Évolution des fréquences et des cooccurrences des entités nommées dans les discours de la presse sur l'intelligence artificielle (2012-2022)

Panos Tsimpoukis, Pierre Ratinaud, Nikos Smyrnaios

Évolution des fréquences et des cooccurrences des entités nommées dans les discours de la presse sur l'intelligence artificielle (2012-2022)

Panos Tsimpoukis[1], Pierre Ratinaud[2], Nikos Smyrnaios[3]

[1]Université de Toulouse, LERASS – panagiotix@gmail.com

[2]Université de Toulouse, LERASS – pierre.ratinaud@univ-tlse2.fr

[3]Université de Toulouse, LERASS – nikolaos.smyrnaios@univ-tlse3.fr

Abstract

In this article we propose an analysis of 13 769 articles in the french national press which discuss artificial intelligence during the period 2012-2022. We apply a textometric analysis and an extraction of named entities using the Spacy library. This combination of methods reveals both the issues put on the agenda by the press on the subject of artificial intelligence and the main stakeholders who dominate the journalistic discourse that refers to artificial intelligence.

Keywords: named entities, textometric analysis, Spacy, artificial intelligence

Résumé

Dans cet article nous proposons une analyse des 13 769 articles de la presse nationale française qui parlent de l'intelligence artificielle pendant la période 2012-2022. Nous y appliquons une analyse textométrique ainsi qu'une extraction des entités nommées un utilisant la bibliothèque Spacy. Cette combinaison de méthodes dévoile tant les thématiques mises en agenda par la presse au sujet de l'intelligence artificielle que les acteurs principaux qui dominent le discours journalistique qui se réfère à l'intelligence artificielle.

Mots clés : entités nommées, analyse textométrique, Spacy, intelligence artificielle

1. Introduction

Le lancement de ChatGPT a généré un formidable engouement dans la sphère publique quant aux applications de l'intelligence artificielle, aux promesses de cette technologie ou aux risques potentiels qu'elle pose à la société. Cependant, le débat public autour de l'intelligence artificielle, au moins dans sa dimension médiatique, a commencé bien avant. Pendant une décennie, plusieurs cadrages ont été mobilisés par la presse afin de mettre en lumière les enjeux économiques, éthiques, de politique nationale et géopolitique internationale résultants du développement de l'intelligence artificielle. Même si la mise en agenda de ces cadrages a contribué à la transformation de l'intelligence artificielle en problème public (Bellon et Velkovska, 2023), il n'en demeure pas moins que le discours médiatique autour de cette technologie est en grande partie surdéterminé par l'activité de l'industrie (Brennen et al., 2018) et par le discours gouvernemental (Dandurand et al., 2022). Dans cet article, nous proposons une analyse simultanée de l'évolution du discours public sur l'IA dans la presse nationale française (2012-2022) et des acteurs qui sont fréquemment mentionnés dans les articles journalistiques. Cette démarche mobilise la méthode de la Classification Hiérarchique Descendante (CHD) de type Reinert et l' extraction d'entités nommées (NER), qui permet d'identifier les acteurs mentionnés dans un texte à travers l'extraction des noms et des

organisations (Bordignon, 2021). Cette combinaison de méthodes permet l'observation de la variété des acteurs mentionnés dans les articles à différentes périodes, ainsi que du contexte discursif dans lequel ces acteurs ont été mentionnés. Nous pourrons ainsi étudier l'évolution non seulement des thématiques et des enjeux abordés dans les articles, mais aussi des acteurs qui sont porteurs de ces enjeux. L'intérêt de cet article est d'évaluer l'efficacité et la pertinence de cette combinaison de méthodes.

2. Analyse textométrique

Dans un premier temps, nous avons collecté dans la base de données Europresse les articles publiés dans neuf journaux nationaux. L'échantillonnage est ici contraint par la base de données : nous avons retenu tous les journaux nationaux disponibles sur la période 20122022. Pour Crépel et Cardon (2022), l'année 2012 marque le commencement de la troisième vague de médiatisation de l'IA, déclenchée par les applications qui exploitent les données de masse et les technologies de Deep Learning. En novembre 2022, l'entreprise OpenAI a mis en ligne la première version de ChatGPT. Cette sortie a engendré une intense communication médiatique autour de l'IA. La Figure 1 présente l'évolution du nombre d'articles dans notre corpus sur la période. Il intègre l'année 2023 que nous n'analyserons pas.

