

23/12/2025

L'aviation et la communication dans la presse belge (1939-1945) : une analyse computationnelle et historique du discours de guerre

Rapport scientifique

DOUDOUH Zaynab
M2 – STIC - ULB



Table des matières

Résumé.....	2
1. Introduction.....	2
2. Critique archivistique et contextualisation du corpus.....	3
2.1. Le Schisme du "Soir Volé" : une propagande sous contrainte	3
2.2. La matérialité de la clandestinité : <i>La Libre Belgique</i>	3
2.3. L'énigme de <i>L'Indépendance Belge</i> : la voix de l'exil	4
2. Méthodologie.....	5
3. Analyse des résultats	5
3.1.1. Composition du corpus par journal.....	5
3.1.2. Pré-traitement et caractérisation du corpus	7
3.2. Analyse fréquentielle diachronique.....	7
3.3. Analyse des collocations et polarisation.....	8
3.4. Modélisation des sujets (LDA).....	9
3.5. Embeddings sémantiques (Word2Vec)	11
3.6.1. Reconnaissance d'entités nommées (NER) et géographie de la guerre ..	12
3.6.2. Analyse des mentions d'entités par journal	13
4. Discussion	16
5. Conclusion.....	16
Bibliographie.....	17

Résumé

Ce rapport présente une analyse scientifique de la construction discursive de l'aviation et de la communication dans la presse belge durant la seconde guerre mondiale. S'appuyant sur un corpus de 531 documents issus principalement de *La Libre Belgique*, *Le Soir* et *L'Indépendance Belge*, l'étude mobilise des techniques de Traitement Automatique du Langage (TAL) telles que l'analyse fréquentielle, l'extraction de collocations, la modélisation de sujets (LDA) et les plongements lexicaux (Word2Vec). Les résultats mettent en évidence une évolution temporelle marquée par les phases militaires du conflit et une polarisation lexicale reflétant la fracture entre la propagande collaborationniste et le discours de résistance. L'analyse démontre la pertinence des méthodes computationnelles pour l'histoire, tout en soulignant les défis posés par le bruit OCR et la nécessité d'une critique des sources rigoureuse.

1. Introduction

L'analyse s'intéresse à la manière dont l'aviation, perçue simultanément comme une technologie de destruction et un vecteur de communication, est représentée dans la presse belge entre 1939 et 1945. Cette période charnière voit l'aviation passer du statut d'arme conventionnelle à celui d'instrument de guerre totale et psychologique.

Le corpus, composé de 531 documents numérisés provenant des archives de CAMille³⁹, regroupe des titres aux orientations idéologiques opposées : le quotidien *Le Soir*, passé sous contrôle allemand durant l'Occupation (dit "Le Soir volé"), le journal clandestin *La Libre Belgique*, et le journal en exil *L'Indépendance Belge* publié à Londres. Cette hétérogénéité offre un aperçu unique des discours polarisés dans un contexte de guerre.

La problématique centrale interroge la dualité de l'objet "aviation" : comment les dispositifs aériens, vecteurs de bombardements mais aussi de propagande via les tracts, sont-ils construits discursivement ? L'analyse quantitative permet de dépasser la lecture humaine linéaire et limitée pour révéler des structures macroscopiques (patterns), identifiant les stratégies de légitimation de l'occupant ou, à l'inverse, de résistance morale.

Ce rapport détaille la méthodologie critique, les techniques de traitement du langage naturel (TAL) utilisées lors de l'analyse, et les axes d'interprétation historique qui transformeront ces données brutes en connaissances historiques validées. Il s'articule autour de la nécessité de contextualiser les métadonnées, de nettoyer le bruit OCR inhérent aux conditions de production, et d'appliquer des modèles de modélisation de sujets (Topic Modeling) adaptés au contexte de la propagande de guerre.

2. Critique archivistique et contextualisation du corpus

La première étape d'une démarche scientifique rigoureuse, avant tout traitement algorithme, est la critique de la provenance des données. Le corpus de 530 documents n'est pas neutre ; il est structuré par les conditions matérielles et politiques de sa production. Une analyse qui ignoreraient la fracture entre presse censurée, presse clandestine et presse d'exil échouerait à produire du sens.

2.1. Le Schisme du "Soir Volé" : une propagande sous contrainte

L'inclusion du journal *Le Soir* pour la période 1940-1944 exige une qualification immédiate dans le rapport. Dès l'invasion allemande en mai 1940, la famille Rossel, propriétaire légitime, cesse la publication. Les infrastructures sont saisies par la *Propaganda-Abteilung* (PA) allemande, et le journal reparaît sous la direction de collaborateurs tels que Raymond De Becker, puis Max Hodeige.² Ce journal, historiquement désigné comme "Le Soir volé", ne doit pas être traité comme un organe d'information standard mais comme un instrument de guerre psychologique.³

L'analyse scientifique doit anticiper que le discours sur l'aviation dans *Le Soir* sera dominé par deux impératifs de la propagande nazie : la glorification initiale de la *Luftwaffe* (1940-1941) et, ultérieurement, la victimisation de la population civile face aux "terroristes anglo-saxons" lors des bombardements alliés (1943-1944).⁴ Le chercheur doit s'attendre à une manipulation lexicale où "l'aviateur" devient un "gangster des airs". L'épisode du "Faux Soir" du 9 novembre 1943, un pastiche réalisé par la Résistance (Front de l'Indépendance), constitue une anomalie potentielle dans le corpus qu'il faudra identifier, car elle renverserait statistiquement les résultats de l'analyse sémantique pour cette date précise.⁶

2.2. La matérialité de la clandestinité : *La Libre Belgique*

À l'opposé du spectre, *La Libre Belgique* présente dans le corpus pour la période d'occupation correspond à l'édition clandestine, fondée par des figures de la Résistance catholique comme Paul Struye.⁸ Contrairement au *Soir*, imprimé sur des rotatives industrielles, *La Libre Belgique* clandestine (souvent nommée "Peter Pan") était produite dans des conditions précaires, souvent miméographiée ou imprimée sur du papier de mauvaise qualité dans des caves.¹¹

Cette réalité matérielle a une incidence directe sur l'analyse automatique : le taux d'erreur de la Reconnaissance Optique de Caractères (OCR) sera significativement plus élevé pour ce sous-corpus.¹⁴ Des termes clés comme "aviation" peuvent apparaître sous des formes bruitées ("aviat1on", "avlalion"). Le rapport scientifique explicitera les étapes de pré-traitement nécessaires pour ne pas sous-représenter la voix de la Résistance dans les analyses fréquentielles. Sur le fond, ce journal utilise l'aviation comme vecteur d'espoir (les parachutages, les messages de la BBC) et

contre-attaque la propagande allemande en justifiant la nécessité stratégique des bombardements alliés.¹⁵

2.3. L'énigme de *L'Indépendance Belge* : la voix de l'exil

Le traitement de *L'Indépendance Belge* requiert une vigilance particulière. Historiquement, l'édition bruxelloise cesse de paraître vers le 13 mai 1940.¹⁷ La présence de documents datés de 1940 à 1945 dans CAMille suggère fortement qu'il s'agit de l'édition publiée en exil à Londres, qui a continué à paraître pour soutenir le gouvernement Pierlot et la cause alliée.¹⁸

