HAL open science

Représentations sociales de l'analytique des apprentissages avec le numérique

Camila Morais Canellas, François Bouchet, Vanda Luengo

To cite this version:

Camila Morais Canellas, François Bouchet, Vanda Luengo. Représentations sociales de l'analytique des apprentissages avec le numérique. 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Marie Lefevre, Christine Michel, Jun 2021, Fribourg, Suisse. pp.354-359. hal-03290068

HAL Id: hal-03290068

https://hal.science/hal-03290068

Submitted on 22 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Représentations sociales de l'analytique des apprentissages avec le numérique

Camila Morais Canellas[1], François Bouchet[1] et Vanda Luengo[1]

Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
{camila.canellas, francois.bouchet, vanda.luengo}@lip6.fr

Résumé. Cette étude vise à identifier les représentations sociales de l'analytique des apprentissages avec le numérique chez différents acteurs de l'éducation. Nous avons analysé les réponses de 286 participants à un questionnaire en utilisant la technique d'Évocation Hiérarchique et les résultats montrent que pour la majorité d'entre eux, les aspects les plus importants pour déterminer le domaine sont : apprentissage, données, analyse, suivi et traces. De plus, nous avons observé des variations entre différents sous-groupes de parties prenantes.

Mots-clé: analytique des apprentissages · représentations sociales · évocation · analyse de similarité · Test d'Indépendance de Contexte (TCI).

Abstract. This study aims to identify the social representations of the Learning Analytics domain among different education stakeholders. We analyzed the responses of 286 participants to a questionnaire using the Hierarchical Evocation technique. Results show that for the majority of them the most important aspects to determine the domain are: learning, data, analysis, tracking and traces. In addition, we observed variations between different subgroups of stakeholders.

Keywords: learning analytics · social representations · evocation · similarity analysis · Test of Context Independence (TCI).

1 Introduction et travaux connexes

L'analytique des apprentissages avec le numérique (traduction française de *Learning Analytics* (LA) d'après [4]) est un domaine scientifique récent lié aux contextes éducatifs dont la première conférence internationale scientifique, a eu lieu en 2011[1]. Parfois, les connaissances produites dans un cadre scientifique entrent en conflit avec les croyances et habitudes antérieures de la société. Avec le temps et les nouveautés produites par le terrain, différents acteurs ou groupes sociaux forment une compréhension commune - une représentation -, de ce qu'est le domaine d'étude, de ce qu'il produit, de la façon dont il pourrait être utile ou néfaste, etc. Le processus d'internalisation (sociale) de nouvelles informations (scientifiques) a fait l'objet d'études dans de nombreux domaines par chercheurs intéressés par la théorie des Représentations Sociales (RS). Par conséquent, lorsque

[1] LAK 2011 : https://tekri.athabascau.ca/analytics/

l'on tente de mettre en œuvre les résultats des LA dans un cadre éducatif, il est important de comprendre à l'avance la nature de ces représentations. Nos questions de recherche sont : 1/ Quelle est la représentation sociale des LA parmi les parties prenantes ? 2/ Quelle est sa structure ? et 3/ Y a-t-il des différences entre les profils des différentes parties prenantes ?

En tant que nouveau domaine d'étude, au cours de ses premières années, la communauté de LA a produit des recherches sur le niveau de compréhension et les attentes envers le domaine, et un certain nombre de réflexions basées sur la bibliométrie ou des revues systématiques. En nous concentrant sur les articles qui s'appuient sur des mots-clés pour tenter d'identifier les limites du domaine, nous pouvons citer : (1) Balacheff et al. [3] qui ont analysé des mots-clés liés aux LA dans des textes produits pendant le *Stellar Grand Challenge* ; (2) Griffiths et al. [7] qui dans le cadre de *Learning Analytics Community Exchange (LACE)* ont identifié 13 thèmes qui ont émergé de leur analyse ; (3) Labarthe et al. [9] qui comparent les LA à des champs connexes, notamment en termes de mots-clés ; et (4) [12] où les auteurs ont utilisé des articles publiés entre 2011 et 2019 afin d'identifier la structure conceptuelle et les progrès thématiques dans les LA en utilisant notamment l'occurrence de mots.

Néanmoins, la définition formelle d'un champ, la pratique réelle de ses chercheurs et les représentations avérées des personnes impliquées d'une manière ou d'une autre sont trois choses potentiellement différentes à considérer.

2 Cadre d'étude et méthode

La théorie des RS se concentre sur l'individu situé dans son contexte historique et sociologique. Les RS sont des images mentales construites à partir d'expériences avec l'objet et de la pratique sociale et du langage [8]. De plus, ils peuvent avoir une dimension prédictive : en connaissant la RS d'un individu ou d'un groupe d'individus, on peut en déduire une tendance à réaliser certaines actions [6].