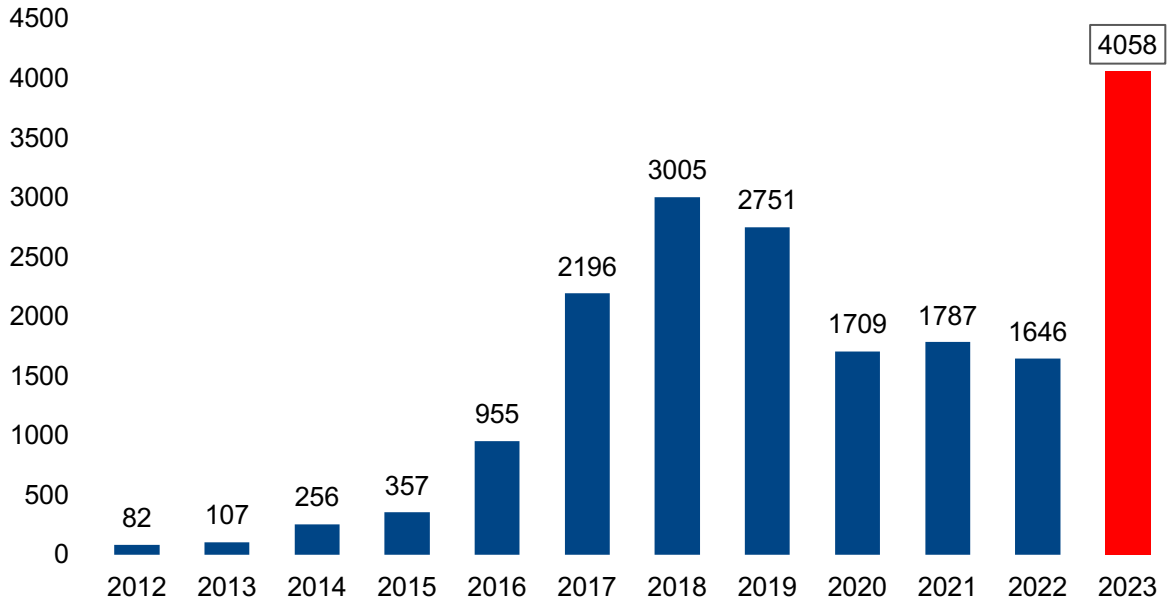

Figure 1 : Evolution des fréquences des articles par année dans le corpus. Le graphique intègre l'année 2023 qui n'est pas analysée.

Nous avons donc fait le choix de nous concentrer sur la période avant-ChatGPT, de façon à comprendre comment l'IA a été introduite dans la sphère publique dans les dix années précédentes.

Les articles extraits ont en commun de contenir les expressions « intelligence artificielle » ou « deep learning ». Le choix de ces termes se justifie par leur utilisation répandue dans les discours journalistiques qui décrivent les technologies de l'intelligence artificielle. Le Tableau 1 rend compte des fréquences d'articles dans chacun des journaux :

Journal	Effectifs d'articles
Les Échos	4072
La Tribune (France)	2826
Le Figaro	2404
Le Monde	2113
Libération	645
La Croix	533
Correspondance économique	495
Aujourd'hui en France	414
L' Humanité	267
	13 769

Tableau 1 : Effectifs d'articles par journal sur la période 2012-2022

Nous pouvons voir dans cette distribution la place occupée par les journaux du champ de l'économie (Les Echos, La Tribune) qui traduit en amont de l'analyse l'importance de cette dimension sur cette période. Au total, ce sont 13 769 articles que nous avons analysés en employant la méthode de la Classification Hiérarchique Descendante (CHD) de type Reinert (Reinert, 1983), implémentée dans le logiciel Iramuteq (Ratinaud et Marchand, 2012). Le corpus représente 10.950.983 occurrences pour 105 126 formes différentes (dont 44 783 hapax, soit 42,60 % des formes et 0,41 % des occurrences). Le corpus a été découpé en 304 738 segments de texte. L'objectif est ici de mettre en évidence les différents « mondes lexicaux » (Reinert, 1990) qui sont mobilisés par la presse pour parler de l'intelligence artificielle ou des thèmes dont le traitement sur cette période a nécessité de faire référence à ces technologies. L'analyse de l'évolution chronologique des fréquences d'apparition de ces thématiques nous amène à délimiter trois sous-périodes dans ce corpus.

2.1. Délimitation des périodes d'étude

La Figure 2 présente l'évolution des fréquences d'expression des classes lexicales issues de la classification. Le dendrogramme est présenté à gauche de la figure. Dans ce graphique, la largeur des colonnes est proportionnelle à la quantité de texte (et donc d'articles ici) de l'année, la hauteur des lignes est proportionnelle à la taille des classes (exprimée en pourcentage de segments de texte classés). Les cases colorées signalent une sur-représentation des segments de l'année considérée dans la classe. Ce graphique permet d'observer certains « décalages » entre les années : le premier décalage survient lors de la période 2013-2015, le deuxième décalage survient lors de la période 2017-2019, et le troisième décalage survient lors de la période 2020-

2021. La division finale du corpus que nous retiendrons (2012-2015, 2016-2019, 2020-2022) a été faite de manière à satisfaire deux conditions. D'une part, chaque période englobe l'un de décalages décrits ci-dessus, ce qui permet d'observer et de qualifier en quoi exactement consiste le changement des classes lexicales -et donc des thématiques- intervenant dans ces moments charnières. Ainsi, la periode 2012-2015 englobe le décalage des classes lexicales observé entre les années 2013-2015, la période 2016-2019 englobe le décalage observé entre les années 2017-2019 et la période 2020-2022 englobe le décalage observé entre les années 2020-2021. D'autre part, cette division permet une répartition équitable des années analysées.

