Institut des sciences cognitives

UQÀM

Peindre Magritte avec des mots : rapport d'une analyse conceptuelle dans l'œuvre de Magritte à l'aide d'un corpus de descripteurs sémiotiques

CHARTRAND, Louis

et

MEUNIER, Jean-Guy

Département de philosophie

LES CAHIERS DE L'ISC

Fondé en 2003, l'Institut des sciences cognitives (ISC) de l'UQAM constitue une unité multifacultaire relevant de la Faculté des sciences humaines et de la Faculté des sciences ainsi que multidépartementale. Il vise à favoriser la recherche, le développement de compétences et la diffusion de connaissances dans le domaine des sciences cognitives. L'ISC est constitué de chercheurs couvrant largement de leur domaine d'expertise le champ des sciences cognitives. Il sert de point de rencontre entre les activités de recherche d'équipes reconnues, et s'ouvre aussi à leurs partenaires externes.

«Les Cahiers de l'ISC» sont l'un des canaux de diffusion de l'ISC.

Les cahiers de l'ISC

ISSN 1925-1076 (Imprimé)

ISSN 1925-1084 (En ligne)

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2015

TABLE DES MATIERES

1. CONTEXTE....4

2. PROBLÉMATIQUE....7

3. MÉTHODE....8

3.1 EXTRACTION ET REPRÉSENTATION DES THÈMES....9

3.2 ANALYSE DES DESCRIPTEURS-CONCEPTS....16

4. RÉSULTATS....18

4.1 EXTRACTION ET REPRÉSENTATION DES THÈMES....18

4.2 ANALYSE DES DESCRIPTEURS-CONCEPTS....20

5. DISCUSSION....24

5.1 EXTRACTION ET REPRÉSENTATION DE THÈMES....24

5.2 ANALYSE DES DESCRIPTEURS-CONCEPTS....28

6. CONCLUSION....31

BIBLIOGRAPHIE....32

Peindre Magritte avec des mots : rapport d'une analyse conceptuelle dans l'œuvre de Magritte à l'aide d'un corpus de descripteurs sémiotiques[1]

Version finale – 2013

CHARTRAND, LOUIS[1, 2, 3] ET MEUNIER, JEAN-GUY[1, 2, 3]

LOCHARTRAND@GMAIL.COM

1. DÉPARTEMENT DE PHILOSOPHIE, UQAM
2. LABORATOIRE D'ANALYSE COGNITIVE DE L'INFORMATION, UQAM
3. INSTITUT DES SCIENCES COGNITIVES, UQAM

RÉSUMÉ

Hébert et Trudel (2011) ont conçu une base de données regroupant toutes les œuvres de Magrittes indexées dans Sylvester et al. (1992-1997), qu'ils ont ensuite analysées suivant un protocole sémiotique rigoureux (Trudel et Hébert, 2011) conçu pour cette occasion. Cette analyse a produit un péritexte de descripteurs associés aux tableaux, que nous avons tenté d'évaluer dans le contexte de l'analyse conceptuelle assistée par ordinateur. L'application de techniques de la méthode LATAO (Lecture et Analyse de Texte Assistée par Ordinateur) a permis de trouver jusqu'à six classes stables de tableau à partir du péritexte, parmi lesquelles quatre expriment clairement des thèmes particuliers (la mer, l'arbre, le corps nu de la femme et le visage) qui sont situables dans le temps. Enfin, nous avons fait l'analyse du descripteur-concept « femme », ce qui a confirmé la polarisation visage / corps nu, a fait ressortir l'utilisation quasi-exclusive de termes suggérant la violence (blessure, mutilation) sur des corps de femmes, de même que le contraste entre la nudité des corps de femmes et le caractère « très habillé » des représentations d'hommes. Nous en concluons donc que, pour l'extraction et la représentation de thèmes comme pour l'analyse conceptuelle, le péritexte et la méthode qui l'a créé constituent respectivement une ressource utile ainsi qu'un outil prometteur.

1. Contexte

L'interprétation des images est l'objet d'une longue tradition pluridisciplinaire, pouvant prendre différentes formes, par exemple, en histoire de l'art, en esthétique et sémiotique. Si le paradigme extérieur à la lumière duquel on tente de comprendre ces images varie, l'exercice reste à peu près le même : en restaurant un contexte historique, philosophique ou sémiotique à l'image, on rend son sens. En cela, l'interprétation des images ressemble à l'analyse des textes, telle que pratiquée dans plusieurs disciplines.

[1] Les auteurs tiennent à remercier Jean-François Chartier et Maxime Sainte-Marie pour leurs contributions à travers de nombreuses discussions et pour leur travail de relecture et de correction, de même que les réviseurs anonymes pour la richesse et l'exhaustivité exceptionnelle des commentaires qu'ils nous ont soumis, et qui alimentent toujours notre réflexion.

Depuis quelques années, l'essor de l'informatique offre des outils qui permettent d'assister les techniques d'analyse développées pour l'interprétation des textes et des images. Ces outils participent du renouvellement des méthodes et outils qu'apportent les *humanités digitales*. Ce domaine de recherche comprend à la fois le développement d'outils d'analyse et l'édition de contenu numérique à partir des œuvres du patrimoine culturel. L'édition vise la valorisation du contenu, ce qui signifie qu'on doit produire un contenu dans un format polyvalent qui convienne à une grande variété d'utilisations. C'est pourquoi les éditeurs fournissent de plus en plus des métadonnées sur les œuvres et sur ce qui les composent.

Dans ce contexte, l'image, par rapport au texte, pose un problème particulier : contrairement à ce dernier, son contenu semble inaccessible au traitement automatisé, que ce soit pour des fins de recherche, d'indexation ou de représentation. En effet, si certains outils informatiques permettent facilement de décrire les caractéristiques de bas niveau, c'est-à-dire les caractéristiques plastiques de l'image (couleur, texture, formes 2D ou 3D, ressemblance à une autre image, etc.), ils arrivent difficilement à faire de la reconnaissance de haut niveau, c'est-à-dire de lier ces représentations de bas niveau à des concepts (Hare et al., 2006). C'est ce qu'on appelle le « fossé sémantique » (*semantic gap*).

C'est en vue de contourner cette difficulté qu'Hébert et Trudel (2011) ont conçu une méthode de description rigoureuse (consignée dans le protocole Trudel et Hébert, 2011), et l'ont appliquée au corpus des œuvres du peintre surréaliste belge René Magritte, tel que catalogué par Sylvester et al. (1992-1997)[2]. Des descripteurs sont ainsi associés à 1 870 œuvres – peintures, sculptures, croquis, affiches, etc. Cette interprétation prend la forme de segments de texte, appelés pour cette raison *péritexte*[3]. Celui-ci est enregistré sur une base de donnée avec les métadonnées des tableaux (titre, année, médium, etc.), que les auteurs ont rendue accessible à travers une interface de recherche et de navigation sur le web.

La méthode de Trudel et Hébert n'est pas la première à permettre l'annotation d'images, mais elle trace sa propre voie. D'une part, elle se distingue des « folksonomies » (descriptions par mot-clé produites par les utilisateur·trices d'une base de données) en ceci qu'elle produit des critères et des normes explicites pour la description, et qu'elle compte sur l'expertise d'annotateur·trices qui ont suivi une certaine formation en sémiotique. D'autre part, elle se distingue d'approches plus structurées (e.g. Tam et Leung, 2001) qui attribuent aux descripteurs un rôle dans l'image et les mettent en relation autour d'évènements, d'actes, etc. en ceci qu'elle n'impose pas *a priori* à

[2] Il s'agit là du catalogue le plus complet de l'œuvre de Magritte à ce jour. Cependant, notre base de données n'inclut que les cinq premiers volumes, la parution du sixième volume étant toujours attendue au moment où nous écrivons ces lignes.

[3] Nous disons *péritexte*, car même si les descripteurs sont des lexèmes, ils sont des expressions synthétiques d'une phrase avec un contenu propositionnel. Par exemple les descripteurs « soleil » et « nuage » sont une abréviation de la phrase « Il y a dans le tableau un soleil et des nuages. ».

l'image une structure particulière. Elle allie donc la souplesse des folksonomies avec la rigueur d'une approche structurée.

Contrairement à la plupart des folksonomies, mais à l'instar des approches de description structurée, la méthode Trudel et Hébert compte sur des directives afin de répondre aux prérogatives d'une annotation précise et adéquate. Cependant, là où la description structurée emploie une ontologie rigide à laquelle on ne peut déroger, l'approche Trudel et Hébert produit des critères explicites qui servent de guide aux personnes qui font la description, mais leur laisse plus d'initiative. Parmi ces critères, on retrouve ceux-là :

1. Les descripteurs doivent référencer des signifiés iconiques figuratifs, c'est-à-dire à des entités (objets, actions, activités, phénomènes, etc.) qui se retrouvent dans l'image qui évoquent directement l'un des cinq sens. Cette condition, cependant, ne spécifie pas si l'objet du descripteur tient son sens des conventions du langage ou de celles de l'image : en effet, Magritte aimait à mêler mots et image dans le but de produire un certain effet (cf. Magritte, 1929). Hébert et Trudel ont donc noté le texte dans les œuvres de Magritte en le mettant entre guillemets ; ils ont aussi ajouté certains caractères (« ~ », « ? ») afin d'exprimer le caractère néologique de l'usage d'un mot ou un doute quant à la catégorisation d'un objet. De plus, afin de désambiguïser certains usages, on retrouve souvent, entre parenthèse, un mot qui vient ajouter de l'information sur l'assignation d'un descripteur (e.g. « feuille_(végétal) »).
2. Les signifiés iconiques représentatifs en question doivent avoir une signification qui leur permet de référer à une autre entité dans le monde – c'est-à-dire qu'ils ne peuvent seulement évoquer leur propre forme, texture ou couleur : ils doivent évoquer autre chose. On notera ainsi la présence d'une pipe, mais pas la texture de la peinture.
3. Le critère fondamental de description est la saillance, et il vaut autant pour déterminer si quelque chose dans l'image doit recevoir un descripteur que pour déterminer le ou les descripteurs qu'on devra lui assigner. La saillance sémiotique se définit comme « l'intensité de présence d'une unité dans le champ perceptif ou dans le champ de présence » (Hébert et Dumont-Morin, 2012). Comme le notent Trudel et Hébert, on peut expliquer qu'un objet soit saillant de différentes façons – la « présence remarquable », les procédés ontologiques (taille, nombre, etc.), les procédés rhétoriques, etc. Dès lors, bien que la saillance ne puisse être mesurée directement[4], elle ne se détermine pas non plus sur la base de la seule intuition de l'expert, puisque celui-ci doit pouvoir justifier ses choix des descriptions. Elle n'est donc pas soumise à l'arbitraire d'une perception incorrigible, et on peut espérer que la variation des descriptions, un problème fréquent dans les ontologies, soit ainsi limitée.

[4] Ce qui ne nous empêche pas de croire qu'il est possible que des techniques utilisées en psychologie et neurosciences, comme l'oculométrie, puissent apporter des données empiriques indicatives de la saillance d'une entité ou d'un type d'entité.

Le souci d'en arriver à une description adéquate – complète, mais respectueuse de la structure de l'image – peut se lire, notamment à travers l'insistance sur la saillance. Pourtant, ce n'est pas en soi que le péritexte est intéressant, mais à cause des usages qu'il permet : une liste exhaustive des tableaux avec leurs descripteurs n'est pas une lecture très digeste. Le site web lui-même suggère certaines utilisations, ne serait-ce que par sa conception : celles-ci ont surtout à voir avec la recherche d'information. Nous croyons que le péritexte pourrait avoir de nombreuses autres utilisations, par exemple en rapport avec l'analyse et la représentation de divers aspects du corpus.

Pourrait-on concevoir un autre usage, comme une forme d'analyse assistée par ordinateur? Une telle perspective serait attrayante, car en raison du fossé sémantique mentionné plus haut, de telles analyses sont impossibles sur les images elles-mêmes. De fait, le péritexte étant une description de celles-ci, et sa structure s'apparentant à des abréviations de phrases, on peut s'attendre à ce que les techniques d'analyse lexicométriques aient des chances de fonctionner sur le péritexte.