La théorie du noyau central. Nous représentons des objets sociaux en organisant des réseaux d'informations autour de connaissances ou de croyances, et c'est cette organisation qui est au cœur des préoccupations de la théorie complémentaire du noyau central [11]. Flament et Moliner [5] ont démontré l'existence de deux éléments dans les RS qui jouent des rôles différents : le noyau central et les éléments périphériques. Par conséquent, pour définir une RS, il est essentiel de connaître son contenu mais aussi sa structure.

Ainsi, cette étude est à la fois quantitative et exploratoire ; elle vise à identifier la RS des LA et sa structure parmi les personnes impliquées dans le domaine de l'éducation. Nous utilisons la technique d'Évocation Hiérarchique [2], basée sur des associations libres, et les résultats sont également examinés via une analyse de similarité. Lors de la collecte de données, un premier moment est dédié aux associations libres et un second à la hiérarchisation de ces associations, permettant d'accéder à une RS et à son hypothèse de centralité. L'interprétation repose sur deux prémisses : 1/ la *saillance* d'un élément, vue comme une fréquence élevée par rapport à la fréquence moyenne de tous les éléments évoqués, qui traduit

un fort consensus des participants et 2/ le *rang* de l'élément, qui dépend du fait que le mot ou l'expression soit classé comme parmi les plus importants. En combinant les deux indicateurs quantitatifs, les quatre cases décrites ci-dessous sont obtenues. L'analyse peut être effectuée à l'aide d'un logiciel [13].

La zone du *noyau central* est considérée comme liée à la mémoire collective et à l'histoire du groupe, elle est cohérente, rigide et stable dans le temps. Elle résiste au changement et n'est pas sensible au contexte immédiat, a tendance à avoir peu d'éléments, à être consensuelle et à définir l'homogénéité du groupe [1]. La *première périphérie* rassemble les éléments périphériques les plus importants. La case des *éléments contrastés* contient des éléments évoqués par peu de personnes, mais qui les considèrent très importants [2]. La *deuxième périphérie* est constituée d'éléments à basse fréquence et qui sont considérés comme peu importants dans le champ sémantique de la représentation.

Nous avons créé un questionnaire[2] diffusé en ligne au sein de groupes liés au domaine de l'éducation. Les réponses ont été recueillies entre mars et juillet 2019. Le questionnaire comprenait deux sections. La première consistait en deux questions permettant d'accéder à la RS de LA et à l'importance relative (classement) de chaque évocation. Les questions de la deuxième section permettaient d'établir des profils basés sur les rôles des participants et sur leur implication et expériences antérieures avec les LA. Avant l'analyse, nous avons préparé les données en suivant plusieurs étapes recommandées dans [13].

Afin de confirmer l'hypothèse de centralité découverte lors de la première phase [10] nous avons utilisé la méthode TCI. Un deuxième questionnaire (disponible sur le même lien : LAFR2_726112.pdf) a été envoyé aux mêmes listes et groupes que le premier, entre juillet et septembre 2020.

3 Résultats

Nous avons obtenu un total de $N = 286$ réponses valides, formant un corpus de 1843 mots ou expressions (598 uniques). La fréquence minimale (visant à réduire le corpus en enlevant des mots très peu cités) a été établie à 10 évocations, réduisant le corpus à 37 mots uniques (cf. Tableau 1). Le rang moyen est de 3, la médiane du nombre de mots demandés par le questionnaire. Compte tenu des caractéristiques de nos données et d'après [14], nous avons fixé la fréquence intermédiaire de coupure comme étant la moyenne des fréquences (29).

Le Tableau 1 affiche les évocations de toutes les parties prenantes participant à notre enquête. La fréquence (Fq.) est le nombre total de fois qu'un mot/une expression a été mentionné et le rang moyen (Rg.) la moyenne de sa position dans la classification faite par les participants du plus (1) au moins (5) important. Certains mots (e.g. "big-data") ont été évoqués en anglais et gardés ainsi.

La case du noyau central contient les éléments les plus fréquents et les plus importants. Il est composé de cinq mots, le plus mentionné étant "apprentissage" (86 fois) et le rang moyen le plus bas dans cette case est pour le mot "traces".