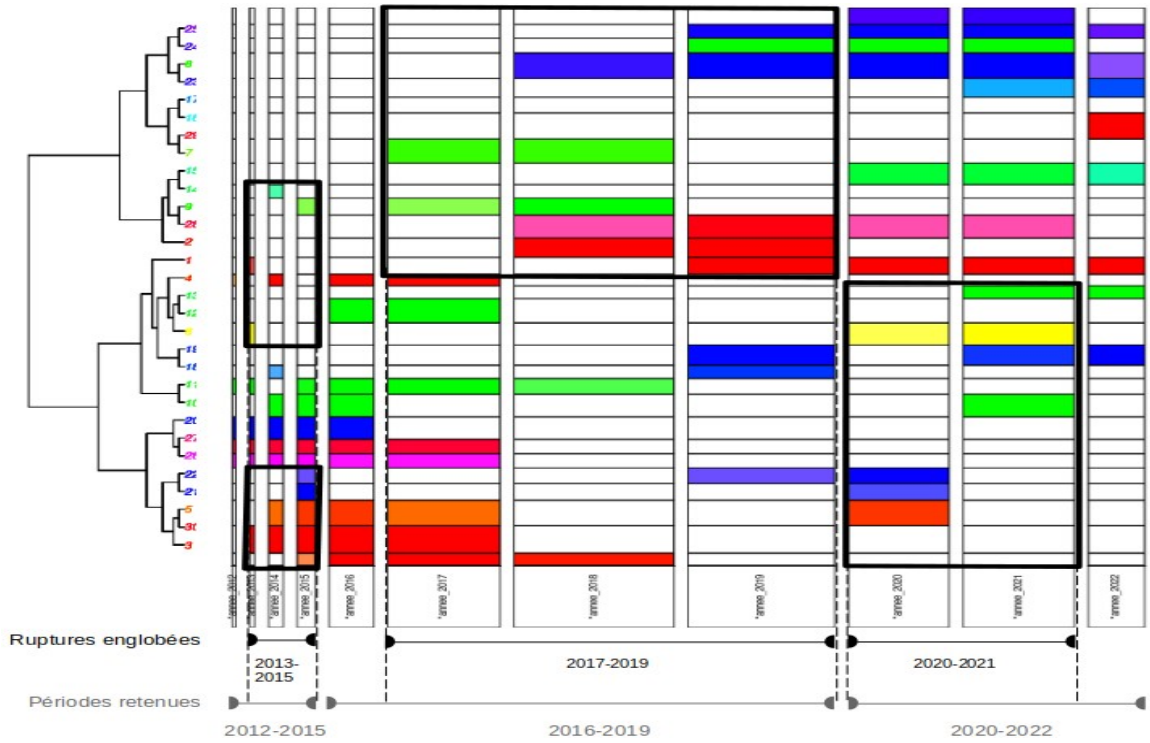

Figure 2 : Aperçu chronologique des classes lexicales pour l'ensemble du corpus (2012-2022). Les cases noires permettent de mieux distinguer les périodes pendant lesquelles nous observons un décalage significatif des différentes classes de discours.

2.2. Période 2012-2015 (1361 articles, 584 342 occurrences)

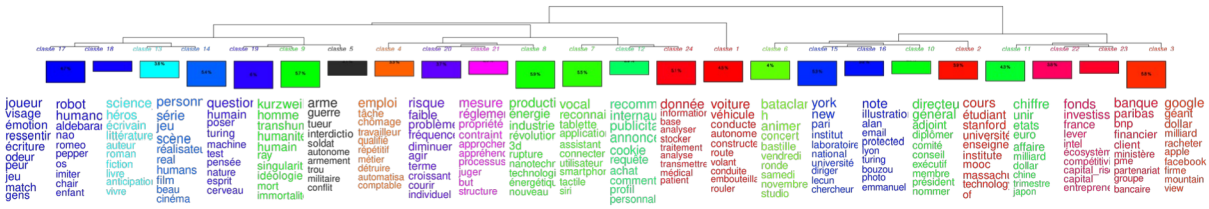

Figure 3 : Dendrogramme de l'analyse CHD de la période 2012-2015

Le discours autour de l'intelligence artificielle pendant la première période se concentre principalement sur les innovations des technologies d'intelligence artificielle. Nous trouvons particulièrement un discours qui concerne des nouvelles autour des voitures autonomes et des assistants vocaux, tels que Alexa et Siri. En 2014 nous trouvons surreprésenté un discours qui concerne les investissements des Big Tech dans des domaines d'innovation comme la réalité virtuelle, la robotique, l'intelligence artificielle ou même la médecine. Ce qui est remarquable pour cette période d'étude, c'est le fait qu'en 2015 devient surreprésenté un discours qui concerne les alertes des scientifiques renommés autour de l'intelligence artificielle. Dès la fin

