

---

# De quoi est-il question dans le discours en art contemporain? La fouille de textes appliquée à l'art contemporain dans les centres d'artistes

**Dominic Forest**

dominic.forest@umontreal.ca  
Université de Montréal, Canada

**Vinh Truong**

congvinhtruong@gmail.com  
Université de Montréal, Canada

**Yvon Lemay**

yvon.lemay@umontreal.ca  
Université de Montréal, Canada

---

## Introduction

Cet article présente les résultats d'une recherche dans le cadre de laquelle nous nous sommes intéressés à l'évolution thématique de l'art contemporain dans les centres d'artistes autogérés du Québec. Nous dressons un portrait du milieu de l'art et de son discours en analysant automatiquement un ensemble de textes d'expositions. Notre approche repose sur des techniques automatisées de fouilles de textes. Notre objectif est donc de cartographier l'évolution des thématiques abordées dans le discours en art contemporain.

Notre approche s'inspire de la notion de *distant reading* développée par Franco Moretti. Selon Moretti, l'analyse des textes devrait être effectuée non pas sur une sélection d'ouvrages (le "canon"), mais sur l'ensemble le plus exhaustif possible de documents. Un échantillon de documents trop limité ou trop spécifique risque d'engendrer une perte d'informations précieuses.

Dans le domaine des arts, K. Bender s'est inspiré des travaux de Moretti afin de développer la théorie du *distant viewing*. Appliqué en histoire de l'art où les monographies se concentrent souvent sur les grands artistes et aux grandes oeuvres, il importe de développer, selon Bender, un "regard" permettant de saisir et

d'analyser un très grand ensemble d'oeuvres d'artistes. Bien qu'il existe des techniques informatiques permettant d'analyser des images et en extraire des mots-clés descriptifs, ces techniques ne font qu'une description visuelle de l'image.

L'image ou l'oeuvre d'art, qu'elle soit figurative ou abstraite, ne se décode pas simplement. L'art visuel nécessite une analyse qui ne se limite pas au niveau descriptif. Interpréter une oeuvre d'art nécessite à priori certaines connaissances, notamment sur le travail et la démarche de l'artiste, le contexte socio-historique de l'oeuvre ou sur l'histoire et la théorie de l'art. Dans ce contexte, le texte d'exposition constitue un médium riche à exploiter pour extraire de l'information sur les oeuvres.

## Le texte d'exposition

Les textes d'exposition en art ont comme objectif de situer et de contextualiser les oeuvres. Ils permettent une meilleure compréhension des oeuvres, du processus créatif de l'artiste, des thèmes, ainsi que des questionnements et des idées évoqués dans les expositions. Dans ce contexte, l'utilisation de techniques de fouille de textes, et plus spécifiquement de modélisation thématique (*Topic Modelling*) (Blei et Lafferty, 2009), permet d'extraire automatiquement les thématiques à partir de textes d'exposition qui n'auraient pu être extraites automatiquement à partir du contenu visuel et qui, à grande échelle, seraient beaucoup difficiles à identifier manuellement. En analysant tous les textes d'expositions disponibles pour une période donnée, nous pouvons poser un regard différent sur le discours en art. L'application de la modélisation thématique sur des textes d'exposition permet, en effet, d'extraire et de structurer automatiquement les principaux thèmes d'un corpus et d'en décrire le vocabulaire spécifique, en plus de permettre la classification d'oeuvres thématiquement comparables. (indépendamment de leur qualités visuelles).

## Modélisation thématique et *Artspeak*

Le langage propre au milieu de l'art contemporain, parfois appelé le *Artspeak* ou *International Art English* (Levine et Rule, 2013), peut être difficile à décoder pour les non-initiés. Il a cependant comme qualité de permettre l'expression d'idées et de concepts abstraits propres à l'art contemporain, afin de lexicaliser la complexité des oeuvres. Comme le mentionne Levine, "IAE is about trying to create a more sensitive language, acknowledging the realities of how things [made by artists] work."