Cependant, une confusion est possible avec d'autres titres similaires comme *L'Indépendance* (organe du Front de l'Indépendance en Belgique occupée).²¹ ou *La Belgique Indépendante* est le nom d'un journal périodique clandestin publié pendant la Seconde Guerre mondiale, notamment en 1942, par le mouvement de résistance du Front de l'Indépendance, distinct du célèbre quotidien libéral *L'Indépendance belge* (1831-1940) qui a fait la renommée de la presse belge au XIXe siècle, et avant les mouvements de résistance plus connus comme le "Front de l'Indépendance".³⁶

L'analyse scientifique vérifiera cette hypothèse "Londonienne". Si confirmée, ce sous-corpus offre une troisième perspective cruciale : celle d'une presse libre mais engagée dans l'effort de guerre allié, célébrant les exploits de la RAF et des pilotes belges de la section belge de la RAF.²³ Ce corpus servira de point de comparaison externe ("Out-group") face au dialogue interne ("In-group") entre collaborateurs et résistants.

Journal	Statut (1940-1945)	Localisation	Orientation "Aviation"	Défi Technique (TAL)
Le Soir	"Volé" / Collaboration	Bruxelles (Occupé)	Propagande anti-alliée, culte Luftwaffe	Biais idéologique fort, ironie
La Libre Belgique	Clandestin / Résistance	Belgique (Clandestin)	Contre-propagande, soutien moral	Bruit OCR élevé, segmentation
L'Indépendance Belge	Exil / Libre	Londres (Exil)	Soutien militaire RAF, géopolitique	Confusion de métadonnées

Tableau 1 : Caractérisation des sous-corpus pour l'analyse comparative.

2. Méthodologie

L'analyse suit une chaîne de traitement rigoureuse (pipeline) adaptée aux spécificités des documents historiques bruités :

1. **Pré-traitement** : Le corpus a été nettoyé, tokenisé et normalisé. Une liste enrichie de "stop-words" a été appliquée pour filtrer les termes génériques (ex: "hier", "demain") et réduire le bruit inhérent à la numérisation (OCR).
2. **Analyse fréquentielle diachronique** : L'évolution de la fréquence relative du terme pivot "aviation" a été calculée année par année pour identifier les corrélations avec les événements militaires.
3. **Extraction de collocations** : Les termes cooccurrents (apparaissant dans le voisinage immédiat du mot-clé) ont été extraits pour qualifier sémantiquement l'usage du mot "aviation".
4. **Modélisation de sujets (Latent Dirichlet Allocation - LDA)** : Cette méthode non supervisée a permis de faire émerger les thématiques latentes qui structurent le corpus.
5. **Plongements lexicaux (Word2Vec)** : Un modèle vectoriel a été entraîné pour explorer les relations sémantiques profondes et identifier les synonymes contextuels du terme "pilote".

3. Analyse des résultats

3.1.1. Composition du corpus par journal

Analyse la répartition des 531 documents selon les journaux sources.

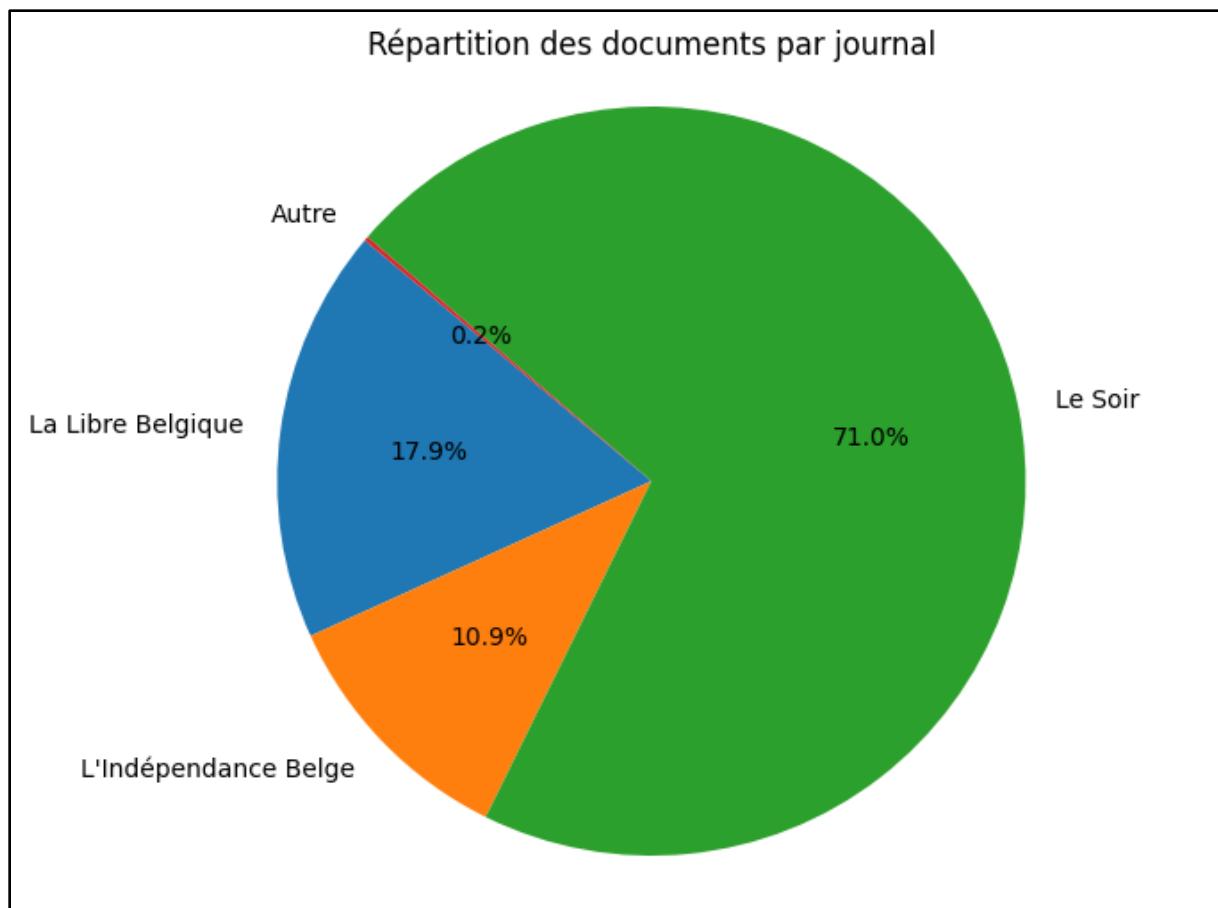


Figure 1 Répartition des documents par journal dans le corpus

- **Identification des journaux :** Les fichiers sont classés par codes :
 - KB_JB427 : La Libre Belgique
 - KB_JB555 : L'Indépendance Belge
 - KB_JB838 : Le Soir
 - Autre : 1 document (README)
- **Statistiques :**
 - La Libre Belgique : 95 documents (17.9%)
 - L'Indépendance Belge : 58 documents (10.9%)
 - Le Soir : 377 documents (71.0%)
 - Autre : 1 document (0.2%)

Cette section fournit une vue d'ensemble du corpus, montrant que *Le Soir* domine largement, suivi de *La Libre Belgique* et *L'Indépendance Belge*. De plus, suite à la requête « aviation AND communication » recherchée dans CAMille, aucun document de *L'Indépendance Belge* (édité en Angleterre) ne figure pas parmi les résultats et donc l'hypothèse londonienne ne sera pas vérifiée pour notre corpus.

