

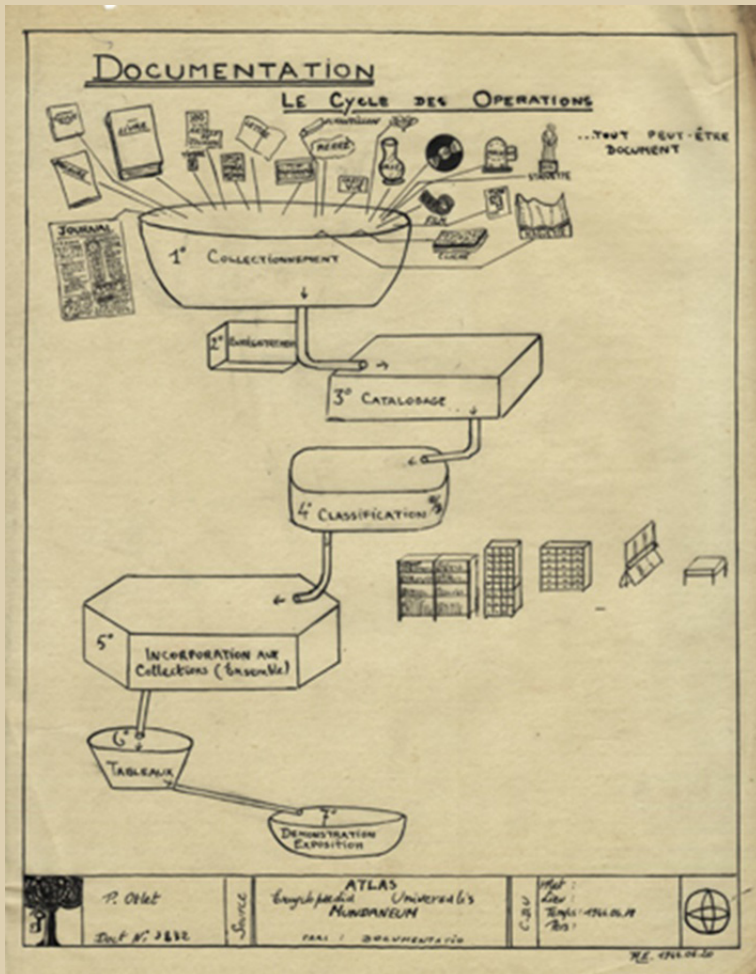
AIDa informazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 3-4

ANNO 41

LUGLIO-DICEMBRE 2023



AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:

Università della Calabria

Direttore Scientifico:

Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:

Fabrizia Flavia Sernia

Comitato scientifico:

Anna Rovella, *Università della Calabria*;

Maria Guercio, *Sapienza Università di Roma*;

Giovanni Adamo, *Consiglio Nazionale delle Ricerche* †;

Claudio Gnoli, *Università degli Studi di Pavia*;

Ferruccio Diozzi, *Centro Italiano Ricerche Aerospaziali*;

Gino Roncaglia, *Università della Toscana*;

Laurence Favier, *Université Charles-de-Gaulle Lille 3*;

Madjid Ihadjadene, *Université Vincennes-Saint-Denis Paris 8*;

Maria Mirabelli, *Università della Calabria*;

Agustín Vivas Moreno, *Universidad de Extremadura*;

Douglas Tudhope, *University of South Wales*;

Christian Galinski, *International Information Centre for Terminology*;

Béatrice Daille, *Université de Nantes*;

Alexander Murzaku, *College of Saint Elizabeth, USA*;

Federico Valacchi, *Università di Macerata*.

Comitato di redazione:

Antonietta Folino, *Università della Calabria*;

Erika Pasceri, *Università della Calabria*;

Maria Taverniti, *Consiglio Nazionale delle Ricerche*;

Maria Teresa Chiaravallotti, *Consiglio Nazionale delle Ricerche*;

Assunta Caruso, *Università della Calabria*;

Claudia Lanza, *Università della Calabria*.

Segreteria di Redazione:

Valeria Rovella, *Università della Calabria*

Editrice: Cacucci Editore S.a.s.

Via D. Nicolai, 39 – 70122 Bari (BA)

www.cacuccieditore.it

e-mail: riviste@cacuccieditore.it

Telefono 080/5214220

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare topics quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE – Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by SerialsSolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti dal comitato scientifico sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 41

N. 3-4 – luglio-dicembre 2023

CACUCCI  EDITORE
BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2023 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.it> e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

Sommario

Contributi

ALESSANDRO ALFIER, Per una rigenerazione teorica dell'archivistica in Italia, a partire dal concetto di documento	9
CAMILLE ARSEL, MARC TANTI, Étude des informations relayées par les médias de presse et sociaux concernant les agents biologiques et chimiques militarisables, dans le conflit russo-ukrainien	27
MARIO CIAMPI, ERIKA PASCERI, GRAZIA SERRATORE, Lo standard FHIR e il Fascicolo Sanitario 2.0. Sviluppo di una Implementation Guide per il Taccuino personale dell'assistito	45
ELEONORA DE LONGIS, Il <i>Corpus Inscriptionum Latinarum</i> dall'analogico al digitale	59
ISABELLA FLORIO, GIORGIA MIGLIORELLI, SARA SANTORSA, Per una Carta dei servizi nel contesto bibliotecario italiano: il caso del CNR	75
ELEONORA LUZI, La digitalizzazione come strumento per acquisire informazioni: l'UPP e la certificazione di processo	95
FRANCESCA PARISI, Dall'esperienza alla memoria. La formazione e la conservazione del carteggio dell'Arma dei Carabinieri riferito al primo conflitto mondiale	119
ANNA ROVELLA, ASSUNTA CARUSO, MARTIN CRITELLI, FRANCESCA M.C. MESSINITI, Knowledge extraction, research projects and archives management	145
MARCIN TRZMIELEWSKI, Les activités informationnelles des professionnels de santé : état de l'art dans une perspective interdisciplinaire et internationale	157

Note e rubriche

CLAUDIO GRIMALDI, Le evoluzioni delle applicazioni di IA nel campo linguistico: quale futuro per il lavoro terminologico?	181
ROBERTO GUARASCI, La valutazione delle fonti documentali. Note in margine al testo di Stefano Moscadelli <i>Dal ricordo al racconto</i>	185

Contributi

Étude des informations relayées par les médias de presse et sociaux concernant les agents biologiques et chimiques militarisables, dans le conflit russo-ukrainien

Camille Arsel*, Marc Tanti**

Abstract: Russia's invasion of Ukraine in February 2022 provoked the largest war in Europe since World War II, raising the risk of use of weapons of mass destruction. The objective of this study is to identify, based on lexicometric and descriptive analyzes of a corpus of data from press and social media, the main actors in communication around this issue from its outbreak until to April 2022. This study also presents the spatial and temporal distribution of the information retained. The methodological approach used here is of interest both in public health and in CIS. Few associated works in fact, both analyzes on social media and press media concerning health issues.

