



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE
LA RECHERCHE SCIENTIFIQUE

**Université des Sciences et de la Technologie Houari
Boumediene**

Faculté d'Informatique

Département SIQ

Spécialité : HPC

Déploiement et Administration d'un Cloud Privé
basé sur OpenStack dans un Environnement de
Cluster de Calcul

BENMAMAR Katia

Amari lina

MEKADEM Radja

Djebassi lina

M2 HPC

2025/2026

Table des matières

TABLE DES MATIERES	2
FIGURES	5
TABLEAU	5
1. CONSTITUTION DU CORPUS	1
1.1 CHOIX DU MEDIA : THE NEW YORK TIMES	1
1.2 SELECTION DES CONFLITS ETUDIES : GAZA ET UKRAINE	1
1.3 METHODE DE COLLECTE DES DONNEES	1
2. NETTOYAGE ET PRETRAITEMENT DU CORPUS	2
2.1 MISE EN MINUSCULES	2
2.2 NETTOYAGE DES CARACTERES	2
2.3 TOKENISATION	2
2.4 SUPPRESSION DES MOTS VIDES (STOP WORDS)	3
2.5 LEMMATISATION (SANS STEMMING)	3
3. ÉTUDE GLOBALE DU CORPUS	3
3.1 STRUCTURATION DES ARTICLES	3
3.2 ANALYSE DESCRIPTIVE GLOBALE	4
4. ANALYSE LEXICALE	4
4.1 VOCABULAIRE DOMINANT-CORPUS GAZA	4
4.2 VOCABULAIRE DOMINANT – CORPUS UKRAINE	5
4.3 COMPARAISON DES VOCABULAIRES DOMINANTS	6
4.4 INTERPRETATION DU BIAIS MEDIATIQUE	6
5.1 ANALYSE DE LA DISPERSION LEXICALE	7
5.1.1 DISPERSION DES MOTS – CORPUS GAZA	7
5.1.2 DISPERSION DES MOTS – CORPUS UKRAINE	7
5.2 MISE EN RELATION AVEC LES FREQUENCES LEXICALES	8
5.3 INTERPRETATION DU BIAIS MEDIATIQUE	9
6 ANALYSE CONTEXTUELLE DES MOTS-CLES (CONCORDANCE ET CONTEXTES LEXICAUX)	9

6.1 ANALYSE CONTEXTUELLE – CORPUS GAZA	9
A) TERME « HAMAS »	9
B) TERME « KILLED » – CORPUS GAZA	9
C) ABSENCE DE CONTEXTES COMMUNS SIGNIFICATIFS	10
6.2 ANALYSE CONTEXTUELLE – CORPUS UKRAINE	10
A) TERME « UKRAINE »	10
B) TERME « KILLED » – CORPUS UKRAINE	10
C) CONTEXTES PARTAGES	11
6.3 SYNTHESE COMPARATIVE ET APPORT DE L'ANALYSE CONTEXTUELLE	11
 7 ANALYSE DE LA VISIBILITE LEXICALE DES APPELS A LA DESESCALADE (L'OCCUPATION)	 11
 7.1 CORPUS GAZA : FREQUENCE ET POIDS LEXICAL DE « CEASEFIRE »	 11
7.2 CORPUS UKRAINE : FREQUENCE ET POIDS LEXICAL DE « CEASEFIRE»	12
7.3 COMPARAISON ET INTERPRETATION DU BIAIS	12
7.4 APPORT A L'ANALYSE DU BIAIS MEDIATIQUE	13
 8 VISUALISATION SYNTHETIQUE PAR NUAGES DE MOTS	 13
 8.1 NUAGE DE MOTS – CORPUS GAZA	 13
ANALYSE	14
Interprétation	14
8.2 NUAGE DE MOTS – CORPUS UKRAINE	14
ANALYSE	15
Interprétation	15
8.3 SYNTHESE COMPARATIVE ET LIEN AVEC LE BIAIS MEDIATIQUE	15
 9. ANALYSE DES COOCCURRENCES LEXICALES PAR CATEGORIE GRAMMATICALE	 15
 9.1 CORPUS GAZA	 16
Interprétation :	17
9.2 CORPUS UKRAINE	17
Interprétation :	17
Conclusion :	17
CONCLUSION DE L'ANALYSE LEXICALE	18
 10 BIAIS INTERNE ENTRE GAZA ET ISRAËL	 18
 10.1 OBJECTIF	 18
10.2 METHODOLOGIE	19
10.3 RESULTATS	19
10.3.1 FREQUENCES LEXICALES	19
10.3.2 COOCCURRENCES ET CADRAGE LEXICAL	19

10.3.3 DISPERSION LEXICALE	19
10.3.4 ANALYSE CONTEXTUELLE	20
10.4 INTERPRETATION DU BIAIS INTERNE	20
10.5 SYNTHESE	20
 11 ANALYSE SEMANTIQUE :	 20
 11.1 ANALYSE SEMANTIQUE – CONCORDANCES ET CONTEXTE LEXICAL	 20
11.1.1 MOT-CLE : <i>ATTACK</i>	21
11.1.2 MOT-CLE : <i>TRUMP</i>	21
CONCLUSION	22
11.2. ANALYSE DES TEMOINS ET ORGANISATIONS (NAMED ENTITIES)	23
11.2.1 CORPUS GAZA	23
11.2.2 CORPUS UKRAINE	23
11.3. ANALYSE DES TEMOIGNAGES ET DISCOURS RAPPORTES	24
11.3.1 TEMOIGNAGES DANS LE CORPUS GAZA	24
11.3.1.1 Répartition des voix et statuts discursifs	24
11.3.1.2 Discours militaire israélien : rationalisation et légitimation	25
11.3.1.3. Civils palestiniens : humanisation émotionnelle mais dépolitisation	25
11.3.1.4. Hamas : parole politique conditionnelle et cadrée	26
11.3.1.5. Témoignages israéliens civils et question des otages	26
11.3.1.6. Institutions internationales et discours moral	27
11.3.2 TEMOIGNAGES DANS LE CORPUS UKRAINE	27
11.3.3 ANALYSE SEMANTIQUE AVEC WORD2VEC(WORD EMBEDDING):	28
11.3.3.1 Entraînement du modèle Word2Vec	28
11.3.3.2 Calcul du score de biais sémantique	28
11.3.3.3 Test sur des phrases représentatives	28
11.3.3.4. Analyse des résultats	29
11.3.3.5. Conclusion	29
 CONCLUSION	 30

Figures

<i>Figure 1 Les 30 mots les plus frequents (non cumulatif) Corpus Gaza</i>	5
<i>Figure 2 Les 30 mots les plus frequents (non cumulatif) Corpus Ukraine</i>	5
<i>Figure 3 Dispersion des mots – Corpus Gaza</i>	7
<i>Figure 4 Dispersion des mots– Corpus Ukraine</i>	8
<i>Figure 5 Nuage de mots – Corpus Gaza</i>	13
<i>Figure 6 Nuage de mots – Corpus Ukraine</i>	14

Tableau

<i>Tableau 1 Fréquences lexicales</i>	19
---------------------------------------	----

1. Constitution du corpus

1.1 Choix du média : The New York Times

Dans ce projet, nous avons choisi de constituer notre corpus exclusivement à partir du journal *The New York Times*. Ce choix repose sur plusieurs considérations méthodologiques et scientifiques.

Tout d'abord, *The New York Times* est un média occidental de référence, bénéficiant d'une forte audience internationale et d'une grande influence dans la construction de l'opinion publique. Il constitue ainsi un terrain d'étude pertinent pour analyser les mécanismes de cadrage médiatique et les biais discursifs potentiels dans la couverture des conflits armés.

Ensuite, le fait de se concentrer sur un seul média permet de limiter les biais structurels liés aux différences de lignes éditoriales, de formats journalistiques ou de styles rédactionnels. Cette homogénéité garantit que les différences observées dans les analyses lexicales sont davantage liées au traitement du sujet qu'à des divergences entre médias.