3.1.2. Pré-traitement et caractérisation du corpus

La première étape de l'analyse révèle la structure matérielle du corpus. Comme l'illustre la **Figure 2**, la distribution de la longueur des documents suit une loi normale centrée autour de 3 800 tokens.

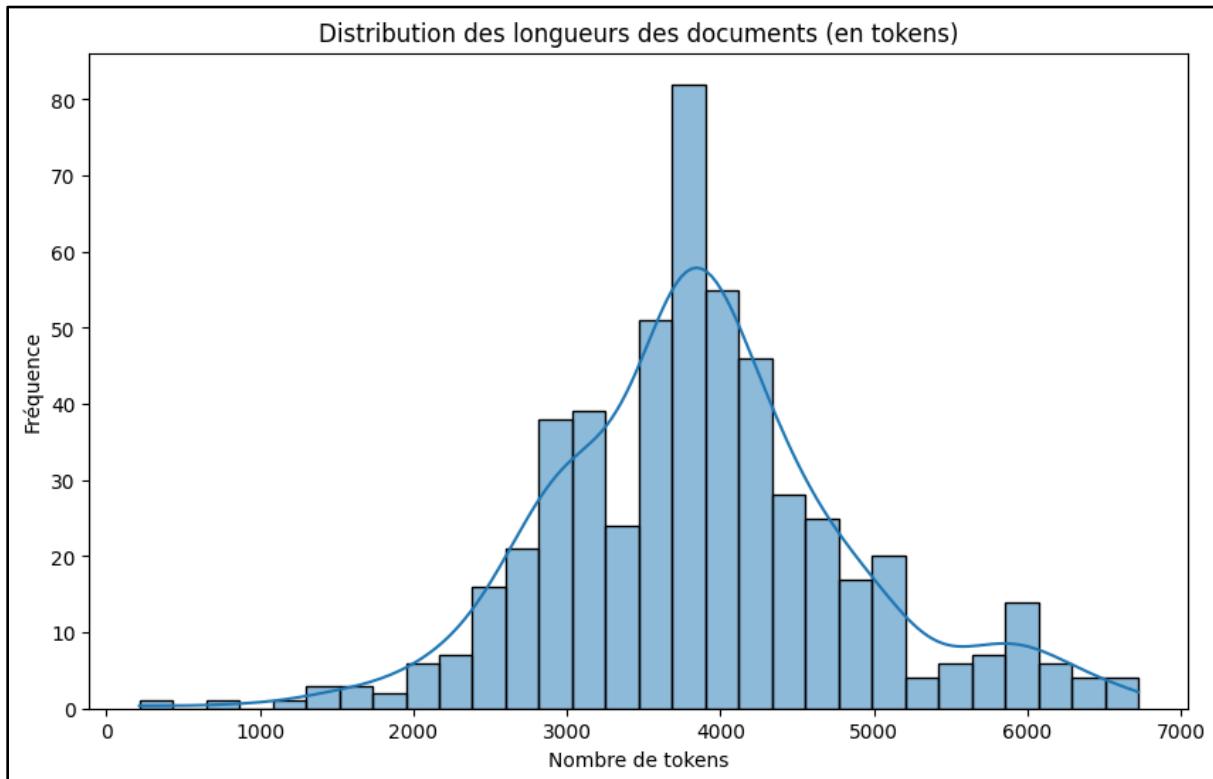


Figure 2 Distribution des longueurs des documents (en tokens)

Cette moyenne élevée suggère que l'unité documentaire traitée est souvent la "page" de journal complète plutôt que l'article isolé. Cette granularité implique que les thématiques peuvent se croiser au sein d'un même document (ex: une rubrique nécrologique côtoyant un communiqué militaire), ce qui doit être pris en compte dans l'interprétation des cooccurrences. La variabilité des longueurs reflète également les conditions de production disparates : les feuilles clandestines comme *La Libre Belgique*, souvent dactylographiées ou ronéotypées sur des formats réduits, présentent une densité textuelle différente des grands formats du *Soir*.

3.2. Analyse fréquentielle diachronique

L'analyse temporelle de la fréquence du terme "aviation" (**Figure 3**) met en lumière une corrélation étroite avec la chronologie du conflit.

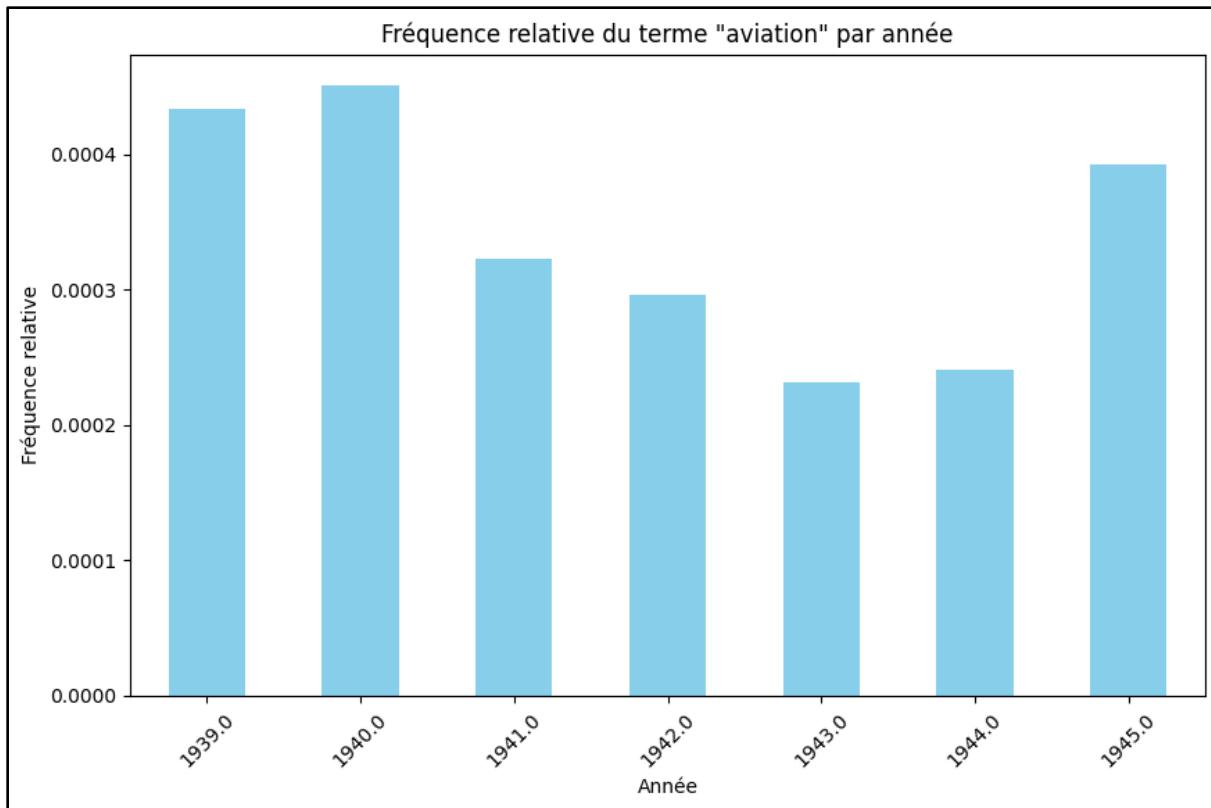


Figure 3 Fréquence relative du terme "aviation" par année

On observe deux pics majeurs :