Dans ce travail, nous nous intéresserons à l'approche de lecture et analyse de texte assistée par ordinateur (LATAO) employée par notre laboratoire, le LANCI (Meunier et al. 2005). Se basant sur l'idée de Firth (1957) qui veut qu'on puisse connaître un mot à son entourage (c'est-à-dire aux mots avec lesquels il apparaît le plus souvent), cette approche considère que les contextes qui sont similaires au niveau du lexique le sont aussi au niveau conceptuel. Dès lors, il est possible de rassembler les contextes similaires sous des thèmes, d'établir des liens entre les mots qui ont des contextes semblables, etc. et de représenter les résultats de façon à souligner ces liens.

D'une part, la LATAO permet ainsi d'explorer un corpus pour y trouver des liens conceptuels qui nous échappent souvent, soit parce que notre contexte culturel nous les rend moins visibles, soit parce que le corpus étudié est si grand qu'il est impossible de le lire attentivement au complet. D'autre part, la LATAO nous permet de faire des mesures empiriques afin de produire des évidences qui peuvent nous aider à confirmer ou à infirmer des hypothèses interprétatives à propos du corpus. Elle vient donc en aide à l'interprétation en fournissant des outils d'exploration qui aident à sa construction, et des outils de justification qui la basent sur des évidences empiriques. Cette approche, adaptée au péritexte, pourrait donc nous permettre de découvrir des aspects inconnus de l'œuvre de Magritte et de rendre l'interprétation de son œuvre, en un certain sens, plus falsifiable.

2. Problématique

Nous avons donc là deux exemples d'utilisation possibles du péritexte, et on pourrait sans doute en trouver d'autres. Mais l'efficacité du péritexte peut varier, aussi faut-il tâcher de l'évaluer.

Dans une recherche préliminaire, Chartrand, Chartier et Meunier (2012) ont évalué le péritexte pour savoir s'il était adéquat pour la recherche d'information telle qu'implémentée dans le site web de la base de données. Ils ont trouvé que le système de recherche dans son ensemble était utile,

mais produisait trop de résultats non-désirés et omettait souvent de produire ceux qui étaient désirés. Ils ont aussi identifié le péritexte comme un facteur important de ces mauvais résultats, car il semblait y avoir une divergence entre le vocabulaire employé par les usager·es dans leurs recherches et le vocabulaire utilisé par le péritexte pour décrire les œuvres. Dès lors, Chartrand et al. concluent que le péritexte n'est pas particulièrement adéquat pour la recherche d'information, et que d'autres méthodes d'annotation pourraient s'avérer plus efficaces à cet effet.

Concernant l'analyse assistée par ordinateur, il est tout à fait possible d'utiliser l'approche LATAO sur le péritexte : adapter le péritexte à un format permettant ce type d'analyse ne pose aucun problème technique. Mais si le péritexte convient à la LATAO par sa forme, qu'en est-il du contenu qui est l'objet d'analyse?

En fait, il est plus que plausible que la maxime de Firth s'applique, dans une mesure ou dans une autre, au vocabulaire pictural décrit par le péritexte : le tableau forme, au même titre qu'une phrase ou un paragraphe, un contexte, et les objets qui s'y trouvent en dépendent intimement. Ainsi, on a peu de chance de trouver un dromadaire dans une nature morte, mais on y retrouvera probablement une pomme et un plateau ou un bol à fruit. Et si on trouve un cheval, il y a de bonnes chances qu'on trouve un cavalier – mais son sens et sa symbolique seront tout autre s'il n'y a pas de cavalier. Les règles d'occurrence et de cooccurrence sur lesquelles repose la LATAO semblent devoir se conserver dans l'utilisation du vocabulaire pictural.

Mieux : *a priori*, le péritexte semble plus adéquat pour la LATAO que pour la recherche d'information. D'une part, puisqu'il n'y a pas de recherche, la divergence avec le vocabulaire qu'emploie spontanément l'usager·e n'est plus un problème – plus besoin que les termes soient exactement les siens ; il suffit que les mots du péritexte soient suffisamment clairs pour être compris. D'autre part, s'il est vrai que la force du péritexte est la description adéquate, cette force est beaucoup plus appréciée dans le contexte de l'analyse, puisque celle-ci a elle aussi comme objectif de représenter le mieux possible certains aspects du corpus. Notre hypothèse est donc que le péritexte est tout à fait adéquat pour une analyse du type LATAO.

Pour vérifier cette hypothèse, nous nous proposons d'appliquer une telle analyse sur le péritexte puis d'en évaluer les résultats.

3. Méthode

On doit d'abord constater que, même dans le cadre plus restreint de la LATAO, les techniques sont beaucoup trop nombreuses pour qu'on les applique toutes. L'expérience accumulée par notre équipe de recherche dans la LATAO nous a suggéré deux sortes d'analyses : une représentation du corpus en général, à travers l'extraction et la représentation des thèmes les plus généraux, et l'analyse de concepts exprimés par un descripteur, qu'on appellera analyse des descripteurs-concepts. L'une et l'autre regroupent en fait plusieurs techniques employées à des fins différentes :

dans le premier cas, extraire et représenter des thèmes qui sont, ensemble, représentatifs de l'ensemble de l'œuvre, et dans le second, analyser des thèmes particuliers choisis par l'interprète.

Notre expérimentation se fera donc en deux parties: (1) extraction et représentation des thèmes et (2) analyse des descripteurs-concepts.

3.1 Extraction et représentation des thèmes

Extraire des thèmes d'un corpus suppose qu'on ait une petite idée de ce que l'on cherche. On conçoit généralement qu'il s'agit de « ce à propos de quoi » une œuvre est (Rimmon-Kenan, 1985). Cependant, comme le note Forest (2002), une telle définition n'est pas suffisamment contraignante pour les fins d'une analyse conceptuelle assistée par ordinateur.

Cependant, lorsqu'on tente de la détailler, on note que cette notion est assez polysémique. Ainsi, certaines définitions en font une sorte de *démonstrandum* – e.g., dans le *Wiktionnaire*, « Sujet, matière, proposition que l'on entreprend de prouver ou d'éclaircir. » ; pour d'autres, c'est une idée sous-jacente à l'œuvre, un principe d'organisation. Pour Rastier (1995), c'est une structure stable d'unités de sens qui se répètent.

Dans l'usage, il est difficile pour une machine de détecter un principe d'organisation ou un concept qui structure un texte ou un tableau, puisqu'elle ne peut en comprendre le sens et le contexte. Cependant, il lui est possible de trouver dans les caractéristiques plus superficielles des œuvres des signes relativement sûrs de la présence d'un thème ou d'un concept.

C'est vrai même lorsqu'un concept n'est pas explicitement nommé. Par exemple, il est certainement possible de parler d'évolution sans utiliser ce mot (Darwin le fait dans *The Origin of Species* ; cf. Sainte-Marie et al. 2011). Il est même possible de changer de vocabulaire d'une fois à l'autre lorsque l'on parle du même concept ou du même évènement (prenons les *Exercices de style* de Queneau, où l'auteur écrit 99 fois la même histoire dans 99 styles différents). Cependant, dans le premier cas de figure, l'absence de mention explicite de la lexicalisation du concept n'empêche pas qu'il soit toujours exprimé par un même vocabulaire ; et dans le second est rarissime car il demande un travail particulier de la part de l'auteur, qui doit réinventer sa façon de lexicaliser son concept à chaque occurrence de celui-ci. Dès lors, on peut faire abstraction de ces cas d'espèce et compter sur la ressemblance du vocabulaire sur plusieurs unités de texte pour indiquer une unité thématique. On peut aussi, pour les mêmes raisons, compter sur la ressemblance du vocabulaire pictural d'un ensemble d'œuvres pour indiquer qu'elles participent d'un même thème.

Il faut cependant retenir que la répétition d'éléments lexicaux est avant tout un *indicateur* de présence d'un thème : la répétition de détails qui appartiennent à une facture stylistique, par exemple, peut se produire même dans des textes traitant de choses très différents. Autrement dit, l'unité de la facture stylistique peut indiquer un thème, mais sa présence n'est en rien une garantie d'unité thématique. Aussi doit-on, dans le cadre d'une LATAO, s'assurer que les régularités identifiées par ordinateur correspondent à des unités thématiques réelles dans le corpus étudié.

Suivant ces considérations, nous adopterons cette définition d'usage : un ensemble de tableaux participe d'un même thème (1) s'ils sont similaires entre eux et dissimilaires aux autres tableaux du corpus[5] quant à leur description dans le péritexte et (2) si les descripteurs fortement associés à cet ensemble ont une cohérence qui permet de supposer une unité ou un principe sous-jacent.

En fait, les deux conditions répondent au même besoin de cohérence : la première nous assure que la cohérence en question est une caractéristique de la structure lexicale du corpus et qu'elle est donc pertinente à celui-ci ; la seconde nous assure que la cohérence perçue dans la structure n'est pas un artefact de la méthode mais qu'elle correspond à un thème détectable. L'idée est donc de détecter les contextes et d'en extraire le vocabulaire propre aux thèmes représentés.

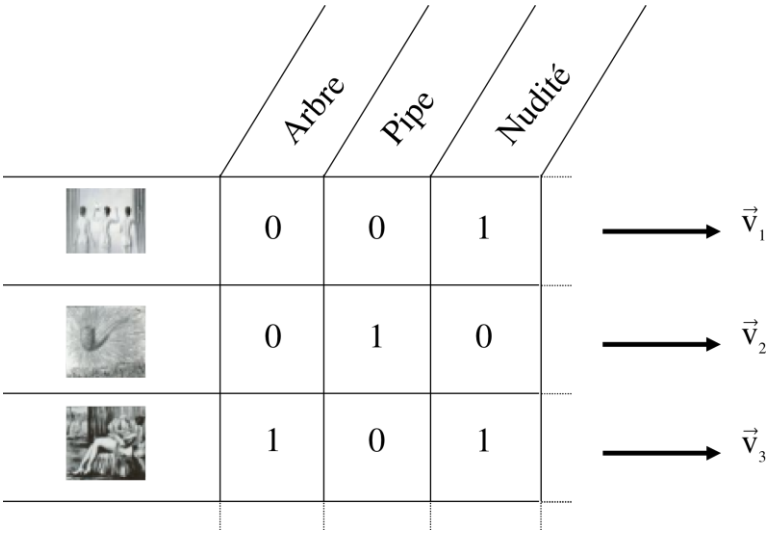


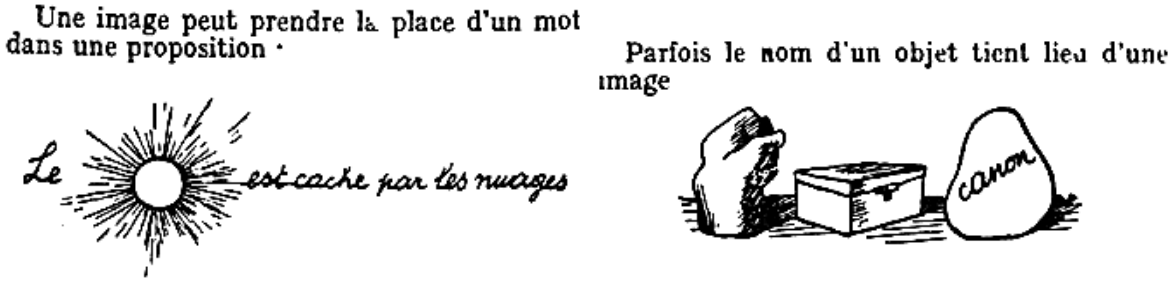
Figure 3.1 : Transformation du péritexte en matrice

Pour ce faire, nous nous sommes inspirés de techniques développées par notre laboratoire (cf. Meunier et al., 2005) et basées sur l'utilisation du modèle vectoriel. Dans celui-ci, chaque descripteur devient une dimension dans un espace vectoriel, et chaque tableau devient un vecteur, dont les coordonnées sont déterminées par la présence ou l'absence d'un descripteur (cf. figure 3.1). On obtient alors une matrice, sur laquelle on peut appliquer diverses opérations mathématiques, par exemple pour décrire les tableaux les uns par rapport aux autres ou pour appliquer des algorithmes de *clustering*.