Keywords: Social media, Press media, Lexicometric analysis, Russian-Ukrainian conflict, Weapons of mass destruction.

1. Introduction - Etat de l'art

L'invasion de l'Ukraine par la Russie en février 2022 a provoqué la plus grande guerre en Europe depuis la Seconde Guerre mondiale. Elle a entraîné des milliers de morts civils et militaires. La presse a rapporté dans les conflits récents, comme celle en Syrie ou en Irak, outre l'usage d'armes conventionnels, l'usage d'armes de destruction massive (ADM) (NATO 2023a).

Ce type d'armes désigne selon l'Organisation des Nations unies, les armes dites non conventionnelles, c'est-à-dire nucléaires, radiologiques, biologiques, chimiques et explosives (NRBCe) (ONU 1946). Pour l'Organisation du traité de l'Atlantique nord, « la prolifération des ADM et de leurs vecteurs pourraient avoir des conséquences incalculables » (NATO 2023b) sur la santé et la sécurité des êtres vivants (humains, animaux, végétaux) à échelle locale, natio-

* Centre d'Épidémiologie et de Santé Publique des Armées (CESPA), Marseille, France. camille.arsel.pro@gmail.com.

** Centre d'Épidémiologie et de Santé Publique des Armées (CESPA), Marseille, France. UMR 1252-SESSTIM-AMU, & Institut Méditerranéen des Sciences de l'Information et de la Communication IMSIC-AMU, Marseille, France. mtanti@gmx.fr.

nale voire mondiale, et dans le cadre du conflit russo-ukrainien des impacts géopolitiques sans précédent.

Selon A. Melin, la problématique du bioterrorisme « attire souvent l'attention dans les médias et les débats scientifiques, ainsi que dans les rapports gouvernementaux » (Melin 2021).

De plus, de nombreuses études ont démontré le rôle central des médias de presse, mais aussi des réseaux sociaux dans la dissémination ultra-rapide de l'information ou de messages gouvernementaux, en cas de crises, d'évènements sanitaires, d'incidents ou d'épidémies, comme par exemple lors de la pandémie de Covid-19 (González-Padilla and Tortolero-Blanco 2020).

D'autres études démontrent l'intérêt d'utiliser les données des médias sociaux, comme Twitter, afin de réaliser une surveillance des maladies infectieuses, comme par exemple le Zika (Masri et al. 2019), la maladie d'Ebola (Hossain et al. 2016) ou encore les pratiques en santé, comme par exemple, l'injection de drogues (Tanti et al. 2018). Ces moyens de communication peuvent s'avérer d'un intérêt en santé publique afin d'anticiper les risques sur la population et la protéger des futures crises.

L'objectif de cet article est, à partir d'une analyse de données issues de médias de presse et de médias sociaux, de déterminer, dans le cadre de l'actuel conflit russo-ukrainien, quels sont les principaux acteurs de la communication et les thématiques véhiculées sur ces médias concernant la problématique des agents biologiques et chimiques militarisables. Ce travail a également pour objectif de présenter la répartition spatiale et temporelle des informations collectées sur ces médias, ainsi que leur divergence, depuis le 24 février 2022, date de l'invasion de l'Ukraine par la Russie jusqu'au mois d'avril de la même année.

2. Matériel et Méthode

Notre étude est une étude exploratoire mixte de données textuelles extraites de médias de presse en ligne (anglophones et francophones) et de médias sociaux (Twitter et forums de discussions).

Pour ancrer de manière théorique nos travaux, selon M.R. Guével, la méthode mixte permet de « combiner les forces des méthodes qualitatives et quantitatives [...] et a l'avantage de permettre l'intégration de plusieurs perspectives » (Guével et Pommier 2012). Concernant la période de collecte, tous les corpus ont été récoltés à partir du 24 février 2022, date de l'invasion de l'Ukraine par la Russie. Les dates de fin de collecte sont différentes selon les sources utilisées, afin de limiter le volume d'informations recueillies. Nous partons de l'hypothèse que cette confrontation de sources permettra de collecter et d'analyser, durant notre période d'étude, les informations émises par les différentes catégories d'acteurs communiquant autour de la problématique :

journalistes, personnalités politiques, forumers, utilisateurs de Twitter (désormais X) ...

2.1. Création du corpus de données

2.1.1. *Corpus « médias de presse »*

Nous avons fait le choix méthodologique de collecter des articles de presse francophones et anglophones. En effet, nous souhaitions avoir un corpus issu de plusieurs États, afin de repérer les similitudes ou divergences de discours en fonction du pays d'origine.

Le corpus a été constitué à partir d'Europresse, un agrégateur de presse locale, nationale, internationale, généraliste et spécialisée. Nous avons complété par une collecte manuelle d'articles anglophones disponibles sur Internet, en utilisant les mêmes requêtes de recherche. Ces derniers ont été traduits grâce au logiciel libre DeepL. Les articles de presse issus d'abonnements payants aux journaux en ligne ont été exclus. Nous avons réalisé, sur Europresse, une recherche avancée rétro-prospective afin d'étendre au maximum la requête pour collecter un nombre d'articles conséquent à partir de la date du début du conflit et éviter un silence documentaire (« informations pertinentes non affichées lors d'une recherche » (Nicolas 2007)).

Nous avons ainsi collecté 385 articles de presse régionale ou nationale francophones. Un nettoyage du corpus selon la pertinence des articles et l'intégration du contexte actuel a permis d'aboutir à 109 articles francophones durant la période du 24 février au 26 avril 2022. Les variables d'intérêts pour chaque texte étaient : la date de parution de l'article, le nom du média de presse et enfin le pays d'origine ou du siège social du média.

Concernant les Médias de presse anglophones, leur collecte s'est effectuée manuellement sur Google du 24 février au 02 mai 2022. Nous avons sélectionné 59 articles. Dans un souci d'harmonisation du corpus, tous les articles ont été traduits en français, et nous avons conservé les mêmes variables d'intérêts que pour les articles francophones.

