

LA CONSTRUCTION DES CORPUS À PARTIR DES INTERFACES NUMÉRIQUES ET CONSTITUTION D'UNE ARCHIVE NUMÉRIQUE AUTOUR DU TOPIQUE « ALGÉRIE »

Karima NABTI

Laboratoire de Recherche EPI Étude pragmatique Inférentielle
Université Alger 2
k.nabti@ensv.dz

&

Lamia KARRAH

Laboratoire de Recherche EPI Étude pragmatique Inférentielle
Université Alger 2
lamia.karrah@univ-alger2.dz

Résumé : Cet article est le résultat d'un travail de recherche mené sur la composition des archives et des corpus numériques dans le cadre du projet de recherche autour de la « Spatialité du topique *Algérie* dans le discours ». Plus précisément, il s'agit de mettre en place un processus de collecte d'une archive numérique pouvant fournir des corpus numériques utilisables par divers domaines d'études, notamment les sciences du langage, les sciences sociales, les sciences politiques, les sciences de la communication, etc. Ainsi, c'est à partir des interfaces numériques, un espace numérique propice à la recension des données largement diffusées, que s'envisage notre réflexion sur la composition d'une archive numérique thématique. La collecte des données a été faite à partir de la combinaison de 120 mots-clés liant le thème "Algérie" à d'autres termes spatialement spécifiques tels que "Algérie-Centre, Algérie-Méditerranée, Algérie-Maghreb", etc. Cette combinaison a généré un ensemble de données numériques regroupant des articles, vidéos, photos, qui à leur tour seront soumis à un tri et une sélection pour en former une archive numérique contenant plusieurs corpus numériques.

Mots-clés : archive numérique, corpus numérique, topique *Algérie*, spatialité

THE CONSTRUCTION OF CORPORA FROM DIGITAL INTERFACES AND CONSTITUTION OF DIGITAL ARCHIVE AROUND THE TOPICAL "ALGERIA"

Abstract: This article is the result of a research work carried out on the composition of archives and digital corpora within the framework of the national research project on the "Spatiality of the topical Algeria in the discourse". More precisely, it is a question of setting up a process of collecting a digital archive that can provide digital corpora usable by various fields of study, notably language sciences, social sciences, political sciences, communication sciences, etc. Thus, it is from the digital interfaces, a digital space conducive to the review of widely disseminated data, that our reflection on the composition of a thematic digital archive is envisaged. Data collection was based on the combination of 120 keywords linking the theme "Algeria" to other spatially specific terms such as "Algeria-Centre", "Algeria-Mediterranean", "Algeria-Maghreb", etc. This combination generated a set of data that could be used to identify and analyse the different types of information. This combination has generated a digital dataset of articles, videos, photos, which in turn will be subjected to a sorting and selection process to form a digital archive containing several digital corpora.

Keywords: digital archive, digital corpus, Algeria topical, spatiality

Introduction

Les espaces numériques présentent un flux constant d'informations. Cette émergence permanente et ce champ en constante expansion a vu naître une instabilité à plusieurs niveaux dans ces environnements numériques, Christine Develotte (2012 : 516), souligne face à ce problème que « la volatilité des données en ligne est en soi une première fragilité des corpus que l'on veut étudier ». D'une part, des changements continus sont observés au niveau de la structure des sites ainsi qu'au niveau des adresses IP. D'autre part, la disparition des informations et publications du Web a été constatée au cours de ces dernières années. Toutefois, plusieurs plans d'archivage ont été mis en place pour conserver la plupart des ressources publiées en ligne (pages, liens hypertextes, contenus multimédias, etc.), permettant ainsi la préservation de ce patrimoine numérique pour les générations futures. Mais, ces actions menées par les chercheurs des différents secteurs n'ont pas pu faire face à la volatilité du Web. Aujourd'hui, plusieurs chercheurs portent un intérêt particulier aux archives numériques et constituent leurs corpus à partir du web. Des travaux de recherche ont tenté d'apporter de nouveaux outils méthodologiques sur la constitution des corpus à partir des espaces numériques, même si le terrain est périlleux, « paradoxalement, la fragilité, l'éphémère, l'incertitude des supports, mais aussi l'accélération et l'obsolescence de l'information qui circulent sur les réseaux ont renforcé la focalisation sur l'archive » (Favier L., 2015 : 9), plusieurs outils ont été mis en place par les différentes disciplines pour pallier à ce déficit et afin de faciliter la tâche aux chercheurs.

