

APs : une approche proxémique pour l'analyse de données sur les réseaux sociaux

Maxime Masson*, Philippe Roose*, Christian Sallaberry*, Rodrigo Agerri**
Marie-Noelle Bessagnet*, Annig Le Parc Lacayrelle*

*LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA)
maxime.masson@univ-pau.fr, christian.sallaberry@univ-pau.fr, marie-noelle.bessagnet@univ-pau.fr, annig.lacayrelle@univ-pau.fr, philippe.roose@univ-pau.fr
** HiTZ, IXA, Université du Pays Basque (UPV/EHU)
rodrigo.agerri@ehu.eus

La proxémique est la science qui étudie l'organisation de l'espace et l'effet des distances dans les relations interpersonnelles (Hall, 1966). Elle se base sur les concepts de zone et de dimension proxémique. Ces dernières sont au nombre de 5 (DILMO) : la Distance, l'Identité, la Localisation, le Mouvement et l'Orientation. La proxémique est généralement appliquée aux interactions physiques (Perez et al., 2020). Cependant, avec l'accroissement important des échanges réalisés via internet et l'essor du contenu généré par les utilisateurs au travers des réseaux sociaux, nous faisons l'hypothèse que l'adapter à ce monde virtuel pourrait se montrer pertinent pour guider l'analyse des interactions se déroulant sur ce type de support.

Nous proposons donc un cadre de travail (APs : *Augmented Proxemic services*) permettant de s'appuyer sur la théorie de la proxémique pour guider l'analyse des interactions sur les réseaux sociaux autour d'un domaine donné. La **Figure 1** présente le cycle de vie de données issues de réseaux sociaux que nous proposons d'étudier. Le domaine d'application que nous expérimentons est le **tourisme**.

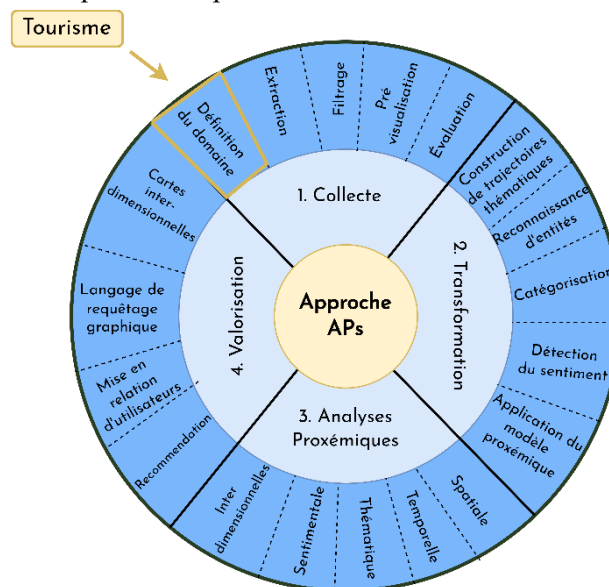


FIG. 1 – Architecture du framework APs (avec application au tourisme)

Notre approche vise cependant à être totalement générique et adaptable à n'importe quel domaine. Le cycle de vie APs se divise en 4 grandes phases distinctes : **la collecte (1), la transformation (2), les analyses proxémiques (3) et la valorisation (4)**. Les utilisateurs sont impliqués à différentes étapes de ces dernières (dans notre cas, nous en avons 2 types : les **professionnels du tourisme** et les **touristes**).

La *collecte* (1) couvre l'ensemble du processus de recherche et de récupération des données. Cela passe par la définition précise du jeu de données cible en fonction des besoins (ex : *le type de tourisme, la zone spatiale et la temporalité ciblée*). C'est l'utilisateur final lui-même qui se charge de ce travail de définition, dans notre cas, un expert du tourisme (office de tourisme). S'en suit la recherche des réseaux sociaux les plus adaptés puis l'extraction en elle-même. Ce dernier point a nécessité la conception d'une **méthodologie générique et indépendante du domaine pour l'extraction de données des réseaux sociaux**. Cette méthodologie repose sur 3 dimensions (*spatiale, temporelle et thématique*). Nous avons ici fait le choix de représenter le thème en tant que vocabulaire hiérarchisé (*ontologie, thésaurus*) et de proposer un fonctionnement itératif et incrémental donnant une place cruciale à l'utilisateur final. En effet, cette méthodologie permet de s'assurer de la qualité des données collectées (*prévisualisation et évaluation*) et d'influencer directement le processus de filtrage via un mécanisme de « *boucles de feedback* ». L'utilisateur final prévisualise et analyse en temps réel les données brutes qui serviront plus tard à répondre à ses besoins.