Enfin, l'utilisation d'un corpus homogène facilite le prétraitement linguistique, la reproductibilité du code et la comparaison directe entre plusieurs conflits traités par une même institution médiatique.

1.2 Sélection des conflits étudiés : Gaza et Ukraine

Deux conflits armés contemporains ont été sélectionnés :

- la guerre à Gaza,
- la guerre en Ukraine.

Ce choix permet de mener une analyse comparative entre deux conflits majeurs, largement couverts par les médias occidentaux, mais impliquant des acteurs géopolitiques et des alliances différentes. L'objectif est d'identifier d'éventuels doubles standards discursifs dans la manière dont les acteurs, les victimes et les responsabilités sont lexicalement représentés.

1.3 Méthode de collecte des données

Nous avons travaillé sur la version texte brute des articles, et non sur leur version HTML.

Le site du *New York Times* étant fortement dynamique et protégé contre le scraping automatique, les bibliothèques classiques telles que *BeautifulSoup* ne permettent pas d'accéder directement au contenu textuel. De plus, les tentatives d'extraction via *Selenium* ont entraîné un blocage automatique du navigateur par le site.

Pour ces raisons, la collecte du corpus a été réalisée manuellement, à partir du texte visible des articles. Cette approche garantit :

- l'accès au contenu journalistique réel,
- le respect des contraintes techniques et légales,
- la qualité linguistique du corpus analysé.

2. Nettoyage et prétraitement du corpus

Le prétraitement constitue une étape essentielle en traitement automatique du langage naturel (NLP). Il vise à réduire le bruit linguistique, à normaliser les données et à garantir la fiabilité des analyses statistiques ultérieures.

2.1 Mise en minuscules

L'ensemble des textes est converti en minuscules afin d'éliminer les variations liées à la casse (par exemple *Israel* vs *israel*). Cette normalisation permet de regrouper correctement les occurrences lexicales identiques.

2.2 Nettoyage des caractères

À l'aide d'expressions régulières, nous avons procédé à :

- la suppression de la ponctuation et des symboles spéciaux,
- la normalisation des espaces,
- la suppression des sauts de ligne inutiles.

Les lettres **et les chiffres** ont été conservés afin de préserver des informations essentielles telles que :

- les dates, les références temporelles (ex. *October 7, 2023*).
- les bilans humains, nombre de victims

2.3 Tokenisation

Le texte nettoyé est ensuite segmenté en unités lexicales appelées *tokens*, correspondant principalement aux mots.

Cette étape est indispensable pour appliquer les traitements linguistiques ultérieurs tels que la suppression des mots vides, la lemmatisation ou le calcul des fréquences.

2.4 Suppression des mots vides (stop words)

Les mots vides de la langue anglaise (*the, and, is, to, of, etc.*) ont été supprimés à l'aide de la liste fournie par la bibliothèque **NLTK**.

Ces mots, bien que très fréquents, n'apportent pas d'information sémantique pertinente et peuvent fausser les analyses statistiques s'ils sont conservés.

2.5 Lemmatisation (sans stemming)

Nous avons choisi d'appliquer **uniquement la lemmatisation**, et non le stemming.

Le stemming produisait des formes trop dégradées, nuisant à l'interprétation :

- Hamas → hama
- Israeli / Israelis → isra

La lemmatisation permet au contraire de regrouper les formes morphologiques tout en conservant le sens linguistique et la lisibilité des résultats.

3. Étude globale du corpus

3.1 Structuration des articles

Le corpus brut a été restructuré en articles individuels à l'aide d'une fonction de segmentation basée sur split. Cette fonction découpe le texte en segments à partir du marqueur **=== METADATA ===**, permettant d'isoler chaque article.

Le principe de fonctionnement peut être résumé par le pseudocode suivant :

Entrée : Texte brut du corpus

1. Découper le texte en segments avec la fonction split sur "=== METADATA ==="
2. Pour chaque segment :
 - Extraire les métadonnées
 - Extraire le texte de l'article
 - Stocker l'article comme une unité indépendante

Sortie : Corpus structuré d'articles

Cette approche permet :

- une analyse article par article,
- une analyse par auteur,
- des comparaisons internes au corpus.

3.2 Analyse descriptive globale

L'analyse descriptive comprend :

- l'identification des auteurs,
- le nombre d'articles par auteur,

Pseudo-code pour l'analyse par auteur :

Entrée : Corpus structuré d'articles

1. Initialiser un dictionnaire `articles_par_auteur` vide
2. Pour chaque article dans le corpus :
 - Récupérer l'auteur (ou "Unknown" si absent)
 - Récupérer le texte de l'article
 - Ajouter le texte dans la liste associée à cet auteur
3. Pour chaque auteur :
 - Afficher le nombre d'articles

Sortie : Dictionnaire `articles_par_auteur` avec listes de textes par auteur

Cette approche permet :

- la taille totale du corpus (nombre de mots),
- la taille du vocabulaire,
- la diversité lexicale.

La diversité lexicale obtenue est d'environ **0,18** pour les deux corpus (Gaza : 0,18398 ; Ukraine : 0,18381), indiquant une forte répétition du vocabulaire.

Cette similarité est cohérente avec :

- un discours journalistique standardisé,
- une couverture focalisée sur un nombre limité de thèmes,
- un corpus issu d'un seul média.

4. ANALYSE LEXICALE

L'analyse lexicale du corpus permet d'identifier les régularités et les asymétries dans l'usage des mots, offrant ainsi une base pour évaluer le cadrage médiatique et la présence de biais.

4.1 Vocabulaire dominant-Corpus Gaza

Résultat observé :

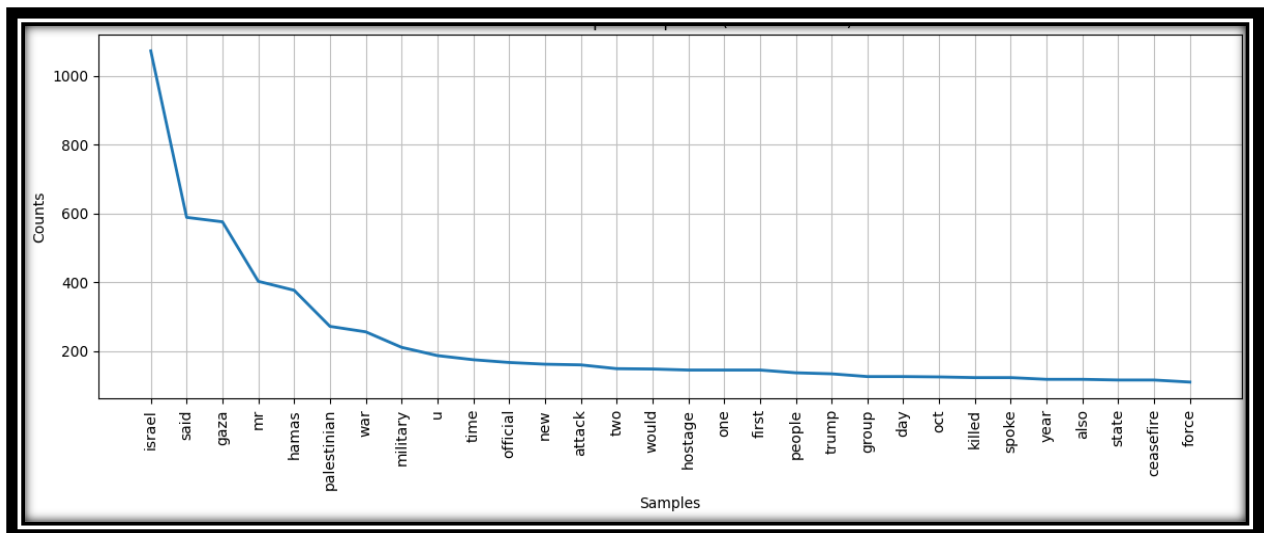


Figure 1 Les 30 mots les plus fréquents (non cumulatif) Corpus Gaza

Les termes les plus fréquents sont :

[('israel', 1073), ('said', 589), ('gaza', 576), ('mr', 403), ('hamas', 377), ('palestinian', 272), ('war', 256), ('military', 211), ('u', 187), ('time', 175)]

L'analyse montre que le corpus est structuré autour de **termes centraux liés aux acteurs du conflit (israel, hamas, palestinian)**, au **champ militaire (war, military)** et au **discours rapporté (said, mr)**. Cette distribution reflète un traitement **institutionnel et conflictuel**, caractéristique d'un discours journalistique centré sur les déclarations officielles et les actions militaires.