1. Contexte d'étude et émergence d'une question de recherche

C'est dans le cadre d'un Projet de Recherche, qui s'articule autour de la « Spatialité du topique "Algérie" dans le discours ». Il est question dans le cadre de ce travail de recherche d'aborder la problématique de l'espace dans le discours et de la localisation spatiale dans le langage et de sa circulation. La circulation des discours est définie par Rosier « Comme un ensemble de mécanismes d'appropriation, de ré-énonciation et de remises en relation discursive relativement organisés entre des espaces discursifs (textes, genres de textes, formations discursives) par des agents de circulation » (Rosier L., 2003 : 69). Notre problématique repose essentiellement sur les notions de l'archive et des corpus numériques. Plus précisément, à travers cet état des lieux que nous dressons, nous voudrions apporter des éclaircissements sur l'intervention des interfaces numériques dans la construction des corpus numériques, notre question de recherche est de voir comment les interfaces numériques interviennent dans la construction des corpus numériques et constitution d'une archive numérique autour du topique « Algérie » ? À cette problématique se greffent d'autres questions, qui serviront au cadrage méthodologique. Quel est l'apport des interfaces numériques dans la construction de l'archive et des corpus numériques ? Comment constituer une archive numérique face à l'instabilité des Web ? Comment constituer une base de données pour le domaine des sciences du langage autour d'une thématique prédéfinie ? Afin de réaliser cette étude, nous sommes parties du postulat que les interfaces numériques offrent une panoplie de données. Ces données peuvent fournir à leur tour une archive numérique constituée de plusieurs corpus numériques exploitables par les différents domaines de recherche.

1. État des lieux : archives vs corpus numériques

Notre objectif de recherche est de dresser d'abord un état des lieux sur les études et recherches qui ont porté sur la constitution des archives et la construction des corpus numériques à partir des sites Web. Pour ce faire, il est important de se pencher sur la question des techniques d'archivage qui ont fait l'objet de recherche de plusieurs études relevant des domaines des humanités numériques et sciences du langage.

1.1 De l'archive à la numérisation

Le conseil international des archives (CIA) qualifie les archives d'une manière générale comme étant « l'ensemble des documents de toute nature, produits ou reçus par une personne physique ou morale, par un organisme public ou privé, résultat de son activité, organisé en conséquence de celle-ci et conservés en vue d'une utilisation éventuelle ». L'archivage est une technique qui permet de réunir un ensemble de documents pour les rendre accessibles. En effet, « l'archive réunit l'ensemble des documents accessibles pour une tâche de description ou une application. Elle n'est pas un corpus, parce qu'elle n'est pas constituée pour une recherche déterminée ». (Rastier F., 2017 : 18). De plus, il est également important de souligner qu'une archive « est considérée comme authentique lorsqu'il s'agit bien du document quelle prétend être et quand elle n'a pas été altérée depuis sa création » (Frey V., 2009 : 121). Incontestablement, l'ère numérique a révolutionné les techniques d'archivage. Cependant, il est important de ne pas confondre les numérisations de documents avec les documents numériques. En effet, un document numérisé est une traduction d'un document imprimé, qui peut être numérisé à partir de n'importe quel support, mais le plus souvent, un document "numérisé" est à l'origine un document papier qui a été "numérisé" par un scanner. Cette numérisation permet de construire une base de données qui forme une archive numérisée accessible dans les centres de recherche, les centres de documentation et les bibliothèques, etc.

Par ailleurs, un document numérique communément appelé *document numérique natif* (Paveau, 2015 a) est un document créé directement sur un ordinateur qui ne contient que des éléments générés par le logiciel qui l'a créé.

Confrontés à la dématérialisation des documents, d'une part, et à leur numérisation, d'autre part, les praticiens des archives ont vu, ces dernières années, leur lexique s'enrichir de nouvelles expressions. L'« archive numérique », également appelée « archive électronique », est une expression pleine d'ambiguïtés. Selon le lexique du site Internet des Archives [...] elle correspond à « tout type de document produit aujourd'hui sous une forme électronique et dématérialisée ». Ce qu'on appelle archive numérique est, par conséquent, toute archive « nativement » (c'est-à-dire dès sa création) numérique. Cette archive est faite à la fois de données (le contenu du document à proprement parler) et de métadonnées (les informations sur le document, telles que sa date de création, son auteur.