La modélisation thématique la complexité des textes ou au niveau d'abstraction des phrases. Il permet, par simple calcul de distribution des mots, de regrouper des textes portant sur des thématiques communes. La modélisation thématique est fréquemment réalisée par l'algorithme LDA (*Latent Dirichlet Allocation*), développé David M. Blei, Andrew Ng et Michael I. Jordan en 2003. Bien qu'il ait été développé, il y a plusieurs années, cette algorithme fait l'objet d'un nombre croissant d'utilisation dans le traitement des textes en humanités numériques (Graham, Milligan et Weingart, 2016). Le principe qui sous-tend cet algorithme est qu'un corpus de textes est composé de nombreux thèmes ("topics") latents; ces thèmes étant exprimés par les différents mots qui le composent. Les thèmes sont donc des regroupements de mots effectués sur la base d'un calcul de probabilité (bayésienne). Au niveau individuel, un document est composé d'un nombre  $k$  de thèmes, lesquels sont composés d'un nombre  $x$  de mots représentés selon différentes proportions. Au niveau macro (au niveau du corpus en entier), tous les documents d'une collection partagent les mêmes thèmes. L'algorithme LDA pré-suppose donc l'existence des thèmes latents qui n'apparaissent pas explicitement dans les documents. Ce principe évoque ainsi l'idée d'une structure cachée du document qui nous permet de dégager une structure de mots sur la base de leur proximité thématique. Les mots du corpus sont tous regroupés dans différents thèmes, tout en ayant la possibilité de se retrouver dans plus d'un thème à la fois.

## Expérimentation

À partir d'une liste exhaustive de galeries d'art membres du RCAAQ (Regroupement des centres d'artistes autogérés du Québec), nous avons constitué un corpus composé de tous les textes d'expositions disponibles sur les sites Web des centres d'artistes (nous définissons ici "exposition" de manière peu restrictive. Il peut donc s'agir de performances, d'événements hors site ou d'expositions individuelles ou collectives). Un total de 3 867 textes d'expositions ont ainsi été trouvés en ligne, datant de 1973 à 2016 et provenant de 49 centres d'artistes autogérés différents. Les données ont été soumises à un processus de modélisation thématique dans l'application *Mallet*.

Divers prétraitements ont été appliqués au corpus initial afin de n'en retenir que le vocabulaire spécialisé en art (suppression des mots fonctionnels, filtrage par seuils statistiques, etc.). Par la suite, le corpus a été soumis à l'algorithme LDA de manière itérative, tout en modifiant le nombre de thèmes afin d'observer les

différences dans les thèmes extraits. Le défi intrinsèque à cette approche réside dans le choix du nombre optimal de thèmes à extraire. Ce paramètre doit être déterminé manuellement. Nous avons donc analysé un par un les thèmes automatiquement extraits et jugé de leur cohérence, en retournant et en remettant en contexte les termes extraits à l'intérieur des textes originaux. Nous avons finalement opté pour un total de 35 thèmes, nous permettant ainsi d'obtenir un maximum de thèmes cohérents et un minimum de thèmes peu significatifs. Les thèmes ont été étiquetés manuellement à partir des mots les plus représentatifs de chaque regroupement.

## Résultats

Le tableau 1 et la figure 1 présente les résultats générés par l'algorithme de modélisation thématique.