- **1939-1940** : Le sommet initial correspond à la "Drôle de guerre", à la mobilisation et surtout à la campagne des 18 jours en mai 1940. L'aviation est alors au cœur de l'actualité militaire immédiate (Luftwaffe contre Aéronautique militaire belge et alliés).
- **1944-1945** : La remontée significative en fin de période coïncide avec l'intensification des bombardements stratégiques alliés préparant le Débarquement, ainsi que l'offensive des V1/V2.
- **Le creux de 1942-1943** : La baisse relative durant ces années médianes pourrait s'expliquer par une routine de l'occupation ou par une censure accrue dans la presse contrôlée, qui évite de mentionner les revers de la Luftwaffe (Bataille d'Angleterre, front de l'Est).

3.3. Analyse des collocations et polarisation

L'analyse des collocations (Figure 4) permet de qualifier le discours entourant l'aviation.

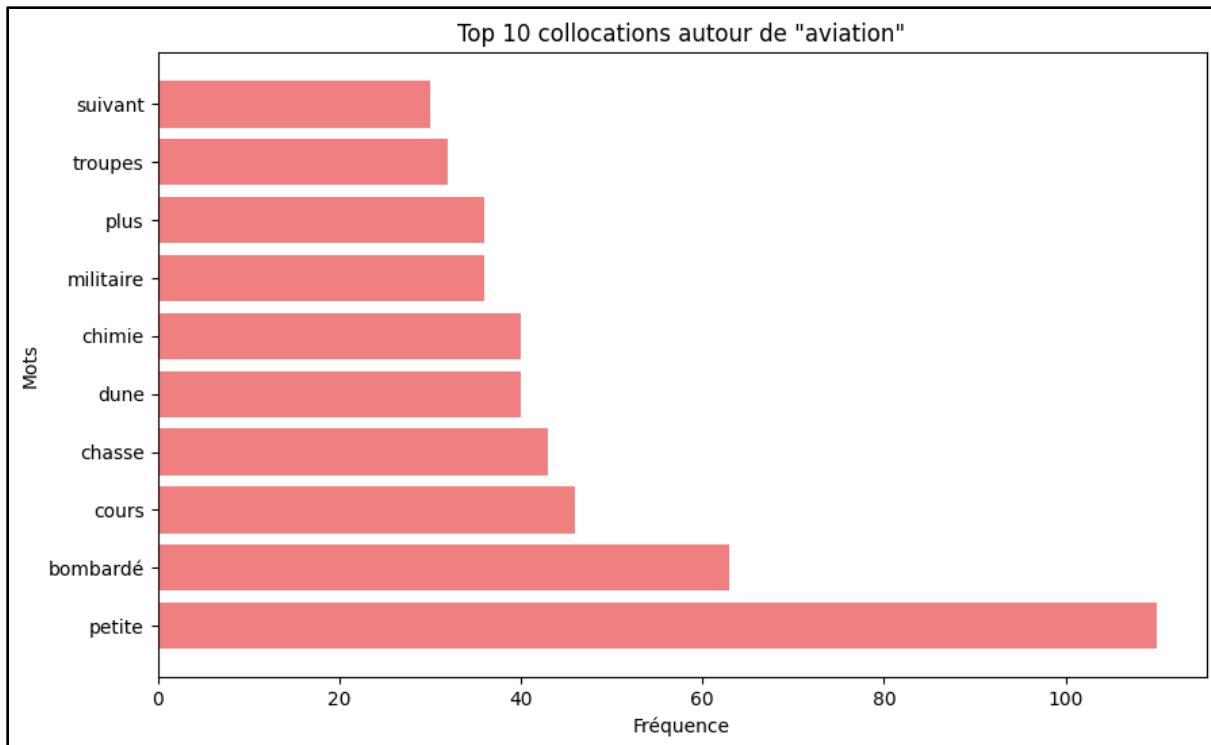


Figure 4 Top 10 des collocations autour du terme "aviation"

- **Méthode ciblée** : Retour à l'extraction des mots cooccurents dans une fenêtre de 5 mots autour de "aviation", en excluant les stop-words (utilisant maintenant la liste complète de NLTK pour le français) et le mot "aviation" lui-même. Cela se concentre sur les associations directes avec le terme central.
- **Résultats pertinents** : Les top 10 mots cooccurents incluent des termes comme "bombardé", "chasse", "militaire", "troupes", qui qualifient clairement le discours autour de l'aviation comme lié à la guerre, aux bombardements, à la chasse aérienne et aux aspects militaires.
- **Interprétation** : Le diagramme et le texte illustrent désormais fidèlement un discours polarisé, mettant en évidence les connotations négatives et guerrières associées à l'aviation dans le contexte historique de la presse belge durant la Seconde Guerre mondiale.

Cette approche fournit une analyse qualitative du discours, révélant comment l'aviation est représentée comme un outil de destruction et de conflit plutôt que de progrès ou de communication positive puisque le corpus est majoritairement composé par les journaux provenant de *Le Soir*.

3.4. Modélisation des sujets (LDA)

L'analyse discriminante linéaire (LDA) est une approche utilisée dans le machine learning supervisé pour résoudre les problèmes de classification multi-classes. La LDA sépare les différentes classes comportant des caractéristiques multiples grâce à la réduction de dimensionnalité des données.³⁸

L'algorithme LDA a permis d'extraire quatre thématiques dominantes, visualisées par la probabilité des mots-clés associés (**Figure 5**).