Barbier J. (2018 :195)

L'archive numérique est au centre de l'attention des différents acteurs intéressés par le Web. Les chercheurs (informaticiens, historiens, linguistes, sociologues, etc.) travaillent en étroite collaboration pour constituer des archives consultables depuis le Web. Sa finalité est de pouvoir exploiter des données et de les rendre accessibles dans le temps à des fins historiques.

Le passage du papier au numérique ne représente pas un simple changement technique de support de la culture, de l'information et du savoir humain, mais une révolution culturelle, épistémologique, anthropologique [...]. Après la modernité, l'hypermodernité s'impose à marche forcée bouleversant notre rapport au monde, au temps, à l'espèce et au très ; l'écran remplace le livre ; le texte est devenu hypertexte.

Mayaffre D. (2010 : 235)

Ce passage à l'ère de l'archivage numérique a nécessité de mettre en place une technicité pour pallier toutes les difficultés rencontrées depuis le début de l'ère numérique. En effet, dans le cadre de notre travail de recherche, sur l'édification d'une archive numérique, nous nous sommes d'abord concentrés sur la collecte des données dans les espaces numériques, tout en prenant en compte les différentes contraintes liées à l'instabilité du Web. Dans cette étape, la recherche est l'opération la plus importante qui consiste à faire interagir un chercheur exprimant ses besoins avec une requête et un système qui traduit ces besoins dans la langue du système, et puis fournir à l'utilisateur une liste de documents qu'il juge pertinents. Il existe trois types de recherche :

- La première consiste à comparer la demande de l'utilisateur avec toutes les informations contenues dans les métadonnées.
- La seconde recherche de mots-clés est la plus utilisée dans les bibliothèques numériques, bien qu'elle ne soit pas approfondie, mais pertinente.
- Le troisième type de recherche, la recherche assistée, consiste à guider l'utilisateur tout au long du processus de recherche. Il fait une requête et le moteur de recherche effectuera sa recherche et lui fournira une liste d'interprétations possibles afin qu'il puisse les vérifier et en sélectionner celle qui l'intéresse.

La méthode de recherche retenue par notre équipe de recherche combine entre la seconde méthode et la recherche booléenne qui consiste à présenter la requête sous la forme d'une expression booléenne de mots clés séparés par des opérateurs booléens (ET, OU, SAUF, etc.). Cette technique de recherche a permis de récolter un ensemble d'articles, à partir de combinaisons de mots-clés (environ 120) soumis au moteur de recherche *Google*, liant le thème «Algérie» à d'autres termes à caractéristiques spatiales, comme "Algérie- centre", "Algérie-Méditerranée", "Algérie-Maghreb", etc. Cette combinaison de mots-clés a généré un ensemble d'articles qui seront soumis à une sélection et un tri pour la constitution d'un corpus numérique.

1.2 Du corpus numérisé au corpus numérique

« Collecter », « rassembler », « réunir », « construire », « constituer » un corpus est l'une des opérations les plus importantes et difficiles dans le domaine des sciences du langage. Il s'agit d'une tâche complexe et une étape décisive dans la recherche scientifique, car cet objet d'étude ne se construit qu'après une longue réflexion méthodologique. En analyse de discours « la notion de corpus est primordial (...), car c'est le corpus qui permet de réaliser une analyse à l'aide de catégories issues de la linguistique et de l'analyse de discours pour répondre à des questionnements qui ne peuvent être formulés que dans la pluralité disciplinaire » (Bonnafous S., 2006 : 221). La notion de corpus a reçu plusieurs acceptions. Selon J.P. Dalbera (2010 : 11), elle

représente « *un ensemble d'éléments sur lequel se fonde l'étude d'un phénomène linguistique* ». Mais avant d'arriver à l'étude des phénomènes linguistiques, il va falloir procéder à l'assemblage des données, et ce en adoptant une démarche empirico-inductive qui consiste à recueillir des données après les avoir observées, et à les catégoriser afin de mettre en place une grille d'analyse qui permettra d'atteindre l'objectif assigné à la recherche, et ce en affirmant ou en infirmant les hypothèses émises préalablement. Plusieurs éléments rentrent directement dans la constitution du corpus : formuler la problématique, s'inscrire dans un contexte et définir le champ interdisciplinaire. De plus, la représentativité et la cohérence sont des critères qui contribuent à la constitution du corpus. En effet,

[...] construire un corpus revient à construire conceptuellement la réalité langagière à étudier en décalage par rapport à son état naturel donné dans sa totalité et différemment de la réalité conceptualisée de l'archive bien qu'elle s'en inspire pour se préciser à travers des problématiques liées à des procédés discursifs instituant une lecture interprétative

