



HAL
open science

Représentations sociales et intelligence artificielle au service de la prévention du suicide

Yannis Haralambous, Philippe Lenca

► **To cite this version:**

Yannis Haralambous, Philippe Lenca. Représentations sociales et intelligence artificielle au service de la prévention du suicide. Céline Kopp Bigault; Michel Walter. La prévention du suicide en Bretagne : connaître, comprendre, agir, L'Harmattan, pp.357-376, 2023, Constellations Psy. hal-04220711

HAL Id: hal-04220711

<https://imt-atlantique.hal.science/hal-04220711>

Submitted on 28 Sep 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Représentations sociales et intelligence artificielle au service de la prévention du suicide

par

Yannis Haralambous et Philippe Lenca

Les travaux précurseurs sur les représentations sociales du suicide (Kopp-Bigault *et al.*, 2016) ont permis de mettre en évidence des croyances, des impacts psychiques collectifs et individuels, des stéréotypes et préjugés qui influencent la façon dont une personne peut percevoir la crise suicidaire, que ce soit sa propre crise ou celle d'autrui. Une récente étude (Jehel *et al.*, 2019) à partir d'une tâche d'association de mots adressée à des personnels soignants et qui leur demandait de fournir trois mots induits par la question « Pour vous, le suicide c'est...? » montre, malgré le faible nombre de répondants (396 personnes), que ces croyances ou préjugés peuvent être inadaptés à la gestion de la crise suicidaire (par exemple il ressort que la consommation d'alcool à la Martinique n'est pas perçue comme un facteur de risque associé au suicide). Ces deux études montrent l'intérêt d'une part de passer à l'échelle par le biais d'un questionnaire en ligne permettant d'adresser un plus grand nombre de personnes, et d'autre part de développer des analyses fondées sur des nouvelles approches autour des représentations sociales, s'inspirant de l'intelligence artificielle, et plus particulièrement du traitement automatique du langage. Ces deux disciplines ont, en effet, d'ores et déjà montré un apport important en santé mentale (Le Glaz *et al.*, 2021).

Représentations sociales et langage

Couvrez le côté linguistique que je ne saurais voir

Depuis leur introduction par Moscovici en 1961, les représentations sociales (abrégées RS dans ce texte) ont été définies et redéfinies dans plusieurs disciplines et plusieurs contextes, au point où, en citant Gaffié, on peut dire que

Notre rencontre s'opère dans l'auberge espagnole (ou roumaine !?) que constitue(nt) les « Représentations Sociales » (RS) : on y trouve ce qu'on y apporte mais on n'est jamais sûr de manger ce qu'on a commandé. C'est la richesse de



Yannis Haralambous, Philippe Lenca. «Représentations sociales et intelligence artificielle au service de la prévention du suicide», pp. 357-376. *La prévention du suicide en Bretagne : connaître, comprendre, agir* (dirigé par Céline Kopp Bigault, Michel Walter), *Constellations Psy*, L'Harmattan, 2023.

cette perspective, mais c'est parfois bien embarrassant. Trouver une définition claire d'une « RS », s'avère ainsi difficile (...)

Tout le monde s'accorde à dire que les RS ont (entre autres) un aspect *constituant* et un aspect *constitué*, le premier étant un processus et le deuxième un « contenu ». Comment ce « contenu » se manifeste-t-il, quel est son périmètre, quelles sont ses traces ?

Selon Jodelet (1989), les RS « circulent dans les discours, sont portées par les mots, véhiculées dans les messages et images médiatiques » ce qui nous permet d'inférer que, en partie au moins, les RS relèvent du langage. Et selon Py (2004), « les spécialistes paraissent d'accord au moins sur ce point : c'est par le discours que (les RS) existent et se diffusent dans le tissu social ».

Dans ces deux citations il est question du milieu dans lequel circulent, sont portées, existent et se diffusent les RS, mais rien n'est dit sur ce qu'elles sont, du point de vue linguistique. Dans l'ouvrage fondateur des RS, *La psychanalyse, son image et son public*, Moscovici (1961) affirmait que

La construction de ces schémas débute par la recherche d'*assertions* significatives, qui paraissent résumer une position importante dans la vision de celui qui émet l'assertion (...) Les éléments de ces communications, ici les assertions, sont mis en *relation*, dégageant de la sorte un modèle structuré de relations *qualifiées* et d'assertions.

ce qui nous rapproche d'une caractérisation linguistique des RS puisqu'une assertion significative est une phrase déclarative (une « position » étant une opinion) et les relations entre les assertions sont des relations discursives entre phrases de ce type. Mais Moscovici est revenu sur ces positions, et ainsi plus tard (Moscovici, 1994) il affirme que « se limiter à la sémantique est gravement insuffisant » – au lieu de cela, il fait le lien entre RS et présuppositions et fait appel à la théorie pragmatique de Grice.

Py (2004) qualifie les RS de « microthéories prêtes à l'emploi » (sans autre explication sur le choix et l'origine du terme de « microthéorie », terme que l'on rencontre surtout en économie) et conclut que les RS « ont comme milieu naturel le discours, qui leur permet de naître, de vivre et d'exister, c'est-à-dire aussi de changer, de s'étioler et de mourir ».

Moliner *et al.* (2002), dont l'ouvrage, par son sous-titre « Pratique des études de terrain », se présente comme un ouvrage pratique avec vocation de se concentrer sur le concret, est caractéristique dans sa manière d'éviter de définir les RS dans un cadre linguistique. L'introduction au troisième chapitre, intitulé « Le recueil des contenus » commence par

L'étude d'une représentation sociale passe, en premier lieu, par l'identification de divers éléments qui la constituent. Il va s'agir pour le chercheur, de dresser la liste des opinions, croyances ou informations que les membres d'un groupe donné partagent à propos de l'objet.

Il y est question d'opinions, croyances ou informations, qui se matérialisent le plus souvent par des phrases déclaratives formant un discours. Mais à la fin de cette introduction, le discours change et on se trouve en face d'une notion parachutée, celle de « contenu » :

Dans ce chapitre, nous passerons en revue les principales sources d'information permettant d'accéder aux *contenus des représentations* (nous soulignons) : les corpus textuels, les discours, les associations de mots, les images.

Il n'est donc plus question de RS mais de « contenu de RS », et si les RS se matérialisent par des séquences de phrases déclaratives, on peut se demander quel est le « contenu » de celles-ci. Serait-ce leur sémantique ? Y aurait-il confusion terminologique entre « sens » et « contenu » ? Il s'avère que non. Le terme de « contenu » est choisi pour faciliter la transition à la section suivante, qui s'intitule « L'analyse des contenus ».