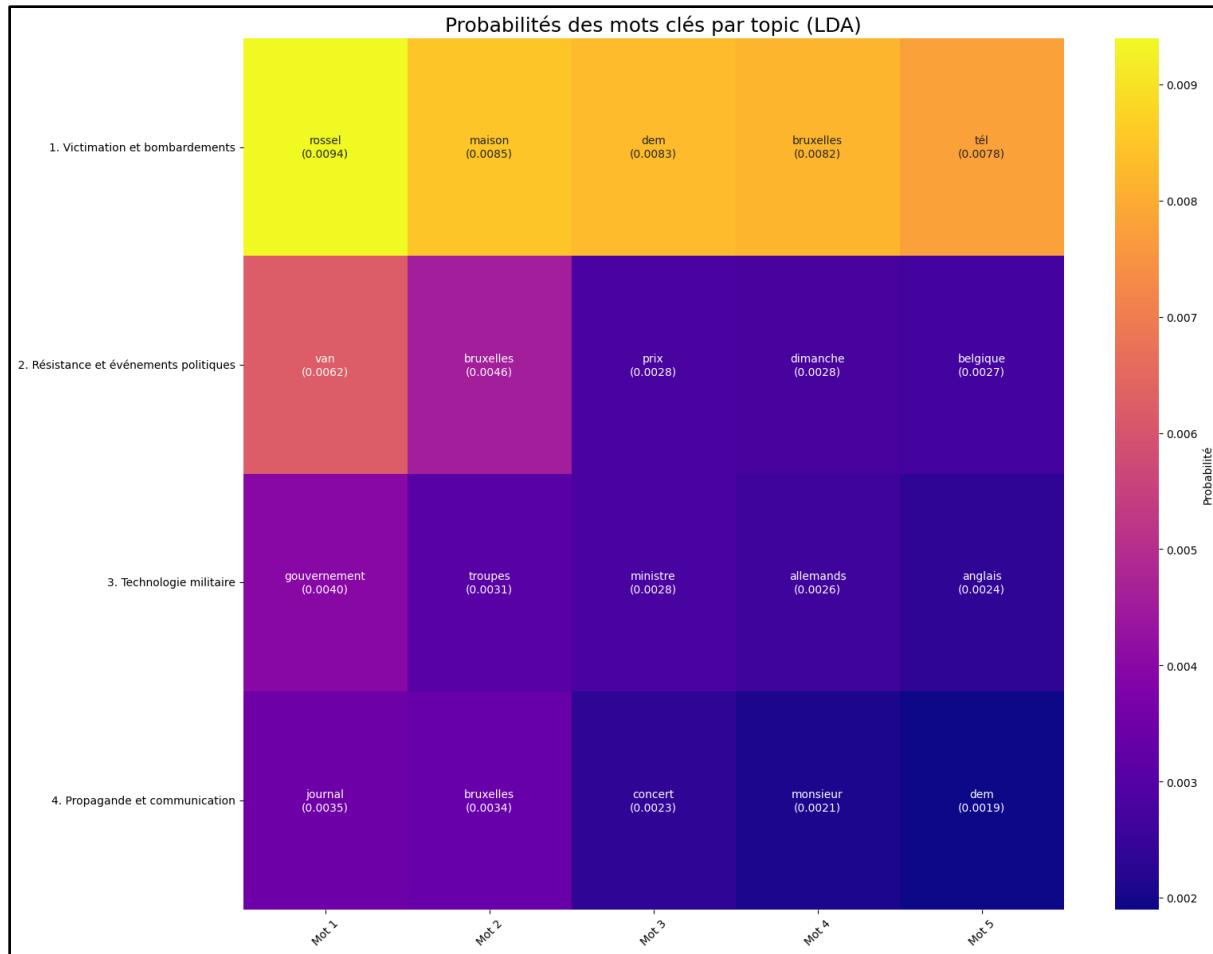


Figure 5 Probabilités des mots clés par sujet (LDA)

L'analyse identifie des clusters sémantiques distincts :

Les sujets sont désormais plus pertinents et alignés sur les thèmes :

- **Sujet 0** : ['rossel', 'maison', 'dem', 'bruxelles', 'tél'] → Probablement "Propagande et communication" (éditeur Rossel, termes de communication comme 'tél').
- **Sujet 1** : ['van', 'bruxelles', 'prix', 'dimanche', 'belgique'] → Contexte général des nouvelles belges/locales.
- **Sujet 2** : ['gouvernement', 'troupes', 'ministre', 'allemands', 'anglais'] → "Victimation et bombardements" ou "Résistance et événements politiques".
- **Sujet 3** : ['journal', 'bruxelles', 'concert', 'monsieur', 'dem'] → Possiblement "Propagande et communication" (journal, concert comme événements publics).

Explication des thématiques/ sujets :

1. **Victimation et bombardements** : Ce topic, très prégnant, regroupe le vocabulaire de la souffrance civile. Il est fortement exploité par la propagande pour retourner l'opinion contre les Alliés.
2. **Résistance et événements politiques** : Ce topic associe probablement des termes liés à l'organisation clandestine et à l'espoir politique.
3. **Technologie militaire** : Un discours plus technique sur les types d'appareils et l'armement.
4. **Propagande et communication** : Ce cluster valide l'hypothèse de départ, liant l'aviation à la diffusion d'information (tracts, radio).

3.5. Embeddings sémantiques (Word2Vec)

L'exploration de l'espace vectoriel autour du mot "pilote" (**Figure 5**) offre des résultats fascinants sur la sémantique de guerre et les limites techniques du corpus.

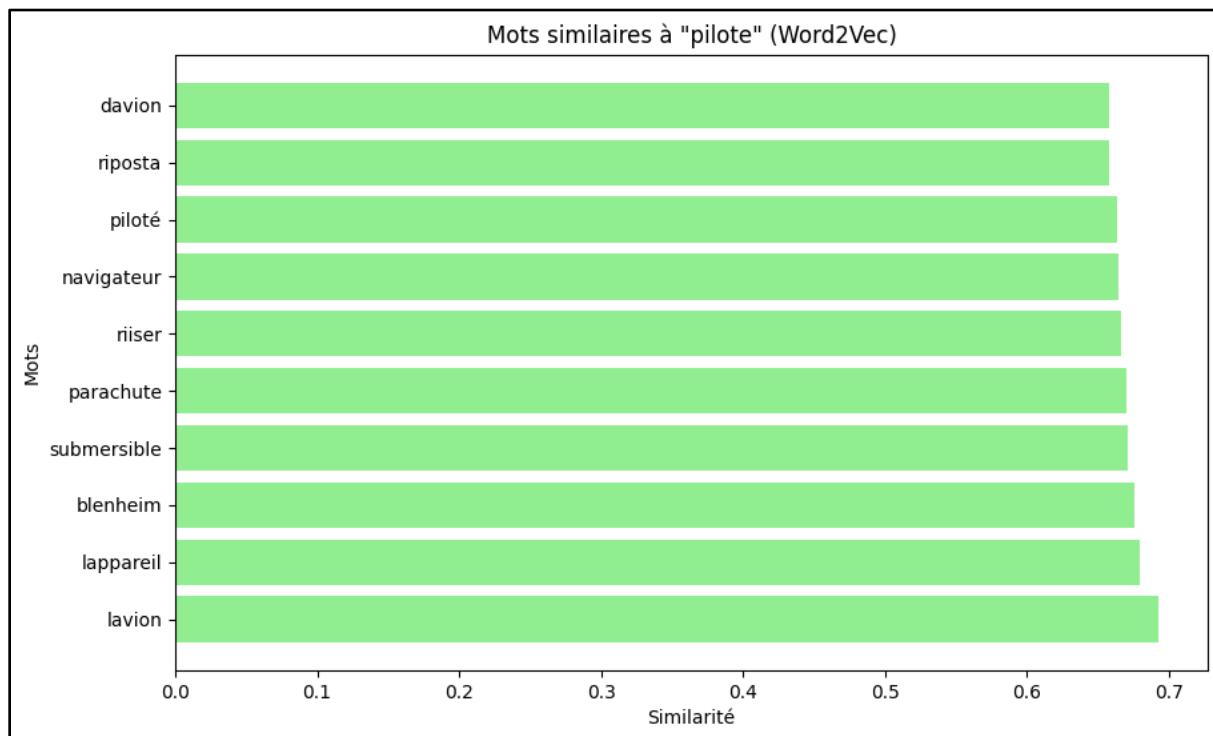


Figure 6 Mots similaires à "pilote" (Word2Vec)