Immune Y. (2023 :28)

Le corpus doit être représentatif, complet et pertinent :

- Représentatif dans la mesure où il regroupe l'ensemble des documents qui proviennent de sources diverses et qui traitent du même thème.
- Complet : il doit contenir l'ensemble des éléments qui nous permettent de l'approcher, de l'analyser, de le décortiquer.
- Pertinent : les documents qui constituent notre objet d'étude doivent traiter de la même thématique. Autrement dit, ils doivent avoir en commun le topique « Algérie » et son rapport à la spatialité.
- Construit : il est important de préciser que le corpus n'est pas un simple recueil de données ou de textes mis à la disposition du chercheur, mais il est selon Mayaffre (2002 : 4) « un objet heuristique. C'est une construction arbitraire, une composition relative qui n'a de sens, de valeur et de pertinence qu'au regard des questions qu'on va lui poser, des réponses que l'on cherche, des résultats que l'on va trouver. »

Contrario aux corpus numérisés, les corpus numériques sont issus directement des interfaces numériques. En effet, les espaces de production des textes en ligne sont désormais considérés comme une machine à texte, ces espaces englobent tous les supports numériques, les sites Web, les blogs, les journaux, les réseaux sociaux... ces textes ou documents numériques appelés aussi « documents numériques » sont des produits « nativement en ligne, sur un site, un blog ou un réseau social, tout lieu numérique accueillant de la production de discours. Il présente des traits de délinéarisation de fil du discours, d'augmentation énonciative, de technogénéricité et de plurisémiotité. » (Paveau A.M., 2015 : 7). Les critères de sélection des corpus numériques sont identiques à ceux fixés préalablement. Cependant, la particularité des corpus numériques, appelés également des technodiscours, réside au niveau de leur « composition, délinéarisation, augmentation, relationalité, investigabilité et imprévisibilité... des formes textuelles modifiées par les environnements connectés » (Mayeur I. et Paveau A.M., 2020 : 9).

D'un point de vue purement technique, le nouveau processus d'archivage devait être envisagé de fond en comble. Alors que la préservation du Web a rapidement été

jugée nécessaire, des équipes pionnières ont rassemblé des connaissances éparses et les ont combinées pour créer des outils de sélection, de collecte et de recherche de corpus d'archives Web futures, « *tous les linguistes de corpus travaillent aujourd'hui sur données numérisées, et tous réfléchissent à des méthodes numériques pour traiter leur objet numérique.* » (Mayaffre D., 2010 : 236). Pour procéder à cette collecte, nous avons dressé une méthodologie de sélection et de tri qui prend en compte plusieurs paramètres liés au péritexte, contenus et la dynamique des sites Web.

-Tri du corpus numérique

Afin de procéder au tri des données, nous avons tenu compte de plusieurs aspects. D'une part, nous savons que les pages Web subissent des changements au cours de leur existence. Ces changements peuvent se manifester à la fois au niveau des contenus, mais aussi au niveau de la structure des liens qui la relie au reste du site. Pour cela, nous avons procédé à la vérification de la disponibilité de tous les liens ainsi que les articles générés, nous avons, également, pris le soin de sauvegarder toute la base de données récoltée, en mentionnant la date, le lien ainsi que l'article généré.

D'autre part, il y a lieu de tenir compte également des critères traditionnels de la formation de corpus qui visent à vérifier trois types de conditions (Charaudeau P., 2009 :37-66) des conditions de signifiante, d'acceptabilité et d'exploitabilité.

- **Condition de signifiante** : selon laquelle, le corpus est constitué d'une étude déterminée (pertinence), portant sur un objet particulier « topique Algérie », une réalité telle qu'elle est perçue sous un certain angle de vue (et non sur plusieurs thèmes ou facettes indépendants, simultanément).
- **Condition d'acceptabilité** : selon laquelle le corpus doit apporter une représentation fidèle (représentativité), sans être parasité par des contraintes externes (régularité). Il doit avoir une ampleur et un niveau de détail adapté au degré de finesse et à la richesse attendue en résultat de l'analyse (complétude).
- **Condition d'exploitabilité** : selon laquelle, les textes qui forment le corpus doivent être commensurables (homogénéité). Le corpus doit apporter suffisamment d'éléments pour pouvoir repérer des comportements significatifs.