[2] LAFR_783534.pdf : `https://github.com/cmc-lip6/socialrepresentations`

Tableau 1: Analyse des évocations hiérarchiques de tous les participants

Freq.	**Rang Moyen** < 3			**Rang Moyen** $\geqslant 3$		
	mot	**Fq.**	**Rg.**	**mot**	**Fq.**	**Rg.**
$\geqslant 29$	apprentissage	86	2,4	évaluation	44	3,3
	données	64	2,7	statistiques	35	3,3
	analyse	62	2,5			
	suivi	49	2,6			
	traces	38	2,3			
[10, 29[	apprenant	26	2,8	numérique	25	3,2
	big-data	24	3,0	enseignement	18	3,2
	personnalisation	22	2,9	adaptation	17	3,5
	parcours	19	2,1	intelligence-artificielle	16	3,7
	recherche	17	2,5	compétences	13	3,0
	formation	15	2,5	amélioration	11	3,2
	activité	15	2,6	contrôle	11	3,4
	compréhension	14	2.8	informatique	10	3,3
	indicateurs	12	2,4	tableau-de-bord	10	3,5
	mémoire	12	2,4	visualisation	10	3,9
	pédagogie	12	2,4			
	résultats	12	2,5			
	individualisation	11	2,2			
	remédiation	11	2,4			
	étudiant	11	2,4			
	connaissance	11	2,8			
	éthique	10	2,3			
	data	10	2,6			
	élèves	10	2,6			
	feedback	10	2,7			

Le mot "données", évoqué 64 fois, est présent dans le noyau même s'il n'est pas agrégé avec le même mot mentionné à l'origine en anglais (10 fois) et en "big-data" (24 fois). Les éléments de la deuxième périphérie ont encore une importance pour la représentation. Seuls deux mots sont présents, "évaluation" et "statistiques". La case des éléments contrastés est composée de 20 éléments, dont la fréquence varie de 26 à 10 et avec un rang moyen de 2,1 à 2,9.

Nous avons également effectué une analyse de similarité, basée sur la théorie des graphes, afin de mieux comprendre visuellement comment les mots sont évoqués ensemble (voir Figure 1). Cette analyse a été effectuée en utilisant Iramuteq[3] avec un seuil d'au moins 12 occurrences pour une meilleure visibilité. Nous avons choisi l'index "arbre maximum" où la taille des nœuds représente la fréquence des mots et l'épaisseur des lignes représente les associations entre eux.

Ensuite, nous avons essayé d'identifier des profils de participants pour vérifier si des différences étaient observables entre des sous-groupes, selon le rôle, la discipline, l'implication et l'expérience avec les LA.

L'analyse des ces sous-groupes nous a permis d'observer que la zone de la première périphérie est souvent peu peuplée (voire vide), ce qui pourrait être le signe d'un fort consensus sur le noyau et, en même temps, d'une opinion assez dispersée des autres éléments formant les RS des LA. De plus, deux sous-groupes, "enseignants" et "sans expérience", semblent avoir des représentations assez différentes par rapport aux autres sous-groupes ainsi que par rapport à la représentation principale du total des participants.

[3] `http://www.iramuteq.org/`

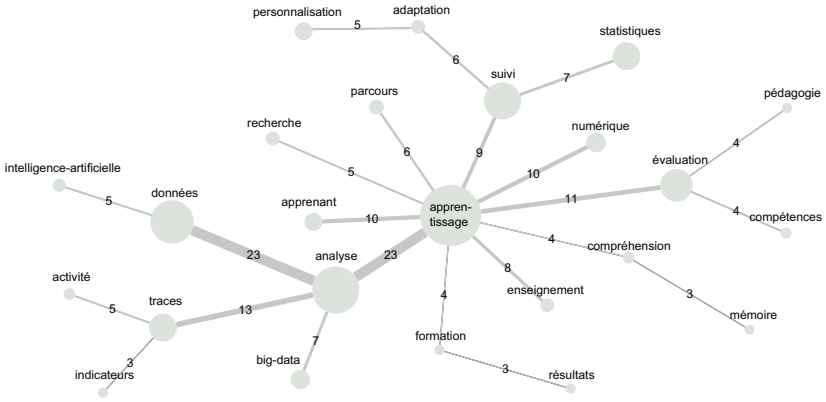

Fig. 1: Analyse de similarité des termes co-occurrents

Le deuxième questionnaire, visant à confirmer ces résultats, a obtenu 61 réponses complètes ($N = 61$). Après avoir calculé les fréquences "oui", nous avons utilisé le test de Kolmogorov–Smirnov pour confirmer que la distribution était normalement distribuée ($p = 0,764$). Nous avons ensuite utilisé un test du Chi-2 pour identifier les évocations statistiquement significatives ($p < .05$) formant potentiellement le noyau central - où l'hypothèse nulle est que l'évocation testée est indépendante du contexte. Les évocations ainsi identifiées sont : "**apprentissage**", "statistiques", "indicateurs", "activité", "**data**", "**analyse**", "**données** et "**traces**". Les résultats montrent que toutes les évocations apparaissant précédemment dans le noyau y figurent à nouveau, sauf "suivi" (situé juste sous le seuil de significativité), avec quelques nouveaux ajouts (mots non en gras ci-dessus).