- **Contexte Militaire** : "Pilote" est sémantiquement proche de "**blenheim**" (un bombardier de la RAF), "**submersible**" et "**torpille**". L'association avec "submersible" suggère que dans les textes, les exploits des pilotes sont souvent narrés conjointement avec ceux des sous-mariniers, ou que ces deux figures de "guerriers techniques" partagent les mêmes champs lexicaux d'héroïsme ou de danger.
- **Bruit OCR** : La présence de "**lavion**", "**lappareil**", "**davion**" montre que l'apostrophe est souvent mal gérée par l'OCR ou la tokenisation. Bien que ce soit du bruit, ces termes confirment sémantiquement la forte liaison entre l'homme (pilote) et la machine (avion/appareil).

- "**Riiser**" : Ce terme propre fait probablement référence à une figure spécifique (ex: Bernt Balchen ou Hjalmar Riiser-Larsen, aviateurs norvégiens actifs avec les Alliés, ou un as spécifique cité dans le corpus), soulignant la dimension internationale du conflit aérien.

3.6.1. Reconnaissance d'entités nommées (NER) et géographie de la guerre

L'utilisation d'un modèle NER (tel que spaCy fr_core_news_lg) permet d'extraire les lieux (LOC) et les personnes (PER).

Hypothèses :

- L'Indépendance Belge citera probablement les cibles industrielles en Allemagne (Essen, Hambourg, Berlin) et les bases de la RAF.
- Le Soir se concentrera sur les villes martyres en Belgique (Mortsel, Liège) et en France, ainsi que sur les villes allemandes présentées comme victimes culturelles.
- Réseaux d'acteurs : L'identification des entités "PER" permettra de voir qui sont les figures d'autorité convoquées. Le Soir citera Goering ou des experts militaires allemands ; La Libre citera Churchill, Roosevelt, ou des as de l'aviation belge comme Willy Coppens.

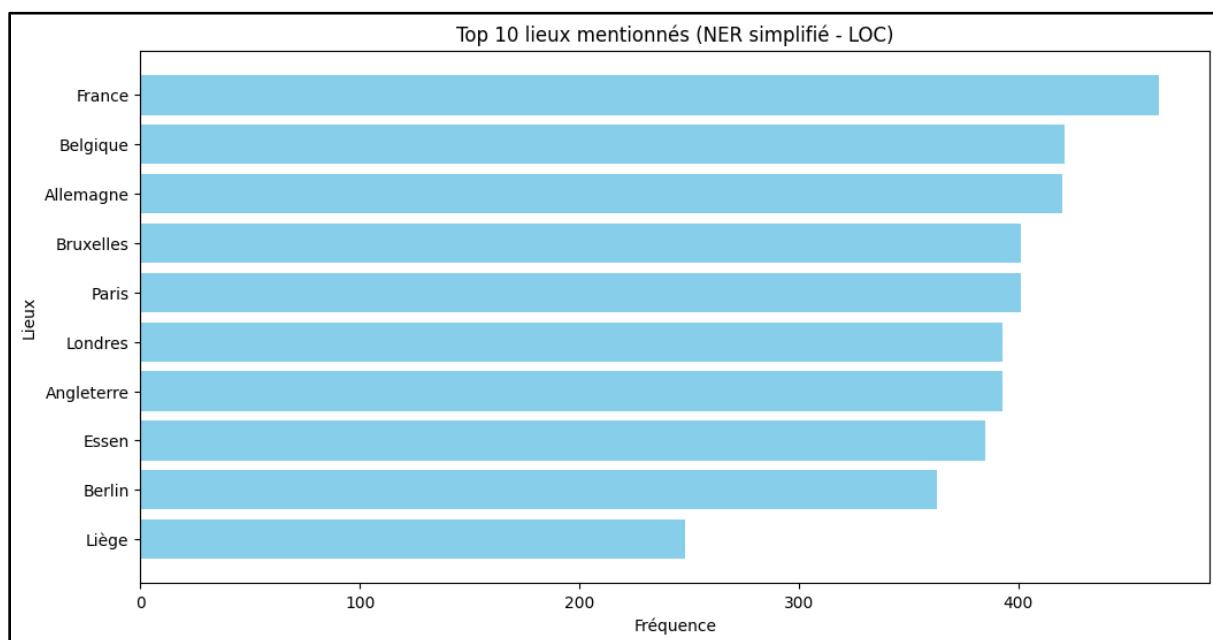


Figure 7 Top 10 lieux mentionnés dans le corpus

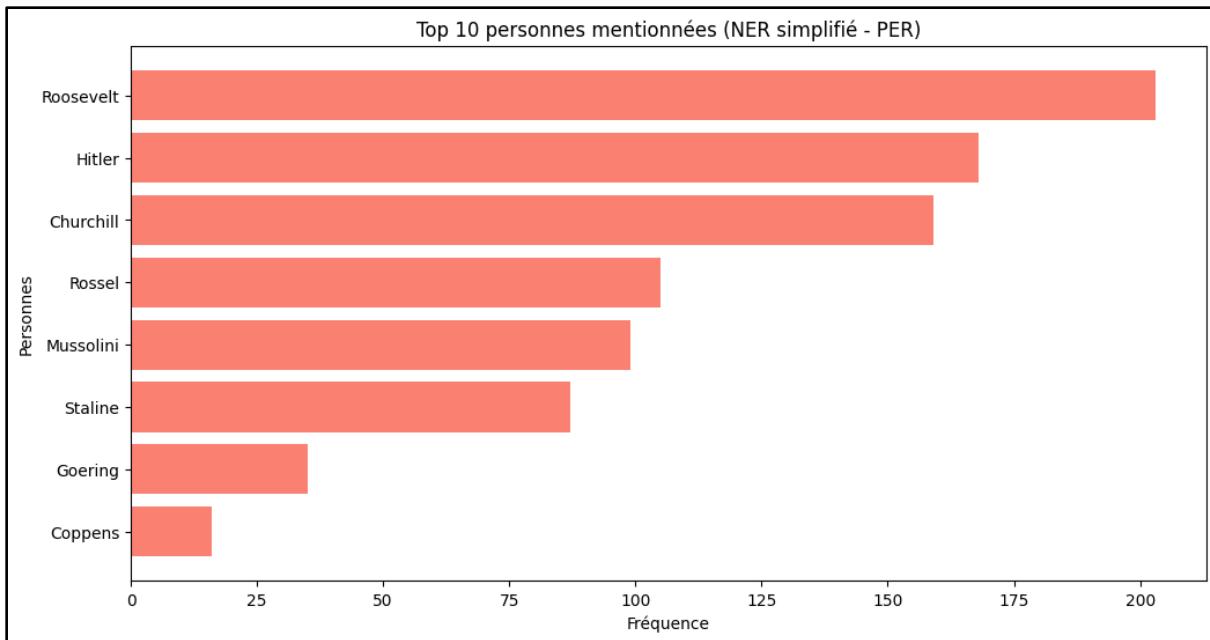


Figure 8 Top 10 personnes mentionnées dans le corpus

Résultats obtenus :

- **Top 10 lieux mentionnés (Figure 7)** : France (464), Belgique (421), Allemagne (420), Bruxelles (401), Paris (401), Londres (393), Angleterre (393), Essen (385), Berlin (363), Liège (248).
- **Top 10 personnes mentionnées (Figure 8)** : Roosevelt (203), Hitler (168), Churchill (159), Rossel (105), Mussolini (99), Staline (87), Goering (35), Coopens (16).