Par ailleurs, en qualité de chercheurs en sciences du langage, nous devons également tenir compte d'autres paramètres de récolte. Ces paramètres dressés par l'équipe de recherche se présentent ainsi :

- Sur le plan méthodologique, prendre en compte que la recherche s'effectue sur les corpus numériques natifs.
- L'accessibilité de l'information à partir d'une connexion locale.
- Tenir compte de l'Algérie en tant que spatialité.
- Tenir compte de la génération des titres selon les items employés comme mots clés.
- Tenir compte de la spatialité médiatique qui rejoint la spatialité géographique.
- Algérie et zones d'influence (influencée, influenceur) : Maghreb, monde arabe...
- Spatialité et temps historique : existence et disparition
- Alger se distingue par une spatialité administrative et des affaires

- Reclasser l'archive avec des items correspondant à des procédés discursifs : sous les procédés se révéleront les lieux qui y sont construits de manière privilégiée ou pas par ces procédés

-Sélection du corpus numérique

Tous ces critères ont permis une récolte ciblée d'un ensemble d'articles exploitables et analysables. Étant notre point d'entrée est clairement défini. La sélection désigne donc l'ensemble des techniques mises en place pour définir ce ou ces points d'entrée. Ces données regroupent à leur tour un ensemble d'articles qui vont former des corpus en fonction de l'objet d'études. Nous rappelons qu'« un corpus au sens général est un ensemble de données, qui peuvent se présenter sous diverses formes, sélectionnées selon des critères précis définis suivant l'objet d'étude. » (Segouat J., 2010 : 81). Il s'agit de transformer les pages du Web en une archive consultable, pour cela une recherche affinée sur les combinaisons *Algérie-espace* ont été fixées et ont permis une récolte d'un ensemble de documents, regroupés en sous thèmes, à partir des sites Web. Les sous thèmes proposés regroupent à leurs tours des ensembles de combinaisons. C'est à partir de ces ensembles qu'une panoplie de documents numériques a été générée. Pour une meilleure lecture, nous avons illustré à travers trois exemples la méthode entreprise pour la récolte des documents à partir de ces combinaisons.

Exemple 1 : la combinaison « Algérie – pays » regroupe un ensemble de combinaisons lié à cette thématique et chacune d'elle permet le recueil d'un ensemble d'articles qui reprends ces deux mots clés : Algérie-nation ; Algérie-continent ; Algérie-terre ; Algérie-Mer ; Algérie-superficie ; Algérie-grandeur

Exemple 2 : la seconde combinaison regroupe un ensemble de documents qui renvoie à tout ce qui est lié au paysage algérien « Algérie-paysage ». Nous avons recensé tous les documents numériques pouvant évoquer ce paysage algérien tout en mentionnant les mots clés des combinaisons suivantes ; Algérie-Alger ; Algérie-Casbah ; Algérie-sahara ; Algérie-montagne ; Algérie-littoral ; Algérie-région ; Algérie-hauts plateaux ; Algérie-ville ; Algérie-urbain ; Algérie-rural ; Algérie-de l'intérieur.

Exemple 3 : la combinaison qui renvoie à la thématique des relations internationales « Algérie- international » regroupe les combinaisons suivantes : Algérie-Méditerranée ; Algérie-Afrique ; Algérie-Europe ; Algérie-Monde ; Algérie – maghreb ; Algérie-monde arabe ; Algérie-Monde Musulman ; Algérie-Tunisie ; Algérie-Maroc ; Algérie-Lybie ; Algérie-sahel ; Algérie-France ; Algérie-Turquie ; Algérie-Moyen-Orient ; Algérie-pays du golf ; Algérie-Palestine ; Algérie-Chine ; Algérie-Russie. Ces combinaisons, à leur tour, sont la source de plusieurs documents numériques reprenant les mots clés qui évoquent les relations de l'Algérie avec les autres pays à plusieurs échelles : relation avec le Maghreb, avec les pays arabes, avec les pays européens et avec les pays du monde. Ces relations peuvent être d'ordre économiques, politiques, touristique. Etc.

2. Constitution d'une archive numérique

L'ensemble des éléments collectés restent accessibles depuis le Web. Cependant, une grande partie a connu une variation suite à l'instabilité du Web et de son évolution dans le temps. Pour cela, nous avons d'une part, classé les données dans des tableaux tout en sauvegardant les documents générés par chaque lien.

2.1 Classement des données pour constituer des corpus numériques

Suite aux étapes précédentes du tri et sélection des articles numériques, nous avons procédé au classement de toutes les données dans des tableaux, dans lesquels nous avons classé les informations ainsi que les documents générés par chaque site. Ces données sont répertoriées dans plus de 120 tableaux reprenant les combinaisons fixées préalablement tout en se basant sur notre objet d'étude. À titre d'illustration, nous n'avons repris que trois exemples d'items « Algérie-pays », « Algérie- centre » et « Algérie- continent » dans les tableaux ci-dessous.