de 2014, Elon Musk et Stephen Hawking expriment des craintes concernant le développement accéléré de l'intelligence artificielle, alors qu'en juillet 2015 est publiée une lettre ouverte cosignée par de nombreux scientifiques (Open Letter, 2015), qui réclament l'interdiction des « robots-tueurs ». Ces déclarations bénéficient d'une grande couverture médiatique, ce qui se reflète à la surreprésentation des classes lexicales qui contiennent le discours autour de ces déclarations.

2.3. Période 2016-2019 (14 549 articles, 6 560 358 occurrences)

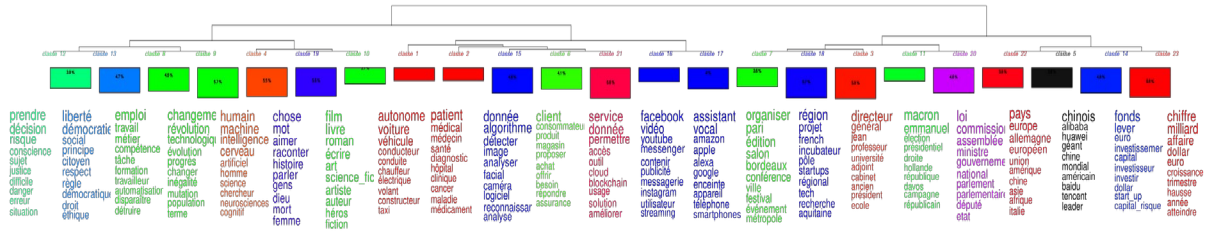

Figure 4 : Dendrogramme de l'analyse CHD de la période 2016-2019

Si la première période est quasi-monopolisée par l'innovation et les investissements des entreprises dans l'intelligence artificielle, la deuxième période est beaucoup plus riche en thématiques traitées par la presse nationale. Cette richesse trouve ses origines dans le fait qu'en mars 2018, le Président Emmanuel Macron a fait un discours au Collège de France autour de l'intelligence artificielle (Élysée, 2018), alors que le mathématicien et député Cédric Villani publie un rapport qui décrit les enjeux de l'intelligence artificielle pour l'économie nationale et européenne ainsi que les enjeux éthiques que ces technologies posent pour la société (Villani, 2018). Au niveau géopolitique, nous remarquons la mise en agenda d'un discours sur les politiques de l'Union Européenne à l'égard du développement des technologies d'intelligence artificielle au niveau européen et de la position de l'UE vis-à-vis des grandes firmes américaines du marché du numérique (GAFAM).

2.4. Période 2020-2022 (7435 articles, 3 806 283 occurrences)

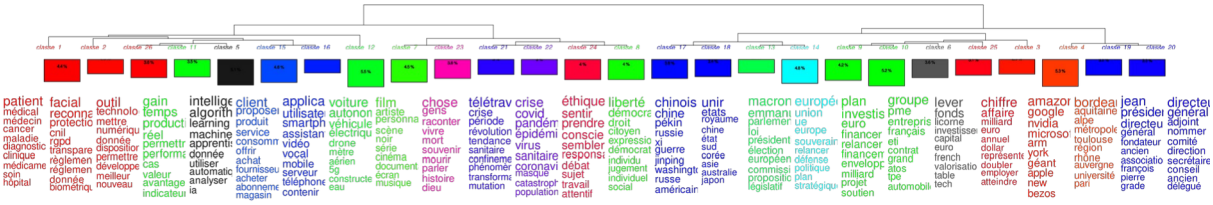

Figure 5 : Dendrogramme de l'analyse CHD de la période 2020-2022

La dernière période étudiée, en termes de thématiques, se trouve en continuité avec la période précédente. Ce qui se différencie significativement, c'est la mise en agenda d'un discours autour de la réglementation des technologies d'intelligence artificielle. En 2020, nous remarquons la surreprésentation d'un discours autour de la réglementation de la reconnaissance faciale suite à l'interdiction de la part de la CNIL (Commission nationale de l'informatique et des libertés) de l'utilisation de la reconnaissance faciale en temps réel dans les voies publiques. Le discours relatif à la reconnaissance faciale était également surreprésenté en 2019 à l'occasion de la mise en place d'une expérimentation de cette technologie au Carnaval de Nice. Toujours en lien avec les technologies de la reconnaissance faciale, nous observons pendant la période 2021-2022 la mise en agenda d'un discours relatif à la protection des données biométriques ainsi qu'un discours concernant la loi « Sécurité Globale », qui donnait le droit à la police d'accéder en direct aux caméras corporelles et aux drones.