4.2 Vocabulaire dominant – Corpus Ukraine

Résultat observé :

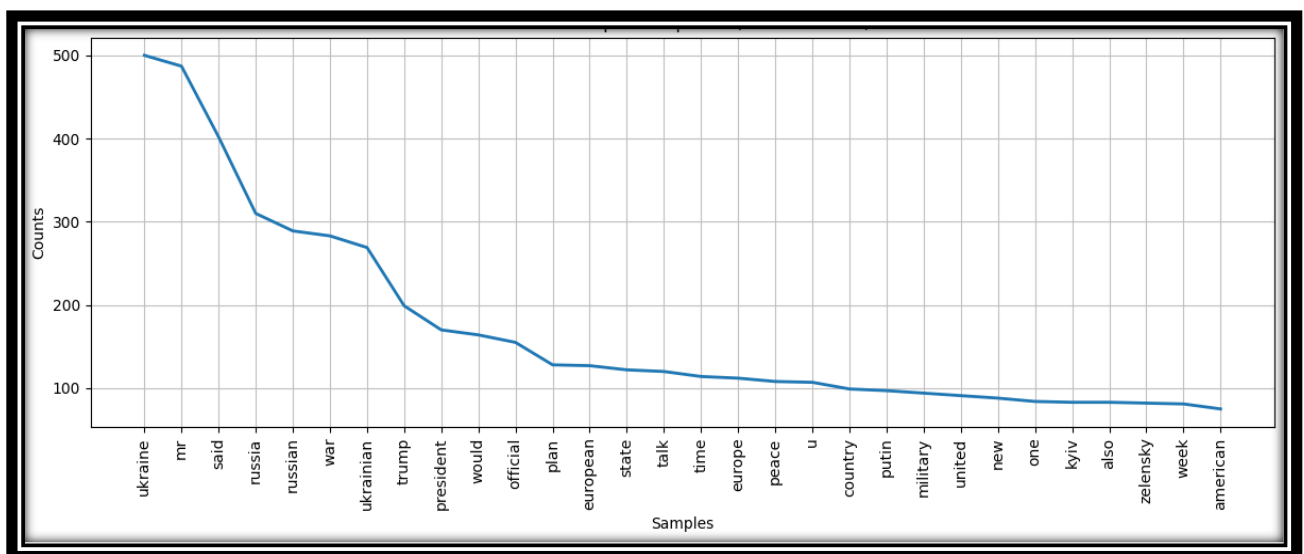


Figure 2 Les 30 mots les plus fréquents (non cumulatif) Corpus Ukraine

Les termes les plus fréquents sont :

('ukraine', 500), ('mr', 487), ('said', 402), ('russia', 310), ('russian', 289), ('war', 283), ('ukrainian', 269), ('trump', 199), ('president', 170), ('official', 155), ('plan', 128), ('european', 127), ('state', 122), ('talk', 120)

L'analyse du vocabulaire dominant du corpus Ukraine met en évidence une forte centralité des acteurs étatiques (*Ukraine, Russia, Russian, Ukrainian*), indiquant un cadrage géopolitique du conflit.

La fréquence élevée des termes *mr* et *said* révèle un recours important au discours rapporté, caractéristique d'un traitement journalistique fondé sur les déclarations officielles.

Par ailleurs, la présence marquée de termes tels que *president, official, state, plan* et *talk* souligne une focalisation sur les décisions politiques, les négociations diplomatiques et les stratégies institutionnelles.

Le conflit ukrainien est ainsi présenté comme un enjeu politique et international, impliquant des acteurs étatiques et des équilibres géopolitiques globaux.

4.3 Comparaison des vocabulaires dominants

Le corpus Gaza est principalement structuré autour d'un champ militaire et d'acteurs directement impliqués dans les combats.

À l'inverse, le corpus Ukraine se distingue par une forte présence de termes politiques, institutionnels et internationaux.

Cette différence suggère un cadrage distinct : militaire pour Gaza, géopolitique pour l'Ukraine.

->Gaza = conflit militarisé, déshumanisé lexicalement

->Ukraine = conflit politisé, humanisé et légitimé moralement

4.4 Interprétation du biais médiatique

L'analyse conjointe des fréquences et de la dispersion lexicale met en évidence un biais discursif implicite.

Le conflit de Gaza est majoritairement cadré comme un affrontement militaire, avec une visibilité limitée des victimes civiles.

À l'inverse, la guerre en Ukraine est présentée comme une crise géopolitique et morale, intégrant plus fortement les dimensions humanitaires et politiques.

5.1 Analyse de la dispersion lexicale

5.1.1 Dispersion des mots – Corpus Gaza

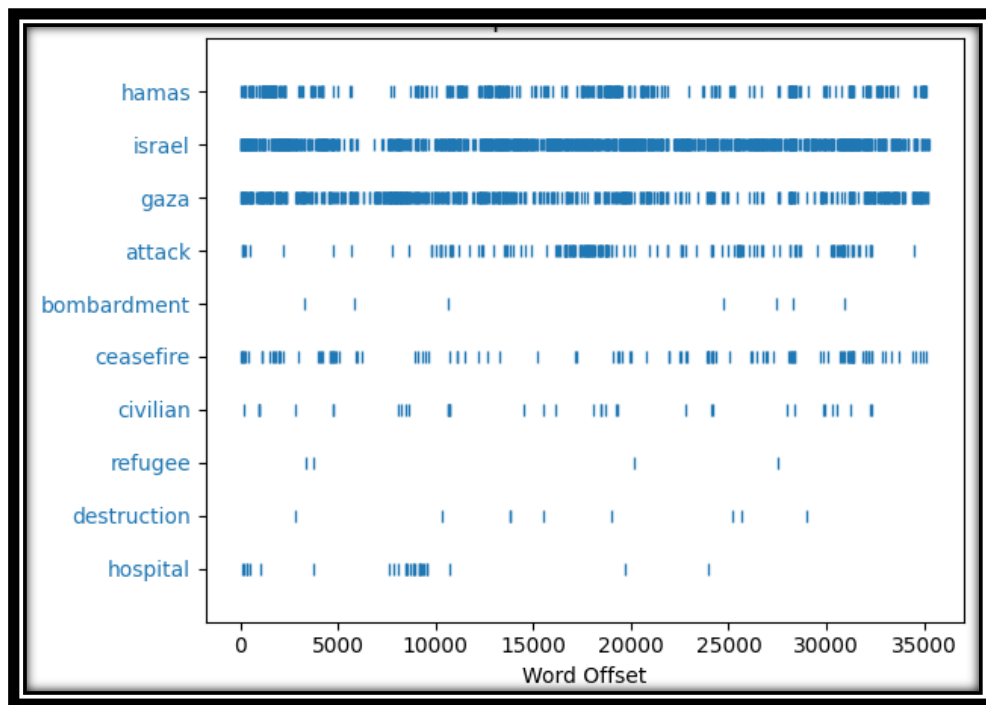


Figure 3 Dispersion des mots – Corpus Gaza

La dispersion montre que les termes *Israel*, *Gaza* et *Hamas* sont répartis de manière relativement homogène sur l'ensemble du corpus. Cette homogénéité indique que ces acteurs constituent des éléments structurants du discours médiatique et sont mobilisés de façon continue dans les articles.