Topic	Mots	% du corpus
Vocabulaire 1	œuvres travail exposition travers artiste manière monde œuvre nature sens	18.09%
Vocabulaire 2	artiste projet travail installation résidence création recherche temps espace pratique	10.61%
Vocabulaire 3	voir fois chose sens œuvre dire forme comment moment semble	9.00%
Vocabulaire 4	artistes exposition art galerie œuvres centre projet pratiques québec collaboration	5.15%
Vocabulaire 5	artiste galerie exposition papier salle dessins dessin œuvre formes installation	4.64%
Vocabulaire 6	espace lieu installation intérieur galerie corps expérience espaces spectateur maison	4.35%
Image	image images regard mouvement réalité vue photographie spectateur photographique vision	3.82%
Sculpture	objets matériaux sculptures objet sculpture éléments installation art formes matière	3.58%
Mise en scène	scène univers artiste culture personnages monde populaire fiction humour fois	3.29%
Vidéo	vidéo film images vidéos cinéma image bande œuvre installation écran	3.16%
Vie	vie monde homme animaux humaine humain mort nature amour animal	3.00%
Photographie	photographie images photographies photographique série photographiques photo exposition photographe portraits	2.94%
Art public	public ville projet espace rue lieu gens montréal quartier intervention	2.65%
Mémoire	histoire mémoire images archives histoires années souvenirs documents temps récits	2.62%
Art sonore	sonore sonores musique sons installation écoute audio système ondes haut	2.28%
Identité culturelle	identité culture politique pays québec guerre politiques société histoire canada	1.88%
Paysage urbain	paysage ville lieux urbain paysages espaces territoire architecture villes sites	1.86%
Peinture	peinture tableaux tableau peintures couleurs peintre motifs couleur pictural motif	1.78%
Nature	paysage terre saint nature québec eau mer nord hiver environnement	1.74%
Corps de la femme	corps femmes peau femme féminin chair vêtements travers propre performance	1.48%
Nouvelles	données internet médias technologies numérique technologie science web technologiques	1.45%
technologies	information	
Vocabulaire 6	arts art montréal canada travail université présenté travaille toronto festival	1.43%
Temporalité	temps eau corps plan moment silence mouvement action instant vitesse	1.42%
Marché de l'art	art valeur politique critique production économie marché artistes artistique conceptuel	1.05%
La textualité	mots texte livre voix textes langage lecture livres langue écriture	0.96%
Lumière	lumière nuit ciel couleur jour obscurité noire boîte blanc bleu	0.93%
Performance	performance performances corps action art actions public vidéo performatif performative	0.92%
Famille	mère père famille fille ans femmes vie enfants grand femme	0.80%
Rêve	rêve lit rêves jardin bête sommeil mais yves proust mimétisme	0.64%
Jeu	sport jeu font goutte artaud siècle joueurs ball jeux carte	0.54%
Junk topic 1	bol ras texte queer instructions populaire maude inhospitalières punk short	0.44%
Films	film guerre films années fast cinéma paris événements politiques musée	0.40%
Junk topic 2	œuvres créatures œuvre terrariums forêt vidéo exposition spectateur ville artiste	0.37%
Voyages	camping tours tour sud visiteurs nord ouest pétrole lisa plaques	0.35%
Junk topic 3	favreau masques machine expe séripop christopher fourrier ve szabo varady	0.35%

Tableau 1. Résultats de la modélisation thématique.

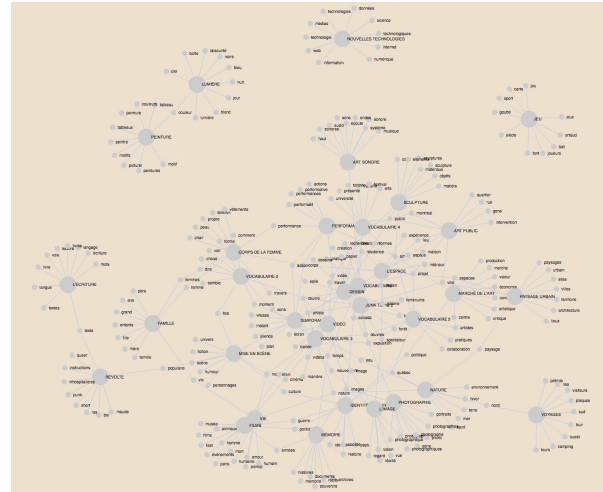


Figure 1. Résultats structurée de la modélisation thématique.