Notons que tous les descripteurs n'ont pas été utilisés. Ainsi, nous n'avons pas retenu les descripteurs renvoyant à du texte peint sur les surfaces des œuvres du corpus. En effet, Magritte

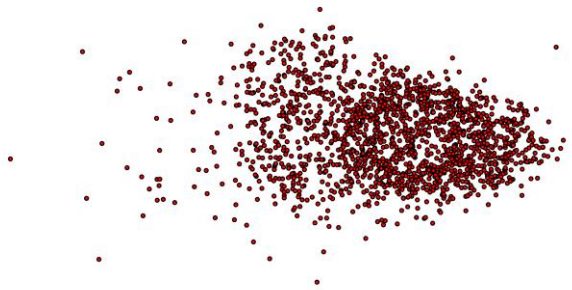
[5] Relatif aux autres tableaux du corpus, car on considère que le œuvres du corpus partagent un mode de représentation et des choix artistiques similaires. Certains artistes ont tendance à se concentrer sur un seul thème à la fois ; d'autres accompagneront leurs œuvres de formes et d'objets qui ne lui sont pas pertinent, de sorte que la ressemblance sera plus forte chez les premiers que les seconds. Cette différence se traduit dans des mesures de similarité différentes, mais cette caractéristique ne pose pas de problème important lorsqu'il y a une certaine uniformité dans les habitudes de l'artiste au travers du corpus. Il s'agit d'ailleurs là d'une condition nécessaire à l'élaboration d'un corpus pour la LATAO (cf. Meunier et al., 2005).

(1929) traite lui-même de la relation entre texte et image dans un article de *La Révolution Surréaliste*, et s'il peut sembler que texte et image puissent souvent jouer le même rôle :



Ils ne sont généralement pas interprétables de la même façon :

Mais surtout, conserver tous les descripteurs aurait probablement pris assez d'espace mémoire pour rendre certaines opérations impossible. Du moins, les opérations qui sont décrites ci-dessous auraient été considérablement plus longues à traiter. Pour ces raisons, nous avons décidé de ne pas utiliser les descripteurs d'objets textuels. Par ailleurs, parce que les représenter sous le modèle vectoriel auraient coûté beaucoup en termes de simplicité du modèle et de computation, nous avons aussi enlevé les marques de doute (« ? ») et de néologisme (« ~ »).



a) Projection Fruchteman-Reingold de la distance Jaccard entre les tableaux

Nbre de partitions (k)	Similarité, hasard	Similarité, *k-means*
2	25%	60,4%
50	0,04%	8,4%
200	0,0025%	2,4%

b) Similarité de Russel-Rao de clusterings *k-means* moyenne sur 10 essais

Figure 3.2 : Homogénéité du corpus et instabilité du clustering automatique

On a donc maintenant une mesure pour évaluer la proportion par laquelle des segments de texte (ici, l'ensemble des descripteurs d'un tableau) portent sur un même thème, ou un même ensemble de thèmes. Cependant, pour des besoins de représentation visuelle qui se clarifieront dans la suite, on a besoin de thèmes discrets. Il nous faut donc tracer quelque part la frontière où débute un thème, au-delà de laquelle les tableaux ne lui appartiennent plus.

L'approche traditionnellement utilisée, et qui est celle de la LATAO, veut qu'on le divise en plusieurs classes à l'aide d'un algorithme de *clustering*, qui produit automatiquement une partition en un nombre déterminé de classes. Or, pour avoir une partition valable, encore faut-il qu'elle reflète la structure du corpus à classifier. Il faut donc que cette structure se porte à la partition – que l'on puisse, pour ainsi dire, la découper à ses articulations naturelles. Le corpus des tableaux de Magritte se porte-t-il à une partition?

Lorsqu'on calcule les distances vectorielles entre chaque paire de tableaux[6] et qu'on les projette sur deux dimensions[7] comme sur la figure 3.2a, on observe un ensemble assez homogène. D'ailleurs, comme le montre le tableau 3.2b, les différentes partitions produites par *k-means* sur notre corpus varient énormément. Il semble donc que ce corpus, du moins comme tel, se porte très mal à la partition, car il n'y a pas d'articulation où l'on peut couper.

On pourrait en déduire que l'extraction et la représentation de thèmes y soit impossible, ou tout au moins méthodologiquement discutable. Cependant, il n'est pas impossible qu'une structure qui permette des partitions apparaisse si on exagère les aspérités dans la structure du corpus. Pour ce faire, nous avons développé une méthode, le *multiclustering*, qui consiste à évaluer la propension qu'ont deux tableaux à être classifiés ensemble par un algorithme de clustering automatique[8]. Cette propension peut être traduite en une distance, qui permet alors de produire une partition avec un algorithme de clustering qui, comme le *k-medoids*[9], travaille à partir des distances plutôt qu'à partir des coordonnées vectorielles.

On peut ensuite évaluer le succès du multiclustering à l'aide d'une projection Fruchteman-Reingold. Si l'opération est probante, on aura alors une projection moins homogène, dans laquelle se

---

6 Par exemple, les distances Jaccard – les mesures sont toutes plus ou moins équivalentes; cf. Albatineh et al. (2006).

7 Utilisant la projection Fruchtemann-Reingold, dont nous utilisons l'implémentation de *NetworkX*, module python par Hagberg et al. (2008). Cette projection utilise un algorithme de forces pour produire une visualisation en 2D d'une structure multidimensionnelle. Chaque point représentant un tableau, on utilise une mesure de distance entre ceux-ci pour déterminer l'attirance ou la répulsion des points entre eux : plus la distance est grande, plus la répulsion l'est, moins l'attirance l'est. Le graphique est le résultat d'une simulation de l'effet de ces forces une fois que le système s'est stabilisé.

8 Pour ce faire, nous avons recourt à la simulation : nous faisons une ou plusieurs partition sur le corpus pour tout k dans [2;N-1], N étant le nombre de segments de texte (ici : de tableaux) dans le corpus.

9 Nous utilisons l'implémentation de la *C Clustering Library*, par de Hoon et al. (2004)

détachent plusieurs amas de points contigus ; sinon, on aura un ensemble plus ou moins uniforme avec un seul groupe important de points.

Par ailleurs, si le *multiclustering* est probant, on peut faire une partition avec *k-medoids.* Mais voilà, cet algorithme demande qu'on spécifie à l'avance le nombre de classes *k* qu'il y aura dans la partition résultante. Aussi, même si on arrive à faire ressortir dans la structure du corpus les articulations où il faut couper, encore faut-il répondre à cette question : combien de classes doit-on produire ?

Pour y répondre, à cette question, il faut d'abord se rappeler que l'on emploie le *clustering* pour faciliter la description du corpus. Le principe de cette méthode remonte au moins jusqu'à Héraclite : pour comprendre un objet, on en distingue les parties et on les décrit. Cependant, toutes les décompositions ne sont pas pertinentes. Suivant l'exemple de Haugeland (1993), si l'on veut comprendre comment fonctionne une télévision, une coupe arbitraire en cubes de 1cm × 1cm ne saurait nous aider. Dans un cube, on pourrait trouver un morceau de transistor, deux condensateurs et demi, et des morceaux de fils qui, déconnectés du réseau, ne mène plus nul part. En revanche, si on découpe aux interfaces entre composantes électriques, on peut établir les liens entre celles-ci et se faire une idée du circuit électronique. Pour la compréhension, il convient donc, selon la formule de Platon, de découper aux articulations.

Même en procédant ainsi, toutes les décompositions ne sont pas égales : si on ne distingue qu'entre le tube cathodique et le reste des composantes, on comprend certes mieux que si on ne faisait aucune distinction, mais pour mieux comprendre le mécanisme de la télévision, on voudra tout de même découper un peu davantage. Ce principe est généralisable : si on découpe plus petit et qu'on fait une description appropriée, qui rend bien compte des interactions entre les parties, on augmente la quantité d'information et notre compréhension de l'objet. En revanche, si on découpe trop petit (par exemple, si on découpait la télévision en atomes), il y aurait beaucoup trop d'information à traiter, et nous n'y comprendrions rien.

Afin de répondre à ces trois contraintes (coupe aux articulations, maximisation du nombre de parties et quantité d'information intelligible), on peut imaginer plusieurs heuristiques. Par exemple, on peut utiliser ce critère : si *k-medoids* est capable de faire une partition sur un corpus, et qu'il peut répéter cette partition autant de fois que nécessaire, c'est qu'il y a une séparation « naturelle » entre les classes de cette partition, et que cette partition est stable. On pourrait donc diviser le corpus en deux, puis en trois, etc. jusqu'à ce qu'il ne soit plus possible d'avoir de partition stable, ou alors jusqu'à ce qu'il y ait trop de classes pour produire une représentation intelligible.

Cependant, dans le cas où la première situation devait se poser avant la seconde – et l'opposé est rarissime, voire inouïe dans la LATAO – on risque de rencontrer un problème pratique. En effet, afin de compenser la variation due à ce que certains de ses paramètres de départ sont fixés au hasard,

on fait généralement plusieurs itérations de *k-medoids*, et on prend le meilleur résultat[10]. Mais évidemment, si on fait 10 000 partitions, on a plus de chances de tomber sur la meilleure partition que si on n'en fait qu'une dizaine. Or il arrive qu'on ne puisse reproduire une partition avec 100 ou 1 000 itérations, mais qu'on puisse le faire avec 10 000 itérations. Mais si ça ne fonctionne pas avec 10 000, doit-on monter encore ? Il est très difficile de déterminer jusqu'à combien d'itérations sont nécessaires pour déterminer si une partition est stable. Et comme chaque itération prend un certain temps à s'accomplir, on risque aussi un temps de calcul indéfiniment long.

Une autre heuristique consiste à profiter de ce que souligne Lewis (2009), à savoir que les humains sont encore les meilleurs lorsqu'il s'agit de trouver de meilleures valeurs de *k*. En effet, lorsqu'on leur demande de déterminer le nombre de classes sur une distribution de points en deux dimensions, les humains ont tendance à converger vers un petit nombre de valeurs. Il est donc logique d'intégrer l'humain dans le processus. Ainsi, on peut demander à l'expérimentateur et à quelques autres personnes de compter le nombre maximum de classes identifiables dans la projection Fruchteman-Reingold des distances entre tableaux obtenues par *multiclustering* – maximum, car, comme on l'a dit plus haut, on cherche à maximiser notre compréhension du corpus.

Dans la mesure où une projection sur deux dimensions de données ayant plus d'un millier de dimensions perd nécessairement beaucoup d'information, on peut se demander si les classes identifiées sur la projection correspondent aux classes qu'on devrait trouver si on était capable de visualiser les vecteurs sans appliquer de méthode de réduction dimensionnelle. Autrement dit, on peut admettre que l'humain puisse, à l'œil, « découper aux articulations », mais comment sait-on que l'on a découpé aux articulations non seulement sur la projection, mais aussi sur les données originales, à dimensions multiples ?

Afin de répondre à cette question, on peut tenter de voir si les partitions que l'humain croit voir lorsqu'elle ou il fixe le nombre de classes correspondent à celles que l'algorithme *k-means* trouve pour le *k* fixé. Pour ce faire, on peut colorer les points selon leur appartenance aux différentes classes sur la projection Fruchteman-Reingold : lorsque les points s'agglutinent autour d'un centre identifié par l'humain en question comme étant une classe, alors on a une confirmation que l'humain sur la projection et la machine sur les données originales ont trouvé à peu près la même classe. Lorsque c'est vrai pour toutes les classes, on a alors une évidence confirmant que le dénombrement des classes par l'humain sur la projection fonctionne aussi sur les données originales. De plus, le fait que le clustering que fait implicitement l'humain pour le dénombrement des classes corresponde à celui de la machine nous donne une raison additionnelle de penser qu'on a découpé aux articulations, et que les classes obtenues reflètent bien la structure du corpus.

[10] Typiquement, l'algorithme choisit la partition la plus compacte. Plus précisément, celle qui a la plus petite erreur quadratique moyenne par rapport au centroïde de la classe.

Une fois la partition produite, on peut tâcher de décrire les classes obtenues. Pour ce faire, nous avons extrait les descripteurs qui étaient les plus associés à chaque classe selon la mesure du χ^2 :

$$\chi^2 = \frac{N\left(N_{11}N_{00} - N_{10}N_{01}\right)}{\left(N_{11}+N_{01}\right)\left(N_{11}+N_{10}\right)\left(N_{10}+N_{00}\right)\left(N_{01}+N_{00}\right)}$$

où *N* est le nombre total de tableaux, N_{11} est le nombre de tableaux de la classe étudiée *C* qui contiennent le descripteur étudié *D*, N_{10} le nombre de tableaux qui contiennent *D* mais ne sont pas dans *C*, N_{01} le nombre de tableaux dans *C* qui ne contiennent pas *D* et N_{00} le nombre de tableaux qui ne sont pas dans *C* et ne contiennent pas *D*.