À l'inverse, certains termes tels que *attack* ou *hostage* apparaissent de manière plus ponctuelle, suggérant qu'ils sont associés à des événements spécifiques plutôt qu'à une narration globale et continue.

-> Cette distribution révèle un cadrage discursif centré sur les acteurs armés et les opérations militaires, où les éléments stratégiques et sécuritaires dominent la structure du récit journalistique.

5.1.2 Dispersion des mots – Corpus Ukraine

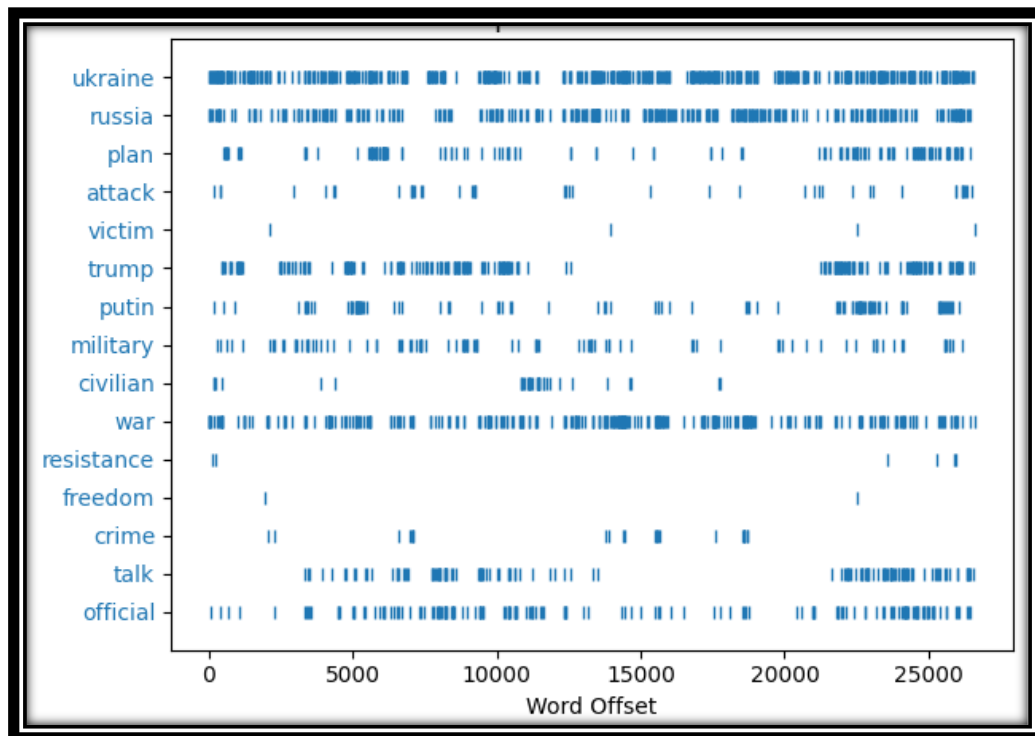


Figure 4 Dispersion des mots– Corpus Ukraine

Les termes *Ukraine* et *Russia* présentent une dispersion homogène, indiquant leur présence constante tout au long du corpus. Les mots liés aux institutions politiques (*president*, *official*, *talk*, *plan*) apparaissent également de manière régulière, soulignant l'importance du discours institutionnel et diplomatique.

Contrairement au corpus Gaza, la dispersion met en évidence une intégration stable des termes politiques dans la narration globale du conflit.

-> Le conflit ukrainien est ainsi structuré comme un récit géopolitique continu, dans lequel les décisions étatiques, les stratégies diplomatiques et les négociations occupent une place centrale dans le discours médiatique.

5.2 Mise en relation avec les fréquences lexicales

Les résultats de dispersion confirment les observations issues de l'analyse fréquentielle.

- Dans le corpus Gaza, les termes les plus fréquents sont *Israel* (1073), *Gaza* (576) et *Hamas* (377), tandis que les termes liés aux civils et aux infrastructures humanitaires sont nettement moins fréquents.
- Dans le corpus Ukraine, les fréquences élevées de *Ukraine* (500), *Russia* (310) et *President* (170), combinées à une dispersion plus régulière des termes *civilian* et *peace*, indiquent un discours plus institutionnel et humanitaire.

5.3 Interprétation du biais médiatique

L'analyse conjointe des fréquences et de la dispersion lexicale met en évidence un biais discursif implicite.

Le conflit de Gaza est majoritairement cadré comme un affrontement militaire centré sur des acteurs armés, avec une visibilité limitée des victimes civiles et des enjeux humanitaires.

À l'inverse, la guerre en Ukraine est présentée comme une crise géopolitique et morale, intégrant plus fréquemment les dimensions humanitaires, politiques et symboliques.

Ce cadrage différencié contribue à une hiérarchisation implicite des souffrances et participe à la construction de représentations asymétriques des conflits dans le discours médiatique.

6 Analyse contextuelle des mots-clés (Concordance et contextes lexicaux)

6.1 Analyse contextuelle – Corpus Gaza

a) Terme « Hamas »

L'analyse de concordance du terme *Hamas* montre que celui-ci apparaît majoritairement dans des contextes militaires et sécuritaires. Les segments de phrases associent fréquemment *Hamas* à des verbes et expressions tels que **attack, targeted, destroy, militant, fighters, ou encore battle against Hamas**.

Ces contextes lexicalisent *Hamas* comme un **acteur militaire central**, souvent présenté comme une entité homogène et menaçante. Les fonctions **similar et common contexts** confirment cette tendance, les termes partageant des environnements lexicaux similaires étant principalement liés à la violence, aux opérations militaires et à la sécurité.

→ Hamas est ainsi construit discursivement comme une cible militaire, intégrée dans une narration stratégique et opérationnelle.

b) Terme « killed » – Corpus Gaza

L'analyse contextuelle du terme **killed** révèle une forte association avec :

- des frappes aériennes,
- des bombardements,
- des descriptions d'actions militaires israéliennes.

Bien que le mot **civilians** apparaisse dans certains contextes, il reste souvent intégré dans des formulations générales (*many civilians have been killed*), sans individualisation ni contextualisation approfondie des victimes.

→ La mort est principalement rapportée comme **conséquence collatérale d'actions militaires**, et non comme un événement humanitaire central structurant le récit.

c) Absence de contextes communs significatifs

Les résultats de **common_contexts** indiquent une absence de contextes lexicaux partagés entre *Hamas, Israel, civilians et Gaza*.

Cette absence suggère une **fragmentation discursive**, où les acteurs armés et les civils ne sont pas systématiquement intégrés dans un même cadre narratif.

→ Cela renforce l'idée d'un **discours déshumanisé**, centré sur les acteurs militaires plutôt que sur les victimes civiles.

6.2 Analyse contextuelle – Corpus Ukraine

a) Terme « Ukraine »

L'analyse de concordance du terme *Ukraine* montre une forte présence dans des contextes géopolitiques et historiques : *invasion, war, territory, NATO, Europe, counteroffensive, peace plan*.

Les fonctions **similar** indiquent que *Ukraine* partage ses contextes avec des termes liés à la diplomatie, aux stratégies militaires étatiques et aux relations internationales.

→ Le conflit est inscrit dans un **récit géopolitique structuré et continu**, dépassant le simple cadre militaire.

b) Terme « killed » – Corpus Ukraine

Contrairement au corpus Gaza, le terme **killed** apparaît fréquemment dans des contextes où les victimes sont **explicitement qualifiées** (*civilians killed, people were killed, victims*), avec des références précises à des lieux, des dates et des bilans humains.

La mort est ici intégrée dans une narration plus détaillée, souvent associée à une **condamnation morale implicite** des attaques.

→ Les victimes civiles occupent une place plus visible et structurante dans le discours.

c) Contextes partagés

Bien que **common_contexts** ne fasse pas apparaître de structures strictement identiques, l'analyse qualitative montre une **cooccurrence régulière** entre les termes **Ukraine, civils, killed et war**, indiquant une intégration discursive plus cohérente entre conflit armé et souffrance humaine.