2.1.1. *Corpus « médias sociaux »*

Nous avons décidé de scinder cette source d'informations en deux avec un sous-échantillon provenant d'un réseau social (Twitter) et un autre provenant de commentaires postés sous les articles, forums ou vidéos de presse. En effet, nous voulions que ce corpus représente deux parties de la population, ceux qui utilisent les réseaux sociaux et ceux qui ne s'en servent pas. Nous avons également fait le choix d'utiliser Twitter car il permet de collecter facilement

des données sans qu'il ne soit nécessaire de faire partie de groupes privés. De plus, comme il permet de commenter les événements personnels ou d'actualité, nous voulions repérer les lanceurs d'alerte. La collecte des commentaires d'articles de presse en langue française a été manuelle. Tandis que l'échantillon de *tweets* a été collecté grâce au logiciel de veille sur les réseaux sociaux Radarly, commercialisé par la société Linkfluence. Radarly permet de capter des messages ou commentaires grâce aux requêtes soumises, en diluant au maximum le bruit documentaire (« accumulation d'une quantité d'informations non pertinentes ») (Nicolas 2007). Pour éviter ce problème, dans cette étude, les *tweets* et leurs réponses ont été recueillis en langues française et anglaise, sans *retweets* (consistant à un republier un *tweet*). Compte tenu de la masse importante de données et d'informations disponibles en ligne, nous avons optimisé les requêtes de recherche pour que la collecte soit pertinente.

Concernant le corpus issu de Twitter, le volume initial de *posts* collecté en anglais et en français a engendré un volume de données trop important, entraînant une perte ou une dilution des informations les plus pertinentes, dans un bruit documentaire. Nous avons donc décidé de nous restreindre qu'à la langue française et de réaliser une restriction temporelle. Nous avons ainsi collecté un échantillon de 5 621 *tweets* bruts francophones du 24 février au 15 avril 2022. Après un nettoyage de la base de données textuelles, avec suppression des doublons et des *posts* illisibles, le corpus final était constitué de 2 019 *tweets*. Les variables d'intérêts pour ce corpus étaient : la date de publication du *tweet*, le pays d'origine du compte ayant publié (si indiqué) et le nombre d'actions d'engagements (nombre de fois où des utilisateurs ont interagi avec le *post* (*like*, *retweet*, partage...)). A noter que nous disposions, grâce à Radarly, d'autres informations sur les *tweets* comme : la tonalité des *posts* (neutre, négative, positive) ; le genre et la profession de l'auteur ; la ville et région de l'auteur etc. Cependant, le nombre important de données manquantes et la véracité des informations personnelles fournies par les utilisateurs étant impossible à vérifier, nous ne les avons exploitées que partiellement.

Concernant les commentaires de forums issus de médias de presse numérique ou vidéos francophones, nous avons pu collecter manuellement un échantillon de 68 commentaires entre le 24 février et le 24 avril 2022.

2.2. Analyses des données

Notre approche méthodologique associe des analyses descriptives à des analyses lexicométriques (étude quantitative du lexique) et comparatives.

2.2.1. Analyse descriptive

Dans un premier temps, nous avons réalisé des analyses descriptives pour chaque corpus collecté à l'aide de l'outil Excel. L'intérêt principal était d'identifier les caractéristiques qualitatives et quantitatives propres aux données de chaque source. L'objectif était de contextualiser les données en repérant les principaux acteurs de la communication sur la thématique. Il s'agissait, par exemple, de mettre en évidence les médias de presse qui ont publié le plus de contenus sur la période donnée. Il était aussi intéressant de connaître le nombre de publications ou *posts* quotidiens durant les périodes de collecte. Les pics informationnels peuvent ainsi être liés ou non à des événements qui se sont produits durant ces périodes. Enfin, nous voulions géolocaliser la provenance des données en fonction des sources, en identifiant les pays émettant le plus d'informations (exemple : le pays d'origine pour les médias de presse ou pour les commentaires sous les articles). Pour Twitter, le pays d'affiliation des utilisateurs a été fourni par Radarly. Les répartitions géographiques par pays ont été cartographiées via l'outil Excel. Grâce à ces analyses, nous avons comparé les résultats des différents types de sources et confronté leurs répartitions spatiales et temporelles.

2.2.2. *Analyses lexicométriques et comparatives*

Nous souhaitons également identifier les principales thématiques dans les différents corpus, concernant les armes B et C ainsi que le bioterrorisme dans le contexte particulier de la guerre en Ukraine. Pour cela, nous avons choisi de réaliser des analyses statistiques multidimensionnelles de textes. Nous avons utilisé le logiciel IRaMuTeQ, développé par P. Ratinaud et le Laboratoire d'études et de recherches appliquées en sciences sociales (LERASS). Ce logiciel d'analyses lexicométriques s'appuie sur R (logiciel de statistique) et sur le langage Python. IRaMuTeQ est une implémentation libre de la méthode d'analyse lexicale par contexte d'un ensemble de segments de texte (méthode ALCESTE) de M. Reinert. Nous avons d'abord prétraité manuellement toutes les bases de données, afin de retirer les informations non utiles (URL, caractères spéciaux, dates...) et donner de la clarté à des termes et harmoniser toutes les données renvoyant au même concept. Tous les corpus ont été traduits en français via le site DeepL. Un reformatage des bases de données textuelles a aussi été nécessaire pour définir des métadonnées. Nos métadonnées pour tous les corpus étaient le type de média, la source et le pays d'origine.

Afin d'identifier les grandes thématiques traitées et connaître leur poids dans le corpus, nous avons réalisé des classifications hiérarchiques descendantes (CHD). Pour cela IRaMuTeQ propose des options de lemmatisation (ramener tous les mots au singulier ou tous les verbes à l'infinitif) et de segmentation de textes de 40 occurrences (nombre de mots tout confondus) en moyenne pour les corpus de médias de presse. Comme les textes sur les mé-

dias sociaux sont limités en caractères, nous avons réalisé des CHD sur textes entiers. Notre but était de créer des classes lexicales, c'est-à-dire des classes de mots appartenant à des rubriques lexicales ayant les mêmes caractéristiques phonologiques, sémantiques, syntaxiques..., sur la base d'un de ces critères de ressemblance, pour pouvoir ensuite décrire le profil de chaque classe lexicale. La projection des CHD se fait sous la forme de dendrogramme, afin d'organiser les données en arborescence en fonction de leurs similitudes. Nous avons pu calculer des Chi 2 d'association, indicateurs de la distance entre deux groupes de « textes » en ce qui concerne la distribution des termes. Le seuil de significativité est fixé à 0,05.