On prend les six descripteurs les plus associés à chaque classe, et on peut les représenter graphiquement avec une représentation en étoile, où les classes (des cercles dont la taille est proportionnelle au nombre de tableaux qu'elles contiennent) sont reliées aux descripteurs qui les décrivent (le ton de gris des lignes indique la mesure par laquelle les classes sont associées à leurs descripteurs : plus la ligne est foncée, plus l'association est forte). Comme il arrive que ce soit par leur absence que certains descripteurs sont associés à une classe, nous représenterons cette particularité en mettant ces descripteurs entre parenthèse.

Enfin, comme la date de composition de la plupart des tableaux est connue, on peut représenter les dates où les tableaux d'une classe ont été produits avec des histogrammes. Pour faciliter la lecture, on a voulu représenter les années où une classe était plus ou moins présente que les autres. Pour ce faire, on calcule, pour chaque année *a*, la proportion $P_a(C)$ de tableaux qui appartiennent à la classe étudiée *C*, ainsi que la proportion $P_a(\neg C)$ de tableaux qui ne lui appartiennent pas. La valeur *y* pour chaque année est la différence entre ces valeurs :

$$y = P_a\left(C\right) - P_a\left(\neg C\right)$$

En somme, notre méthode pour l'extraction et la représentation de thèmes se résume ainsi :

1. À partir des données vectorielles, regrouper les œuvres corpus en classes :
    a. Exagérer les distances avec la technique de *multiclustering*
    b. À partir d'une projection Fruchteman-Reingold, estimer le nombre de classes
    c. Faire le clustering avec un algorithme *k-medoids*
2. Représenter les classes avec la représentation en étoiles
3. Représenter leur distribution temporelle avec des histogrammes

3.2 Analyse des descripteurs-concepts

Pour la dernière partie de l'expérimentation, la question est plutôt de savoir si on peut approfondir un concept, tel que défini par un descripteur. Pour ce faire, nous pouvons appliquer plusieurs opérations qui peuvent nous apporter des informations pertinentes sur le concept exprimé.

Dans un premier temps, on peut étudier les caractéristiques générales du sous-corpus en question, qui est l'ensemble des tableaux qui sont associés au descripteur que l'on choisit d'étudier. En l'occurrence, nous avons choisi de faire deux mesures qui nous sont relativement familières. Ainsi, une fois déterminé l'ensemble en question, nous représentons la distribution du descripteur-concept dans le temps selon la méthode qui a été décrite à la fin de la présente section. Ensuite, nous calculons le TF-IDF de tous les descripteurs du sous-corpus, et en extrayons ceux dont les valeurs sont les plus importantes. Le TF-IDF d'un descripteur *i* se calcule ainsi:

$$\mathrm{tfidf}_i = \frac{N_i}{\sum_{t \in T} N_t} \times \log \frac{|D|}{|\{d \in D | i \in d\}|}$$

où *D* est l'ensemble des classes, *T* l'ensemble des descripteurs et N_t le nombre d'occurrences d'un terme *t* dans le corpus. Ainsi, d'une part, plus un descripteur est présent dans une classe, plus le TF-IDF est grand, et d'autre part, plus il y a d'occurrences de ce terme dans tous le corpus, plus il baisse. Ainsi, comparativement au χ^2, le TF-IDF favorise les descripteurs qui s'expriment presque exclusivement dans la classe étudiée, même s'ils représentent peu d'occurrences dans celle-ci. On pourrait donc dire que le TF-IDF représente surtout l'importance de la classe pour l'expression du descripteur, davantage que l'inverse.

Une seconde étape consistera à répéter les opérations d'extraction et de représentation de thèmes sur le sous-corpus de tous les tableaux qui contiennent le descripteur étudié, selon la méthode décrite dans la section 3.2. Ce procédé peut nous donner une idée des sous-thèmes qui caractérisent le descripteur-concept.

Enfin, nous produirons un arbre conceptuel en s'inspirant de très près de la technique décrite par Sainte-Marie et al. (2011) dans l'étude du concept d'évolution dans *The Origin of Species* de Darwin. Cette technique a été conçue afin de répondre à une problématique particulière 2011): celle de l'étude d'un concept qui se manifeste dans l'objet étudié, mais sans que le descripteur qui en serait l'expression canonique ne soit mentionné. Dans un tel cas, l'étude du descripteur-concept se révèle insuffisante, de sorte que Sainte-Marie et al. ont dû créer une technique pour représenter un réseau de liens conceptuels qui se rattachent au descripteur concept. Ainsi, ayant observé que le terme « évolution » et ses proches parents « evolve » et « evolved » n'apparaissaient au total que 14 fois dans la 6e édition de *The Origin of Species* et une ou deux fois seulement dans les précédentes, Sainte-Marie et al. en ont déduit que le concept devait se trouver ailleurs, énoncé autrement. Ils ont donc développé une heuristique partant de l'idée selon laquelle le voisinage du

mot canonique doit se retrouver dans des contextes où le concept est énoncé en l'absence de son expression canonique.

Ceci dit, même dans le cas où le concept s'exprime toujours et seulement là où son descripteur se trouve, on peut penser que l'arbre conceptuel, nous renvoyant à des contextes où se trouvent des termes fortement associés au descripteur en question, pourrait enrichir le sens de ces descripteurs-là, et donner une sorte de ramification du sens du descripteur-concept étudié.

La technique se décrit comme suit :

1. Déterminer les seuils s_0 et s_1.
2. Relever l'ensemble *C* de toutes les classes qui contiennent le mot analysé M_0, et où l'indice TF-IDF de M_0 est supérieur au seuil s_0.
3. Relever tous les mots M_i de toutes les classes c_k de l'ensemble *C* dont l'indice TF-IDF dans c_k dépasse le seuil s_1.
    a. Sur le graphe, faire des nœuds pour le mot M_0 et chaque mot M_i, et relier ces derniers à M_0.
4. Pour chaque M_i, relever tous les mots M_{ij} de toutes les classes c_k où l'indice TF-IDF pour M_i dépasse le seuil s_1.
    a. Sur le graphe, faire des nœuds pour chaque mot M_{ij}, et pour chacun, le relier au mot M_i auquel est lié la classe c_k d'où il a été tiré.
5. Répéter l'étape 4 avec les mots M_{ij} obtenus.
    a. Répéter jusqu'à ce qu'on aboutisse à un mot qui a déjà subi le traitement de l'étape 4 ou jusqu'à ce qu'il n'y ait plus de mot auquel on puisse faire subir ce traitement.
6. Souligner (en trait pointillé) les connexions où deux mots se renvoient l'un à l'autre mutuellement

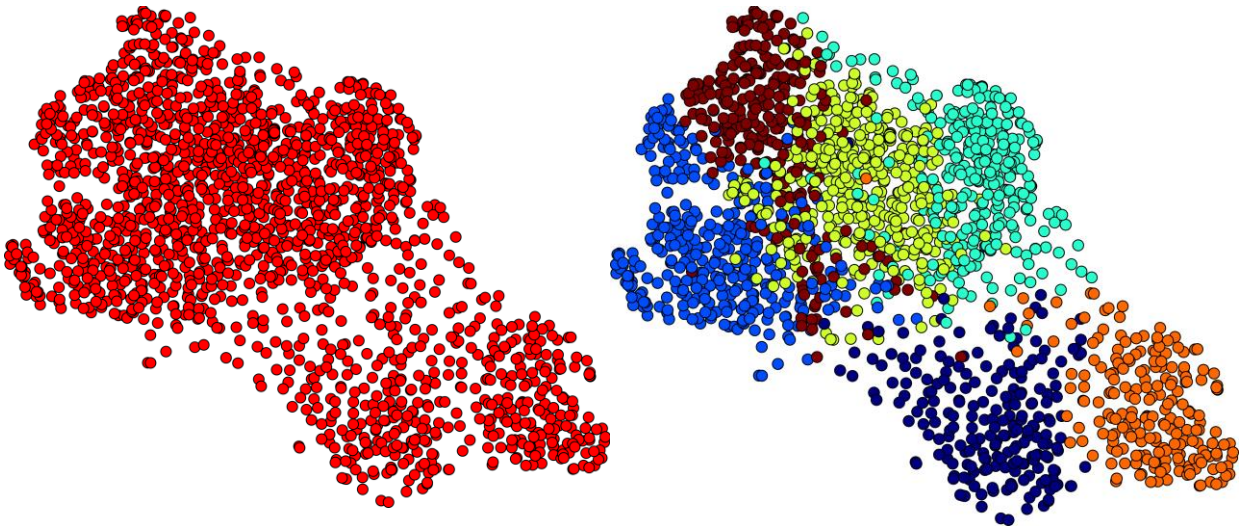
Dans l'algorithme original, s_0 et s_1 étaient fixés à -1, de sorte qu'ils ne jouaient aucun rôle. Cependant, comme cet algorithme a été conçu pour détecter les contextes d'un mot presque absent, il faut ajouter un seuil pour éviter que l'arbre conceptuel ne soit trop fourni au point où la lecture du graphe devient impossible – ce qui arrive souvent avec un descripteur très usité. Dans notre cas, nous avons choisi d'enlever les connexions pour un TF-IDF inférieur à 0,01.

Par ailleurs, il convient de mentionner que la partition sur laquelle se fait l'arbre conceptuel ne saurait être stabilisée par *multiclustering* : le nombre de classes issues d'une telle technique est trop petit, de sorte que les classes seraient trop grandes et exprimeraient un contexte très vague. Dans notre cas, nous appliquons 100 000 fois l'algorithme *k-means* de Jones et al. (2001–) avec $k = 100$ et prenons la meilleure partition. Enfin, nous représentons l'arbre sous la forme d'un graphe dirigé, et nous représenterons la force des associations TF-IDF par le ton de gris des flèches reliant

les descripteurs entre eux. Et lorsque les associations vont dans les deux sens, nous le soulignons l'association avec un large trait pointillé.

4. Résultats

4.1 Extraction et représentation des thèmes



a) Projection simple b) Partition de 6 classes

Figure 4.2 : Projection Fruchteman-Reingold des distances après multiclustering

Concernant l'extraction et la représentation des thèmes, on avait dit que son succès dépendait du succès de la première étape, le multiclustering, qui doit exagérer les distances afin de permettre une partition stable. Comme on peut le voir en figure 4.2a, cette étape nous donne un corpus beaucoup moins homogène, dont on peut penser qu'il puisse produire une partition stable. On peut également observer six sous-ensembles se former : les deux du haut se démarque assez bien, alors qu'en bas, on voit assez bien un centre un peu plus dense, deux autres endroit un peu plus dense en bas et à droite, et on peut croire que l'appendice à gauche se séparerait bien du reste. De fait, avec 100 000 itérations de *k-medoids*, on obtient presque toujours[11] la partition de la figure 4.2b, qui correspond aux partitions qu'on pouvait prédire avec la figure 4.2a.

[11] Faute de temps, nous n'avons pas de calcul exact de cette certitude. Cependant, avec 10 000 itérations, la partition en question revenait environ quatre fois sur cinq, donc nous estimons que cette certitude doit se situer au-delà de 99,99%.