6.3 Synthèse comparative et apport de l'analyse contextuelle

L'analyse contextuelle confirme et approfondit les résultats issus des analyses fréquentielles et de dispersion :

- Corpus Gaza :
 - Les acteurs armés sont au centre du récit.
 - Les civils et les victimes apparaissent de manière marginale et peu contextualisée.
 - Le discours privilégie un cadrage militaire et sécuritaire.
- Corpus Ukraine :
 - Les civils et les victimes sont davantage intégrés dans la narration.
 - Le conflit est présenté comme une crise géopolitique et morale.
 - Les dimensions humanitaires et politiques sont lexicalement visibles.

→ L'analyse par contexte révèle ainsi un **biais discursif qualitatif**, où les mêmes notions (guerre, morts, civils) ne sont pas construites de la même manière selon le conflit traité.

7 Analyse de la visibilité lexicale des appels à la désescalade (l'occupation)

7.1 Corpus Gaza : fréquence et poids lexical de « ceasefire »

Dans le corpus Gaza, le terme *ceasefire* apparaît **116 fois**, ce qui correspond à un **taux d'occupation lexicale de 0,34 %** du corpus.

Ce pourcentage reste faible lorsqu'il est comparé aux termes centraux du discours :

- **Israel** : 1073 occurrences (3,12 %),
- **Gaza** : 576 occurrences (1,67 %),
- **Hamas** : 377 occurrences (1,10 %),
- **attack** : 160 occurrences (0,46 %).

Les termes associés aux infrastructures humanitaires et aux conséquences civiles du conflit présentent également une visibilité très réduite :

- **hospital** : 28 occurrences (0,08 %),
- **bombardement** : 7 occurrences (0,02 %).

→ **Malgré une présence non nulle**, le terme *ceasefire* demeure marginal dans le corpus Gaza, indiquant que les appels à la désescalade occupent une place secondaire par rapport au cadrage militaire et sécuritaire du conflit.

7.2 Corpus Ukraine : fréquence et poids lexical de « ceasefire »

Dans le corpus Ukraine, le terme *ceasefire* apparaît seulement **4 fois**, avec un **taux d'occupation lexicale de 0,015 %**, soit une visibilité extrêmement faible.

Cependant, cette faible fréquence s'inscrit dans un environnement lexical différent :

- **Ukraine** : 500 occurrences (1,90 %),
- **Russia** : 310 occurrences (1,18 %),
- **military** : 94 occurrences (0,36 %),
- **civilian et victim** apparaissent, bien que faiblement, de manière régulière dans le discours.

Par ailleurs, des termes politiques et diplomatiques tels que **Trump** (199 occurrences) et **Putin** sont fortement présents, indiquant que la question de la guerre est intégrée dans un cadre géopolitique et institutionnel plus large.

→ Le discours sur l'Ukraine accorde peu de place explicite au terme **ceasefire**, mais intègre davantage la question de la guerre dans une narration diplomatique et politique globale.

7.3 Comparaison et interprétation du biais

La comparaison des deux corpus met en évidence un contraste paradoxal :

- **Gaza** :
 - *Ceasefire* est plus fréquent qu'en Ukraine (116 vs 4 occurrences),
 - mais reste **lexicalement marginal** face à la domination des termes militaires et sécuritaires.
- **Ukraine** :
 - *Ceasefire* est quasi absent,
 - mais le discours inclut davantage de références aux civils, aux victimes et aux enjeux diplomatiques, ce qui inscrit indirectement la question de la paix dans le cadre narratif.

Analyse

La forte visibilité des termes *Russia* et *Ukraine* confirme la centralité des **acteurs étatiques** dans la narration du conflit.

Contrairement au corpus Gaza, on observe une présence marquée de termes relevant du **champ politique et diplomatique** (*president, official, plan, talk*), ainsi que du lexique de la résolution (*peace*).

La dimension militaire est présente (*war*), mais elle coexiste avec une structuration discursive orientée vers les **négociations, les stratégies politiques et les équilibres internationaux**.

Interprétation

Le nuage de mots reflète un cadrage **géopolitique et institutionnel** du conflit ukrainien, dans lequel la guerre est inscrite dans un récit politique global et non uniquement militaire.

8.3 Synthèse comparative et lien avec le biais médiatique

La comparaison des deux nuages de mots confirme visuellement les analyses précédentes :

- Corpus Gaza :
 - dominance des acteurs armés et du lexique militaire,
 - faible visibilité des civils et des thématiques humanitaires.
- Corpus Ukraine :
 - centralité des acteurs étatiques et politiques,
 - présence explicite du lexique diplomatique et de la paix.

→ Cette asymétrie visuelle renforce l'hypothèse d'un **biais de cadrage médiatique**, où le conflit de Gaza est majoritairement représenté comme une confrontation militaire, tandis que la guerre en Ukraine est présentée comme une crise géopolitique légitime et négociable.

9. Analyse des cooccurrences lexicales par catégorie grammaticale

Pour approfondir la représentation discursive des acteurs dans les deux conflits, nous avons extrait les **adjectifs, verbes et noms** co-occurents avec chaque acteur à partir des phrases où ils apparaissent. Cette analyse permet de révéler les traits, actions et thèmes associés à chaque groupe dans le discours médiatique.

Entrée :

- Texte du corpus (texte)
- Liste de mots-clés représentant un acteur (keywords)

Procédure :

1. Découper le texte en phrases.
2. Initialiser des compteurs pour les adjectifs, verbes et noms.
3. Pour chaque phrase :
 - a. Vérifier si elle contient un des mots-clés.
 - b. Si oui, identifier chaque mot et sa catégorie grammaticale (POS).
 - c. Pour chaque mot :
 - Si c'est un adjectif et qu'il n'est pas dans la liste d'exclusion, l'ajouter au compteur d'adjectifs.
 - Si c'est un verbe, l'ajouter au compteur de verbes.
 - Si c'est un nom, l'ajouter au compteur de noms.
4. Retourner les compteurs pour adjectifs, verbes et noms.
5. Afficher les 10 cooccurrences les plus fréquentes pour chaque catégorie.

Sortie :

- Listes des adjectifs, verbes et noms les plus fréquemment associés à chaque acteur.

9.1 Corpus Gaza

Acteurs palestiniens (Gaza, Hamas, Palestinian)

- **Adjectifs dominants** : *Israeli, military, humanitarian, international, southern*
→ Les adjectifs associés aux Palestiniens reflètent principalement un cadrage militaire et externe, avec une attention marginale à l'aide humanitaire.
- **Verbes dominants** : *said, has, was, is, had, were, have, be, been*
→ Le discours est largement centré sur le récit rapporté, les déclarations officielles et les actions des acteurs militaires.
- **Noms dominants** : *Israel, war, Israeli, Palestinians, officials*
→ Les noms indiquent que les Palestiniens sont souvent évoqués en relation avec Israël et le conflit militaire, avec une faible visibilité des civils ou des infrastructures humanitaires.

Acteurs israéliens (Israel, IDF, soldier)

- **Adjectifs dominants** : *Israeli, military, Palestinian, southern, humanitarian*
→ Le cadrage met l'accent sur Israël militaire, mais inclut également des références aux Palestiniens comme entité.

- **Verbes dominants** : *said, has, was, is, have, had, be, were, are*
→ Le discours rapporté est central, reflétant le rôle de l'Israël institutionnel et militaire dans le récit.
- **Noms dominants** : *Gaza, Israeli, Hamas, war, mr., officials*
→ Israël est présenté en interaction directe avec Gaza et Hamas, soulignant le cadre militaire du conflit.

Interprétation :

L'analyse révèle un **cadrage militaire et sécuritaire**, où les Palestiniens sont associés à des lieux et groupes armés, tandis qu'Israël est au centre du récit comme acteur structurant.