Enfin, nous avons comparé les différents corpus médias de presse (francophones, anglophones) et médias sociaux (Twitter et forums) concernant les principaux acteurs de la communication sur la thématique des agents B et C. Nous avons également analysé les pics informationnels et la géolocalisation des données. Cela nous a permis d'identifier les points de similitudes ou de divergences entre les sources et de visualiser l'évolution de ces thématiques dans le temps.

3. Résultats de l'étude

3.1. Analyse descriptive du corpus « média de presse »

Le corpus médias de presse a été scindé en deux entre les journaux francophones et anglophones. Les dynamiques temporelles, informationnelles et acteurs de la communication sont présentés ci-dessous, pour chacun des deux sous-corpus, alors que la répartition géographique des données est présentée pour l'ensemble du corpus.

3.1.1. *Dynamique temporelle et acteurs de la communication de la presse francophone*

Entre le 24 février et le 26 avril 2022, nous avons retenu 109 articles de presse. Sur cette période, cinq pics informationnels sporadiques ont été détectés les 11, 22, 24 mars et 7, 12 avril avec respectivement 10, 9, 13, 7 et 10 articles publiés. Le 24 mars se caractérise par le nombre de publications qui correspond à lui seul à 12 % du corpus (journée où l'OTAN déclarait répondre si la Russie avait recours à l'arme chimique). L'ensemble de ces pics représente plus de la moitié des articles du corpus francophone (51 %).

Les principaux médias francophones ayant traité le sujet des armes B et C dans le conflit russo-ukrainien sont deux quotidiens nationaux français : “Le Monde” (9 articles) et “Le Figaro” (7 articles) ; ainsi que deux quotidiens régionaux “Le Parisien” et “Ouest-France” (6 articles chacun). Vingt-six pour cent des articles provenaient de ces quatre journaux, sur un total de 46. Concernant les pays d’origine de ces médias, la France se place largement en tête avec 90 articles (soit 83 % du corpus), devant la Belgique (6 articles) et le Canada (4 articles).

3.1.2. Dynamique temporelle et acteurs de la communication de la presse anglophone

Comme pour la presse francophone, il y a eu cinq pics informationnels les 11, 25 mars et les 6, 12, 14 avril avec respectivement 3, 4, 3, 7 et 4 articles sur la période. Sur 59 articles anglophones, ces pics représentent 34 % des publications dont 12 % le 12 avril 2022 (lendemain d’une probable attaque chimique russe à Marioupol en Ukraine).

Concernant les médias anglophones, l’agence de presse russe Tass a publié le plus avec huit articles sur la période du 24 février au 02 mai 2022. Les médias britanniques “The Guardian” et “BBC News” ont quant à eux édité trois publications chacun. Les 31 autres organes de presse ont publié en moyenne 1,45 articles sur la période. Les pays qui ont publié davantage dans ce corpus sont le Royaume-Uni, les États-Unis, la Russie et l’Ukraine avec respectivement 15, 14, 8 et 5 articles sur les 59 au total.

3.1.3. Répartition spatiale et temporelle du corpus

Sur un total de 168 articles de presse francophones et anglophones, 53,5 % étaient issus de France, soit 90 publications. Les pays d’où provenaient la plupart des autres articles étaient le Royaume-Uni (10,1 %) les États-Unis (10,1 %) et la Russie (4,7 %).

Les pics d’informations suivaient globalement une même évolution temporelle entre les médias francophones et anglophones (Fig. 1). Néanmoins, les pics anglophones sont moins marqués que ceux de la presse française les 11, 24 mars et 7 avril. Inversement, le 12 avril, la presse anglophone publie davantage que la presse francophone (respectivement 12 % et 9 %).

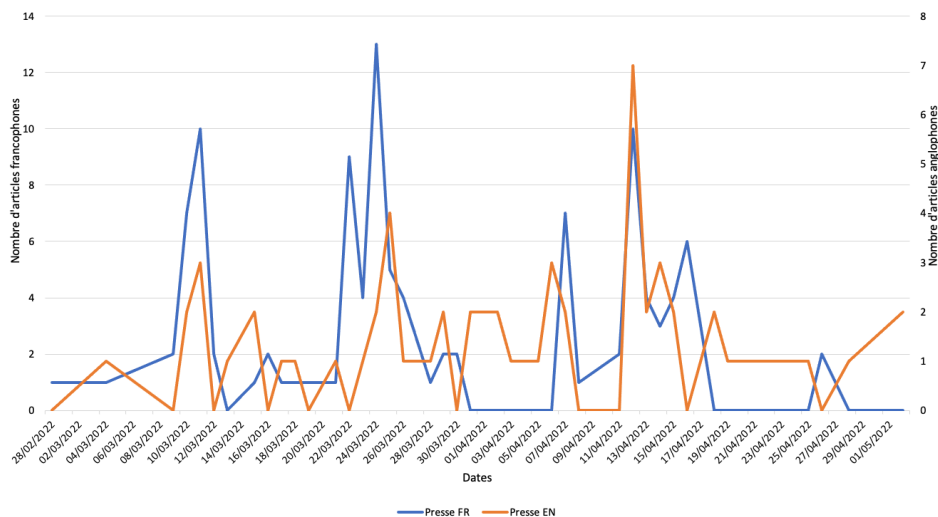


Figure 1. Comparaison de la répartition par jour des médias de presse francophones et anglophones.

3.2. Analyse descriptive du corpus « média sociaux »

Le corpus « médias sociaux » a été scindé en deux pour les analyses, avec d'une part les données issues du réseau social Twitter et d'autre part les commentaires d'articles ou de vidéos de presse.

3.2.1. Dynamique temporelle et acteurs de la communication dans les commentaires

Entre le 24 février et le 24 avril 2022, nous avons collecté 42 commentaires postés sous des articles de presse en ligne et 26 commentaires provenant de vidéos YouTube de médias de presse. Nous avons identifié quatre pics majeurs d'informations les 12, 22, 25 mars et le 12 avril avec respectivement 14, 9, 7 et 6 publications sur notre échantillon de 68 commentaires. Ces pics représentaient 53 % de tous les commentaires du public postés dans les espaces de discussion. Les commentaires provenaient à 94 % de France, le reste d'Australie et de Chine.

Les utilisateurs peuvent commenter et donner leurs avis sur les espaces de discussion d'articles ou de vidéos de presse (en ligne). Les quotidiens les plus utilisés pour communiquer sont français (avec « Le Parisien » 11 posts, « La Dépêche » 6 posts et « Le Figaro » 6 posts) et européens (avec le journal « 20 Minutes » 10 posts). Ceux-ci comptabilisent 49 % de tous les commentaires issus de 17 sources.