Une analyse des résultats obtenus met en lumière que la notion de thèmes se manifeste de différentes manières dans notre corpus. Ainsi, certains thèmes sont composés de mots-clés spécialisés et très caractéristiques des textes d'expositions (installation résidence création recherche temps espace, par exemple). Par ailleurs, d'autres regroupements thématiques renvoient sans équivoque à une discipline ou à une technique artistique (vidéo film images vidéos cinéma image bande œuvre). Les résultats du modèle thématique obtenu à partir de notre corpus de textes en art contemporain peuvent être synthétisés de la manière suivante :

1. les mots (les thèmes étiquetés « vocabulaire ») couramment utilisés (comme les verbes voir, faire, sentir) dans les textes d'exposition ou dans le discours artistique
2. les thèmes en art contemporain (révolte, famille, mémoire, rêve)
3. les disciplines ou techniques en art (peinture, sculpture, vidéo, photographie)
4. les thèmes incohérents (« junk topics ») (c'est-à-dire les regroupements sans cohérence apparente).

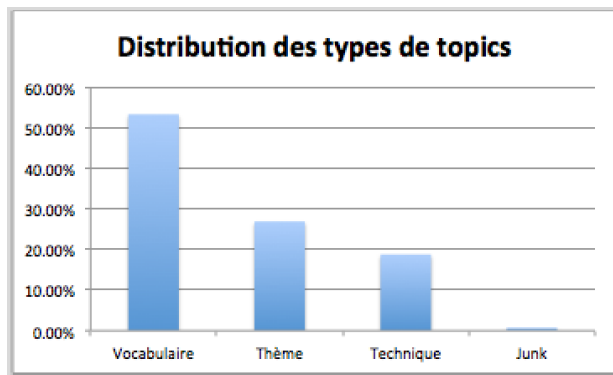


Figure 2. Distribution des types de thèmes.

Nous constatons la prépondérance des thèmes traitant du vocabulaire en art contemporain par rapport aux regroupements par sujets ou par technique (figure 2). En effet, la majorité des documents de notre corpus partagent un vocabulaire de base commun, un lexique propre aux textes d'exposition, constituant en quelque sorte le jargon du milieu. En excluant les mots vides, les textes d'exposition semblent donc composés surtout de mots sur le travail général de l'artiste en art (artiste, exposition, travail, projet, création, etc.). Le reste du corpus est consacré aux thématiques et à la technique.

Nous avons cherché à analyser les résultats du modèle thématique en diachronie afin d'observer l'évolution des thèmes dans le temps. À cette fin, nous n'avons conservé que les documents ayant été publiés depuis les 16 dernières années, puisque nous ne disposons de très peu de documents dont la publication était antérieure à l'an 2 000. Nous présentons ici que certains thèmes pour lesquels une variation chronologique significative a été notée. Ainsi, les thématiques (figure 3) du "marché de l'art" et la "nature" ont fait l'objet d'un traitement croissant, alors qu'il en va inversement pour les thèmes du "rêve", de la "famille", de la "temporalité" et du "corps de la femme".

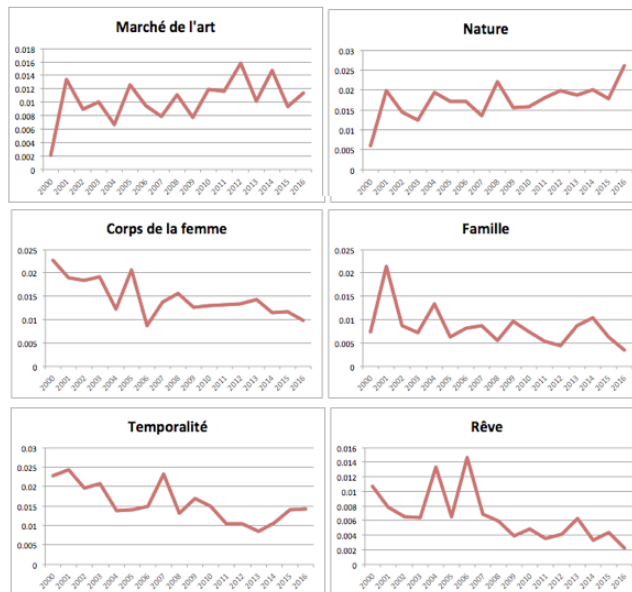


Figure 3. Thèmes à variation chronologique significative.