3. Analyse des entités nommées

L'analyse textométrique nous a permis de suivre l'évolution des thématiques abordées par la presse nationale au sujet de l'intelligence artificielle au cours d'une période de dix ans. Cependant, ce type d'analyse ne fournit que peu d'indices sur les personnages ou les institutions, autrement dit les acteurs, qui occupent une position importante dans les articles.

Dans ce but, nous avons extrait dans une première étape toutes les entités nommées (lemmes associés aux noms propres) de chacune des trois périodes étudiées dans la partie précédente. Ces entités nommées incluent des personnes (PER), des organisations (ORG), des lieux (LOC) et d'autres entités diverses (MISC). Nous les avons extraits à partir de la bibliothèque Spacy (Honnibal et al., 2020). Au total, sur les trois périodes, Spacy a identifié 130 076 entités nommées différentes, qui représentent 875 163 occurrences, dont 46 118 personnalités (PER) et 25 424 organisations (ORG). Mais cette extraction n'est bien sûr pas parfaite. Ainsi, par exemple, nous trouvons des entités comme *s*, *v*, *d*, *j*. Notre hypothèse est que Spacy a extrait des initiales de prénoms mentionnées dans les textes sous la forme *P. Nom*. En plus des faux positifs, une autre limite doit être pointée : un même individu peut apparaître sous différentes formes (*Prénom Nom*, *P. Nom*, *Nom par exemple)*. Après l'extraction, nous avons calculé la fréquence des occurrences de ces entités dans les trois périodes respectivement. Le Tableau 2 présente les scores de spécificités des 15 entités les plus sur-représentées sur chaque période.

Période 2012-2015		**Période 2016-2019**		**Période 2020-2022**	
Entités	**Score de spécificités**	**Entités**	**Score de spécificités**	**Entités**	**Score de spécificités**
google	144,60	alphago	66,05	covid-19	inf
turing	60,92	amazon	47,63	covid	250,35
alan turing	49,25	google	45,16	lire	109,62
internet	48,74	orange bank	44,82	joe biden	104,83
nbic	48,20	facebook	39,82	ukraine	81,66
ray kurzweil	42,87	alexa	35,84	arm	80,32
watson	34,09	gemalto	35,70	ue	59,21
stanford	32,83	uber	33,04	stellantis	51,46
singularity university	29,95	cédric villani	32,05	biden	48,53
bruno bonnell	29,33	mounir mahjoubi	30,94	ovhcloud	46,56
coursera	27,13	echo	30,94	zoom	45,32

ibm	26,37	google home	27,11	commission	44,59
stephen hawking	25,67	station f	26,00	tsmc	38,77
web	25,22	villani	23,66	wuhan	35,22
laurent alexandre	24,81	internet	23,31	gpt-3	31,85

Tableau 2 : La liste des entités nommées présentant les scores de spécificité les plus élevés suite à l'analyse des Spécificités sur Iramuteq.

Dans la section suivante, nous décrirons quelques entités nommées dont le score de spécificité dépasse la valeur 10 pour chaque période. Ces descriptions incluront également des entités nommées qui ne sont pas répertoriées dans le Tableau 2.

3.1. Période 2012-2015

Pendant cette première période, les entités les plus spécifiques dans le corpus sont en lien étroit avec le monde technologique. On y trouve des machines comme Watson -l'ordinateur d'IBM- qui est capable de répondre aux questions, l'assistante vocale d'Apple Siri, le robot humanoïde Nao et son entreprise de construction Aldebaran, mais aussi Enigma, la machine du chiffrement utilisée par les Nazis pendant la Seconde Guerre mondiale. Nous trouvons également des personnalités comme Alan Turing -le mathématicien qui a déchiffré Enigma et qui a proposé le « test de Turing »-, Raymond Kurzweil -un ingénieur de Google dont le discours sur le transhumanisme a été très médiatisé-, Bruno Bonnell -entrepreneur dans le secteur de la robotique qui est devenu plus tard chargé du plan France 2030-, Stephen Hawking, Larry Page -le cofondateur de Google-, Alexandre Laurent -entrepreneur et futurologue- et Yann Le Cun, directeur à l'époque du laboratoire d'intelligence artificielle FAIR chez Facebook. Au niveau des termes, nous remarquons Internet, Web, NBIC, une abréviation pour les mots « Nanotechnologies, biotechnologies, informatique et sciences cognitives ». Finalement, des institutions comme Google, Stanford, MIT, Singularity University, Silicon Valley, Harvard, NSA (Agence Nationale de Sécurité des États-Unis), NASA, Apple, Facebook, MIT, Boston Dynamics, Baidu, Université Stanford se trouvent parmi les entités nommées ayant un score de spécificités élevé.