Tableau 1 : les documents numériques reprenant les items « Algérie-pays »

Thème	Date	Média numérique	Titres et liens en ligne
Algérie – Pays	5/10/2019	Echoroukonline	L'Italie considère l'Algérie pays « sûr » https://www.echoroukonline.com/litalie-considere-lalgerie-pays-sur
	15/05/2019	Franceinfo	Crise en Algérie : un pays malade de sa caste dirigeante et de son pétrole https://www.francetvinfo.fr/monde/afrique/algerie/crise-en-algerie-un-pays-malade-de-sa-caste-dirigeante-et-de-son-petrole_3442879.html
	13/8/ 2019	TSA	L'Algérie, pays au stress hydrique « élevé » https://www.tsa-algerie.com › lalgerie-pays-au-stress-hydrique-eleve

Tableau 2 : les documents numériques reprenant les items « Algérie-centre »

Thème	Date	Média numérique	Titres et liens en ligne
Algérie- centralité	6 /8/2019	Algérie 360	Alger/ Trottoirs défoncés, saleté et laisser-aller, triste décor de la capitale : La Blanche noircit chaque jour https://www.algerie360.com/alger-trottoirs-defonces-salete-et-laisser-aller-triste-decor-de-la-capitale-la-blanche-noircit-chaque-jour/
	18 /5/2017	L'Echo d'Algérie	Vendredi, journée sans voiture à Alger centre https://www.lechodalgerie.dz/vendredi-journee-sans-voiture-a-alger-centre/
	29/05/2017	Autobip	La circulation automobile interdite à Alger centre et Bab el Oued tous les vendredis soir https://www.autobip.com/fr/actualite_auto/la_circulation_automobile_interdite_a_alger_centre_bab_el_oued_tous_les_vendredis_soir/4343

Tableau 3 : les documents numériques reprenant les items « Algérie-continent »

Thème	Date	Média numérique	Titres et liens en ligne
Algérie continent	/	lexpressiondz.com	"Le poids de l'Algérie dans le continent est important" https://www.vitamedz.com/fr/Algerie/le-poids-de-l-algerie-dans-le-6304895-Articles-0-18300-1.html
	25/05/2016	algeria-watch.org	Les Algériens ont la meilleure espérance de vie du continent africain https://algeria-watch.org/?p=13374
	11/03/2019	emb-argelia.pt	Sécurisation du continent et lutte contre le trafic d'armes "L'Algérie: un partenaire stratégique pour l'Otan" https://www.algerie360.com/securisation-du-continent-et-lutte-contre-le-trafic-darmes-lalgerie-un-partenaire-strategique-pour-lotan/

2.2 Les articles générés par les liens et les titres

Après avoir recensé tous les articles générés par les différentes interfaces numériques, nous avons pris le soin de sauvegarder l'ensemble des données. Cette sauvegarde va permettre aux chercheurs de consulter cette batterie de corpus. Cette dernière est constituée d'un ensemble d'articles numériques exploitable par les différents domaines de recherche. Nous avons renseigné dans le tableau ci-dessous, deux exemples d'articles ainsi que les informations liées à la mise en ligne de ces articles (date d'émission, média, support numérique, lien de disponibilité de l'article).

Source et titre	Article
L'Italie considère l'Algérie pays «sûr» Lien en ligne : https://www.echroukouonline.com/litalie-consider-lalgerie-pays-sur date : 5/10/2019 source : Echroukouonline	<p>Le gouvernement italien a considéré officiellement l'Algérie pays « sûr » ainsi qu'un certain nombre de pays, comme la Tunisie, le Maroc et le Sénégal.</p> <p>De ce fait, il sera procédé au rapatriement des migrants clandestins algériens qui ne cessent de rallier les côtes italiennes, notamment Lampedusa.</p> <p>Ainsi, les immigrants n'auront presque plus de chance de bénéficier de droit d'asile ou d'intégration dans ce pays européen, excepté des homosexuels qui pourraient en bénéficier eux qui prétendent être persécutés dans leurs pays d'origine, à savoir l'Algérie et la Tunisie.</p> <p>À ce titre, le ministre italien des Affaires étrangères, Luigi Di Maio a indiqué lors d'une conférence de presse, vendredi, avoir signé décret qui réduit à quatre mois le délai de rapatriement pour des migrants originaires de certains Etats africains et européens considérés des « pays sûrs ».</p> <p>Selon le chef de la diplomatie italienne, il y a une douzaine de pays qui figurent sur la liste de ceux qui doivent être insérés dans ce décret interministériel (Affaires étrangères, Justice, Intérieur), entre autres l'Algérie, le Maroc, la Tunisie, l'Albanie, la Bosnie-Herzégovine, la</p>