3.2.2. Répartition spatiale et temporelle, actions d'engagement sur Twitter

Sur la période du 24 février au 15 avril 2022, 2 019 *tweets* ont été retenus. La moyenne était de 39,5 *tweets* journaliers avec une médiane de 24 *tweets* par jour et un maximum global de 159 *tweets* enregistrés le 9 mars (contexte : allégations russes sur la présence de laboratoires américains d'armes biologiques en Ukraine). Le nombre de publications a augmenté durant trois périodes : entre le 6 et le 14 mars avec un pic de données le 9 mars ; entre le 21 et le 27 mars avec un pic le 25 mars à 142 *tweets*/jour ; et enfin entre le 11 et le 13 avril avec un pic de 143 *tweets*/jour le 12 avril. De plus, 15 421 actions d'engagement ont été opérées sur la période du 18 mars au 15 avril 2022. Nous avons observé un lien entre le nombre de *tweets* publiés et les actions d'engagement, sur un même jour ou à quelques jours d'intervalle. Par exemple, le 28 mars illustre le maximum d'actions d'engagement, soit quelques jours après la vague de *tweets* publiés entre le 20 et le 26 mars.

Sur 2 019 publications, 682 contenaient l'information sur le pays d'origine du *tweet*, soit 34 %. Nous avons identifié trois zones géographiques depuis lesquelles les posts ont été publiés : l'Europe de l'Ouest (France, Royaume-Uni, Suisse, Belgique) avec 75 % de publications, l'Amérique du Nord (États-Unis et Canada) avec 17 % et l'Europe de l'Est (Russe, Biélorussie, Ukraine) avec 2 %. Concernant la démographie des utilisateurs de Twitter dans ce corpus, 68,7 % sont des hommes dont 16,4 % ont entre 55 et 64 ans.

3.3. Analyse comparative des corpus « médias sociaux » et « médias de presse »

Le nombre de publications journalières issues des médias sociaux et des médias de presse suivent la même tendance temporelle. Sur la période du 24 février au 1er mai 2022, les pics informationnels sont quasiment similaires pour la presse internationale et les médias sociaux. Ces pics sont situés entre le 8 mars et le 13 mars, entre le 21 et le 26 mars ainsi qu'entre le 11 et le 16 avril (Fig. 2).

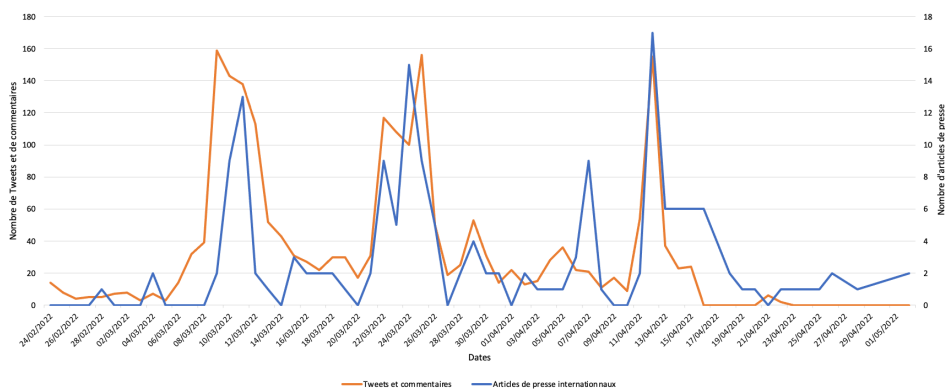


Figure 2. Répartition des publications des médias sociaux et de presse par jour.

3.4. Analyse lexicométrique des corpus

3.4.1. Analyse lexicométrique du corpus « médias de presse »

Les 168 articles de presse francophones et anglophones ont été divisés en 3 020 segments de texte afin d'analyser les formes actives (mots en usage sans formes supplémentaires ou mots de liaisons). La classification hiérarchique descendante (CHD) a permis d'analyser 92,4 % de la totalité des segments de textes et d'identifier trois classes lexicales. Le poids de chaque classe oscille entre 28,3 % et 38,6 % des segments classés. Ces trois thématiques majeures se regroupent en deux dimensions distinctes (Fig. 3). La première dimension représente le risque biologique, avec les allégations concernant les laboratoires de recherche américains en Ukraine travaillant sur des pathogènes (classe 3 : 33,1 % des segments classés). La seconde dimension se rapporte au risque chimique avec d'une part, la vision tactique et les déclarations politiques sur ces risques

(classe 1 : 28,3 %) et d'autre part, les suspicions d'attaques chimiques russes antérieures et actuelles sur des villes ukrainiennes (classe 2 : 38,6 %).

L'analyse factorielle des correspondances (AFC) effectuée sur la CHD nous permet de conforter graphiquement ces résultats. L'AFC illustre l'opposition des discours sur la thématique des agents B (classe 3, bleue) versus celle des agents C (classes 1 et 2, rouge et verte).

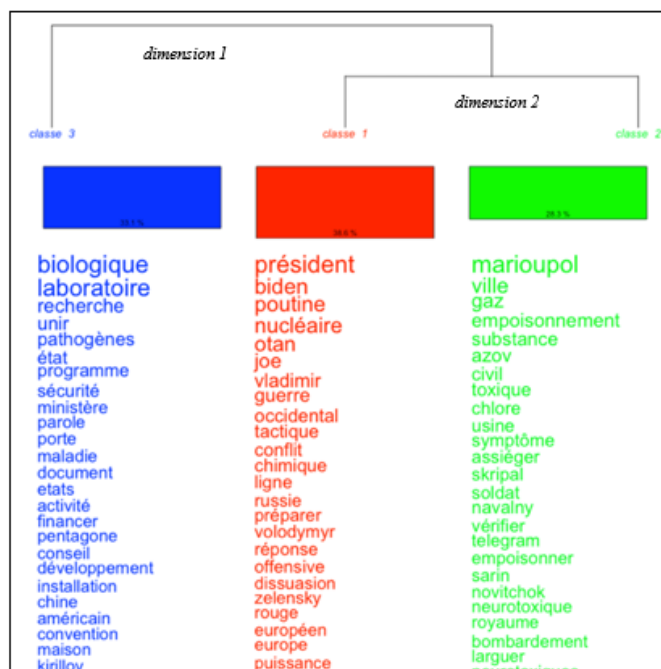


Figure 3. Dendrogramme du corpus Médias de presse (IRaMuTeQ).