Analyse(s) de contenu

L'*analyse de(s) contenu(s)* est une méthode expérimentale d'extraction de topiques utilisée en sciences humaines depuis les années 70. Basée sur la classification hiérarchique descendante, l'analyse de composantes principales et l'analyse factorielle de correspondances, elle permet, selon Mucchielli (1996) d'« expliciter le ou les sens qui sont contenus (dans les documents textuels) et/ou la ou les manières dont (ces documents) parviennent à faire effet de sens ».

Si l'analyse de contenu s'applique à des corpus textuels quelconques et permet d'en extraire du sens, comment s'applique-t-elle à l'étude des RS ?

À cela Moliner *et al.* (2002, p. 86) répond en indiquant, d'abord, qu'il s'agit d'analyse de contenu « des productions discursives ou textuelles relatives à un objet ». On se focalise donc sur des corpus autour de l'objet dont on veut étudier les RS – l'analyse de contenu de ces corpus permettrait alors d'extraire des structures représentant les RS autour de cet objet. Notons que Moliner *et al.* avertit « le chercheur s'intéressant aux représentations sociales qu'il doit être conscient du fait qu'en mobilisant les techniques d'analyse de contenu, il s'engage dans une voie étroite. »

Il enchaîne ensuite sur différents types d'analyse de contenu, énumérés par ordre de complexité croissante : l'analyse thématique, qui part du principe que les thèmes détectés dans le corpus autour d'un objet, sont révélateurs des RS sur cet objet ; l'analyse propositionnelle du discours, qui va découper le texte en propositions et assembler ces propositions en modèles argumentatifs ; l'analyse stylistique, partant du principe que « le passage par la forme n'est qu'une manière indirecte d'atteindre un autre niveau » (Bardin, 2001) ; l'analyse sémantique où des sèmes (structures élémentaires de signification) sont rassemblés en sémèmes ; l'analyse structurale, inspirée de la sémantique structurale de Greimas, où l'on dégage des relations sémantiques entre mots ou groupes de mots et on raisonne sur ces relations ; et enfin l'analyse des

relations par opposition, où on se focalise sur les relations entre systèmes matériels et systèmes symboliques.

Parmi ces méthodes, l'analyse propositionnelle du discours est sans doute la plus proche des préoccupations et méthodes actuelles de l'intelligence artificielle. Elle s'applique à des textes quelconques, de préférence des textes longs, de manière à dégager tout un système de propositions, reliées par des relations rhétoriques.

Questionnaires, association de mots

Mais qu'en est-il des enquêtes qui visent à déterminer les RS sur un sujet donné ? Faut-il demander aux sujets interrogés de rédiger des textes entiers autour du sujet en question, afin de les soumettre à ce type d'analyse ? Ce serait une méthode vouée à l'échec. Dans les enquêtes on se sert plutôt de *questionnaires*. Néanmoins, s'il s'agit de découvrir des traces de RS qui échappent peut-être même à la conscience des interrogés, les questions posées devraient laisser une certaine marge à l'expression libre, on est donc quasiment d'utiliser, du moins en partie, des questions à réponses libres.

Pour éviter l'analyse fastidieuse de textes provenant de réponses libres d'une part et le caractère restreint des réponses à des questions fermées d'autre part, les enquêtes de RS utilisent massivement la méthode de l'*association de mots*, introduite en psychanalyse par Carl Jung en 1910 (Jung, 1910). Cette méthode, que Moliner *et al.* (2002) dans la citation donnée *infra* qualifie de « source d'information » au même titre que les corpus textuels, est inspirée de la psycholinguistique. En effet, lors de la verbalisation de concepts (une étape du processus sous-jacent à l'acte de parole) on passe du concept non-verbal au lemme et ensuite au lexème, et à proximité de chaque lemme il existe des *lemmes associés*, qui peuvent être des collocations ou des antonymes (Warren, 2013, p. 44–45), leur existence permet d'expliquer certains phénomènes d'erreurs d'énonciation.

On peut néanmoins se poser la question : si je m'intéresse aux RS autour d'un sujet s et que je récupère par divers protocoles les mots associés les plus fréquents w_1, \dots, w_n , quelles informations ceux-ci m'apportent-ils sur les RS ? Admettons que je récupère les 100 mots les plus fréquemment associés à s , comment en déduire des éventuelles RS autour de s ?

D'un point de vue linguistique on constate que si s et les w_i sont des mots ou, plus généralement, des groupes nominaux, alors on obtient des relations binaires entre s et les w_i (les « associations »). Mais on ignore *a priori* la nature de ces associations. Il suffit d'observer les mots suivants qui ont été générés en tant que mots induits à partir du mot « voiture » sur le site wordassociations.net : « fourgon, carrosserie, volkswagen, accident, pneu, garage, locomotive, location, constructeur... », on y trouve des relations binaires de toute sorte.

La nature de l'association entre s et chacun des w_i correspond au *verbe* du triplet sujet-verbe-complément, au *prédicat* de la proposition de base. Si une RS peut être décrite par un ensemble de *phrases* ou *propositions*, il nous manque ici les verbes/prédicats de ces phrases/propositions.

La méthode des schèmes cognitifs de base

Il existe un modèle de questionnaire qui combine associations libres de mots et détection de la nature de celles-ci : il s'agit de la méthode des *schèmes cognitifs de base* de Guimelli and Rouquette (1992). Dans cette méthode on part d'un terme inducteur s et on demande au sujet d'induire des « cognèmes » associés w_i . Ensuite on présente au sujet 28 relations possibles R_j en posant à chaque fois la question : « dans quelle mesure pensez-vous que $s R_j w_i$? ». Ainsi, R peut prendre les valeurs « est la même chose que », « peut être défini comme », « est le contraire de », « a pour exemple », « appartient à la même classe que », « est une action qui a pour objet », etc. Les réponses autorisées sont « oui », « non » et « peut-être ».

Cette méthode est donc ouverte au niveau des lemmes (qui peuvent être librement choisis par les interrogés) et fermée au niveau des relations entre terme inducteur et terme induit. Elle présente l'avantage de permettre la description de natures d'association complexes puisque des réponses positives à plus d'une question sont autorisées. Et elle enchaîne des opérations cognitives relevant de l'imagination créative à des opérations relevant du raisonnement et de la logique : le sujet fournit dans un premier temps des mots associés de toute sorte mais doit, dans un deuxième temps, justifier ses « choix » en invoquant des relations logiques entre terme inducteur/stimulus et termes induits/réactions.

Si cette méthode est très riche en informations, elle aussi caractérisée par une certaine lourdeur : pour le (très faible) nombre de dix mots associés fournis par l'interrogé, on lui demandera de répondre à 280 (!) questions ; et si les réponses construites proposées peuvent être abstraites, voire même surréalistes lors du premier passage, elles deviennent vite répétitives et artificielles lorsqu'on les repasse pour les autres mots associés.