3.2. Période 2016-2019

Dans la deuxième période de notre analyse, nous remarquons une « emprise » des entités nommées ayant un grand score de spécificités par les grandes entreprises numériques et d'innovation et les applications qu'elles lancent. Nous y trouvons Orange Bank, Google, Amazon, Facebook, Uber, Gemalto -racheté par le groupe électronique Thalès-, DeepMind, Apple, Station F (incubateur des start-up), mais aussi le logiciel de jeu AlphaGo, Alexa, Google Home, Google Assistant. Une autre particularité de cette période est la grande fréquence des personnes politiques, comme Cédric Villani, qui a dirigé le rapport « Donner un sens à l'intelligence artificielle », Mounir Mahjoubi -ancien président du Conseil national du numérique-, Emmanuel Macron, Axelle Lemaire -secrétaire d'État chargée du numérique de 2014 à 2017- et le Comité Consultatif National d'Éthique. Nous trouvons aussi des entrepreneurs comme Mark Zuckerberg, Stéphane Richard -directeur d'Orange-, et l'organisation patronale MEDEF. Finalement, il est intéressant de noter que nous retrouvons

des termes comme conférence, blockchain, fintech, internet, brexit, SNCF, Davos, Crédit Mutuel. Ces termes illustrent l'incarnation du terme de l'intelligence artificielle auprès des multiples facettes de l'économie.

3.3. Période 2020-2022

Pendant la dernière période, nous observons la montée des entités qui sont liées, d'une part avec des sujets d'actualité, d'autre part avec des institutions et des personnages européens. En termes de sujets d'actualité, ce qui prédomine est l'actualité pendant la période de COVID-19. Il est indicatif que quatre sur quinze entités ayant le plus grand score de cette période se réfèrent à la période de la pandémie (Covid-19, Covid, Zoom, Wuhan). Plus loin dans la liste, nous trouvons aussi des entités comme Moderna, Post-COVID et l'OMS (Organisation mondiale de la Santé). Il est remarquable de constater aussi la présence de l'entité « Ukraine », qui met en lumière les débats autour de l'utilisation de l'intelligence artificielle dans le contexte du conflit entre la Russie et l'Ukraine. Toujours au niveau de l'actualité, nous trouvons des entités nommées comme ARM, Huawei, TSMC, NVIDIA, ou Intel, qui renvoient à la pénurie des semi-conducteurs dans l'industrie automobile, qui sont aussi nécessaires pour le développement des applications d'intelligence artificielle. Au niveau de personnages, nous trouvons Joe Biden, Jean Castex, Arvind Krishna -directeur de l'IBM-, Cédric O' -secrétaire d'État à la Transition numérique et aux Communications électroniques de 2020 à 2022-, la présidente de la Commission Européenne Ursula von der Leyen, Boris Johnson et le chargé du marché intérieur, de la politique industrielle, du numérique, Thierry Breton. La grande fréquence de ces personnes se justifie par les consultations qui ont eu lieu pendant cette période pour la réglementation de l'intelligence artificielle. Ceci se confirme également par la présence fréquente des entités nommées comme Commission, Bruxelles, DSA (Digital Services Act).

Pour compléter cette première analyse, nous noterons que les entités qui renvoient à la science-fiction, comme Steven Spielberg, Terminator, Frankenstein, Starwars, Hollywood, Ex Machina, Isaac Asimov, Stanley Kubrick ou HAL, sont surreprésentées pendant la première période, mais leur proportion diminue dans la deuxième et la troisième période. Il semble que l'imagination cède progressivement la place à des applications concrètes d'intelligence artificielle.

4. Visualisation des entités nommées des personnes et des organisations

Dans l'objectif de visualiser les relations, dans chacune des périodes, entre les entités nommées des personnes (PER) et des organisations (ORG) détectées, nous avons produit leur graphe de cooccurrences. Le principe consiste à compter les apparitions simultanées d'entités dans un même article avant de sommer les résultats obtenus sur l'ensemble des articles d'une même période. Nous avons fixé une fréquence minimum pour les entités sélectionnées, proportionnelle à la taille de chacune des périodes (3 pour la période 2012-2015 (1361 articles), 30 pour la période 2016-2019 (14 549 articles) et 15 pour la période 2020-2022 (7435 articles)). Nous avons aussi exigé un minimum de 3 co-occurrences entre les entités nommées. De cette manière, les graphes résultants visualisent les entités nommées qui apparaissent de manière fréquente dans les articles de chaque période et qui cooccurrent plus de 3 fois avec d'autres entités nommées dans les articles. Les graphes qui en résultent font apparaître les principaux acteurs (personnes et organisations) qui ont été mentionnés dans les articles de presse qui parlent de l'intelligence artificielle.