3.4.2. Analyse lexicométrique du corpus « médias sociaux »

La CHD sur les 2 087 commentaires d'articles et *tweets* a été réalisée sur les textes entiers étant donné la restriction en nombre de caractères imposée par Twitter. Deux classes lexicales ont été identifiées dans lesquelles 90,3 % de la totalité des textes sont classés. La première classe, ayant un poids de 73,5 % dans le corpus, évoque les allégations d'attaques chimiques réalisées sous faux-drapeaux entraînant une possible intervention de l'OTAN. Le poids de la seconde classe représente 26,4 %. Celle-ci met en évidence des discours complotistes sur la présence de laboratoires de recherche d'armes biologiques en Ukraine, financés par les États-Unis (Fig. 4).

Enfin, nous avons repéré lors du nettoyage des bases de données textuelles un grand nombre d'adhérents aux théories du complot (décrivant les événe-

ments comme le résultat d’actions planifiées et dissimulées par un petit groupe de protagonistes).

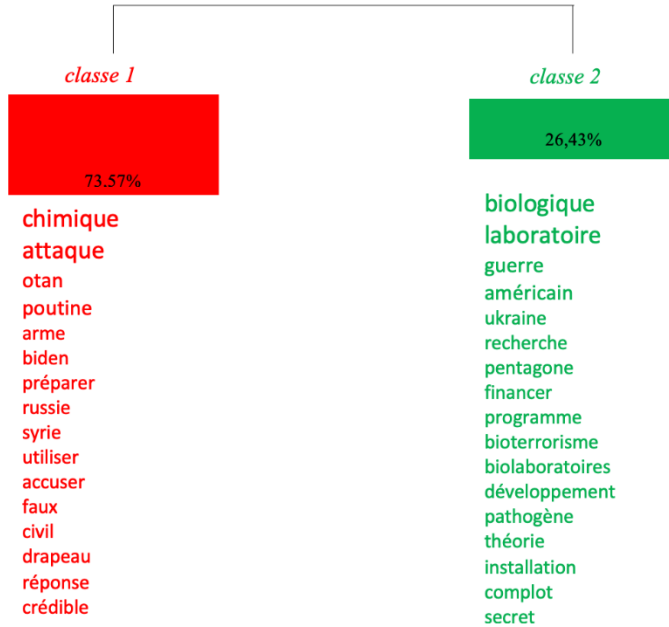


Figure 4. Dendrogramme du corpus Médias sociaux (IRaMuTeQ).

3.4.3. Analyse lexicométrique comparative des corpus « médias sociaux » et « médias de presse »

L’AFC réalisée sur l’ensemble des corpus textuels illustre le positionnement des sources les unes par rapport aux autres, concernant les données textuelles émises (Fig. 5). Le premier facteur (58,8 % de la masse du corpus) sépare nettement les corpus médias de presse (tous pays confondus) au corpus média social – Twitter. En revanche, les forums de discussions ont une position plus centrale, certaines formes actives se rapprochent de celles des médias de presse (commentaires de vidéos), d’autres de celles des réseaux sociaux (commentaires d’articles). Le second facteur (25,9 %) montre une distinction entre les corpus de presse francophone et anglophone. Grâce au concordancier d’IRaMuTeQ, nous avons mis en évidence une discordance entre les discours dits « pro-russes » et ceux « pro-ukrainiens ». Les dimensions lexicales des articles de presse provenant de Russie, de Chine, du Liban et du Brésil ne s’opposent pas aux publications occidentales (France, Ukraine, Allemagne,

Canada, Royaume-Uni, États-Unis...). En revanche il y a un antagonisme concernant le sens de ces discours.

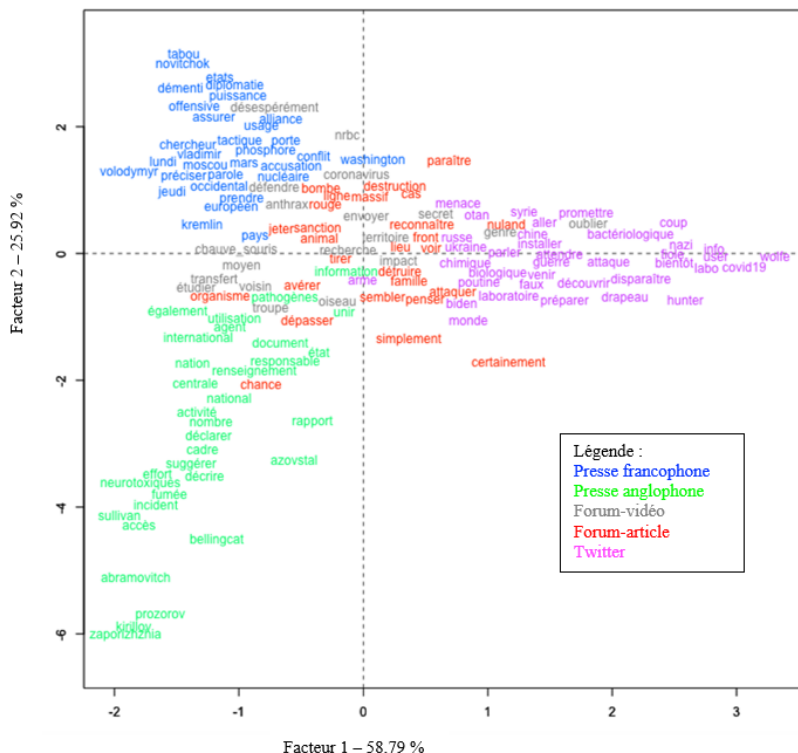


Figure 5. Plan factoriel issu de l'AFC sur le tableau lexical agrégé de l'ensemble des sources (IRaMuTeQ).

4. Discussions

Les principaux résultats de notre étude mettent en évidence une similitude des pics informationnels entre les différentes sources médiatiques, concernant les thématiques des armes B et C dans le cadre du conflit russo-ukrainien. Par exemple, nous visualisons un pic d'informations le 11 avril, lorsque tous les médias relayaient l'hypothèse d'une attaque chimique intentionnelle de la part de la Russie sur la ville de Marioupol (Ukraine). A ce jour, nous n'avons aucune conclusion officielle publiée concernant cette possible attaque. On note également une forte activité médiatique sur la période du 9 au 12 mars : lorsque la Russie affirmait au conseil de sécurité de l'ONU qu'il existait des laboratoires américains développant des armes biologiques sur le territoire ukrainien. Les

personnalités politiques américaines ont rapidement contredit ces allégations et ont affirmé que ces laboratoires étaient dédiés à la recherche sur les moyens d'atténuer les menaces B contre les êtres vivants. Nous avons aussi constaté que l'ensemble de ces informations a fait réagir les utilisateurs de médias sociaux, exprimant leurs préoccupations et interrogations sur ces plateformes.