Notons que dans la description de la méthode des schèmes cognitifs de base nous avons omis une étape intermédiaire : l'étape de justification de l'association de mots, où l'enquêteur demande aux sujets de justifier leurs choix de mots associés en rédigeant un petit texte de quelques lignes. Voici comment Guimelli (2003) justifie la nécessité de cette étape :

En fait ces schèmes opératoires peuvent être qualifiés de « prédiscursifs » (Jodellet, 1972) dans le sens où ils sont probablement à l'origine d'une grande partie de la production discursive des sujets. C'est en effet en fonction de l'activation de tels ou tels connecteurs appartenant à tel ou tel schème que va s'articuler la mise en forme du discours tenu à propos de l'objet de représentation. Il est clair cependant qu'au moment de la procédure d'association verbale, le sujet n'a pas

une conscience claire du (ou des) connecteur(s) qui ont déterminé sa réponse. Ce qui est clair pour lui, c'est la réponse qu'il fournit, non les processus cognitifs qui en sont à l'origine ou qui l'ont générée. Or, l'étape de justification va fournir au sujet l'occasion d'explicitier ces processus. Dès lors, cette étape peut être considérée comme une étape de clarification de la relation inducteur/induit, au cours de laquelle les sujets vont prendre conscience (au moins en partie) des déterminants de leurs réponses. Ce qui va rendre plus facile l'exercice suivant (troisième étape de la procédure) qui consistera, comme on va le voir, dans l'inventaire des relations intervenant entre l'inducteur et l'induit.

Il précise d'ailleurs qu' « en l'absence de cette étape de justification les sujets sont en grande difficulté pour réaliser cet exercice ». Nous craignons que cette étape rajoute de la lourdeur à la tâche.

La structure des représentations sociales

Abric (1976, 2003) a proposé une structure tripartite pour les RS : le *noyau*, la *périphérie* et la *zone muette*. Il reparle de « contenu » mais en le distinguant de la « structure », en disant que « deux contenus identiques peuvent correspondre à deux représentations sociales différentes ».

L'exemple par lequel Abric (2003) illustre la différence entre noyau et périphérie est celui d'une enquête sur les RS du travail dans deux groupes de jeunes (un groupe de « qualifiés » et un groupe de « non-qualifiés »). Dans les deux cas on trouve les thématiques d' « accomplissement personnel », de « reconnaissance sociale », de « gagner sa vie » et de « contraintes », sauf que les deux premières font partie du noyau des RS du premier groupe et de la périphérie des RS du deuxième groupe, et vice-versa pour les deux dernières.

En ce qui concerne la zone muette, il s'agit d'« éléments de la représentation qui ont un caractère *contrenormatif* », l'exemple donné est celui des électeurs du FN qui, lors des sondages de la campagne présidentielle de 2002, n'ont pas voulu s'identifier en tant que tels. Guimelli (1998) va plus loin en disant que la zone muette est constituée de *schèmes dormant*, dont il existe deux types : les *non activés* (car triviaux, redondants ou inutiles) et les *non exprimables* (car contre-normatifs). Ce sont ces derniers, le « côté obscur », qui intéressent le plus les enquêteurs.

Dès son introduction en 1992 (Guimelli and Rouquette, 1992), la méthode de schèmes cognitifs de base était dotée d'un paramètre numérique, la *valence*, qui permettait d'évaluer l'importance des différentes RS obtenues. Guimelli (2003) applique la valence au repérage du noyau des RS. L'exemple qu'il donne (Guimelli, 1993) est celui d'une étude sur les RS du « groupe social idéal ». Quatre éléments thématiques ont été dégagés : l'amitié (relations positives entre les membres du groupe), l'égalité (pas de chef), la convergence des opinions et l'appartenance à un même milieu social. Des études précédentes ont montré que les deux premiers font partie du noyau et les deux autres de la périphérie. Les valeurs de valence obtenues (0.5 pour les deux premiers,

0.41 et 0.42 respectivement pour les deux autres) ont permis de valider cette caractérisation de manière expérimentale.

Pour décrire la zone muette, les associations de mots ne suffisent pas puisque, par définition, les mots induits qui permettraient d'y accéder sont tus. Une solution à ce problème, d'une simplicité déconcertante, est proposée par Guimelli and Deschamps (2000) dans une étude sur les gitans. Ils demandent aux interrogés de fournir d'abord cinq mots associés aux gitans selon leur propre avis, et ensuite cinq mots « que donneraient à leur avis les « Français en général » », ce qu'ils appellent un *contexte de substitution*. Ils ont ainsi obtenu les thèmes positifs « famille » et « liberté » avec la même fréquence dans les deux contextes, alors que les thèmes (que l'on qualifierait aujourd'hui de « politiquement incorrects ») de « voleurs » et de « gens sales » ont été produits avec des fréquences significativement supérieures dans le contexte de substitution que dans le contexte personnel (242% et 275%, respectivement), et ceci malgré le caractère anonyme de l'enquête.

On voit donc que la méthode des schèmes cognitifs est tout à fait indiquée pour explorer la structure des RS, avec un seul bémol : si l'on souhaite explorer également la zone muette, cela aurait comme conséquence d'alourdir davantage le questionnaire en multipliant le volume de questions par deux.

Méthodes de l'intelligence artificielle

La représentation des connaissances

Pour arriver aux phantasmes actuels de robots intelligents impossibles à distinguer des humains, l'intelligence artificielle (dorénavant abrégée IA dans ce texte) a parcouru un long chemin. Le domaine de l'IA qui nous intéresse dans ce texte est celui du traitement du langage naturel et de la représentation des informations et des connaissances.

Commençons par quelques mises au point terminologiques : en IA on considère que les *données* sont des suites de nombres (représentés dans la mémoire de la machine ou sur les autoroutes de l'information par des bits); lorsqu'on attache une sémantique à ces suites de nombres (en indiquant, par exemple, que tel nombre rationnel représente une température à tel endroit et à tel moment), alors ils deviennent des *informations*; enfin, lorsqu'on détecte des tendances ou des règles parmi les informations, règles qui permettent de faire des prédictions (lorsqu'on obtient, par exemple, par induction la prédiction que le soleil se lèvera bel et bien encore une fois le lendemain), alors on dit qu'on a extrait des *connaissances* à partir des informations.

Les humains ont depuis toujours utilisé les langages naturels pour représenter les informations et les connaissances. Ils ont même inventé des méthodes pour obtenir de nouvelles informations/connaissances à partir d'autres informations/connaissances : principalement l'*inférence déductive* (si

je possède l'information que « Paris est une capitale » et la règle que « toutes les capitales sont des villes » je peux en inférer que « Paris est une ville », une nouvelle information que je rajoute à ma « base de connaissances », c'est le procédé usuel qui fait avancer la science), mais aussi d'autres modes de raisonnement comme l'induction et l'abduction.

