distinction entre les individus ayant spontanément beaucoup utilisé la multimodalité et ceux qui ont plutôt utilisé le geste et n'ont recouru à la multimodalité qu'en cas d'échec de la communication. L'axe 2 oppose GESTE – CONCURRENCE avec PAROLE – COMP-D. Nous interprétons cet axe comme une opposition Action/Dialogue, l'Action étant représentée par l'utilisation fréquente du geste, parfois même en parallèle de la parole (concurrence), et le Dialogue étant associé à l'utilisation fréquente de la parole, y compris pour annoncer l'action gestuelle (complémentarité dialoguée).

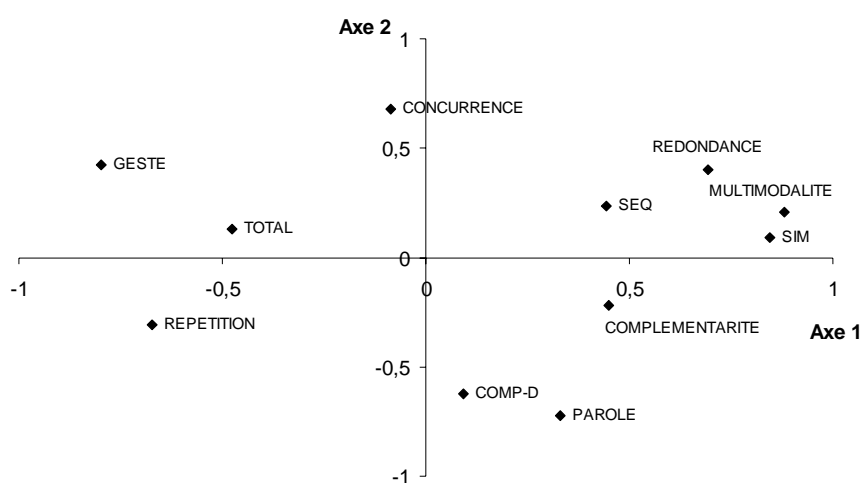


Figure 2 : Nuage des variables dans le Plan 1-2.

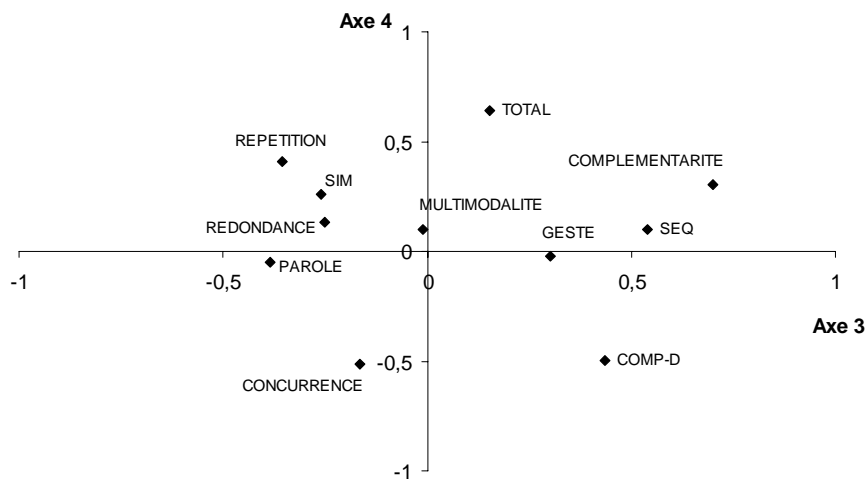


Figure 3 : Nuage des variables dans le Plan 3-4.

L'axe 3 oppose quant à lui COMPLEMENTARITE – COMP-D – SEQ avec la PAROLE. Les constructions complémentaires et/ou séquentielles peuvent donc

correspondre à une économie de parole pour les individus ayant peu utilisé la modalité verbale. Enfin, l'axe 4 oppose TOTAL – REPETITION, qui traduisent une certaine inefficacité de l'interaction, avec COMP-D – CONCURRENCE, qui peuvent être vus comme des modes d'interaction optimisés, dans la mesure où les exemples de concurrence que nous avons observés concernaient une utilisation parallèle des deux modalités, c'est-à-dire une manière de faire deux choses à la fois.

Cette analyse du nuage des variables nous permet de présenter le nuage des points moyens des Adultes et des Enfants en fonction des axes tels que nous venons de les interpréter. La Figure 4 nous montre que les Adultes sont situés du côté de la multimodalité spontanée et du dialogue, alors que les enfants sont du côté du geste et de l'action. Les enfants ont tendance à recourir à la multimodalité en cas d'échec dans la communication ou bien pour effectuer simultanément deux actions concurrentes. La Figure 5, même si elle montre que Adultes et Enfants s'opposent moins sur les axes 3 et 4, semble confirmer que les enfants cherchent à économiser la parole et à optimiser l'interaction en utilisant les modalités de façon complémentaire ou concurrente. Les adultes restent du côté de la parole et d'une interaction un peu moins performante.

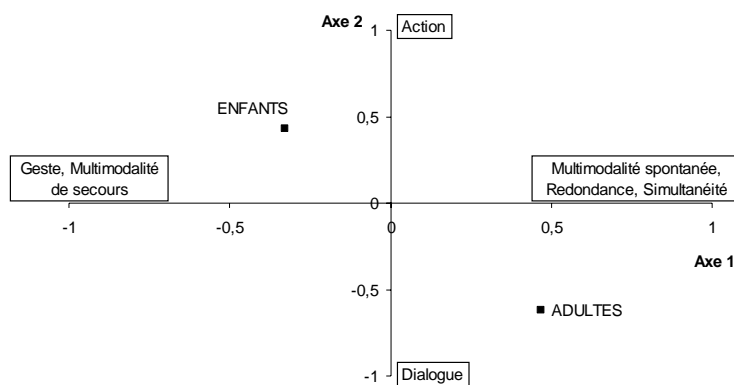


Figure 4 : Nuage des points moyens des Adultes et des Enfants dans le plan 1-2.