La *transformation* (2) concerne les multiples modifications et enrichissements appliqués sur les données précédemment collectées afin d'augmenter leur valeur ajoutée et de les préparer pour les prochaines étapes. Cet enrichissement peut se faire, par exemple, en extrayant les entités nommées pertinentes du texte pour le domaine d'application (lieux, entités temporelles, activités pour le tourisme) ou encore en construisant des trajectoires thématiques (correspondant aux enchaînements des activités des visiteurs dans le cas du tourisme). Nous catégorisons également les posts (ex : *type de pratique touristique*) et détectons la polarité de ces derniers (*le sentiment*). Toutes les informations ainsi extraites et enrichies permettent d'instancier notre modèle proxémique. Dans notre modèle proxémique, la localisation n'est plus géographique mais thématique (*concepts d'une ontologie*) et l'orientation devient le sentiment dominant du post étudié. Cette phase est purement technique sans implication directe de l'utilisateur final.

Les *analyses proxémiques* (3) permettent de produire une batterie d'indicateurs bruts pour le domaine ciblé. Elles sont obtenues en combinant les dimensions proxémiques entre elles. Le choix des dimensions proxémiques à utiliser et la combinaison de ces dernières afin d'obtenir des analyses pertinentes sont guidés directement par les besoins et attentes de l'utilisateur final. Par exemple, dans notre cas d'étude du tourisme, il serait possible d'analyser la satisfaction moyenne des visiteurs pour chaque activité touristique en assimilant leurs flux sur un réseau social (*ensembles de publications*) à un mouvement (M) composé de multiples localisations (L, les publications). Chaque localisation étant positionnée dans une ou plusieurs zones de l'espace thématique étudié (*ici l'espace des activités touristiques*) et associé à une orientation (O, le sentiment dominant exprimé). On obtient une modélisation de type LMO (*localisation, mouvement, orientation*). Pour analyser la satisfaction par

type de visiteur (*locaux, externes, etc.*) on pourrait ajouter à cette modélisation la dimension identité (I). La dimension distance (D) est modulable et doit être redéfinie en fonction des besoins utilisateurs. On pourrait l'utiliser pour modéliser le lien entre activité touristique et météo ou période de l'année (une distance faible dénoterait une forte corrélation). Par exemple, entre l'activité « *se baigner à la plage* » et la météo « *soleil* ». Cette approche proxémique permet de prendre du recul par rapport au domaine d'étude et de proposer un cadre d'analyse facilement réutilisable et adaptable à n'importe quel cas d'usage.

La *Valorisation* (4) permet de visualiser les résultats des analyses précédentes pour des utilisateurs finaux tels que les offices de tourisme ou les touristes eux-mêmes. Par exemple, pour les professionnels du tourisme, des cartes thématiques multidimensionnelles visualisent les tendances et associations de thèmes sur les réseaux sociaux. Ou encore, les touristes peuvent utiliser un système de recommandation (*activités, lieux, itinéraires*).

Finalement, nous envisageons d'utiliser la proxémique comme base pour un langage (DSL) de requêtage de corpus issu des réseaux sociaux dédié à l'utilisateur final. Chaque dimension proxémique servirait de métaphore pour abstraire des concepts de plus bas niveau.

Références

Hall E. T. (1966). *The hidden dimension*, Anchor, vol. 609.

Perez P., Roose P., Cardinale Y., Dalmau M., Couture N., et al. (2020). *Mobile Proxemic Application Development for Smart Environments*. 18th International Conference on Advances in Mobile Computing & Multimedia, Chiang Mai, Thailand. pp.94.