L'IA a donc très tôt cherché à formaliser ces notions pour être capable de manipuler des versions formalisées des informations et des connaissances, pour effectuer des inférences, pour stocker les résultats dans des dépôts informatiques et pour communiquer avec ceux-ci à travers des requêtes. Aux débuts de l'IA, le langage formel utilisé pour la représentation des connaissances et le calcul des inférences a été la *logique du premier ordre*, originaire d'Aristote mais formalisée par Frege au XIX^e siècle.

Depuis les années 50, l'IA a vu émerger plusieurs langages de représentation des connaissances, certains plus puissants que la logique du premier ordre (par exemple, les *logiques modales*, qui font intervenir d'autres gadgets, comme la possibilité/nécessité, la connaissance d'un fait par un agent, les aspects déontiques, etc.) et certains moins puissants, pour faciliter les calculs et garantir des réponses rapides aux requêtes envoyées aux bases de connaissances. Nous renvoyons le·la lecteur·rice à Marquis *et al.* (2014) pour un panorama général des langages de représentation des connaissances et allons, par la suite, nous concentrer à un système spécifique, qui a l'avantage d'exister et d'être soutenu par le monde industriel, ce qui constitue une garantie de longévité et de résilience : le *Web sémantique*.

Pour citer Aimé and Arnould (2021, p. 57) :

Au commencement était le Web, et le Web était en Tim Berners-Lee. Il était au commencement en lui. Tout par lui a été fait, et sans lui n'a été fait rien de ce qui existe (...),

(citation qui est suivie, 26 pages plus loin, par un titre de section très évocateur : « Nous sommes tous des documents »).

La paraphrase de l'incipit de l'Évangile selon Saint-Jean se réfère au Web sémantique, ultime contribution de Berners-Lee (qui est à l'origine du WWW). Il s'agit d'un assemblage de normes et de technologies, tels que RDF et les ontologies, dont nous parlerons dans la section suivante et dont le·la lecteur·rice intéressé·e trouvera une description complète dans Aimé and Arnould (2021).

On peut à ce stade se poser la question suivante : si les humains utilisent le langage naturel comme principal outil de communication, comment représenter celui-ci à travers les technologies du Web sémantique ?

Langage naturel et représentation des connaissances

Le passage du langage naturel au langage formel se fait en deux étapes : de l'*énoncé* on fait (ou non) abstraction des aspects pragmatiques et on garde des *phrases*, et de celles-ci on arrive à des *propositions*. Par phrase on entend

(Oswald, 1980) une « entité linguistique abstraite, purement théorique, en l'occurrence un ensemble de mots combinés selon les règles de la syntaxe, ensemble pris hors de toute situation de discours », alors que « ce que produit un locuteur, ce qu'entend un auditeur, ce n'est donc pas une phrase, mais un énoncé particulier d'une phrase ». Une proposition est « l'invariant de signification commun à tous les énoncés déclaratifs synonymes » (Delcomminette, 2019, p. 9). Ainsi l'énoncé « Je vous ai compris », que le général De Gaulle formule du balcon du Palais gouvernemental général d'Alger le 4 juin 1958, peut être sorti de son contexte pour devenir une simple phrase « Je vous ai compris », dont on peut imaginer des phrases (quasi-)synonymes « J'ai compris ce que vous me demandez », « Message reçu », etc., qui peuvent toutes être représentées par la formule logique $\text{comprendre}(\text{moi}, \text{vous})$ où comprendre est un prédicat binaire dont les arguments sont le sujet moi et le complément d'objet direct vous.

Il y a différentes méthodes pour passer d'une phrase à une formule logique, comme la sémantique formelle de Montague (Haralambous, 2015). Notons néanmoins que la phrase que nous avons utilisée en guise d'exemple est simpliste sur le plan syntaxique. On peut facilement imaginer des phrases déclaratives qui requièrent des prédicats d'arité supérieure à 2, comme, par exemple, « le garçon donne la pomme à la fille » (arité 3), « voyager en avion est plus sûr que manger des crevettes » (arité 4), « Alice, Ben, Clara, Daniel et Élise forment une famille » (arité 5), etc.

Dans son système, Berners-Lee a fait le pari de n'utiliser que des relations binaires, ce qui facilite les calculs puisqu'à la base de tout on trouve toujours la même structure archétypale : le *triplet* sujet-prédicat-objet, qui correspond à la structure syntaxique sujet-verbe-complément, que nous connaissons depuis l'école primaire. Ainsi, la technologie intitulée RDF est basée sur cette idée de triplet, un *triplet RDF* étant un triplet sujet-prédicat-objet.

Deux questions se posent : « des triplets de quoi ? » et « que se passe-t-il quand la phrase déclarative requiert un prédicat d'arité supérieure à 1 ? ».

Pour répondre à la première question, le sujet ou l'objet d'un triplet peut être des chaînes textuelles (appelées « littéraux ») ou, de préférence, des références vers des objets décrits dans des ressources contenant des triplets RDF. À nous de faire la correspondance entre ces objets et le monde réel (une opération appelée *interprétation*). Ces ressources peuvent être placées partout sur le Web, leurs emplacements sont représentés par des codes appelés IRI. Si je veux donc parler de Paris, la capitale de la France, j'utilise un IRI qui pointe vers une ressource qui décrit des villes, et Paris en particulier. Une telle ressource est, par exemple, Wikidata, un site comprenant quelques 97 millions d'objets (en février 2022). Les prédicats sont également des objets de ce type.

Concernant les prédicats d'arité supérieure à 2, Berners-Lee se sert d'une astuce appelée *réification*. Ainsi, pour la phrase « le garçon donne la pomme à la fille » il va créer un objet de type « énoncé » (*statement*), il va déclarer

que le sujet de cet énoncé est l'objet correspondant au garçon, le prédicat est l'objet correspondant au verbe « donner », l'objet sera l'objet correspondant à la pomme, cela fait déjà trois triplets (allant de l'objet « énoncé » vers les objets « sujet », « prédicat », « objet »), et on rajoute un quatrième triplet qui représente le complément d'objet indirect relatif à la préposition « à ». Par le biais de cette astuce on peut formaliser n'importe quelle phrase déclarative, le seul inconvénient étant le nombre de triplets RDF nécessaires.

En ajoutant des nouvelles notions aux triplets RDF (les notions de classe et d'instance, des opérateurs d'un type de logique appelé *logique descriptive*, etc.), on arrive à une autre technologie du Web sémantique, OWL, dont les objets sont appelés des *ontologies*.

Les ontologies (que certains appellent *ontologies computationnelles* pour les distinguer de la sous-discipline de la philosophie de même nom, qui est l'étude de l'ὄν, c'est-à-dire de l'être) sont donc des dépôts de triplets RDF dont les sommets peuvent être dispersés partout sur le Web et peuvent être interprétés de manière utile comme des entités du monde réel.