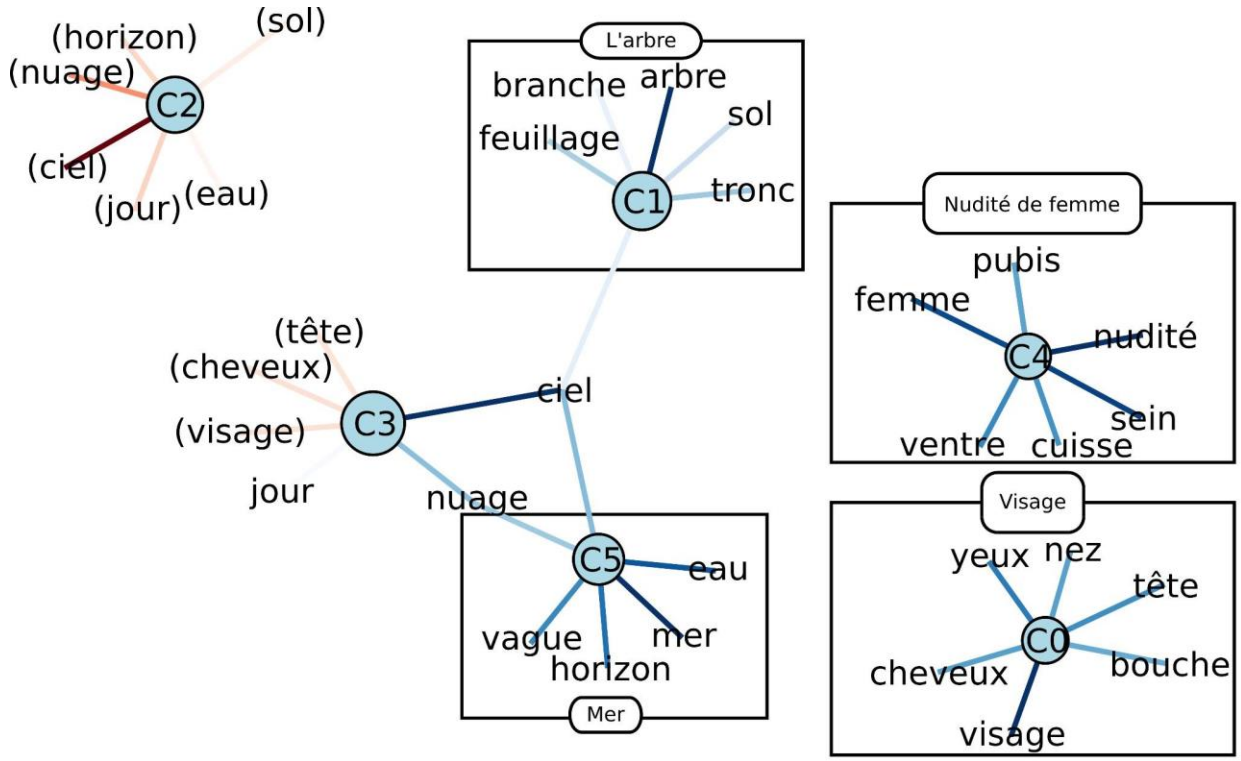


Figure 4.3 : Représentation en étoile

La représentation en étoile est illustrée par la figure 4.3. Rappelons qu'une ligne plus foncée indique un χ^2 plus fort, et que les descripteurs entre parenthèse indiquent une association avec l'absence du descripteur.

Deux classes ne semblent être associées à aucun thème. La classe C2 se démarque par l'absence des descripteurs les plus usités dans le péritexte. Quant à la classe C3, un examen approfondi révèle qu'elle ne semble se caractériser que par la présence de descripteurs omniprésents (ciel, nuage, jour) et l'absence de descripteur associés aux autres classes. Les descripteurs qui leur sont associés ne permettant pas de supposer une cohérence quelconque, on en conclut qu'elles ne remplissent pas la deuxième condition (cf. section 3.2).

En revanche, les autres classes font ressortir des thèmes dominants dans l'œuvre : ainsi, les classes C1 et C5 sont associées à des descripteurs qui appartiennent aux thèmes de l'arbre et de la mer. Les classes C4 et C0, quant à elles, sont fortement associées au corps, alors que l'une se caractérise par la nudité féminine et l'autre par le visage.

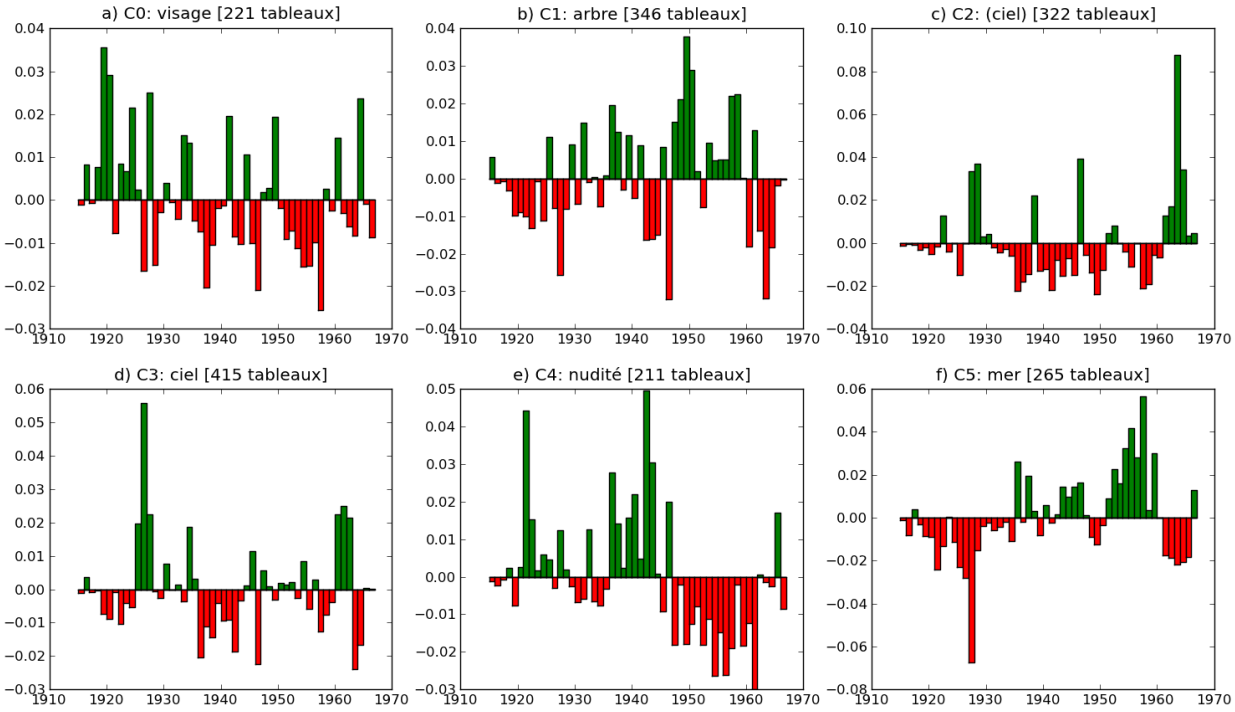
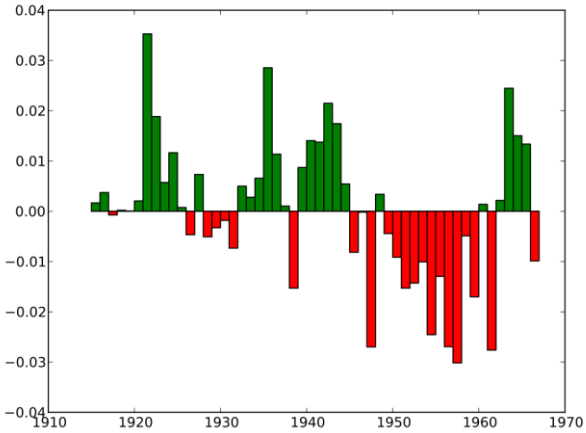


Figure 4.4 : Distribution temporelle de chacune des 6 classes, relativement aux autres

Enfin, la distribution temporelle de ces classes (figure 4.4) nous apprend qu'il y a des tendances remarquables dans la distribution temporelle des classes. Ainsi, le thème de la nudité, très présent jusqu'en 1945, devient soudainement beaucoup moins important, et le thème de « visage » semble lui aussi devenir moins important au fil des années. En revanche, les thèmes de la mer et de l'arbre suivent une tendance inverse, alors que Magritte les utilise de plus en plus jusqu'en 1960.

4.2 Analyse des descripteurs-concepts

Pour cette partie, il nous fallait choisir un descripteur-concept. Pour notre usage, le choix pourrait être arbitraire : l'objectif ici est d'appliquer notre méthode et d'explorer le corpus – n'importe quel choix susceptible de contribuer à la connaissance du corpus et de l'œuvre de Magritte peut donc convenir. En ce sens, notre choix du descripteur « femme » pour la présente analyse est arbitraire.



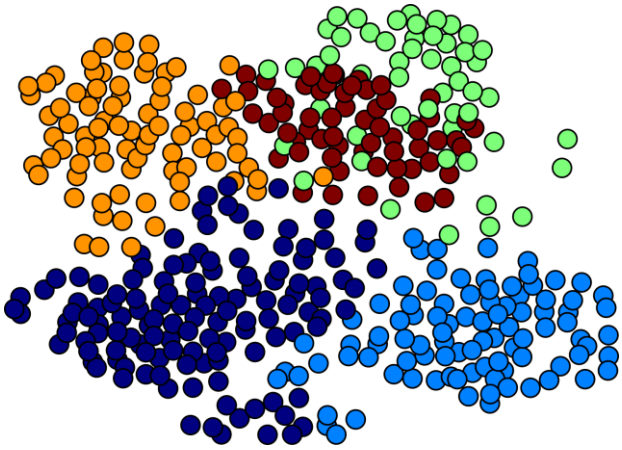
a) Distribution temporelle

femme	0.023408
mutilation	0.002975
décolleté	0.000786
chemisier	0.000786
position_allongée	0.000730
emboîtement	0.000674
bassin_(anatomie)	0.000674
torse_(sculpture)	0.000617
mode	0.000617
jupe	0.000505

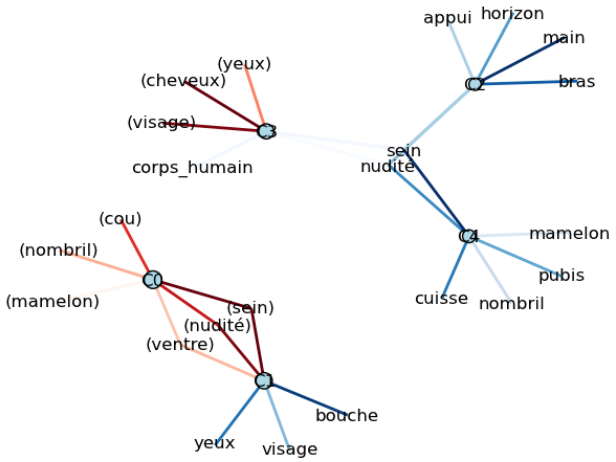
b) Descripteurs ayant les plus haut TF-IDF

Figure 4.5 : Caractéristique du sous-corpus formé par les tableaux associés au descripteur « femme »

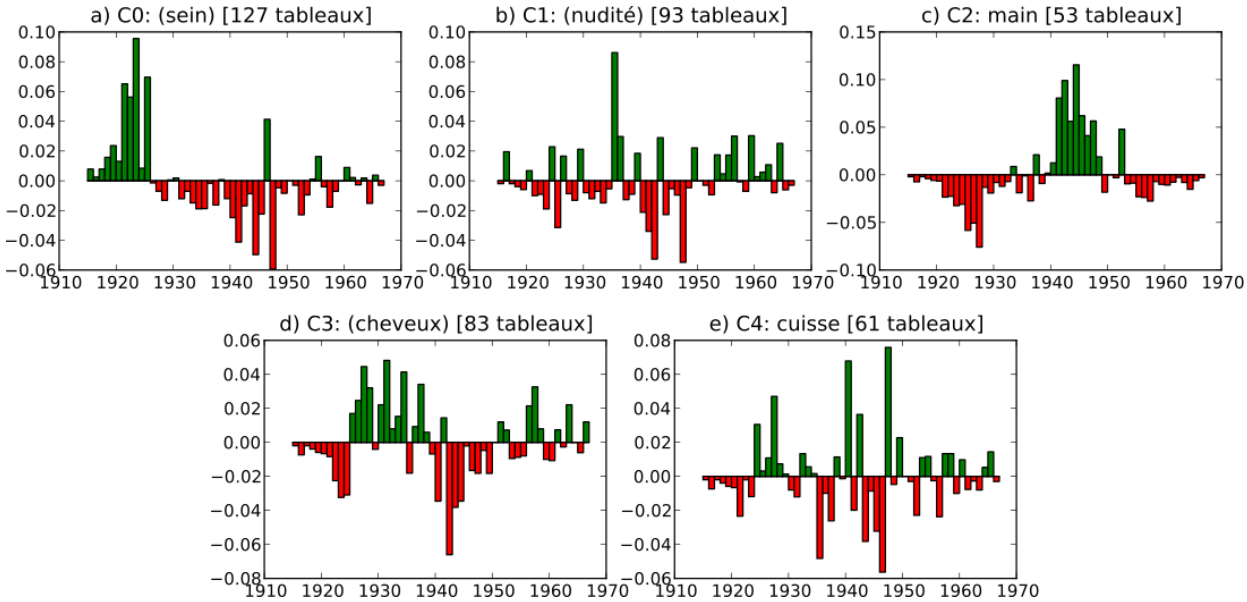
La première étape de l'analyse de descripteurs concepts concernait la représentation des caractéristiques générales du sous-corpus, en l'occurrence la distribution temporelle (figure 4.5a) et les plus hauts TF-IDF associés (tableau 4.5b). Concernant la distribution temporelle, on constate qu'elle est presque identique à celle du thème de la nudité dans la partie précédente. Les TF-IDF surprennent un peu plus, notamment vu le très haut indice qu'on retrouve pour « mutilation ».