Pendant la période 2012-2015 (Figure 6.a.), les acteurs souvent mentionnés relèvent du monde technologique et numérique, alors que le monde politique est complètement absent. Pendant la

période 2016-2019 (Figure 6.b.) nous observons un enrichissement spectaculaire des acteurs énoncés dans les articles. Nous trouvons dans le graphe un pôle des acteurs du secteur numérique, un pôle des acteurs de l'Union Européenne et un pôle des acteurs de la politique nationale. La présence renforcée de ces trois pôles illustre l'existence d'un discours qui tourne autour de la stratégie nationale et européenne pour le développement de l'intelligence artificielle, mais aussi autour de la réglementation des technologies qui relèvent de l'IA. Une nouveauté de cette période est l'apparition d'un pôle qui relève du secteur de l'aviation et de la défense. Ce pôle met en valeur un ensemble d'acteurs qui sont énoncés très probablement dans des articles qui concernent l'emploi de l'intelligence artificielle dans le secteur militaire. Lors de l'analyse lexicale, nous n'avons pas repéré une classe de discours surreprésentée concernant cette thématique, ainsi nous ne sommes pas en mesure de savoir dans quel contexte exactement ces acteurs sont énoncés. Cependant, l'existence de ce pôle, malgré l'absence d'une classe lexicale distincte qui concerne ce sujet, démontre que l'exploration des entités nommées enrichit l'analyse lexicale et qui peut attirer notre attention sur des aspects que nous n'avions pas explorés auparavant. Pendant la période 2020-2022 (Figure 6.c.) nous remarquons l'existence d'un grand pôle, dont l'acteur central est la CNIL (Commission nationale de l'informatique et des libertés). Les entités nommées qui cooccurrent avec cette institution (Thales, Orange, Atos, Capgemini, mais aussi des acteurs de la politique nationale), démontrent que ce pôle découle d'un discours qui concerne la protection des données personnelles et les réglementations mises en place à ce propos. Finalement, nous observons le renforcement d'un pôle avec des acteurs internationaux et d'un autre pôle avec des acteurs de l'industrie des microprocesseurs, ce dernier démontrant la consolidation d'un discours autour de l'activité économique accrue de ce secteur de l'industrie.

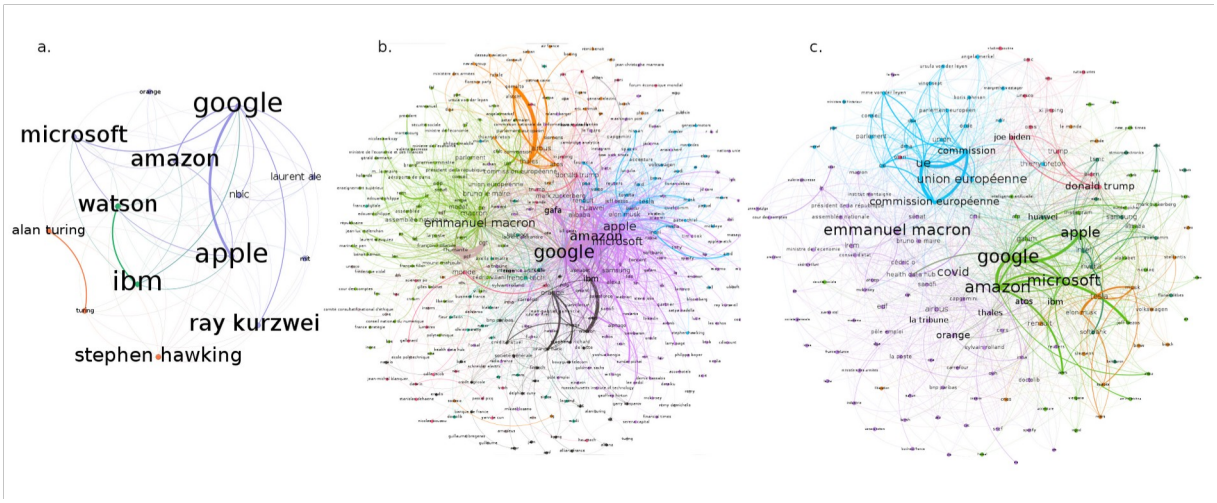

Figure 6 : De gauche à droite : Visualisation des cooccurrences des entités nommées des personnes et organisations apparues dans les articles de la période 2012-2015 (a), 2016-2016 (b) et 2020-2022 (c).