	<p>Serbie, le Cap Vert, le Ghana, le Kosovo, la Macédoine du Nord, le Monténégro, l'Ukraine et le Sénégal.</p> <p>Luigi Di Maio a fait savoir qu' « fonds rapatriements spécifique de 50 millions d'euros était prévu » pour accélérer les processus de rapatriement de migrants y compris algériens.</p> <p>Interrogé sur l'éventuelle persécution et la violation des droits des migrants homosexuels algériens et tunisiens, le responsable italien a précisé que de tels cas sont pris en charge par les lois et la constitution italienne, ouvrant ainsi la possibilité à cette catégorie d'obtenir un asile politique à condition de présenter un dossier attestant être persécuté dans son pays d'origine.</p>
<p>L'Algérie, pays au stress hydrique « élevé »</p> <p>https://www.tsa-algerie.com › lalgerie-pays-au-stress-hydrique-eleve</p> <p>13/8/ 2019 TSA</p>	<p>L'Algérie est classée à la 29e place des pays en fonction du degré de stress hydrique subi, selon le dernier classement établi par le think tank américain World Resources Institute.</p> <p>Selon cette étude publiée mardi dernier, la 29e place de l'Algérie classe le pays dans la catégorie rouge des pays au stress hydrique « élevé », avant-dernière catégorie (pourpre) des pays au stress hydrique « extrêmement élevé ».</p> <p>D'autres pays voisins de l'Algérie ou du pourtour méditerranéen figurent dans la même catégorie, tels que la Tunisie qui occupe la 30e position, l'Espagne (28e), l'Italie (44e), l'Égypte (43e) et le Maroc (22e).</p> <p>Le Qatar, Israël et le Liban occupent respectivement les trois premières places des pays au stress hydrique « extrêmement élevé ». Au total, 17 pays se trouvent dans cette catégorie, à l'image de la Libye (6e), de l'Iran (4e), de l'Arabie Saoudite (8e) ou de la Jordanie (5e).</p> <p>Le stress hydrique fait peser de graves menaces sur les vies humaines, les moyens de subsistance et la stabilité des entreprises. La croissance démographique, le développement socio-économique et l'urbanisation augmentent la demande en eau, tandis que le changement climatique peut rendre les précipitations et la demande plus variables.</p> <p>Les données du projet « Aqueduct » révèlent que 17 pays, qui concentrent près d'un quart de la population mondiale, sont confrontés à des niveaux de stress hydrique « extrêmement élevés ». Plus de 80 % de la ressource disponible chaque année en surface et dans les nappes phréatiques est pompée pour être consommée pour l'alimentation, l'agriculture ou encore l'industrie.</p>

Conclusion

Tout au long de cet article, nous avons tenté d'établir le processus de création et les dimensions constitutives d'une archive numérique pour faciliter l'accès aux données numériques par les chercheurs de différents domaines de recherche (sciences sociale, communication, sciences politiques, sciences humaines, etc.) Il était question dans un premier temps de définir le cadre dans lequel s'inscrit cette recherche, à savoir le projet de recherche à travers lequel nous avons essayé de faire l'état des lieux et ce en partant de la définition des notions d'archive et de corpus ainsi que l'ensemble des critères de constitution de ces derniers. Aborder la problématique de la constitution d'une base de données numérique à partir des interfaces numériques consiste pour nous à répertorier tous les articles générés par le moteur de recherche Google. Ce qui nous a permis de collecter une base de données spécifiques à la notion de l'espace tout en la liant au topique

« Algérie ». Il s'agit de répertorier tous les articles générés par le Web en réponse à une combinaison de 120 mots-clés reliant le sujet "Algérie" à d'autres termes spécifiques à l'espace (par exemple "Algérie-Pays, Algérie-Méditerranée, Algérie-Espace", etc.) Cette démarche nous a permis de collecter un ensemble d'articles, discours, vidéos, images, etc. Ces données brutes ont été, sélectionnées, triées, catégorisées pour créer une archive numérique... Ceux-ci forment à leur tour un ensemble de corpus numériques regroupés par des sous-thèmes qui répondent aux besoins du chercheur. À savoir la modélisation à partir des descriptions, schèmes cognitives, schèmes de structuration sémantico-pragmatiques, l'ensemble de ces derniers permettra l'interprétation et l'organisation discursive de ces données.