a) Projection Fruchteman-Reingold des distances multiclusterisées et de la partition du sous-corpus en cinq classes



b) Représentation en étoile de la partition en cinq classes



c) Distributions temporelles

Figure 4.6: Représentation des thèmes du sous-corpus formé par les tableaux associés au descripteur « femme »

La seconde étape consistait à produire une représentation des thèmes similaire à celle décrite dans la section 3.2, mais seulement pour le sous-corpus étudié. On constate ainsi que si le thème de nudité ressort nettement (cf. descripteurs de la classe C4, figure 4.6b), on retrouve aussi, encore

une fois, le thème du visage (classe C1). S'ajoute un troisième thème, celui du bras et de la main (classe C2). Comme en section 4.2, on retrouve deux classes que l'on ne saurait décrire comme des thèmes. Cependant, sur les distributions temporelles (figure 4.6c), on constate que l'une de ces classes, celle qui se caractérise par l'absence de termes comme « sein », « nudité », etc. se situe surtout au début de son œuvre. Encore sur les distributions temporelles, on constate que les femmes représentées dans les années 40 étaient probablement plus souvent représentées avec leurs bras et leurs mains, puisqu'il y a à cette date une sur-représentation notable de la classe C2.

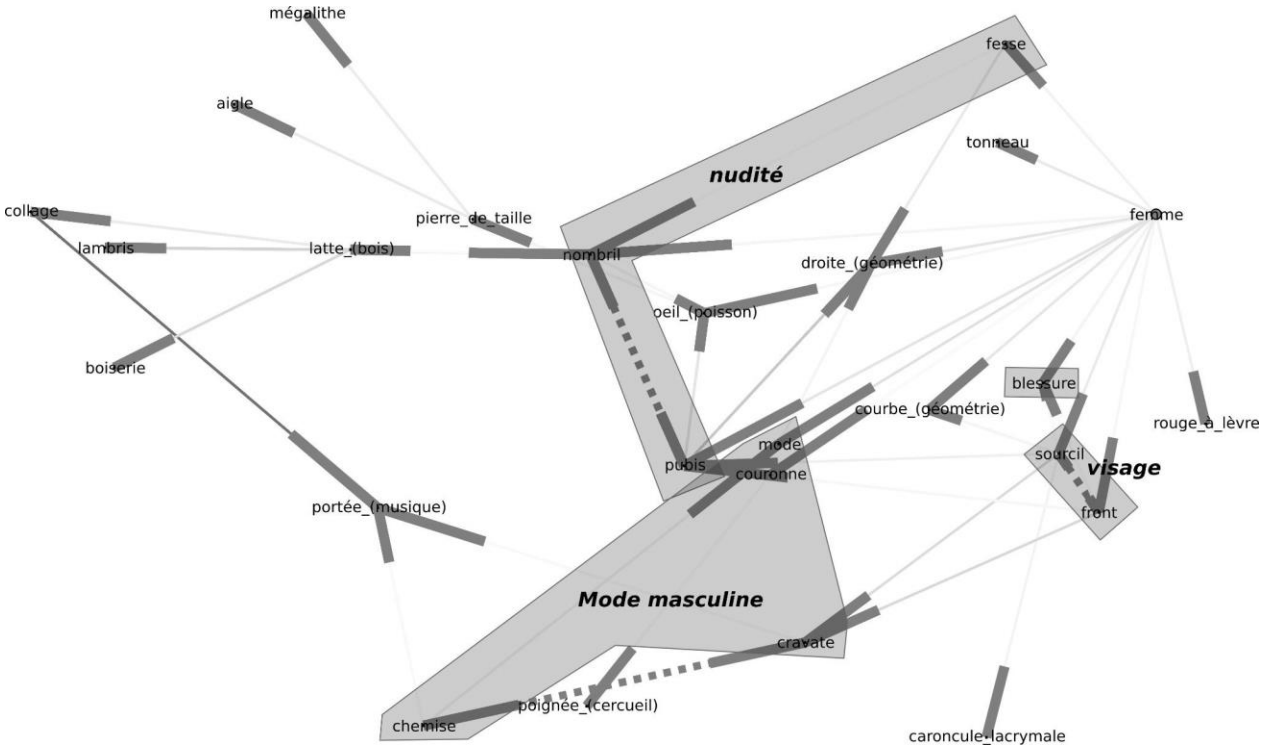


Figure 4.7: Arbre conceptuel pour « femme »

Enfin, nous avons produit l'arbre conceptuel qui est montrée sur la figure 4.7. Rappelons que les lignes plus foncées indiquent un TF-IDF plus élevé, et que les traits pointillés dénotent une association qui va dans les deux sens.

Parmi ce que nous appellerons par commodité les « associations de premier ordre », soit celles qui sont directement associées à « femme », on remarque le retour de thèmes connus, notamment la nudité (fesse, nombril, pubis) et le visage (front, sourcil). On y note également la présence du descripteur « blessure », qui est sans doute à analyser de pair avec celle « mutilations » dans la figure 4.5b. Les descripteurs qui apparaissent dans les associations de second et troisième ordre divergent un peu plus de ce que l'on a déjà vu : on retrouve ainsi des descripteurs liés à la mode masculine (cravate, chemise), ou plus loin, à la pierre (pierre_(matière), mégalithe), le bois (lambris, boiserie, latte_(bois)) et une référence aux collages de partitions de musique qui se trouvent dans quelques œuvres (collage, portée_(musique)).

5. Discussion

5.1 Extraction et représentation de thèmes

Le premier constat que l'on peut faire est que l'étape du *multiclustering* semble avoir fonctionnée (figure 4.2a). D'une part, bien que 100 000 itérations soient nécessaires pour assurer la stabilité du résultat, le seul fait que l'on obtienne un résultat stable suffit à démontrer qu'il y a là un motif réel. D'autre part, que les ensembles reconnaissables à l'œil nu sur la projection Fruchteman-Leibler correspondent aux ensembles trouvés par *k-medoid* (figure 4.2b) nous donne une raison supplémentaire de croire que les motifs obtenus sont appropriés pour caractériser la forme de la structure du corpus.

On note ensuite que sur les six classes obtenues, il y en a deux qui sont à écarter, car les descripteurs qui leur sont associés ne permettent pas de supposer une cohérence quelconque, de sorte qu'elles ne remplissent pas la deuxième condition (cf. section 3.2). Les autres, en revanche, en ont (« mer », « visage », « nudité de femme » et « arbre »).

On a vu en section 3.1 que, pour qu'une classe soit identifiée comme un thème, il faut qu'elle corresponde à un concept valide et identifiable. On peut cependant se demander si les classes évincées ne correspondent pas à un concept valide, mais qui n'est pas identifiable avec notre méthode, et dont l'unité sous-jacente échappe à notre représentation[12]. La chose n'est certes pas impossible – il y a en tout cas une raison pour laquelle *k-means* a produit ces classes. Cependant, il faut se rappeler que la machine et l'humain ne fonctionnent pas de la même façon. Là où l'humain interprète plus volontiers les caractéristiques qui sont présentes, la machine donne autant de poids à l'absence qu'à la présence d'un descripteur. Là où l'humain admet qu'un tableau puisse participer de plusieurs thèmes, notre algorithme découpe l'espace des tableaux en catégories exhaustives et mutuellement exclusives. Notre objectif, en produisant des classes comme on le fait, n'est donc pas d'aller chercher tous les thèmes ni de ne produire que des classes qui correspondent à des thèmes, mais plutôt de produire une décomposition qui fait ressortir certains thèmes que l'on peut à leur tour examiner. L'objectif en est un d'exploration et de fouille, non pas de description. En ce sens, d'une part, il est normal que plusieurs classes ne produisent pas de thèmes, et d'autre part, il est acceptable que l'on choisisse de ne pas approfondir certaines classes dont le principe n'est pas évident, pour des raisons d'économie de temps et d'énergie.

Ceci dit, nous trouvons tout de même plusieurs classes qui, elles, semblent à première vue posséder les qualités nécessaires à produire des thèmes. Ces classes, toutes représentatives qu'elles soient de la structure du corpus, sont-elles représentatives de son contenu ?

Une critique que l'on pourrait ici formuler vient d'un biais qu'on pourrait qualifier de « méréologique ». On observe que certains objets, comme le visage, semblent se décomposer en

[12] C'est l'inquiétude qu'a soulevée un réviseur anonyme.

d'autres objets : le nez, la bouche, etc. et s'accompagnent d'autres objets qui ne peuvent manquer qu'à la suite d'une opération un peu dramatique, comme les cheveux, le cou, et souvent d'autres parties du corps. Que des groupes des mots qui sont tous des parties du visage ou de la tête soient toujours ensemble, il n'y a là rien de surprenant. Mais le fait qu'un objet comme le visage soit composé de tant de parties reconnaissables et nommables crée un biais qui fait que les tableaux où sont dessinés des visages ont plus de chances de se retrouver ensemble que des tableaux qui contiennent des pommes, qui sont décomposables en peu d'éléments.

Cependant, il n'est pas accidentel que l'artiste ait mis assez de soin pour rendre saillant des parties d'objet, et il n'est pas accidentel non plus que nous ayons des mots pour désigner ces parties. Pour reprendre l'exemple du visage, les études en oculométrie montrent depuis les années 30 (Buswell, 1935) que lorsqu'une œuvre présente des visages, le regard s'y porte naturellement. Pas surprenant alors que leur représentation et notre langage reflètent cette saillance. On peut même arguer que ce « biais méréologique » en est un qui rend le péritexte plus fidèle encore à notre perception des tableaux.

Passons à l'analyse des classes elles-mêmes – on veut savoir si elles correspondent bien à une thématique abordée par l'artiste, condition nécessaire à ce que la classe forme un thème.

On peut déjà commencer par les classes dont nous avons dit plus haut qu'elles ne semblent pas manifester de thème particulier, selon les critères mentionnés en section 3.1. La classe C2 (figure 4.3), par exemple, est caractérisée par des associations négatives, c'est-à-dire par l'absence de caractérisation à certains descripteurs. Ceux-ci ont ceci en commun de pouvoir faire partie du fond d'un tableau– le ciel, l'horizon, les nuages, etc. L'absence de ces descripteurs peut vouloir dire quelque chose pour un algorithme de classement, mais pour un humain, elle est beaucoup trop ambiguë – elle peut signifier que ce qui est représenté est dans un intérieur sans fenêtre, dans un espace abstrait, ou même simplement que l'artiste n'a pas pris la peine de faire un fond. On peut en dire autant de la classe C3 (figure 4.3) qui semble quant à elle caractérisée non pas par la présence d'un fond, mais par l'absence de têtes de personnes – difficile de dire à quoi peut correspondre cette classe.

Continuant avec les autres classes, la nudité de la femme (classe C4, figure 4.3), manifestement, attire un regard concupiscent, au moins de la part de l'artiste. En effet, Magritte lui-même admettait que l'érotisme était un thème majeur dans son œuvre, allant jusqu'à louer Freud pour avoir défendu que « l'activité artistique n'est qu'une autre activité de substitution à l'activité sexuelle » (Saint-Martin, 2010, 67-70).

a) Les Amants (1926)

b) Le fils de l'homme (1964)

c) La reproduction interdite (1937)

d) Les regards perdus (1927/1928)

e) L'assassin menacé (1926)

Figure 5.2 : Visages chez Magritte

De fait, les thèmes découverts ne sont pas sans précédents dans la littérature, même si il arrive qu'ils soient oubliés. Saint-Martin s'étonne que, malgré la longue liste de ceux et celles qui reconnaissent la prépondérance de la sexualité chez Magritte, les commentateurs évitent de souligner le thème et se parent de précautions excessives pour l'aborder : « On croirait se retrouver devant un canular d'étudiant, visant à ridiculiser le pansexualisme ! ».