5. Conclusion

Cette recherche nous a permis de repérer tant les thématiques mises en agenda par la presse au sujet de l'intelligence artificielle que les acteurs qui semblent avoir alimenté le discours autour de cette technologie au cours des dix dernières années. Nous constatons en effet une cohérence entre les acteurs et les sujets abordés par la presse, ce qui apporte une autre granularité à l'analyse textométrique des articles tout en mettant en lumière les acteurs principaux mis en avant par les discours journalistiques. Une amélioration de la bibliothèque Spacy en ce qui

concerne la reconnaissance des entités nommées pourrait apporter une plus grande finesse à ce type d'analyse.

Bibliographie

Atif J., Burgess J. P. et Ryl I. (2022). *Géopolitique de l'IA. Les relations internationales à l'ère de la mise en données du monde*. Paris : Le Cavalier Bleu.

Bellon A. et Velkovska J. (2023). L'intelligence artificielle dans l'espace public : du domaine scientifique au problème public. *Réseaux*, 240 (4), 31–70. https://doi.org/10.3917/res.240.0031

Bordignon F. (2021). *Le véhicule autonome dans les discours médiatisés : présentation des méthodes d' analyse d' articles de presse et de tweets*. https://enpc.hal.science/hal-03506806

Brennen A. J. S., Howard P. N. et Nielsen R. K. (2018). An Industry-Led Debate : How UK Media Cover Artificial Intelligence. *Reuters Institute for the Study of Journalism Fact Sheet*, *December*, 1–10. DOI: 10.60625/risj-v219-d676

Crépel M. et Cardon D. (2022). Robots vs algorithmes. *Réseaux*, 232-233 (2), 129–167. https://doi.org/10.3917/res.232.0129

Dandurand G., Blottière M., Jorandon G., Gertler N., Wester M., Chartier-Edwards N., Roberge J. et McKlve F. (2022). Training the News : Coverage of Canada's AI Hype Cycle (2012 – 2021). *Shaping 21st-Century AI*. https://espace.inrs.ca/id/eprint/13149/

Élysée. (2018). *Discours du Président de la République sur l'intelligence artificielle*. Élysée.

https://www.elysee.fr/emmanuel-macron/2018/03/29/discours-du-president-de-la-republique-surlintelligence-artificielle

Honnibal M., Montani I., Van Landeghem S. et Boyd A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. *Zenodo*. https://doi.org/https://doi.org/10.5281/zenodo.1212303

McCombs M. (2005). A Look at Agenda-setting : Past, present and future. *Journalism Studies*, 6 (4), 543–557. https://doi.org/10.1080/14616700500250438

Nocetti J. et Seaman J. (2019). L'affaire Huawei. Un miroir de la guerre technologique sinoaméricaine. In T. de Montbrial (Ed.), *Ramses 2020. Un monde sans boussole ?* Paris : Institut Français des Relations Nationales, 294–297. https://www-cairn-info.gorgone.univ-toulouse.fr/un-monde-sans-boussole--9782100801138page-294.htm

Open Letter. (2015). *Autonomous Weapons Open Letter : AI & Robotics Researchers*. Future of Life Institute. https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/

Ratinaud P. et Marchand P. (2012). Application de la méthode Alceste à de "gros" corpus et stabilité des "mondes lexicaux" : analyse du "CableGate" avec Iramuteq. In *Actes Des 11e Journées Internationales d'analyse Statistique Des Données Textuelles*, Liège, 835–844.

Ratinaud P., Smyrnaios N., Figeac J., Cabanac G., Fraisier O., Hubert G., Pitarch Y., Salord T. et Thonet T. (2019). Structuration des discours au sein de Twitter durant l'election présidentielle francaise de 2017 : Entre agenda politique et représentations sociales. In *Réseaux : communication, technologie, société*, 2-3 (214-215), 171-208. https://doi.org/10.3917/res.214.0171

Reinert M. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Cahiers de l'analyse Des Données*, 8 (2), 187–198.

Reinert M. (1990). Alceste une méthodologie d'analyse des données textuelles et une application : Aurelia De Gerard De Nerval. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 26 (1), 24–54. https://doi.org/10.1177/075910639002600103

Villani C., Schoenauer M., Bonnet Y., Berthet C., Cornut A.-C., Levin F. et Rondepierre B. (2018). *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*. Mission Villani sur l'intelligence artificielle, 2018. hal-01967551

Sebbah B., Bousquet F. et Cabanac G. (2023). Le journalisme scientifique à l'épreuve de l'actualité « Tout Covid » et de la méthode scientifique : les journalistes scientifiques soudain au centre de la production de l'information. *Les cahiers du Journalisme*, 2 (8-9). DOI : 10.31188/CaJsm.2(8-9).2022.R119