De surcroît, les grandes thématiques qui font l'objet de cette étude sont traitées d'une manière différente en fonction des médias sélectionnés. Par exemple, nous avons observé qu'une grande partie des utilisateurs de Twitter prônaient des discours complotistes durant le conflit concernant les activités des laboratoires biologiques en Ukraine. Autre exemple, certains discours exposent la théorie du faux-drapeau : des attaques chimiques réalisées par l'Ukraine pour ensuite accuser les forces armées russes et entraîner une intervention de l'OTAN. En revanche, au sein du corpus médias de presse, nous avons mis en évidence les nombreux discours factuels ou expertises de personnalités politiques internationales et de professionnels du domaine NRBC débattant sur les allégations d'utilisations d'armes C et sur les activités des bio-laboratoires américains en Ukraine, repris par les journaux.

Comme nous venons de le voir, les discours diffèrent en fonction du média sélectionné et des contributeurs. Les analyses lexicométriques réalisées au cours de cette étude révèlent une analogie entre les champs lexicaux employés par les différents pays. Mais ils soulignent aussi un antagonisme dans le sens des discours formulés. Par exemple, les discours provenant de Russie et de Chine s'opposent à ceux des pays occidentaux. Il serait donc intéressant dans une prochaine étude de réaliser des analyses sémantiques sur ces données afin d'étudier le sens et la signification des textes. Il est important de noter que dans le contexte particulier de ce conflit, les événements évoluent en permanence, ce qui entraîne une possible labilité des thématiques identifiées, dans le temps.

La désinformation et le contenu invérifiable se propagent extrêmement vite sur les réseaux sociaux (Rosenberg, Syed, et Rezaie 2020) et en particulier en contexte d'infodémie (propagation rapide d'informations exactes et inexactes au cours d'une épidémie (Gardère 2021)) ou de guerre. L'analyse des médias sociaux reste donc complexe car la véracité des données n'est pas garantie (anonymat des profils et/ou biais de prévarication). En effet, des études récentes quantifient le pourcentage élevé de désinformations en santé (Kouzy et al., 2020), fakes news et théories complotistes (Shahsavari et al. 2020) sur les médias sociaux, particulièrement en période de crises sanitaires. Dans le cadre de notre étude, ces fakes news et théories complotistes génèrent un important bruit informationnel qui nuit à la lisibilité de l'information en rapport avec le conflit.

Nous remarquons toutefois que de nombreux internautes se font le relais des médias de presse et repostent rapidement leurs contenus valorisant ainsi des informations plus factuelles. Nos résultats illustrent ce phénomène avec la

similarité des pics informationnels et la proximité temporelle entre la publication d'un *post* et le nombre d'actions d'engagement qu'il génère.

De plus, au sein de nos échantillons, nous n'avons pas repéré de lanceurs d'alerte (personne physique qui révèle ou signale, de manière désintéressée et de bonne foi, un crime ou un délit, une violation grave et manifeste d'un engagement international) concernant l'usage d'agents B et/ou C durant ce conflit.

Nous savons que les médias sociaux, et en particulier les réseaux sociaux, conditionnent les échanges entre les personnes (Proulx, Millette, et Heaton 2012). Nous avons notamment remarqué lors de l'analyse des *tweets* et des commentaires de forums de discussion que les acteurs échangent et débattent avec plus d'ardeur, souvent sous couvert d'anonymat. Il est important de noter que le format des *tweets* avec des caractères limités entraîne aussi une restriction d'informations, les utilisateurs vont à l'essentiel. Par rapport aux médias de presse (journalistes et experts), les utilisateurs de Twitter et forumers représentent davantage la population générale, l'utilisation qu'ils font de ces médias est donc différente.

Enfin, nous avons identifié des limites méthodologiques à notre étude. Les médias dits « classiques » suivent une ligne éditoriale, par conséquent le contenu est orienté (Charon 2010). De plus, le logiciel de veille sur les réseaux sociaux Radarly collecte aléatoirement un grand échantillon de *tweets* ; nous ne connaissons pas son algorithme de recueil de données, ce qui soulève des questions sur l'exhaustivité, la standardisation et la tendance des informations recueillies.

De plus, un biais de représentativité lié aux langues sélectionnées pour la collecte de données a également été repéré. Nous avons fait le choix méthodologique de réaliser un recueil en français et anglais, ce qui exclut toutes les informations provenant par exemple des articles ou posts en langues russe ou ukrainienne. De même, avec l'exclusion des abonnements payants aux médias de presse, cette étude ne peut prétendre à l'exhaustivité des informations recueillies. Aussi, certaines sources d'informations sont plus ou moins présentes et utilisées selon les nations et les différentes catégories au sein de celles-ci. Par exemple, Facebook ou Twitter sont préférés dans la grande majorité des pays européens, hormis en Russie, Biélorussie, Ukraine, Moldavie, Estonie et Lettonie qui utilisent en majorité VK.

En conclusion de cette discussion, nous pouvons dire que l'approche méthodologique de notre étude présente des intérêts à la fois en santé publique et en sciences de l'information et de la communication (SIC). En effet, peu de travaux de recherche en santé et en SIC associent dans leur approche, à la fois des analyses sur les médias sociaux et sur les médias de presse. L'autre originalité de notre travail réside sur son sujet en lui-même : un travail de recherche sur une thématique généralement confidentielle, celle des armes de destruction massive. Enfin notre étude présente des intérêts stratégiques pour les états,

pour la détection de signaux faibles en émergence, par l'analyse de données issues de médias sociaux et de presse.

5. Conclusions

Dans notre étude, nous avons analysé des données textuelles issues de différentes sources médiatiques de presse et de réseaux sociaux sur la thématique des ADM de type Biologique et Chimique dans le cadre de l'actuel conflit russo-ukrainien. Nous avons pu identifier les principaux acteurs de la communication sur ce sujet, sur une période déterminée. Nous avons présenté la répartition spatiale et temporelle des informations émises sur ces médias. Des analyses lexicométriques et comparatives ont permis d'identifier les thématiques principales traitées et les discordances dans les informations émises par les différents médias. En conclusion générale, au regard de la « lutte contre les menaces biologiques ou chimiques, l'éventuelle utilisation des nouvelles technologies à des fins malveillantes demeure un sujet de préoccupations majeure » pour les nations (NATO 2021). Aussi, notre approche méthodologique présente pour celles-ci un réel intérêt pour détecter sur ces nouvelles technologies de l'information et de la communication (NTIC) des signaux faibles en émergence concernant ces menaces. Nos futurs travaux pourront envisager comme autre source de collecte, le Darknet, une partie « cachée » et anonyme du web, comprenant du contenu illégal et dissident et qui ne répond à aucune réglementation ou censure.