Mais le plus important est le fait qu'à l'aide des ontologies on construit des *bases de connaissance* qui possèdent un *processus d'inférence*. Ce processus permet de répondre à des requêtes complexes dont la réponse n'est pas forcément un triplet qui se trouve explicitement dans l'ontologie. Pour reprendre l'exemple de Paris étant une capitale et toute capitale étant une ville, on peut par inférence répondre par l'affirmative à la question « est-ce que Paris est une ville ? ». Sur le plan technique cela revient à avoir des classes d'objets appelées Capitales et Villes, un objet Paris représentant Paris, un triplet « Paris appartient_à Capitales » et un triplet « Capitales \sqsubseteq Villes » où le symbole \sqsubseteq (réminiscent de la relation ensembliste \subseteq) représente la relation « est une sous-classe de ».

Retenons de cette section le fait que, selon l'approche du Web sémantique, toute phrase déclarative peut être représentée dans une base de connaissances sous forme d'un ou plusieurs triplets sujet-prédicat-complément (quitte à passer par certaines astuces de type réification, classes, etc.).

Notons également que le Web sémantique est encore loin de devenir une réalité, lire par exemple O'Reilly (2021), un texte qui compare l'enthousiasme autour du Web sémantique avec la bulle financière de la crise des tulipes, dans les Pays-Bas des années 1634–1637.

Représentations sociales et intelligence artificielle

Arrivé·e à ce stade le la lecteur·rice ne pourra que constater la ressemblance flagrante entre le principe de base des schèmes cognitifs de base (inducteur-relation-induit) et les triples RDF. Cette ressemblance n'a pas échappé à Aimé and Arnould (2021, p. 193) et les a conduits à déclarer que « toute ontologie est une représentation sociale » :

En tant que reconstruction d'une « réalité » perçue, telle que se l'approprie un ensemble d'individus, les ontologies peuvent être considérées comme des artefacts issus d'une pensée sociale. Nonobstant, ces ontologies computationnelles, en tant quantités formelles, peuvent également être considérées comme des artefacts issus d'une pensée rationnelle. En résumé, nous posons ici l'hypothèse que les ontologies computationnelles sont des artefacts d'une pensée sociale rationalisée.

Sans aller aussi loin, nous allons nous servir de cette analogie entre schèmes cognitifs de base et unités élémentaires du Web sémantique. Notre but sera d'appliquer des méthodes de l'IA, et en particulier des méthodes du traitement automatique des langues (dorénavant TAL), à l'exploration des RS. Le sous-domaine du TAL qui s'occupe directement du type de questions qui nous intéresseront est celui des *relations sémantiques*, cf. par exemple Nastase *et al.* (2021).

Première étape de la méthode des schèmes cognitifs de base

Dans la première étape de la méthode des schèmes cognitifs de base on définit l'« inducteur » et on demande aux participants de produire des « associations de mots ». Si nous avons mis ces termes entre guillemets c'est parce que leur nature linguistique n'est jamais proprement définie. Guimelli (2003) décrit cette première étape ainsi :

On présente aux sujets un terme inducteur A et on leur demande de donner, par écrit et le plus rapidement possible, trois *mots ou expressions* (nous soulignons) qui leur viennent à l'esprit en relation avec le terme A.

Fraïssé (1999, p. 192) est plus précise :

On se donne à l'origine un vocabulaire fini V composé d'éléments appelés cognèmes. (...) De façon opérationnelle, ce sont des *termes isolés, des syntagmes, ou encore les items utilisés dans les questionnaires*. (nous soulignons)

Dans une perspective structuraliste, le *cognème* est défini (Rouquette and Rateau, 1998) comme l'*unité élémentaire de la cognition* :

Ces éléments, qu'on appelle alors des cognèmes, sont en général indexés par des *mots ou des syntagmes* (nous soulignons) prélevés dans le discours spontané ou provoqué des individus.

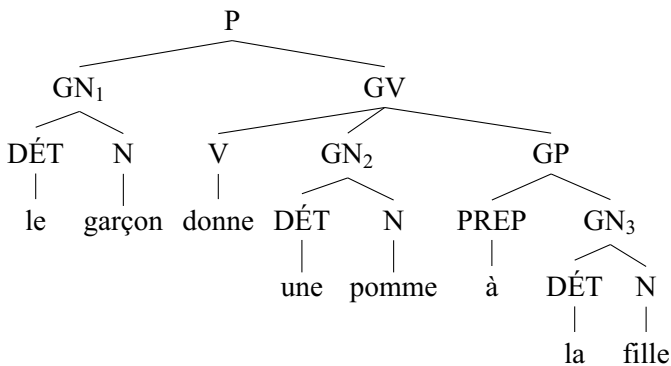
Reprenons les termes utilisés par Fraïssé (1999) : un *terme*, objet d'étude de la terminologie, est la représentation lexicale d'un concept dans un domaine de connaissances donné. Appelé *unité terminologique* dans Depecker (2005), un terme y est défini comme « une désignation (*unité linguistique*) liée à un concept déterminé » (nous soulignons). Dans ce texte, le problème de la nature linguistique du terme est évité par la qualification d'« unité linguistique ».

On trouve une définition plus précise de la nature linguistique du terme dans Pavel (2001, p. 17) :

L'unité terminologique est l'étiquette d'un concept dans un arbre conceptuel. Ce peut être un mot ou un syntagme, un symbole, une formule chimique ou mathématique, une appellation scientifique en latin, un acronyme, un sigle ou l'appellation officielle d'un emploi, d'un organisme ou d'une entité administrative.

Dans cette liste on retrouve des entités linguistiques et non-linguistiques (les formules) qui se comportent syntaxiquement comme des syntagmes, et on retrouve la notion de syntagme également dans l'énumération de types de cognèmes (Fraïssé, 1999). Qu'est-ce qu'un syntagme (également appelé *constituant*)?

Un syntagme est une séquence de mots consécutifs dans une phrase. Cette séquence n'est pas arbitraire (ainsi, par exemple, dans « le garçon donne une pomme à la fille », « une pomme » est un syntagme mais pas « pomme à »). Dans l'incapacité de donner une définition formelle du syntagme, les linguistes se servent de l'arbre syntaxique par constituants, obtenu à partir d'une grammaire formelle décrivant la langue en question. L'arbre syntaxique par constituants de notre exemple est :



Dans cet arbre on dit qu'un sommet A domine un sommet B si A se trouve plus haut que B et s'il existe un chemin (une suite d'arêtes) qui va de A à B. Un sommet *terminal* est un sommet qui se trouve à la base de l'arbre, c'est donc forcément un « mot ». On définit un *syntagme* comme une séquence maximale de sommets terminaux dominés par un même sommet. Ainsi les mots de « la pomme » sont dominés par GN₂, c'est un syntagme ; par contre, les mots de « pomme à » sont dominés par GV mais ne sont pas les seuls à l'être, il faudrait prendre le groupe verbal complet : « donne la pomme à la fille ».