9.2 Corpus Ukraine

Acteurs ukrainiens (Ukrainian, Kyiv, heroic resistance)

- **Adjectifs dominants** : *Russian, military, European, social*
→ Les adjectifs reflètent à la fois la présence de l'adversaire et un cadrage politique/international.
- **Verbes dominants** : *said, was, had, is, be, have, has*
→ Le discours rapporté est prépondérant, mais certaines actions décrivent des stratégies ou résistances.
- **Noms dominants** : *Ukraine, Ukrainians, officials, war, Russia*
→ Les noms mettent en avant le pays et sa population, ainsi que les aspects institutionnels et politiques.

Acteurs russes (Russian, Moscow, soldier)

- **Adjectifs dominants** : *Ukrainian, military, Eastern, European*
→ Les adjectifs présentent l'adversaire et un contexte géopolitique.
- **Verbes dominants** : *said, has, was, had, is, are, been*
→ Les verbes décrivent principalement des actions ou déclarations institutionnelles.
- **Noms dominants** : *Ukraine, war, Russia, forces, Trump*
→ Les noms reflètent la confrontation militaire et l'inscription du conflit dans un contexte diplomatique et international.

Interprétation :

Dans le corpus Ukraine, les termes liés aux civils et aux acteurs institutionnels sont plus fréquents et dispersés, ce qui traduit un **cadrage géopolitique et humanisé**, où la guerre est intégrée dans un récit diplomatique global.

Conclusion :

Cette analyse confirme que le **cadrage médiatique diffère selon le conflit** : Gaza est représenté de manière militarisée et déshumanisée, tandis que l'Ukraine est cadrée dans un récit

géopolitique et humanisé, avec une attention relative aux victimes et aux stratégies institutionnelles.

Conclusion de l'analyse lexicale

L'ensemble des analyses lexicographiques – vocabulaire dominant, dispersion lexicale, nuages de mots, fréquence des termes et cooccurrences par catégorie grammaticale – met en évidence des **différences nettes de cadrage médiatique** entre les deux conflits étudiés.

- **Corpus Gaza :**

Le discours est **majoritairement militarisé et sécuritaire**, centré sur les acteurs armés (Israël, Hamas) et les lieux du conflit (Gaza). Les civils et les dimensions humanitaires sont lexicalement marginalisés. Les verbes et adjectifs associés renforcent un cadrage opérationnel et stratégique, tandis que les termes rapportés témoignent d'une dépendance au discours institutionnel.

→ **Biais implicite** : focalisation sur les forces militaires et le cadre sécuritaire, avec une représentation déshumanisée des Palestiniens.

- **Corpus Ukraine :**

Le discours présente un **cadrage géopolitique et institutionnel**, où les acteurs étatiques (Ukraine, Russia) et les enjeux diplomatiques occupent une place centrale. Les civils et les victimes sont davantage intégrés dans le récit, et la narration prend en compte les stratégies politiques et les négociations internationales.

→ **Biais implicite** : le conflit est humanisé et légitimé moralement, avec une représentation structurée des victimes et du rôle des institutions.

- **Synthèse comparative :**

Gaza = conflit **militarisé, centré sur les acteurs armés**, récit fragmenté et déshumanisant.

Ukraine = conflit **politico-diplomatique et humanisé**, récit structuré et cohérent, intégrant civils et dimensions institutionnelles.

Cette analyse lexicale fournit ainsi une **base solide pour l'analyse sémantique**, qui pourra explorer non seulement la fréquence et la distribution des mots, mais surtout **le sens, les relations et les cadres discursifs** dans lesquels ces termes sont mobilisés.

10 biais interne entre Gaza et Israël

10.1 Objectif

Cette analyse vise à identifier les asymétries de représentation au sein du corpus Gaza, en comparant la visibilité lexicale, les cooccurrences et le cadrage discursif des acteurs **Israël** et **Gaza / Palestiniens**. L'objectif est de détecter les biais internes au traitement médiatique d'un même conflit.

10.2 Méthodologie

- Extraction des fréquences des termes associés à chaque acteur :
 - Israël : Israel, IDF, soldier, Israeli, officials
 - Palestiniens : Gaza, Hamas, Palestinian, civilian
- Analyse des adjectifs, verbes et noms co-occurents avec chaque acteur.
- Étude de la dispersion lexicale pour évaluer la répartition des termes tout au long du corpus.
- Analyse contextuelle (concordances) pour identifier les cadres narratifs : militaire, humanitaire, diplomatique, etc.

10.3 Résultats

10.3.1 Fréquences lexicales

Terme	Occurrences	% du corpus
<i>Israel</i>	1073	3,12 %
<i>Gaza</i>	576	1,67 %
<i>Hamas</i>	377	1,10 %
<i>Palestinian</i>	272	0,79 %
<i>civilian</i>	36	0,10 %
<i>IDF</i>	9	0,03 %

Tableau 1 Fréquences lexicales

→ Israël apparaît près de **2 fois plus souvent que Gaza**, et les termes liés aux civils palestiniens sont très minoritaires.

10.3.2 Cooccurrences et cadrage lexical

- **Israël** : fortement associé à des adjectifs et noms militaires (Israeli, military, officials) et des verbes de déclaration (said, has, was).
- **Palestiniens / Gaza** : souvent mentionnés à travers des noms géographiques ou groupes armés (Gaza, Hamas), avec peu de mise en valeur des civils.
- Les termes humanitaires (hospital, civilians) restent marginalisés.

10.3.3 Dispersion lexicale

- Israël est mentionné de manière continue dans tout le corpus, ce qui en fait un **acteur central du récit**.
- Gaza et Hamas sont répartis de façon homogène mais moins dominante, tandis que les civils apparaissent de manière ponctuelle, généralement liés à des événements spécifiques (attaques, bombardements).

10.3.4 Analyse contextuelle

- Les concordances montrent que **Israël** est présenté dans des contextes stratégiques ou officiels : opérations militaires, déclarations gouvernementales.
- Les Palestiniens / Gaza sont associés aux actions militaires et aux conflits ponctuels, rarement à des récits humanitaires ou moraux.
- Les cooccurrences entre “Israel” et “civilians” sont rares, tandis que Gaza et “civilian” apparaissent uniquement dans des mentions collatérales.

10.4 Interprétation du biais interne

- Le corpus **favorise Israël** comme acteur central et structurant.
- Gaza et les Palestiniens sont majoritairement cadrés comme **entités territoriales ou groupes armés**, avec une faible visibilité des civils et de l’impact humanitaire.
- Le discours crée un **biais interne** : il hiérarchise implicitement les acteurs, donne plus de légitimité et d’attention aux actions israéliennes et marginalise la souffrance palestinienne.

10.5 Synthèse

- L’analyse lexicale confirme un **cadrage asymétrique interne au conflit** Gaza-Israël : Israël est central et actif, Gaza/Palestiniens sont périphériques et déshumanisés.
- Ce biais interne complète l’analyse comparative Gaza-Ukraine : non seulement Gaza est traité militarisé par rapport à Ukraine, mais à l’intérieur du même corpus, Israël domine la narration tandis que les Palestiniens sont relégués au second plan.

11 Analyse sémantique :

11.1 Analyse sémantique – Concordances et contexte lexical

L’analyse sémantique permet d’examiner les contextes précis dans lesquels certains mots-clés sont utilisés, révélant le cadrage discursif et les biais implicites dans le traitement médiatique des conflits. Pour cela, nous avons extrait des **concordances** avec une fenêtre de mots autour du terme cible.

11.1.1 Mot-clé : *attack*

Corpus Gaza :

Les occurrences du mot *attack* montrent une forte association avec le Hamas et les actions militaires en Israël.

Exemples :

- “in retaliation for a Hamas attack in nearby Rafah that it ...”
- “taken hostage during the Hamas-led attack on Oct. 7, 2023 ...”
- “was ignited by the Hamas-led attack on Israel on Oct. 7 ...”