Le visage est un cas plus étrange, car il est plus souvent abordé pour son absence, alors que Magritte aime à le cacher. On a souvent dit que *Les Amants* (figure 5.2a) renvoyaient au drame du suicide de sa mère, car celle-ci, selon Scutenaire (qui était un ami de Magritte), aurait été retrouvée avec un tissu sur la tête (Viederman, 1987). Cependant, Magritte utilise d'autres procédés pour cacher les visages, comme en témoignent *Le fils de l'homme* (figure 5.2b) et *La reproduction interdite* (figure 5.2c). Par ailleurs, lorsqu'il montre des visages, il tend à employer des procédés pour les rendre plus saillant et plus reconnaissables – par exemple, dans *Les regards perdus* (figure 5.2d), où le visage est mis en évidence par un redoublement.

Certes, le visage voilé peut évoquer le suicide de la mère, mais l'effet d'un visage caché tient surtout de sa saillance. Comme l'a montré Buswell (1935), le regard se porte en premier sur le visage ; lorsqu'il s'avère absent, le choc de cette absence est plus grand que pour un objet moins saillant. De plus, comme le notent Graziano et Kastner (2011) à partir de recherches précédentes, on a tendance à attribuer la conscience à un personnage à partir de son attitude face à une situation, et tout particulièrement en regardant les yeux et l'expression faciale. Cacher le visage correspond donc à une forme de déshumanisation, qui est un trope très usité chez Magritte.

Le soin apporté à ce que les autres visages soient visibles participe souvent de la même préoccupation. Le fait que les personnages de Magritte dirigent généralement leur regard hors champs, et fixent rarement le centre d'intérêt du tableau, leur donne l'apparence d'être insensibles au contexte dans lequel ils se trouvent. Or c'est un genre de situation où, selon Graziano et Kastner, le cerveau refuse de reconnaître ce qui est présenté comme un être conscient. Même dans les situations les plus dramatiques (e.g. *L'assassin menacé*, figure 5.2e), les personnages gardent une expression absente. Dans le but de déshumaniser les figures, il peut sembler important de rendre les visages saillants afin d'éviter qu'un spectateur charitable ne prête à un visage moins visible une expression plus appropriée au contexte. En ce sens, le véritable thème du visage est peut-être, paradoxalement, celui de la déshumanisation (cf. Hauck, 2012); en tout cas, il s'agit bien d'un thème réel, pervasif chez Magritte, et non pas du fruit d'un artefact.

L'arbre, quant à lui, est l'objet ou l'outil de toutes sortes de métamorphose. Il n'est pas seul en cela, mais le vocabulaire de la nature prend une place importante chez Magritte. De fait, le thème de l'animal chez Magritte a récemment reçu une certaine attention (cf. Trudel, 2012; Robinson, 2012), mais Magritte exploite encore davantage les représentations de végétaux – par exemple, dans les métamorphoses et hybridations qu'il fait subir aux objets. C'est d'ailleurs peut-être parce qu'il se porte bien à toutes sortes de jeux et de manipulations que l'arbre est la figure la plus présente d'un ensemble d'objets qui inspire considérablement Magritte.

Enfin, on parle assez souvent de la mer en référence au traumatisme du suicide de la mère (e.g. Viederman, 1987), en invoquant le rapport symbolique entre la mer et la maternité, présent dans de nombreuses cultures, l'homonymie entre « mère » et « mer » en français, et, bien sûr, le fait qu'elle se soit noyée. Nous reviendrons plus loin sur le fait que les tableaux participant de ce thème se trouvent davantage en fin de carrière ; pour le moment, il suffit de reconnaître que le thème en question a été abordé dans la littérature, témoignant de son importance dans l'œuvre de Magritte.

Bref, il nous semble que nous avons de bonnes raisons de croire que les quatre thèmes sont « réels », et l'hypothèse selon laquelle ils constitueraient des artefacts de la méthode nous semble la moins probable.

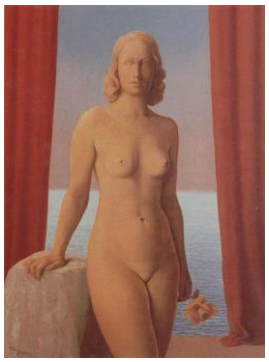
5.2 Analyse des descripteurs-concepts

La première chose qui transparaît de nos résultats provenant de l'analyse des descripteurs-concepts est la confirmation de certains éléments qu'a pu montrer la représentation des thèmes. Ainsi, comme que l'on aurait pu s'y attendre, la distribution temporelle de « femme » (figure 4.6b) est presque identique à celle du thème « nudité de femme », la représentation en étoile (figure 4.3) de la partition fait ressortir assez nettement les thèmes de nudité et de visage, et l'arbre conceptuel (figure 4.7) fait ressortir des descripteurs des thèmes du visage et de la nudité. La première et la troisième observation vient confirmer l'hypothèse selon laquelle les femmes sont presque toujours représentées nues chez Magritte, alors que la seconde montre que le sous-corpus formé par « femme » a une structure relativement continue à celle du corpus de toutes les œuvres.

Aux deux thèmes mentionnés s'ajoute le thème de « main », qui laisse supposer que les femmes peintes dans les années 40 le sont dans des postures différentes, qui laissent voir leurs mains. C'est ce que semble confirmer l'observation des tableaux. Il est, à vrai dire, assez intéressant de constater que cette période est transitoire : dès 1946, on sent que les images de femmes deviennent moins présentes dans les œuvres de Magritte (figure 4.5a), et cette tendance s'accentue dans les années 50. Parallèlement, les thèmes « arbre » et « mer » deviennent plus importants; le dernier tout particulièrement, qui domine dans les années 50 au moment où Magritte peint beaucoup moins de femmes.

Les interprètes font souvent référence au suicide de la mère de Magritte, qui aurait causé un traumatisme qui aurait suivi ce dernier toute sa carrière, influençant considérablement son œuvre (e.g. Viederman, 1987; Saint-Martin, 2010; Gubar, 1987). Bien que Viederman (1987) prétende que l'œuvre de Magritte a commencé à refléter ce traumatisme à la période où Magritte a développé son propre style (1926-1930), personne, à notre connaissance, ne semble avoir noté d'évolution dans la façon dont il s'est exprimé.

a) L'âge du plaisir (1946)



b) La fleur du mal (1946)

c) La mémoire (1945)

Figure 5.3 : Visages de femmes des œuvres des années 40

Bien sûr, Magritte avait pour habitude de jouer avec les dates de ses tableaux, de sorte qu'il peut être difficile d'interpréter l'évolution du style de Magritte à travers le temps. Cependant, cet état de fait, bien qu'il doive informer notre interprétation, ne saurait disqualifier l'analyse des tendances observées dans le temps. De fait, si tendance il y a – si la représentation d'un tableau dévie suffisamment de ce que donnerait le hasard pour qu'on puisse considérer cette différence comme significative – on peut la comprendre de deux façons. Ainsi, il se peut que les manipulations que Magritte a fait sur les dates soient trop rares ou décalent trop peu les œuvres de leur date réelle pour que le bruit ainsi induit ne vienne briser la tendance dans les résultats. Il se peut aussi que Magritte ait pipé les dés de façon à faire ressortir une interprétation particulière.

Le deuxième cas nous semble improbable : Magritte, bien qu'il aimait voir les gens tenter d'interpréter ses œuvres, n'avait pas beaucoup de sympathie pour les interprétations finies. Il encourageait un processus interprétatif continu, et ses interventions visaient généralement à éviter qu'il se stabilise sur une conclusion. En ce sens, s'il a manipulé les dates de ses œuvres, c'est presque certainement pour éviter qu'on ne donne d'étiquette trop rigide à une période ou à une autre. Pour comprendre des phénomènes comme l'apparition des mains dans les tableaux des années 40 et le désintéressement du peintre pour la représentation des corps de femmes dans les années qui s'en suivent, nous croyons donc que la première stance interprétative est à préférer en l'absence d'indices suggérant le contraire.

De fait, on commence à voir avec le début de la Seconde Guerre mondiale l'apparition de toiles où figurent des personnages, et des femmes en particulier, qui sont impliqués dans leur contexte. On imagine mal le Magritte des années formatives, tout occupé qu'il était à objectiver tous ses sujets, dessiner une femme visiblement amoureuse, comme dans *L'âge du plaisir* (figure 5.3a). Même lorsqu'il fait usage de ses procédés usuels de déshumanisation, comme la pétrification dans *La fleur du mal* (figure 5.3b), il joue sur l'intelligence d'une expression pétrifiée pour déjouer l'effet même de ce procédé et nous plonger dans le malaise d'être observé par quelque chose qui ne devrait pas être humain.

Dans ces circonstances, il n'est pas banal que les femmes commencent à être dessinées avec leurs mains. En effet, elles sont l'organe par lequel on interagit avec notre environnement, et ainsi, elles traduisent une bonne partie de l'attitude qu'on peut avoir envers notre environnement. Dès lors, elles étaient superflu et détournaient l'attention lorsque Magritte tentait de représenter des personnages inanimés et « plastifiés », mais lorsqu'il veut faire paraître le contraire, elles lui permettent de manifester l'attitude du personnage face son environnement. Les mains sont un moyen, pour Magritte, de montrer que son sujet a une âme.

Par la suite, Magritte se désintéresse de peindre des corps de femmes, et s'intéresse plutôt à des thèmes naturels, comme l'arbre et la mer. Est-ce qu'il aurait fait la paix avec le suicide de sa mère, et en serait venu à appliquer sa méthode de déconstruction du réel et de reconstruction du mystère sur l'ensemble du vivant? Ou, au contraire, se serait-il soudainement senti menacé par ces

femmes animées qu'il en était venu à peindre, et en serait-il venu à chasser toute vie de ses tableaux, jusqu'à tenter de domestiquer la nature elle-même pour qu'elle ne devienne qu'un objet? On pourrait aussi, plus simplement, attribuer le besoin de rendre un peu d'humanité aux sujets de ses tableaux à une réaction à la guerre ou à l'occupation nazie. Qu'importe – notre rôle ici n'est pas de déterminer les raisons pour lesquelles ce phénomène survient, mais plutôt de se donner les moyens de le décrire.

En poursuivant l'analyse du descripteur-concept, on réalise que le terme avec l'indice TF-IDF le plus haut, après « femme » lui-même, est « mutilation » (figure 4.5b) – ce qui signifie que ce procédé est appliqué souvent aux femmes, mais jamais à d'autres sujets. « Mutilation » n'est d'ailleurs pas le seul descripteur à suggérer la violence qui est faite aux corps de femmes dans l'œuvre de Magritte – comme nous l'avons noté plus haut, l'arbre conceptuel de la figure 4.7 indique que le mot « blessure » est fortement associé à « femme », et correspond généralement à une blessure sanguinolente à la tête comme dans *La mémoire* (figure 5.3c). La violence appliquée aux corps de femmes par le biais de mutilations et de représentation de blessures peut se lire, par exemple, comme une réponse au traumatisme de la mort de sa mère (e.g. parce qu'à un certain niveau il se voyait comme le meurtrier de sa mère – cf. Viederman, 1987), ou alors comme un schème, un trope ou une attitude participant d'une culture plus large (e.g. culture pornographique), ou bien même comme une manifestation des deux (e.g. Gubar, 1987).

Enfin, on a mentionné en section 4.3 que l'arbre conceptuel (figure 4.7) renvoyait à d'autres ensembles de descripteurs qui évoquent certains thèmes, au sens large du terme. Difficile de déterminer le lien que le bois ou la musique peuvent avoir avec les femmes chez Magritte, mais le thème du vêtement masculin évoque un contraste intéressant entre la nudité des femmes et l'habillement des hommes. Certes, l'habit permet à Magritte de rendre les hommes interchangeables, et ainsi de les déshumaniser, mais comme ces habits renvoient à une certaine position sociale, cette déshumanisation ne rend pas les hommes vulnérables. Cet état de fait contraste fortement avec la vulnérabilité de la femme nue, et ce contraste est évoqué dans l'opposition homme (très) habillé/femme nue qui se dessine dans l'arbre conceptuel.