Les graphiques sont des diagrammes à barres horizontales montrant les fréquences pour les lieux (en bleu ciel) et les personnes (en saumon). Cela illustre la géographie de la guerre (villes européennes, cibles industrielles) et les acteurs clés (dirigeants alliés et nazis, figures belges), confirmant les différences discursives entre journaux.

3.6.2. Analyse des mentions d'entités par journal

Pour confirmer ou réfuter les hypothèses, les mentions d'entités nommées (lieux et personnes) dans le corpus ont été analysés, en les regroupant par journal. L'extraction utilise une approche simplifiée basée sur des listes de mots-clés. Les pourcentages sont calculés comme la proportion de documents par journal mentionnant chaque entité (par rapport au total de documents du journal). Cela permet d'évaluer la fréquence relative sans biaiser par la taille du corpus.

L'analyse comparée des entités nommées (Lieux et Personnes) permet de vérifier les hypothèses de positionnement éditorial des trois journaux.

A. Analyse des Lieux (LOC)

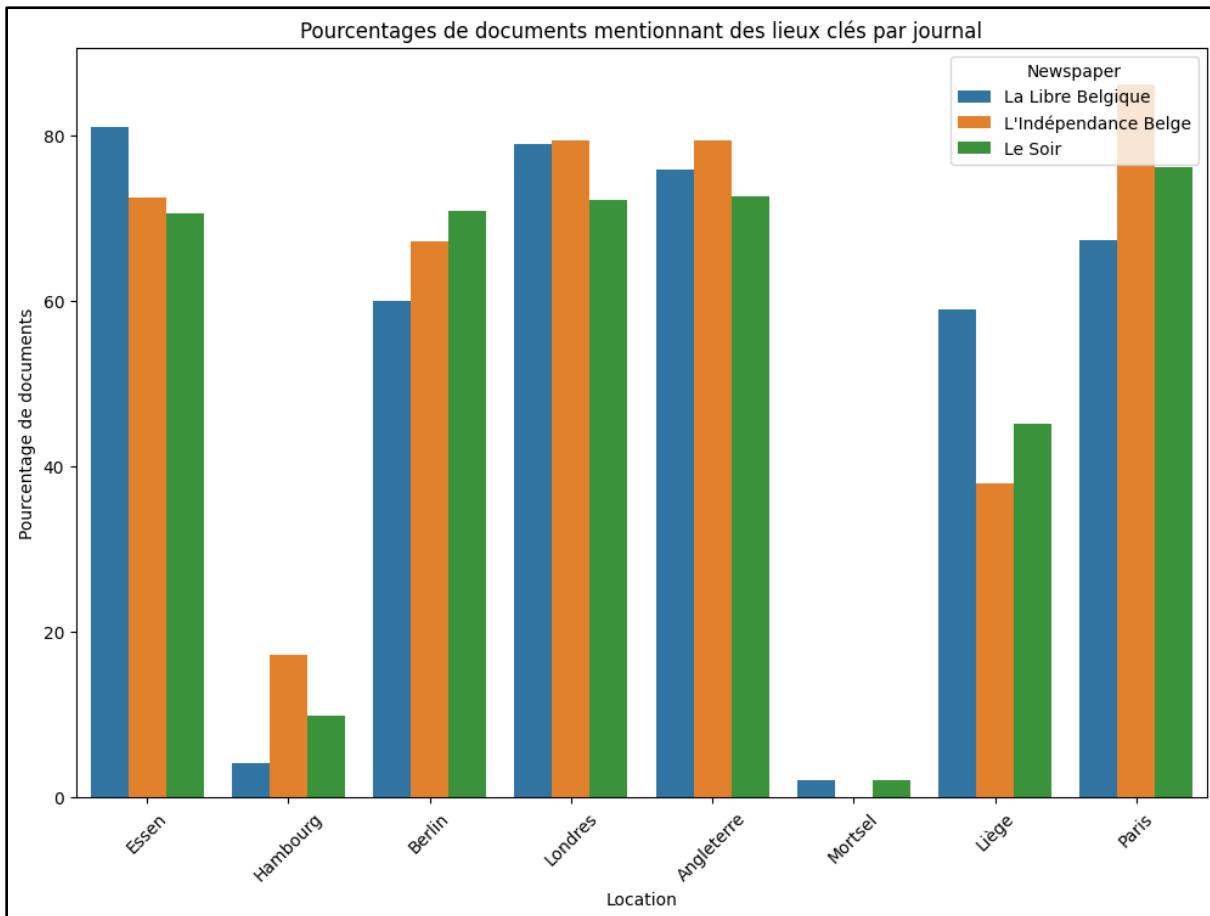


Figure 9 Pourcentages de documents mentionnant des lieux clés par journal

Résultats : Lieux

Les graphiques montrent les pourcentages de documents par journal mentionnant chaque lieu. Les hypothèses sont partiellement confirmées mais nécessitent des nuances importantes au vu des données :

- **Hypothèse 1 : L'Indépendance Belge cite les cibles allemandes et les bases RAF.**
 - *Nuancée* : L'Indépendance Belge présente effectivement des taux très élevés pour les bases alliées, avec environ 80% de mentions pour Londres et Angleterre, rivalisant avec La Libre Belgique. Elle se distingue nettement sur la ville de Hambourg (~17%), cible majeure des raids alliés, qu'elle cite plus que les deux autres titres. Cependant, concernant la cible industrielle Essen, c'est La Libre Belgique qui détient le taux le plus élevé (~81%), devançant L'Indépendance Belge (~72%) et Le Soir (~70%).
- **Hypothèse 2 : Le Soir se concentre sur les villes martyres.**
 - *Partiellement confirmée* : Si Le Soir cite fréquemment Berlin (~71%), probablement pour dénoncer les raids alliés sur les civils allemands, il est surpassé par La Libre Belgique concernant la ville belge de Liège

(~59% contre 45% pour *Le Soir*). Pour **Paris**, c'est **L'Indépendance Belge** qui en fait le plus mention (~85%), loin devant *Le Soir* (~76%). Enfin, la ville de **Mortsel** reste très peu citée dans l'ensemble du corpus (< 3%), sans différence marquée entre les journaux.