Notons qu'il a été montré (Garrett, 1967) que la notion de syntagme a une réalité psychologique : un son de clic a été inséré dans un enregistrement et il a été demandé aux sujets de l'expérience d'indiquer l'endroit auquel ils l'ont perçu. Il s'est avéré qu'ils choisissent systématiquement des frontières de syntagmes pour indiquer l'emplacement du clic.

Mais quels sont vraiment les cognèmes fournis par les interrogés ?

Les cognèmes obtenus lors d'une étude sur le suicide

Une étude sur les RS du suicide, menée par Jehel *et al.* (2019) en Martinique, a fourni 3 115 cognèmes différents (la différence s'entend au niveau graphémique, les erreurs de frappe n'ayant pas été corrigées), repartis de la manière suivante :

- 2 037 cognèmes consistant en un seul mot, dont
 - 1 389 noms communs (68%),
 - 569 adjectifs (28%),
 - 64 verbes à l'infinitif (3%),
 - 12 adverbes (affectivement, complètement, inexorablement, mentalement, moralement, physiquement, psychiquement, psychologiquement, quotidiennement, socialement, souvent, subitement) (<1%),
 - 3 verbes à l'indicatif (souffre, souffrent, tourne) (<1%);
- 1 078 cognèmes multi-mots, dont
 - 875 groupes nominaux (81%),
 - 98 groupes prépositionnels (à bout de force, d'amour, en colère, etc.) (10%),
 - 57 groupes verbaux avec des verbes à l'infinitif (mettre en lumière son mal, perdre pied, quitter l'autre par « obligation », etc.) (5%),
 - 34 groupes verbaux avec des verbes à l'indicatif (se sent exclu, touche tout le monde, s'en sort plus, etc.) (3%),
 - 14 phrases complètes (c'est dommage, on n'a pas pu sauver la personne, on voit noir, etc.) (1%),

Si l'on regroupe les noms communs avec ou sans déterminant (un nom commun sans déterminant est un groupe nominal si on considère la grammaire du style télégraphique) ainsi que les verbes à l'infinitif, alors on constate que 75% des cognèmes sont des groupes nominaux.

En deuxième lieu arrivent les 569 adjectifs : nous observons qu'ils peuvent être expliqués par des ellipses : (c'est) insupportable, (celui qui se suicide est) lâche, (le suicide est) irrévocable, etc.

Les groupes prépositionnels représentent soit des expressions (à bout de souffle, en marge de la société, etc.) soit des constructions elliptiques : (suicide) d'amour, (ils sont) dans l'impasse. Il en est de même des verbes à l'indicatif : (le suicidé) se sent exclu, (le suicide) touche tout le monde. Il est évident que dans tous ces cas la marge d'interprétation est importante, ce qui justifie la deuxième étape de la méthode des schèmes cognitifs de base.

Apport de l'IA

On peut maintenant se poser la question du traitement automatisé (ou semi-automatisé) des cognèmes induits. Après une correction des éventuelles erreurs de frappe et une ré-écriture des cognèmes qui ne répondent pas au critère donné ci-dessus (syntagme nominal, avec ou sans déterminant initial), on utilisera un parseur syntaxique pour reconnaître la structure syntaxique du cognème et en extraire les mots pleins.

À l'aide de ceux-ci on va construire une ontologie *ad hoc*, c'est-à-dire qu'on va d'abord les considérer comme des sommets d'un graphe d'ontologie, puis fusionner les synonymes (dans le contexte du domaine de connaissances de l'étude) et introduire les relations hiérarchiques inhérentes aux cognèmes (en se rapportant à une ontologie existante, comme WikiData), ainsi que d'autres relations, appelées *ontologiques* (Nastase *et al.*, 2021, p. 26) c'est-à-dire quasiment toujours réalisées, comme la relation entre « espoir » et « désespoir » qui est une antonymie. On distingue les relations *ontologiques* des relations *idiosyncratiques*, qui sont occasionnelles (la relation entre « plage » et « pollution » est occasionnelle puisque toutes les plages ne sont pas polluées, ni le contraire).

Ayant acquis cette ontologie *ad hoc* on a préparé le terrain pour la désignation des natures de relations entre inducteur et induit par les participants.

Deuxième étape de la méthode des schèmes cognitifs de base

À partir des 28 relations prédéfinies¹ de la méthode des schèmes cognitifs de base, on introduit de nouvelles arêtes dans l'ontologie *ad hoc*.

Nous avons déjà évoqué la lourdeur de ce type de questionnaire. Les relations ontologiques introduites dans l'étape précédente peuvent nous permettre d'alléger la procédure. En reprenant l'exemple de la paire de cognèmes « espoir » / « désespoir » et en prenant le premier comme un inducteur et le deuxième comme induit, on trouvera dans toute ontologie générale qui se respecte une relation d'antonymie entre les deux. Inutile donc de poser la question « Est-ce que « espoir » est le contraire de « désespoir »? », de même que la question absurde « Est-ce que « espoir » est la même chose que

1. Notons que dans la définition des relations des schèmes cognitifs de base donnée dans Guimelli (2003) il y a deux catégories dont la différence n'est pas claire : le *schème de voisinage* (classe incluse, etc.) et le *schème de composition* (tout à la partie, etc.). La confusion est accentuée par les exemples donnés : dans le premier cas l'exemple est « ville / rue » (la ville inclut des rues) et dans le deuxième cas, « nature / rivière » (les rivières sont des parties de la nature). Mais être partie de, n'est-il pas aussi être inclus dans, et réciproquement, du moins au sens métaphorique ? Pour clarifier ces catégories nous avons pris des hyperonymes dans le premier cas (« véhicule / bus ») et des holonymes dans le deuxième (« main / doigt »), en espérant que cela ne dénature pas outre mesure la méthode des schèmes cognitifs de base.

« désespoir » ? ». En omettant ces deux questions on réduit le volume du questionnaire, on évite le ridicule des trivialités, et on peut se concentrer sur des questions beaucoup plus sensées comme « Est-ce que « espoir » est une cause de « désespoir » ? », etc.

Par des procédés similaires on pourrait omettre des questions incongrues comme « Est-ce que « espoir » est une action pour laquelle on utilise l'outil « désespoir » ? », pour cela il suffirait de poser une contrainte du type « un outil est un objet matériel », ce qui omettrait les questions dont la partie « outil » serait une entité immatérielle comme l'espoir.

Cette approche peut alléger le questionnaire, mais comporte aussi des risques. En effet, elle se base sur l'hypothèse que les schémas cognitifs de l'interrogé sont conformes à la structure de l'ontologie dont on se sert pour valider ou invalider des questions du questionnaire. Cela peut ne pas être le cas. Doit-on exclure la possibilité d'un participant pour qui « espoir » et « désespoir » signifient la même chose ? On dit communément que les mots « peuvent perdre leur sens », cela pourrait être le cas pour ces deux mots. Tout incongrue qu'elle serait, une réponse positive à la question « Est-ce pour vous « espoir » est la même chose que « désespoir » ? » pourrait être extrêmement utile de par son incongruité-même.