Les discours sur les techniques telles que la sculpture, l'art sonore ou les technologies numériques sont à la hausse depuis 2 000. Nous constatons que le discours sur la peinture est relativement stable, alors que celui sur la photographie et l'art public sont en baisse après avoir connu un intérêt marqué au début du 21<sup>e</sup> siècle (figure 4).

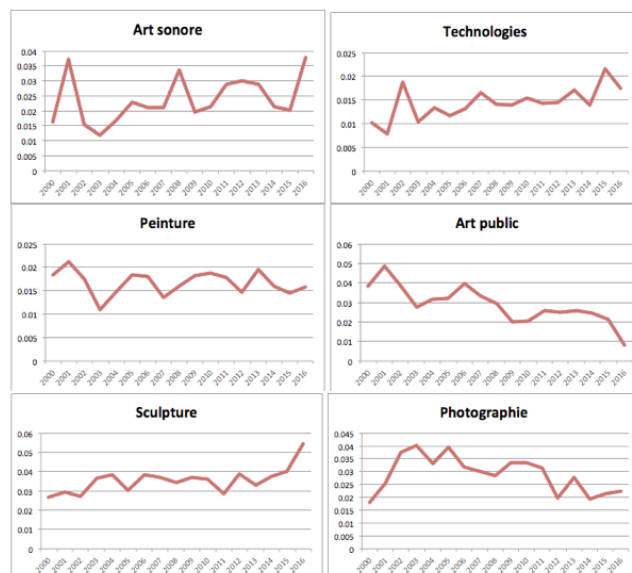


Figure 4. Les techniques à variation chronologique significative.

## Conclusion

Dans le cadre de ce projet, nous avons appliqué la modélisation thématique sur un corpus de textes d'ex-

positions en art contemporain dans les centres d'artistes autogérés du Québec. Les résultats témoignent de l'hétérogénéité des thématiques abordées. Ainsi, certaines thématiques concernent des techniques particulières, alors que d'autres reflètent le vocabulaire spécialisé en art contemporain. Une analyse diachronique des résultats nous a permis de constater l'évolution des thématiques abordées dans les textes produits par les centres d'artistes autogérés du Québec. Cette approche basée sur la notion de "distant reading", appliqué en histoire de l'art, nous a permis d'observer d'un point de vue macro à la fois les tendances d'un milieu artistique et le vocabulaire communément utilisé dans les textes d'exposition.

2016, à [https://www.canopycanopycanopy.com/issues/16/contents/international\\_art\\_english](https://www.canopycanopycanopy.com/issues/16/contents/international_art_english)

## Bibliography

- Bender, K.** (2015). Distant Viewing in Art History. A Case Study of Artistic Productivity. *Journal of Digital Art History*, (1), 101-109.
- Blei, David M.** (2012). Probabilist Topic Models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y. et Jordan, M. I.** (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3), 993-1022.
- Blei, D. M. et Lafferty, J. D.** Topic Models. In *Text mining: classification, clustering, and applications*. CRC Press, pp. 71-93.
- Graham, S., Milligan, I., & Weingart, S. B.** (2016). *Exploring big historical data: The historian's microscope*. Imperial College Press.
- McCallum, Andrew.** (s.d.). MALLET homepage. Machine Learning for Language Toolkit. Repéré 22 avril 2016, à <http://mallet.cs.umass.edu/>
- Réseau art actuel - Répertoire des membres.** (s.d.). Réseau Art Actuel. Repéré 20 mai 2016, à <http://www.rcaaq.org/html/fr/membres/index.php>
- Rhody, L. M.** (2013). Topic Modeling and Figurative Language. *Journal of Digital Humanities*. Repéré 21 mai 2016, à <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/>
- Triple Canopy – International Art English by Alix Rule & David Levine.** (s.d.). Triple Canopy. Repéré 30 octobre