4 Discussion et limitations

Même s'il n'y a pas de consensus sur la taille optimale des échantillons pour étudier les RS, plus le groupe de participants est grand, plus les résultats sont stables. La taille de notre échantillon est plus grande ($N = 286$) que ce qui est parfois recommandé [14], mais ce n'est pas le cas de certains sous-groupes que nous avons analysés pour lesquels les résultats doivent être considérés avec prudence et nous envisageons donc de reconduire l'enquête auprès d'autres acteurs, notamment parmi des apprenants.

Dans cette étude, nous avons mis en lumière les représentations sociales de différents acteurs de l'éducation au sujet des LA. Nous pensons que la connaissance des RS des LA pourra aider les décideurs dans leurs résolutions et favoriser le dialogue entre les différentes parties prenantes pouvant être impactées par le domaine. Les variations entre les différentes parties prenantes, à confirmer,

pourraient potentiellement signifier qu'il est difficile de comprendre ce qu'est le domaine sans en faire l'expérience.

Références

1. Abric, J.C.: L'organisation interne des représentations sociales : système central et système périphérique. In: Guimelli, C. (ed.) Structures et transformations des représentations sociales, chap. 2, pp. 73–84. Delachaux et Niestlé, Paris (1994)
2. Abric, J.C.: La recherche du noyau central et de la zone muette des représentations sociales. In: Abric, J.C. (ed.) Méthodes d'étude des représentations sociales, chap. 3, pp. 59–80. Erès, Saint-Agne (2003)
3. Balacheff, N., Lund, K.: Multidisciplinarity vs. multivocality, the case of "learning analytics". In: LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge. pp. 5–13. No. April, Association for Computing Machinery, Leuven, Belgium (2013)
4. Cherigny, F., Kechai, H.E., Iksal, S., Lefevre, M., Labarthe, H., Luengo, V.: L'analytique des apprentissages avec le numérique - Groupes thématiques de la Direction du numérique pour l'Éducation (DNE-TN2). Tech. rep. (2020)
5. Flament, C.: Structure et dynamique des représentations sociales. In: Jodelet, D. (ed.) Les représentations sociales, chap. 9, pp. 224–239. Presses Universitaires de France, Paris, 7 edn. (1989)
6. Franco, M.L.P.B.: Representações Sociais, Ideologia e Desenvolvimento da Consciência. Cadernos de Pesquisa **34**(121), 169–186 (2004)
7. Griffiths, D.: Visions of the Future, Horizon Report. Tech. rep., Learning Analytics Community Exchange (2015)
8. Jodelet, D.: Représentations sociales : phenomènes, concept et théorie. In: Psychologie sociale, chap. 13, pp. 361–382. Presses Universitaires de France, Paris, 7 edn. (1998)
9. Labarthe, H., Luengo, V., Bouchet, F.: Analyzing the relationships between learning analytics, educational data mining and AI for education. In: S. Béland, M. C. Desmarais, N.L. (ed.) Proceedings of Learning Analytics Workshop at ITS. pp. 10–19. [Electronic medium], Montréal (2018)
10. Lo Monaco, G., Piermattéo, A., Rateau, P., Tavani, J.L.: Methods for Studying the Structure of Social Representations: A Critical Review and Agenda for Future Research. Journal for the Theory of Social Behaviour **47**(3), 306–331 (2017)
11. Moliner, P., Abric, J.C.: Central core theory. In: Sammut, G., Andreouli, E., Gaskell, G., Valsiner, J. (eds.) The Cambridge Handbook of Social Representations, chap. 6, pp. 83–95. Cambridge University Press, Cambridge (2015)
12. Papamitsiou, Z., Giannakos, M.N., Ochoa, X.: From Childhood to Maturity: Are We There yet? Mapping the Intellectual Progress in Learning Analytics during the Past Decade. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK'20. pp. 559–568. Association for Computing Machinery, Frankfurt, Germany (2020)
13. Scano, S., Junique, C., Pierre Vergès: Ensemble de programmes permettant l'analyse des évocations. EVOC2000. (2002)
14. Wachelke, J., Wolter, R.: Critérios de construção e relato da análise prototípica para representações sociais. Psicologia: Teoria e Pesquisa **27**(4), 521–526 (12 2011)