L'attitude à adopter va dépendre de plusieurs facteurs : le type d'étude (n'y a-t-il que les résultats statistiques qui importent ou s'intéresse-t-on aux « signaux faibles » émanant d'individus spécifiques ?), la taille de l'échantillon (un échantillon de 10 individus ne sera pas traité de la même manière qu'un échantillon de 1 000 individus), le contexte du cognème inducteur (un cognème inducteur dans le domaine émotionnel aura plus de chances d'avoir des réponses inattendues qu'un cognème technique), la présence ou non d'une zone muette, etc.

Afin de limiter le volume du questionnaire on pourrait avoir recours au *machine learning* : en créant un modèle de prédiction de réponse on peut évaluer la probabilité d'une réponse vis-à-vis des autres réponses du participant (ainsi que de ses réponses à d'autres parties du questionnaire et, plus généralement, son profil). On pourrait évaluer la significativité d'une réponse statistiquement très rare et se servir d'un seuil pour évacuer dynamiquement toutes les réponses improbables. D'autre part, on peut imaginer un élément d'interface qui afficherait toutes les réponses omises, en supposant que l'interrogé/e est conscient/e de l'absence de ses réponses ou de l'inadaptation des réponses affichées et souhaite disposer d'alternatives. Il peut aussi être envisagé de prévoir un champ d'entrée libre, qui n'apparaîtrait que si les réponses à toutes les questions, sans exception, sont négatives.

Perspectives

Nous avons envisagé *supra* des alternatives à la procédure standard de la méthode des schèmes cognitifs de base, telle qu'elle a été définie par Guimelli

and Rouquette (1992). On peut se demander dans quelle mesure l'outil informatique peut contribuer à l'amélioration de cette méthode, tout en gardant son essence.

Cela est d'autant plus important que la saisie sur smartphone ou tablette peut être bien plus pénible que le fait de cocher des cases sur papier.

On peut apporter une multitude de modifications à la procédure, en utilisant des algorithmes de TAL. Pour chacune il faudra évaluer les risques d'effets secondaires et la conformité de la procédure ainsi obtenue à la méthode originale.

Au niveau des cognèmes

On peut imaginer un parseur syntaxique qui calcule l'arbre syntaxique du cognème et alerte le participant quand il s'écarte de la structure de groupe nominal (avec ou sans déterminant initial). Ce parseur peut interagir avec un correcteur orthographique qui propose des versions alternatives lorsque certains mots du cognème ne sont pas détectés dans un dictionnaire donné (dictionnaire comportant toutes les formes fléchies).

Le parseur de cognème pourrait également envisager d'autres types d'information comme le temps des verbes : autorise-t-on des cognèmes avec des verbes dans des temps autres que le présent de l'indicatif et l'infinitif ?

D'autre part le parseur pourrait détecter des ambiguïtés et demander des précisions en temps réel. Ainsi la réponse « voler » pourrait déclencher une question « “voler” au sens de l'aviation ou au sens du délit ? » et insérer cette information sous forme de métadonnée dans les résultats du questionnaire. Une telle intervention doit être envisagée uniquement quand les deux sens sont suffisamment différents et le sens souhaité ne peut être inféré du contexte. Des algorithmes de désambiguïsation peuvent être utilisés dans ce but.

Les risques encourus par des interventions de ce type ne se limitent pas uniquement au risque d'altération du cognème souhaité par le participant. On est également exposés à un risque de réduction de la spontanéité et de l'imagination créative du participant : en assistant à des tentatives de correction ou de reformulation/explication des cognèmes produits, elle/il peut se sentir observé/e et/ou jugé/e, ce qui peut conduire à une rationalisation de ses productions, à les faire entrer dans un moule.

Les effets des interventions de ce type et leur perception de la part des utilisateurs doivent être évalués. De même, toute intervention de ce type doit être désactivable par l'utilisateur : si elle/il souhaite écrire « différence » à la manière derridéenne, ou « dyscours » à la manière de Jean-François Jeandillou (1985), le système doit l'autoriser à le faire, tout en lui indiquant l'irrégularité de la forme.

Au niveau des relations

On assiste à un problème bien différent. En effet, si elle garantit une liberté quasi-absolue pour les cognèmes, la méthode des schèmes cognitifs de base est *très restrictive* au niveau des relations : elle en propose 28, et pas une de plus.

Une perspective d'amélioration a déjà été envisagée *supra* : l'omission de certaines questions dont la réponse est évidente, pour diminuer le volume du questionnaire et donc la charge mentale de l'utilisateur.

Une autre possibilité serait la « naturalisation » des énoncés des questions. Dans la méthode originale on appelle A le cognème inducteur et B le cognème induit, et on demande à l'utilisateur/rice de faire, mentalement, les substitutions de A et de B par les cognèmes correspondant. Cette approche (inévitabile dans le cas du support papier) permet de garder une syntaxe correcte dans toutes les questions. L'outil informatique peut remplacer A et B par les cognèmes correspondants, mais cela implique une étape d'adaptation, le plus fréquemment l'ajout d'un déterminant : si les cognèmes sont « espoir » et « désespoir », pour être inséré dans la phrase « A est la cause de B », le premier doit être précédé du déterminant « l' » et la préposition « de » doit devenir « du ». Cette opération ne doit pas être effectuée lorsqu'il s'agit d'un nom propre, avec toujours des exceptions : pas de déterminant pour « de Gaulle », mais un déterminant pour « Abbé Pierre »...

Au-delà de ces petites modifications qui préservent l'esprit de la méthode de Guimelli & Rouquette, on peut se demander si la liste des 28 relations peut être mise en cause. On pourrait imaginer une sélection adaptative de relations, par étapes : on soumet le questionnaire avec la liste la plus générale de relations à un premier échantillon d'utilisateurs, puis on élimine les relations qui ont été reçues la même réponse par tous les participants (ou quasiment tous), on prend celles qui ont eu le plus de variation de réponse (donc, dans le vocabulaire de la théorie de l'information, celles qui ont une entropie maximale) et on les spécialise, c'est-à-dire qu'on les démultiplie par affinage. En répétant cette étape plusieurs fois, on arrive à un ensemble de questions très fines, adaptées aux inducteurs prévus et au domaine de connaissances sous-jacent.

Ce procédé gagnerait certainement en précision dans l'exploration des RS, mais perdrait peut-être en rappel. La méthode mérite d'être évaluée.

D'autres méthodes d'adaptation de l'ensemble des questions peuvent être imaginées, selon le but de l'étude : s'agit-il d'affiner la compréhension (et la représentation en langage naturel) des RS ? Ou plutôt de « ratisser large » pour accéder à des RS imprévues ? Sait-on à l'avance si la population interrogée a des RS très homogènes ou, au contraire, très différenciées ?