Il faut noter que l'arbre conceptuel est un peu différent des autres outils employés ici: il repose en effet sur un clustering en un nombre de classes *k* très large, et la partition qui en résulte ne saurait donc être stable. En ce sens, l'arbre conceptuel n'est pas apte à représenter un concept dans un corpus ; aussi l'utilise-t-on plutôt pour sonder l'espace des liens qu'entretiennent les concepts entre eux. Alors que les distributions temporelles, les classements TF-IDFs et les représentations en étoile basées sur des classes stables peuvent avoir valeur d'évidence dans l'analyse conceptuelle, les arbres conceptuels ne sauraient qu'assister la découverte, car les termes représentés pourraient ne l'être qu'à cause d'un artefact de la méthode – par exemple, d'une variable d'initialisation déterminée au hasard.

Nous pouvons donc conclure deux choses. D'une part, les méthodes d'analyse conceptuelle appliquées ici nous ont permis de d'appuyer l'opinion de celles et ceux qui voient dans la représentation que Magritte fait des femmes une certaines violence, en montrant la forte association avec « blessure » et surtout « mutilation » ; elles ont aussi permis de découvrir quelque chose de nouveau sur l'évolution de la façon qu'a Magritte de représenter les femmes, passant par une phase de « réhumanisation » de celles-ci avant de s'en désintéresser progressivement à partir de 1946. On peut donc dire que l'analyse conceptuelle des descripteurs-concepts donne des fruits. D'autre part, l'arbre conceptuel laisse miroiter des horizons d'analyse que nous n'avons pas étudiés : on peut donc croire que le péritexte pourrait permettre de mieux comprendre le concept étudié avec l'aide de techniques d'analyse plus poussées.

6. Conclusion

Notre question était de savoir si, à l'aide des technologies d'analyse d'information par ordinateur, le péritexte pouvait remplir des fonctions de d'analyse, de représentation et de recherche d'information.

Concernant la recherche d'information, Chartrand, Chartier et Meunier (2012) ont montré que le système de recherche (compris comme le péritexte + le logiciel de recherche et son implémentation) est pénible à utiliser. En revanche, il semble que le péritexte soit bien adapté à des activités de forage conceptuel, comme l'extraction et la représentation des thèmes et l'analyse de descripteurs-concepts. En effet, notre méthode d'extraction des thèmes nous a donné quatre thèmes qui représentent bien une bonne partie du corpus et qui rendent quelques unes des idées essentielles sous-jacentes à l'œuvre de Magritte. L'analyse, de son côté, nous a permis de consolider certaines intuitions, comme celle de la violence dans la représentation des femmes, et a permis de découvrir des aspects moins connus du concept étudié, par exemple les tentatives de Magritte dans les années 40 de représenter des femmes plus expressives.

Certes, ces méthodes n'ont pas épuisé leur potentiel, mais les résultats obtenus sont encourageants, en ce sens qu'ils dévoilent des caractéristiques du corpus qui sont tout à fait pertinentes à l'étude des œuvres de Magritte.

Notre équipe a souvent insisté sur le fait que (e.g. Meunier et al., 2005) que les techniques d'analyse conceptuelle assistée par ordinateur sont des outils dans les mains du chercheur. La plupart des outils développés jusqu'à maintenant (Meunier et al., 2005 ; Sainte-Marie et al., 2011) étaient surtout destinés à aider le chercheur dans la découverte d'éléments pertinents à l'analyse – en cela, ils participaient du contexte de découverte. Certains outils développés pour cette étude – les distributions temporelles et l'heuristique pour former des partitions stables pour la représentation en étoile – nous donnent en revanche une représentation robuste et stable de certains aspects du péritexte. On peut donc concevoir un usage de ces méthodes pour confirmer ou

infirmer des hypothèses sur le corpus étudié : l'analyse conceptuelle assistée par ordinateur entre donc dans le contexte de justification.

Ce genre d'usage permet d'imaginer l'application des techniques d'analyse conceptuelle assistée par ordinateur en des domaines qui ne pouvaient jusqu'à maintenant compter que sur l'analyse qualitative pour justifier leurs hypothèses. Certaines disciplines des sciences humaines, comme la psychanalyse et l'histoire de l'art, semblent souvent condamnées à une situation épistémologique insoluble, où une hypothèse consiste en une interprétation qui ne peut être justifiée que par d'autres interprétations. Si une discipline comme la psychanalyse peut, en faisant des prédictions quant aux statistiques textuelles et à la structure du corpus, offrir sa cou au couperet du falsificationnisme, elle peut s'affranchir de ses limites épistémologiques et donner un outil plus solide pour analyser ce qui passe souvent sous le radar des autres sciences.

Bibliographie

Buswell, G. T. 1935. « How people look at pictures: a study of the psychology and perception in art.»

Firth, John R. 1957. « A Synopsis of Linguistic Theory 1930–1955 ». *Studies in Linguistic Analysis*, p. 1-32.

Forest, Dominic. 2002. « Lecture et analyse de textes philosophiques assistées par ordinateur : application d'une approche classificatoire mathématique à l'analyse thématique du Discours de la méthode et des Méditations métaphysiques de Descartes ». Mémoire de maîtrise, Montréal, Canada : Université du Québec à Montréal.

Graziano, Michael S. A. et Sabine Kastner. 2011. « Human consciousness and its relationship to social neuroscience: A novel hypothesis ». *Cognitive Neuroscience*, vol. 2, no 2, p. 98−113.

Gubar, S. 1987. « Representing pornography: Feminism, criticism, and depictions of female violation». *Critical Inquiry*, vol. 13, no 4, p. 712−741.

Hagberg, Aric, Daniel A. Schult et Pieter J. Swart. 2008. « Exploring network structure, dynamics, and function using NetworkX ». Dans *Proceedings of the 7th Python in Science Conference*, sous la dir. de Gaël Varoquaux, Travis Vaught, et Jarrod Millman, p. 11−15. Pasadena, CA, USA : Los Alamos National Laboratory (LANL). En ligne. <http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=960616>. Consulté le 31 août 2012.

Hare, Jonathon S., Paul H. Lewis, Peter G. B. Enser et Christine J. Sandom. 2006. « Mind the Gap: Another look at the problem of the semantic gap in image retrieval ». Dans *Multimedia Content*

Analysis, Management and Retrieval 2006, sous la dir. de Edward Y. Chang, Alan Hanjalic, et Nicu Sebe, vol. SPIE V, p. 607309‑1. SPIE and IS&T. En ligne. <http://eprints.soton.ac.uk/261887/>.
Hauck, Nicholas. 2012. « L'archive absurde : le mystère de la déshumanisation magrittienne ». Dans Université de Toronto.
Haugeland, John. 1993. « Mind Embodied and Embedded ». Dans *Mind and Cognition: 1993 International Symposium*, p. 233–267. Academica Sinica.
Hébert, Louis et Guillaume Dumont-Morin. 2012. « Dictionnaire de sémiotique générale ». Dans *Signo*, sous la dir. de Louis Hébert. Rimouski, Québec. En ligne. <http://www.signosemio.com/documents/dictionnaire-semiotique-generale.pdf>.
Hébert, Louis et Éric Trudel. 2011. *Magritte: toutes les oeuvres, tous les thèmes*. Rimouski, Québec. En ligne. <http://www.signosemio.com/signodb/login.php>.
De Hoon, M. J. L., S. Imoto, J. Nolan et S. Miyano. 2004. « Open source clustering software ». *Bioinformatics*, vol. 20, no 9, p. 1453−1454.
Lewis, J. M. 2009. « Finding a Better k: A psychophysical investigation of clustering ». Dans *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, sous la dir. de N. A. Taatgen et H. van Rijn, p. 315−320.
Magritte, René. 1929. « Les mots et les images ». *La Révolution Surréaliste*, no 12.
Meunier, Jean-Guy, Ismaïl Biskri et Dominic Forest. 2005. « Classification and categorization in computer assisted reading and analysis of texts ». Dans *Handbook on categorization in cognitive science*, sous la dir. de Claire Lefebvre et Henri Cohen, p. 955−978. New York : Elsevier.
Rastier, François. 1995. « La sémantique des thèmes-ou le voyage sentimental ». *Rastier, F. éd. L'analyse thématique des données textuelles-L'exemple des sentiments, Paris, Didier*, p. 223–249.
Rimmon-Kenan, Shlomith. 1985. « Qu'est-ce qu'un thème? »
Robinson, Vanessa. 2012. « Les hybrides de Magritte et Ponge, ou le développement de l'animal moderne ». Colloque Magritte : perspectives nouvelles, nouveaux regards, Université de Toronto.
Sainte-Marie, Maxime, Jean-Guy Meunier, Nicolas Payette et Jean-François Chartier. 2011. « The concept of evolution in the Origin of Species: a computer-assisted analysis ». *Literary and Linguistic Computing*, vol. 26, no 3, p. 329−334.
Saint-Martin, F. 2010. *L'immersion dans l'art: comment donner sens aux œuvres de 7 artistes: le Maître de Flémalle, O. Leduc, Magritte, Mondrian, Lichtenstein, Rothko, Molinari*. Presses de l'Université du Québec.

Sylvester, D., S. Whitfield, M. Raeburn et L. Cawthra. 1992. *René Magritte: catalogue raisonné*. Sous la dir. de. D. Sylvester. Anvers : Fonds Mercator/Menil Foundation.

Tam, A. M. et C. H. C. Leung. 2001. « Structured natural-language descriptions for semantic content retrieval of visual materials ». *Journal of the American Society for Information Science and Technology*, vol. 52, no 11, p. 930−937.

Trudel, Éric. 2012. « Les figures animalières chez Magritte ». Dans *Magritte : perspectives nouvelles, nouveaux regards*. Université de Toronto.

Trudel, Éric et Louis Hébert. 2011. « Protocole d'analyse de la base de données « Magritte. Toutes les œuvres, tous les thèmes » ». En ligne. <http://www.magrittedb.com/Protocole_d'analyse.pdf>.

Viederman, M. 1987. « René Magritte: Coping with loss—reality and illusion ». *Journal of the American Psychoanalytic Association*, vol. 35, p. 967−998.

Les Cahiers de l’ISC - No 4 – Hiver 2015

Peindre Magritte avec des mots : rapport d'une analyse conceptuelle dans l’œuvre de Magritte à l’aide d’un corpus de descripteurs sémiotiques

CHARTRAND, Louis et MEUNIER, Jean-Guy

Hébert et Trudel (2011) ont conçu une base de données regroupant toutes les œuvres de Magrittes indexées dans Sylvester et al. (1992-1997), qu'ils ont ensuite analysées suivant un protocole sémiotique rigoureux (Trudel et Hébert, 2011) conçu pour cette occasion. Cette analyse a produit un péritexte de descripteurs associés aux tableaux, que nous avons tenté d’évaluer dans le contexte de l’analyse conceptuelle assistée par ordinateur. L'application de techniques de la méthode LATAO (Lecture et Analyse de Texte Assistée par Ordinateur) a permis de trouver jusqu'à six classes stables de tableau à partir du péritexte, parmi lesquelles quatre expriment clairement des thèmes particuliers (la mer, l'arbre, le corps nu de la femme et le visage) qui sont situables dans le temps. Enfin, nous avons fait l'analyse du descripteur-concept « femme », ce qui a confirmé la polarisation visage / corps nu, a fait ressortir l'utilisation quasi-exclusive de termes suggérant la violence (blessure, mutilation) sur des corps de femmes, de même que le contraste entre la nudité des corps de femmes et le caractère « très habillé » des représentations d'hommes. Nous en concluons donc que, pour l'extraction et la représentation de thèmes comme pour l'analyse conceptuelle, le péritexte et la méthode qui l'a créé constituent respectivement une ressource utile ainsi qu'un outil prometteur.

Les Cahiers de l’ISC – No 4, Hiver 2015

Peindre Magritte avec des mots : rapport d'une analyse conceptuelle dans l’œuvre de Magritte à l’aide d’un corpus de descripteurs sémiotiques - CHARTRAND, Louis et MEUNIER, Jean-Guy

ISSN 1925-1076 (Imprimé)

ISSN 1925-1084 (En ligne)