Analyse du biais interne Gaza-Israël :

- Le terme *attack* est systématiquement lié aux attaques attribuées à des acteurs palestiniens (Hamas) et aux réponses militaires israéliennes.
- Israël apparaît comme l'acteur central du récit, alors que Gaza/les Palestiniens sont cadrés principalement via des événements de violence.
- Les civils palestiniens sont rarement individualisés dans les contextes, ce qui révèle un **biais interne déshumanisant** dans le corpus Gaza.

Corpus Ukraine :

Les occurrences du mot *attack* sont intégrées dans un cadre plus géopolitique et militaire, avec des références aux victimes civiles et aux stratégies étatiques.

Exemples :

- “least 50 civilians. The attack came at the start of ...”
- “of 89 soldiers in the attack; Ukraine said hundreds had ...”
- “military if China were to attack Taiwan. But in remarks ...”

Analyse du biais externe Gaza-Ukraine :

- Dans le corpus Ukraine, les attaques sont contextualisées politiquement et humanitairement, avec des mentions de civils et de victimes.
- Comparativement au corpus Gaza, la violence dans Gaza est plus centrée sur Israël comme acteur dominant, alors que la violence en Ukraine est décrite dans un récit géopolitique et moralement légitimé.
- Cela révèle un **biais externe** : le conflit de Gaza est cadré militairement et déshumanisé, tandis que l'Ukraine est humanisée et politisée.

11.1.2 Mot-clé : *Trump*

Corpus Gaza :

Les occurrences de *Trump* apparaissent principalement dans le cadre des plans et interventions politiques liés à Israël et Gaza.

Exemples :

- “a formidable obstacle to the Trump administration’s plan to ...”
- “The Trump plan for Gaza is predicated ...”
- “President Trump offers guarantees that Israel will ...”

Biais interne Gaza-Israël :

- Trump est mentionné en relation directe avec Israël et la gestion militaire ou diplomatique du conflit.
- Les actions ou discours concernant Gaza ou les Palestiniens apparaissent indirectement ou périphériquement, renforçant le **biais interne** : Israël au centre, Palestiniens en marge.

Corpus Ukraine :

Les occurrences de *Trump* sont liées à des plans de paix, des déclarations politiques ou des stratégies internationales.

Exemples :

- “As Trump Pushes to End Ukraine War ...”
- “Mr. Trump’s 28-point peace plan ...”
- “scrambling for influence as Mr. Trump and his team make pronouncements ...”

Biais externe Gaza-Ukraine :

- Dans le corpus Ukraine, Trump est intégré à un récit diplomatique et international, qui vise à stopper le conflit contrastant avec son rôle centré sur Israël dans le corpus Gaza.
- Cela illustre encore une fois la différence de cadrage : militarisé et centré sur Israël dans Gaza, politico-diplomatique et humanisé dans Ukraine.

Conclusion :

L’analyse des concordances confirme que le cadrage médiatique du conflit de Gaza est fortement centré sur Israël et les actions militaires, marginalisant les Palestiniens et les dimensions humanitaires. Comparativement, le conflit en Ukraine est intégré dans un récit géopolitique et humanisé, illustrant un **biais externe clair** dans la couverture médiatique entre les deux conflits.

11.2. Analyse des témoins et organisations (Named Entities)

L'extraction des **entités nommées** (PERSON, ORG) permet d'identifier les acteurs humains et institutionnels présents dans les articles, révélant le cadrage médiatique et les biais implicites.

Entrée : Liste d'articles (articles)

Procédure :

1. Pour chaque article :
 - a. Analyser le texte avec un outil de reconnaissance d'entités nommées (NER).
 - b. Extraire les entités de type PERSON (personnes).
 - c. Extraire les entités de type ORG (organisations).
 - d. Afficher l'article, la liste de personnes (ou "Aucun témoin PERSON"), et la liste d'organisations (ou "Aucun témoin ORG").

Sortie : Liste des entités PERSON et ORG identifiées pour chaque article

11.2.1 Corpus Gaza

- **Exemples d'individus identifiés** : Jordan, Ronen Bergman, Adam Rasgon, Benjamin Netanyahu, Majdi Ahmed, Yasser Abu Shabab, 'Rafah',...
- **Exemples d'organisations identifiées** : New York Times, Hamas, Al Aqsa TV, The Palestinian Authority, Trump, etc.

Analyse du biais interne Gaza-Israël :

- Les individus cités sont souvent **journalistes occidentaux**, responsables politiques israéliens ou acteurs militaires, tandis que les Palestiniens apparaissent moins individuellement ou par des groupes (Hamas, Islamic Jihad).
- Les organisations dominantes mettent en avant **Israël et les institutions internationales** (New York Times, Trump, comité de défense), alors que les acteurs palestiniens sont cités collectivement.
- Conclusion : le corpus Gaza présente un **biais interne déshumanisant les Palestiniens**, centrant l'attention sur Israël et ses décisions militaires ou diplomatiques.

11.2.2 Corpus Ukraine

- **Exemples d'individus identifiés** : Volodymyr Zelensky, Vladimir V. Putin, Matthew Mpoke Bigg, Michael D. Shear, Keir Starmer, Emmanuel Macron, Trump, etc.
- **Exemples d'organisations identifiées** : New York Times, NATO, United Nations, Kremlin, Pentagon, etc.

Analyse du biais externe Gaza-Ukraine :

- Les articles sur l'Ukraine citent **fréquemment des chefs d'État et des diplomates**, donnant une visibilité politique et humanitaire plus large.
- Les organisations sont souvent liées à la **géopolitique internationale**, ce qui contraste avec la focalisation militaire et nationale des articles sur Gaza.
- Conclusion : par rapport à Gaza, le conflit ukrainien est présenté de manière **plus humanisée et internationalisée**, intégrant les dimensions diplomatiques et humanitaires.

11.3. Analyse des témoignages et discours rapportés

L'analyse des témoignages vise à étudier qui parle, comment, et avec quelle légitimité discursive, en procédant à l'extraction automatique des citations directes et à l'identification des locuteurs.

Cette approche permet de révéler :

- des biais internes, au sein du conflit Gaza–Israël,
- des biais externes, lors de la comparaison entre les corpus Gaza et Ukraine.

Entrée : Texte complet d'un article

Procédure :

1. Découper le texte en phrases.
2. Initialiser une liste résultats vide.
3. Pour chaque phrase :
 - a. Si la phrase contient des guillemets :
 - Chercher un locuteur (nom propre suivi de « said », « told », « added », ou « according to »).
 - Si un locuteur est trouvé, l'enregistrer.
 - Ajouter la phrase et le locuteur (ou None si absent) dans la liste résultats.
4. Retourner la liste résultats.

Sortie : Liste de dictionnaires contenant la citation et le locuteur associé.

11.3.1 Témoignages dans le corpus Gaza

11.3.1.1 Répartition des voix et statuts discursifs

L'extraction automatique des témoignages dans le corpus Gaza–Israël met en évidence une **asymétrie forte dans la distribution des voix**. Trois catégories de locuteurs dominant :

1. Responsables militaires et politiques israéliens
2. Civils palestiniens
3. Responsables de Hamas
4. Institutions internationales et diplomatiques

11.3.1.2 Discours militaire israélien : rationalisation et légitimation

Les témoignages de responsables militaires israéliens sont **fréquents, clairement identifiés et fortement contextualisés**.

Par exemple, le colonel Ben-Anat décrit l'opération militaire contre Hamas comme :

"It's like catching a mouse."

"We do our best to destroy Hamas only, without harming the civilians."

Ce type de métaphore :

- **déshumanise l'ennemi** (réduction à un animal),
- **normalise la violence** comme un acte technique,
- présente l'armée israélienne comme rationnelle et morale.

Le discours insiste également sur la discipline et la prudence militaire :

"We'll think 10 times before we do something."

->**Biais interne**: La parole israélienne est construite comme **stratégique, responsable et légitime**, bénéficiant d'un fort capital de crédibilité médiatique.

11.3.1.3. Civils palestiniens : humanisation émotionnelle mais dépolitisation

À l'inverse, les témoignages palestiniens sont **massivement émotionnels**, centrés sur la peur, la survie et la souffrance quotidienne :

"I am afraid to go out one night and meet a tank."