Références bibliographiques

- Charon, Jean-Marie. 2010. "De la presse imprimée à la presse numérique. Le débat français." *Réseaux* 160-161 (2-3): 255-81. <https://doi.org/10.3917/res.160.0255>.
- Gardère, Elizabeth. 2021. "De l'infodémie à l'infopandémie. L'OMS face aux (des)informations virales." *Les Cahiers du numérique* 17 (3-4): 199-204. <https://doi.org/10.3166/LCN.2021.018>.
- González-Padilla, Daniel A., and Leonardo Tortolero-Blanco. 2020. "Social media influence in the COVID-19 Pandemic." *International Brazilian Journal of Urology : official journal of the Brazilian Society of Urology* 46 (Suppl 1): 120-24. <https://doi.org/10.1590/S1677-5538.IBJU.2020.S121>.
- Guével, Marie-Renée, et Jeanine Pommier. 2012. "Recherche par les méthodes mixtes en santé publique : enjeux et illustration." *Santé Publique* 24 (1): 23-38. <https://doi.org/10.3917/spub.121.0023>.

- Proulx, Serge, Mélanie Millette, et Lorna Heaton. sous la direction de. 2012. *Médias sociaux, enjeux pour la communication*. 1^{ère} ed. Presses de l'Université du Québec. <https://doi.org/10.2307/j.ctv18pgp4w>.
- Hossain, Liaquat, Derek Kam, Fiona Kong, Rolf T. Wigand, and Terry Bosso-maier. 2016. "Social media in Ebola outbreak." *Epidemiology and Infection* 144 (10): 2136-43. <https://doi.org/10.1017/S095026881600039X>.
- Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour. 2020. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter." *Cureus* 12 (3): e7255. <https://doi.org/10.7759/cureus.7255>.
- Masri, Shahir, Jianfeng Jia, Chen Li, Guofa Zhou, Ming-Chieh Lee, Guiyun Yan, and Jun Wu. 2019. "Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic." *BMC Public Health* 19 (1): 761. <https://doi.org/10.1186/s12889-019-7103-8>.
- Melin, Anders. 2021. "Overstatements and Understatements in the Debate on Synthetic Biology, Bioterrorism and Ethics." *Frontiers in Bioengineering and Biotechnology* 9. <https://doi.org/10.3389/fbioe.2021.703735>.
- NATO. 2023a. "Maîtrise des armements, désarmement et non-prolifération à l'OTAN." 08 mars 2023. https://www.nato.int/cps/fr/natohq/topics_48895.htm.
- NATO. 2023b. "Les armes de destruction massive." 07 mars 2023. https://www.nato.int/cps/fr/natohq/topics_50325.htm.
- NATO, Commission des Sciences et des Technologies. 2021. "Menaces biologiques : progrès technologiques et spectre du bioterrorisme pour l'Après-Covid-19." Rapport. <https://www.nato-pa.int/download-file?filename=/sites/default/files/2022-01/024%20STCTTS%2021%20F%20r%C3%A9v.%201%20fin%20-%20MENACES%20BIOLOGIQUES.pdf>.
- Nicolas, Laetitia. 2007. "Bruits et silences - Réflexion sur l'interprétation des données de terrain et sur leur archivage." *Bulletin de l'AFAS. Sonorités* 30 : 2-10. <https://doi.org/10.4000/afas.1531>.
- ONU. 1946. "Résolutions adoptées sur les rapports de la première commission de l'Assemblée générale des Nations Unies." <https://www.un.org/french/documents/ga/res/1/fres1.htm>.
- Rosenberg, Hans, Shahbaz Syed, and Salim Rezaie. 2020. "The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic." *Cjem* 22(4): 418-21. <https://doi.org/10.1017/cem.2020.361>.

- Shahsavari, Shadi, Pavan Holur, Tianyi Wang, Timothy R. Tangherlini, and Vwani Roychowdhury. 2020. "Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news." *Journal of Computational Social Science* 3 (2): 279-317. <https://doi.org/10.1007/s42001-020-00086-5>.
- Tanti, Marc, Perrine Roux, Bruno Spire, et Patrizia Carrieri. 2018. "Étude des acteurs de la communication des salles de drogues par une analyse de corpus multivarié." *Revue française des sciences de l'information et de la communication*, no. 14 (septembre). <https://doi.org/10.4000/rfsic.4397>.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 41

N. 3-4 – luglio-dicembre 2023

Contributi

ALESSANDRO ALFIER, *Per una rigenerazione teorica dell'archivistica in Italia, a partire dal concetto di documento*

CAMILLE ARSEL, MARC TANTI, *Étude des informations relayées par les médias de presse et sociaux concernant les agents biologiques et chimiques militaires, dans le conflit russo-ukrainien*

MARIO CIAMPI, ERIKA PASCERI, GRAZIA SERRATORE, *Lo standard FHIR e il Fascicolo Sanitario 2.0*

ELEONORA DE LONGIS, *Il Corpus Inscriptionum Latinarum dall'analogico al digitale*

ISABELLA FLORIO, GIORGIA MIGLIORELLI, SARA SANTORSA, *Per una Carta dei servizi nel contesto bibliotecario italiano: il caso del CNR*

ELEONORA LUZI, *La digitalizzazione come strumento per acquisire informazioni: l'UPP e la certificazione di processo*

FRANCESCA PARISI, *Dall'esperienza alla memoria*

ANNA ROVELLA, ASSUNTA CARUSO, MARTIN CRITTELLI, FRANCESCA M.C. MESSINITI, *Knowledge extraction, research projects and archives management*

MARCIN TRZMIELEWSKI, *Les activités informationnelles des professionnels de santé : état de l'art dans une perspective interdisciplinaire et internationale*

Note e rubriche

CLAUDIO GRIMALDI, *Le evoluzioni delle applicazioni di IA nel campo linguistico: quale futuro per il lavoro terminologico?*

ROBERTO GUARASCI, *La valutazione delle fonti documentali*



mundaneum

In copertina

Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-329-1



9 791259 653291

ISSN 1121-0095



9 770112 100950