Références bibliographiques

- Barbier, J. & Mandret-Degeilh, A. (2018). Les archives numériques et numérisée S. Dans : J. Barbier & A. Mandret-Degeilh (Dir), *Le travail sur archives: Guide pratique* (pp. 195-222). Paris: Armand Colin.
- Bonnafous, S. (2006). L'analyse de discours », dans un ouvrage coordonné par S. Olivesi, *Science de l'information et de la communication, objet, savoir, discipline*, Ed. PUG.
- Charaudeau, P. (2009). Dis-moi quel est ton corpus, je te dirai quelle est ta problématique, *Corpus* [Online], 8 | Online since 01 July 2010, connection on 15 May 2023. URL: <http://journals.openedition.org/corpus/1674>; DOI: <https://doi.org/10.4000/corpus.1674>
- Dalbera, J-P. (2010). Le corpus entre données, analyses et théorie, *Corpus*, en ligne : <http://corpus.revues.org/10>
- Develotte, C. (2012). L'analyse des corpus multimodaux en ligne : état des lieux et perspectives, in *Actes du 3^e Congrès Mondial de Linguistique Française*, consulté le 3/3/2016 à l'adresse suivante : http://www.shs-conferences.org/articles/shsconf/abs/2012/01/shsconf_cmlf12_000213/shsconf_cmlf12_000213.html DOI : [10.1051/shsconf/20120100213](https://doi.org/10.1051/shsconf/20120100213)
- Favier, L. & El Hadi, W. (2015). Introduction - L'archivage numérique des savoirs: Perspectives européennes. *Les Cahiers du numérique*, 11, 9-14. <https://www.cairn.info/revue--2015-1-page-9.htm>
- Frey, V. (2009). La préservation de l'authenticité dans un environnement numérique. *Documentation et bibliothèques*, 55(3), 121-132. <https://doi.org/10.7202/1029099a> <https://hal-univ-paris13.archives-ouvertes.fr/hal-01163507/document>.
- Immune, Y. (2023). Spatialité du topique « Algérie » dans le discours : réflexions méthodologiques et problématisation, *Aleph* [En ligne], Vol 10 (2), mis en ligne le 31 mars 2023, consulté le 04 mai 2023. URL : <https://aleph-alger2.edinum.org/8230>
- Mayaffre, D. (2006). Corpus et Web-corpus. Réflexion sur la corporalité numérique. *Les cahiers de praxématique*, Montpellier : Presses universitaires de la Méditerranée, 2010, 54-55, pp.233-248. fhal-00916279f.
- Mayaffre, D. & Viprey, J-M. (2002). La cooccurrence. Du fait statistique au fait textuel, N°12, P4. Corpus in <https://journals.openedition.org/corpus/2183> ,
- Mayeur, I. & Paveau, M-A. (2020). Présentation. Les devenir du texte numérique natif, *Corela* [En ligne], HS-33, mis en ligne le 23 septembre 2020, consulté le 04

- mai 2023. URL : <http://journals.openedition.org/corela/11749>; DOI : <https://doi.org/10.4000/corela.11749>
- Paveau, M-A. (2015a). En naviguant en écrivant. Réflexions sur les textualités numériques ». In *Faire texte. Frontières textuelles et opérations de textualisation*, par Jean-Michel Adam.
- Rastier, F. (2017). Corpus numériques et accès à la culture, *Questions Vives* [En ligne], N° 28, mis en ligne le 06 novembre 2018, consulté le 06 avril 2023. URL : <http://journals.openedition.org/questionsvives/2421>; DOI : <https://doi.org/10.4000/questionsvives.2421>
- Rosier, L. (2003). Du discours rapporté à la circulation des discours : l'exemple des dictionnaires de "critique ironique", in https://www.academia.edu/20185485/Du_discours_rapport%C3%A9_%C3%A0_la_circulation_des_discours_lexemple_des_dictionnaires_de_critique_ironique
- Segouat, J. & al. (2010). Corpus de langue de signes : premières réflexions sur leur conception et leur représentativité, In : *L'exemple et le corpus : Quel statut ?*, Travaux linguistiques du CerLiCO, N° 23, PUR.