"Now I can hear the shooting."

"The hardest thing is living with the feeling that all you can do is wait for death."

Ces voix :

- expriment une expérience humaine intense,
- mais restent déconnectées de toute capacité d'action politique.

Les civils palestiniens sont décrits comme :

- victimes,
- déplacés,
- affamés,
- privés de choix.

-> **Effet discursif** :

Gaza est humanisée sur le plan émotionnel mais **dépolitisée**. Les Palestiniens parlent de ce qu'ils subissent, **jamais de ce qu'ils décident**.

11.3.1.4. Hamas : parole politique conditionnelle et cadrée

Les témoignages attribués à des responsables de Hamas (ex. Husam Badran) apparaissent sous une forme **fortement encadrée et conditionnelle** :

“Hamas was hit hard, but it wasn’t defeated.”

“Without that, talking about these matters would be nonsensical.”

La parole de Hamas est :

- rare,
- contextualisée comme stratégique,
- souvent suivie d’un rappel de la violence du groupe.

-> **Biais interne** :

Hamas est présenté comme **acteur politique illégitime**, même lorsqu’il adopte un discours diplomatique.

11.3.1.5. Témoignages israéliens civils et question des otages

Un élément particulièrement révélateur du biais interne est la **faible présence de témoignages directs d’otages israéliens** dans les extraits analysés.

Malgré l’importance centrale de la question des otages dans le discours politique israélien, leurs voix apparaissent :

- rarement,
- de manière indirecte,

- sans développement narratif équivalent à celui des civils palestiniens.

-> **Résultat analytique clé :**

Les otages sont **un enjeu discursif**, mais **pas des sujets parlants**.

Ils servent à justifier l'action militaire, plutôt qu'à produire un récit humain autonome.

11.3.1.6. Institutions internationales et discours moral

Les témoignages d'organisations internationales (ONU, ONG, diplomates) introduisent un **contre-discours normatif** :

"An accelerated process of dehumanization of Palestinians."

"Apparent summary execution."

Cependant, ces prises de parole :

- restent marginales,
- n'influencent pas la structure dominante du récit.

L'analyse des témoignages révèle un **biais interne structurel** :

- Israël est construit comme acteur stratégique doté de parole et de rationalité.
- Gaza est représenté comme **espace de souffrance**, sans souveraineté discursive.
- Les civils palestiniens sont audibles, mais **jamais décisionnaires**.
- Les otages israéliens sont symboliquement centraux mais narrativement absents.

11.3.2 Témoignages dans le corpus Ukraine

Dans le corpus Ukraine, les témoignages sont majoritairement attribués à des **responsables politiques et diplomatiques clairement identifiés**, tels que :

- Emmanuel Macron :

"There are elements in this plan that should be discussed, negotiated, and improved."

- Richard Fontaine (think tank)
- *Ivo Daalder* (ancien ambassadeur à l'OTAN)
- *Yaroslav Yurchyshyn*, membre du Parlement ukrainien

Ces témoignages sont :

- précis,

- contextualisés,
- reliés à des fonctions politiques explicites.

Les citations s'inscrivent dans un **cadre discursif rationnel et diplomatique**, centré sur la négociation, la souveraineté et la responsabilité internationale.

->**Biais externe :**

- Gaza est raconté à travers :
 - la violence,
 - la survie,
 - le chaos humanitaire.
- L'Ukraine est racontée à travers :
 - la politique,
 - la diplomatie,
 - la légitimité internationale.

->Cela révèle un biais externe majeur :

le conflit ukrainien est humanisé et politisé, tandis que Gaza est militarisé et dépolitisé.

11.3.3 Analyse sémantique avec Word2Vec(word embedding):

11.3.3.1 Entraînement du modèle Word2Vec

- Le modèle a appris des représentations vectorielles des mots à partir du corpus.
- Chaque mot est représenté par un vecteur numérique reflétant son contexte et ses co-occurrences avec les autres mots du corpus.

11.3.3.2 Calcul du score de biais sémantique

- Pour chaque phrase, on calcule le vecteur moyen de tous ses mots.
- On mesure ensuite la similarité cosinus entre ce vecteur et les vecteurs des mots cibles : "*palestinian*" et "*israel*".
- Les valeurs possibles varient de -1 à 1 :
 - Valeurs proches de 1 → forte proximité sémantique
 - Valeurs proches de 0 → faible proximité ou neutre
 - Valeurs négatives → éloignement sémantique

11.3.3.3 Test sur des phrases représentatives

Phrases test :

1. test_pal : « **source new york time title israel take journalist combat zone gaza** »
 2. test_isr : « **update date conflict israel hamas war article israel colonel projected confidence described battle** »
- La phrase "test_pal" obtient un score de **0,019** pour *palestinian* et **0,412** pour *israel*. Bien que la phrase mentionne Gaza, la majorité des mots sont sémantiquement plus proches du vocabulaire associé à Israël. L'orientation générale de cette phrase est donc vers Israël.
 - La phrase "test_isr" obtient un score de **0,121** pour *palestinian* et **0,562** pour *israel*. Elle est centrée sur le conflit militaire, avec de nombreux mots liés à Israël. La proximité avec le mot cible *Israel* est très forte, tandis qu'elle est faible pour *Palestinian*.

11.3.3.4. Analyse des résultats

- Les scores montrent que le modèle capte l'influence sémantique globale de la phrase, et pas seulement la présence d'un mot clé.
- Même si la première phrase mentionne Gaza, elle contient de nombreux mots associés à Israël, ce qui explique le score plus élevé pour ce mot.
- La deuxième phrase est clairement centrée sur Israël, reflété par le score élevé pour ce mot.
- Cette approche permet de quantifier l'inclinaison médiatique ou la focalisation sur un acteur dans un corpus.

11.3.3.5. Conclusion

- L'utilisation de Word2Vec et de la similarité cosinus permet de mesurer objectivement la proximité sémantique d'un texte avec des mots cibles.
- Les scores obtenus peuvent servir à cartographier l'orientation des articles ou des phrases, utile pour l'analyse des médias et la détection de biais implicite.
- Les résultats doivent être interprétés comme des mesures sémantiques et non comme des jugements politiques ou moraux.

CONCLUSION

Cette étude avait pour objectif d'analyser, à l'aide des outils du traitement automatique du langage naturel (NLP), le cadrage médiatique de deux conflits armés contemporains — Gaza et l'Ukraine — dans un même média de référence, *The New York Times*. En s'appuyant sur un corpus homogène et des analyses lexicales, contextuelles et sémantiques, ce travail visait à mettre en évidence les biais discursifs et les asymétries de représentation du discours journalistique.

Les résultats révèlent une **dualité médiatique marquée** dans le traitement de ces deux conflits. Bien qu'ils impliquent tous deux des violences armées et des enjeux humanitaires majeurs, leur représentation diffère profondément. Le conflit de Gaza est majoritairement cadré comme un affrontement militaire immédiat, dominé par les acteurs armés et les opérations sécuritaires, avec une visibilité limitée et dépersonnalisée des civils palestiniens. À l'inverse, la guerre en Ukraine est présentée comme une crise géopolitique et morale, intégrant fortement les dimensions politiques, diplomatiques et humanitaires, ainsi qu'une représentation plus structurée des victimes civiles.

Cette asymétrie ne relève pas seulement d'une différence thématique, mais révèle une **hiérarchisation implicite des conflits et des souffrances**, où les mêmes notions — guerre, morts, civils — sont construites différemment selon le contexte. L'analyse interne du corpus Gaza montre en outre un déséquilibre discursif, Israël apparaissant comme l'acteur central et légitimé, tandis que les Palestiniens sont majoritairement représentés comme des victimes passives ou des entités territoriales. Les analyses sémantiques confirment que cette dualité repose sur une structuration discursive profonde, illustrant le rôle central du langage médiatique dans la construction des représentations des conflits contemporains.