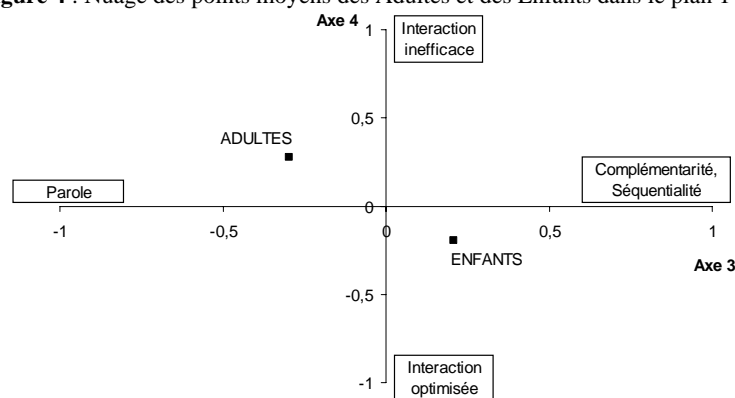


Figure 5 : Nuage des points moyens des Adultes et des Enfants dans le plan 3-4.

4. Conclusion

Dans l'exemple précédent, l'ACP a permis de mettre en évidence des résultats que nous aurions eu peu de chances de trouver avec d'autres méthodes statistiques. Ces résultats donnent des indications sur le comportement de jeu des adultes et des enfants : l'ACP a montré que les Adultes se sont davantage positionnés du côté de la conversation, alors que les enfants étaient plutôt du côté de l'action et du geste. Nous avons également relevé une tendance des enfants à produire des commandes concurrentes : ils dissocient parfois le canal verbal et le canal gestuel pour réaliser deux actions à la fois. Ces résultats permettent d'engager de nouvelles réflexions et peuvent servir de support de créativité pour imaginer de nouveaux scénarios de jeu.

Face au nombre important de variables recueillies, l'ACP permet de résumer celles-ci afin d'en dégager des profils comportementaux. Ceux-ci sont susceptibles d'apporter des informations nouvelles, de plus haut niveau d'abstraction que celles portées par chaque variable considérée individuellement. La principale limite de cette méthode est son étape subjective lors de la phase d'interprétation des axes (composantes principales). La subjectivité peut cependant être réduite en confrontant les résultats d'interprétation de plusieurs juges et en convergeant vers une interprétation commune. Il semble aujourd'hui peu probable que cette phase d'interprétation puisse être automatisée ; il restera toujours une part d'expertise dans la conduite de l'analyse. Nous pensons que la technique de l'ACP peut être bénéfique dans de nombreuses situations d'analyse de données complexes, et mérite d'être diffusée plus largement au sein de la communauté travaillant sur l'exploitation de corpus d'interaction.

Remerciements

Cette étude a été partiellement financée par le projet EU/HLT NICE (IST-2001-35293). Nous remercions M. Wolff et C. Le Quiniou, (Paris-5) pour leur contribution.

Références

- Buisine, S., Martin, J.C. (2005). Children's and adults' multimodal interaction with 2D conversational agents. CHI'05 International Conference on Human Factors in Computing Systems (short paper), pp. 1240-1243, ACM Press.
- Devillers, L., Martin, J.C., Cowie, R., Douglas-Cowie, E., Batliner, A. (2006). Proceedings of the Workshop on Corpora for Research on Emotion and Affect: In conjunction with LREC'2006.
- Dybkjaer, L., & Bernsen, N.O. (2002). Natural interactivity resources - Data, annotation schemes and tools. Proceedings of LREC'02.
- Kipp, M. (2001). Anvil - A generic annotation tool for multimodal dialogue. Proceedings of Eurospeech'2001, pp. 1367-1370.

- Martin, J.C., Den Os, E., Kuhnlein, P., Boves, L., Paggio, P., & Catizone, R. (2004). Proceedings of the Workshop on Multimodal Corpora: Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces: In conjunction with LREC'2004.
- Martin, J.C., Grimard, S., & Alexandri, K. (2001). On the annotation of the multimodal behavior and computation of cooperation between modalities. Proceedings of International Conference on Autonomous Agents Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, pp. 1-7.
- Martin, J.C., Kühnlein, P., Paggio, P., Stiefelhagen, R., Pianesi, F. (2006). Proceedings of the Workshop on Multimodal Corpora from Multimodal Behaviour Theories to Usable Models: In conjunction with LREC'2006.
- Maybury, M., & Martin, J.C. (2002). Proceedings of the Workshop on Multimodal Resources and Multimodal Systems Evaluation: In conjunction with LREC'2002.
- Wolff, M. (2003). Apports de l'analyse géométrique des données pour la modélisation de l'activité. In J.C. Sperandio & M. Wolff (Eds.), *Formalismes de modélisation pour l'analyse du travail et l'ergonomie*, pp. 195-227. Paris: Presses Universitaires de France.
- Wolff, M., Visser, W. (2005). Méthodes et outils pour l'analyse des verbalisations : Une contribution à l'analyse du modèle de l'interlocuteur dans la description d'itinéraires. *Activités*, 2, pp. 99-118.