Enfin, l'adaptation des questions peut faciliter la comparaison des résultats d'un questionnaire soumis aux mêmes personnes à des moments différents, par exemple avant et après un certain événement (par exemple, une formation), à T+6 mois, etc. De même, les questions peuvent être adaptées à des facteurs

externes comme des périodicités (les saisons, l'heure de la journée, etc.). Ce n'est qu'en analysant un grand nombre de résultats que l'on peut en déduire des tendances corrélées avec des facteurs externes, pour ensuite essayer de les compenser en adaptant les questions.

Conclusion

La méthode des schèmes cognitifs de base a été en avance de son époque : les outils utilisés (cognèmes, relations induisant des triplets cognème inducteur-relation-cognème induit) préfiguraient déjà les outils de l'IA alors que, à notre connaissance, la méthode n'a pas encore, à ce jour, été utilisée sur support informatique. Son adaptation à l'outil informatique présente certains risques mais, avant tout, ouvre des nouvelles perspectives à l'étude des RS, un sujet dont l'actualité n'a cessé de croître.

Bibliographie

- ABRIC J.-C., *Jeux, conflits et représentations sociales*, Thèse de doctorat, Université de Provence, 1976.
- ABRIC J.-C., La recherche du noyau central et de la zone muette des représentations sociales, dans ABRIC J.-C. (éd.), *Méthodes d'étude des représentations sociales*, Érès, 2003.
- AIMÉ X., ARNOULD F., *Modélisation ontologique & psychologies*. Éditions matériologiques, 2021.
- BARDIN L., *L'analyse de contenu*, PUF, 2001.
- DELCOMMINETTE S., Proposition, phrase, énoncé chez Platon, Aristote et les stoïciens, dans *Proposition, phrase, énoncé*, ISTE Éditions, 2019 : 9-27.
- DEPECKER L., Contribution de la terminologie à la linguistique. *Langages*, 157, 2005 : 6-13.
- FRAÏSSÉ C., *La représentation sociale de la médecine naturelle. Étude des liens entre structure de représentation et pratiques sociales*, Thèse de doctorat, Université Aix-Marseille 1, 1999.
- GAFFIÉ B., Confrontations des représentations sociales et construction de la réalité, *Journal International sur les Représentations Sociales*, 2, 2005 : 6-19.
- GARRETT M.F., *Syntactic Structures and Judgments of Auditory Events*. PhD thesis, University of Illinois, 1967.
- GUIMELLI C., Locating the central core of social representations : towards a method, *European Journal of Social Psychology*, 23, 1993 : 555-559.
- GUIMELLI C., *Chasse et nature en Languedoc. Étude de la dynamique d'une représentation sociale chez les chasseurs languedociens*, L'Harmattan, 1998.

- GUIMELLI C., Le modèle des schèmes cognitifs de base (SCB) : méthodes et applications, dans ABRIC J.-C. (éd.), *Méthodes d'étude des représentations sociales*, Érès, 2003.
- GUIMELLI C., DESCHAMPS J.-C. Effets de contexte sur la production d'associations verbales. Le cas des représentations sociales des gitans, *Cahiers Internationaux de Psychologie Sociale*, 47-48, 2000 : 44-54.
- GUIMELLI C., ROUQUETTE M.-L., Contribution du modèle associatif des schèmes cognitifs de base à l'analyse structurale des représentations sociales, *Bulletin de Psychologie*, 45, 1992 : 196-202.
- HARALAMBOUS Y., Les mathématiques de la langue : l'approche formelle de Montague, *Quadrature*, 98, 2015 : 9-19.
- JEANDILLOU J.-F., Queneau ou le *dys*discours de la linguistique dans *La linguistique fantastique*, Denoël, 1985 : 211-221.
- JEHEL L., DUPONT E., LENCA P., KOPP-BIGAULT C., The ReSCU-M study : Comprehension of suicide social representation in Martinique Island. Perspective for prevention and communication strategies, dans *International Association for Suicide Prevention Caribbean Regional Symposium*, Trinidad, Trinidad and Tobago, 2019.
- JODELET D., Représentations sociales : un domaine en expansion, dans JODELET D. (éd.), *Les représentations sociales*, PUF, 1989 : 31-61.
- JODELET F., L'association verbale, dans FRAISSE P., PIAGET J. (éd.), *Traité de psychologie expérimentale, VII. Langage, communication et décision*, PUF, 1972 : 97-153.
- JUNG C.G. The Association Method, *American Journal of Psychology*, 31, 1910 : 219-269.
- KOPP-BIGAULT C., WALTER M., THEVENOT A., The social representations of suicide in France. An inter-regional study in Alsace and Brittany, *Int. J. Soc. Psychiatry* 62, 2016 : 737-748.
- LE GLAZ A., HARALAMBOUS Y., KIM-DUFOR D.-K., LENCA P., BILLOT R., TAYLOR R., MARSH J., DEVYLDER J., WALTER M., BERROUIGUET S., LEMEY C., Machine learning and natural language processing in mental health, *Journal of Medical Internet Research*, JMIR Publications, 23 (5) 2021 : e15708.
- MARQUIS P., PAPINI O., PRADE H. (éd.), *Représentation des connaissances et formalisation des raisonnements*, Cepaduès Éditions, 2014.
- MOLINER M., RATEAU P., COHEN-SCALI V., *Les représentations sociales. Pratique des études de terrain*, Presses Universitaires de Rennes, 2002.
- MOSCOVICI S., *La psychanalyse son image et son public*, PUF, 1961.
- MOSCOVICI S., Social representations and pragmatic communication. *Social Science Information*, 33, 1994 : 163-177.
- MUCCHIELLI A., *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*, Armand Colin, 1996.

- NASTASE V., SZPAKOWICZ S., NAKOV P., Ó SÉAGDHA D., *Semantic Relations Between Nominals*, Morgan & Claypool Publishers, 2021.
- O'REILLY T., Why it's too early to get excited about Web3, <https://www.oreilly.com/radar/why-its-too-early-to-get-excited-about-web3/>, 2021.
- DUCROT O., Analyse de textes et linguistique de l'énonciation, dans DUCROT O. (éd.), *Les Mots du discours*, Éditions de Minuit, 1980 : 7-56.
- PAVEL S., *Précis de terminologie*, Bureau de la traduction du Canada, Hull, 2001.
- PY B., Pour une approche linguistique des représentations sociales, *Langages*, 154, 2004 : 6-19.
- ROUQUETTE M.-L., RATEAU P., *Introduction à l'étude des représentations sociales*, Presses Universitaires de Grenoble, 1998.
- WARREN P., *Introducing Psycholinguistics*, Cambridge University Press, 2013.