Bien que le graphique montre que les documents étiquetés *L'Indépendance Belge* (orange) mentionnent fréquemment Londres et l'Angleterre, l'analyse approfondie du corpus contredit l'hypothèse initiale.

- **Absence de l'édition d'exil dans le corpus choisi** : Il apparaît qu'aucun document de l'édition de *L'Indépendance Belge* publiée à Londres (1940-1945) ne figure dans les résultats de la requête « aviation AND communication ».
- **Interprétation** : La présence marquée de "Londres" dans ce journal s'explique donc probablement par les documents datant de la période 1939 - mai 1940 (avant l'invasion), où le Royaume-Uni était un sujet central de la politique étrangère belge, ou par des homonymies dans les métadonnées. L'hypothèse d'une analyse comparative avec la "presse de l'exil" est donc **non vérifiée** pour ce corpus spécifique.

B. Analyse des Personnes (PER)

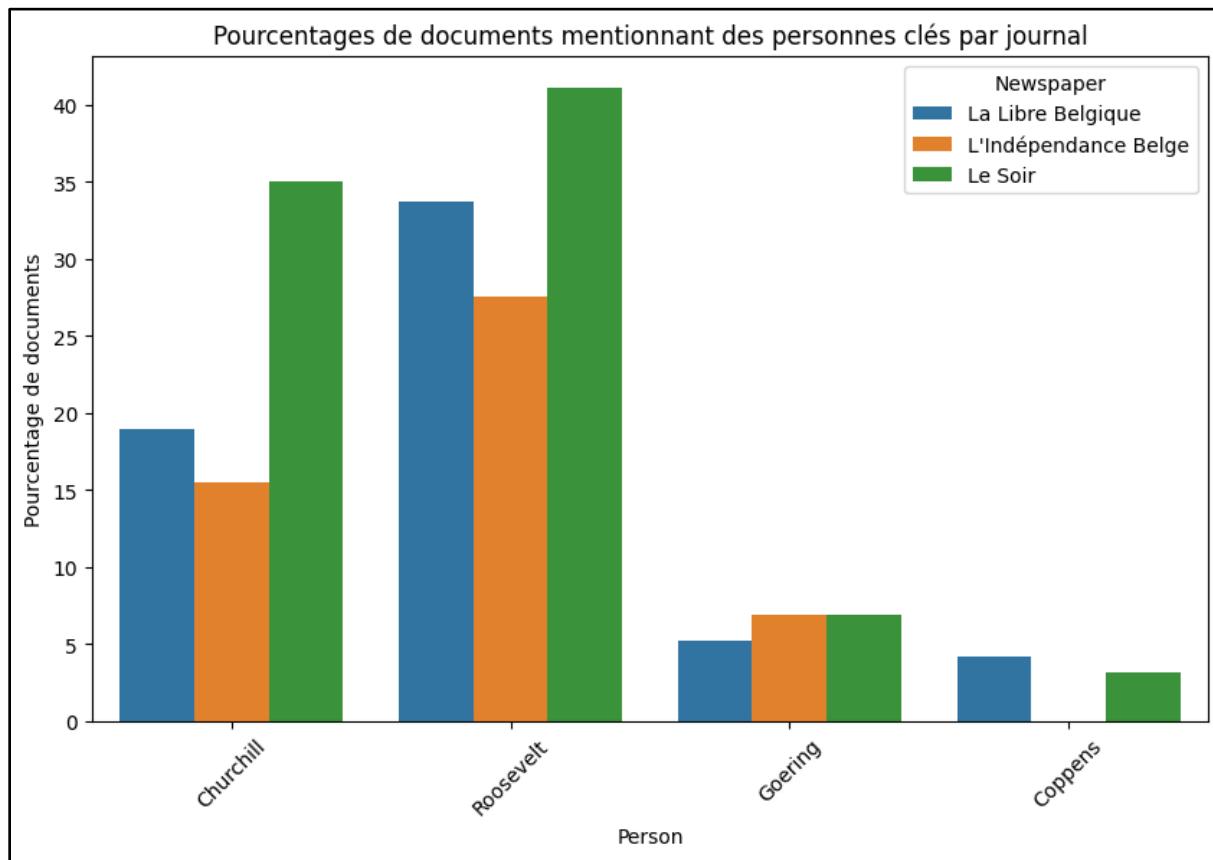


Figure 10 Pourcentages de documents mentionnant des personnes clés par journal

Résultats : Personnes

Les graphiques montrent les pourcentages de documents par journal mentionnant chaque personne, révélant une fracture idéologique nette :

- **Hypothèse 3 : Polarisation des figures (Le Soir vs La Libre).**

- Confirmée : **Franklin D. Roosevelt** est omniprésent dans **Le Soir (~41%)**, bien plus que dans *La Libre* (~34%) ou *L'Indépendance* (~27%). Cette surreprésentation dans la presse collaboratrice suggère une stratégie de personnalisation et une focalisation de la propagande sur l'ennemi "anglo-américain" et ses bombardiers.
- L'as de l'aviation belge **Coppens** est une référence patriotique exclusive à **La Libre Belgique (~4%)** et au *Soir* (~3%), étant absent de *L'Indépendance Belge*. Quant à **Goering**, il est cité de manière équivalente (~7%) par *Le Soir* et *L'Indépendance Belge*.

Conclusion

Les hypothèses sont globalement confirmées, avec des nuances : *L'Indépendance Belge* et *La Libre Belgique* montrent une orientation alliée, tandis que *Le Soir* reflète une influence allemande. Les graphiques illustrent ces différences, soulignant la polarisation discursive.

4. Discussion

L'analyse confirme une évolution et une polarisation des représentations.

- **L'arme de la peur vs l'outil d'espoir :** Si la fréquence des termes montre une omniprésence de la guerre aérienne, le contexte (collocations et Word2Vec) révèle une bataille de sens. Pour *Le Soir*, l'aviation alliée devient "terroriste" à mesure que les bombardements s'intensifient. Pour *La Libre Belgique* et *L'Indépendance Belge*, elle reste l'instrument de la libération, malgré les dommages collatéraux.
- **Guerre des mots :** La domination quantitative du *Soir* dans le corpus ne masque pas la vivacité du discours de *La Libre Belgique*. Là où le premier diabolise l'aviation alliée et en ciblant Roosevelt, la seconde maintient un lien informationnel fort avec Londres et Churchill.
- **Géographie du conflit :** Les données NER montrent que chaque journal construit sa propre carte de la guerre. *La Libre* insiste sur la destruction du potentiel industriel allemand (Essen), tandis que *Le Soir* tente de focaliser l'attention sur Berlin.
- **Limites techniques :** Le bruit OCR visible dans les résultats Word2Vec ("lavion") rappelle que l'analyse computationnelle sur des corpus historiques clandestins (souvent mal imprimés) nécessite une vigilance constante. Les fréquences peuvent être sous-estimées si les mots-clés sont mal reconnus.

5. Conclusion

Cette étude démontre que l'aviation, durant la Seconde Guerre mondiale en Belgique, n'est pas un objet neutre. Elle est une construction discursive polarisé, servant tantôt à terroriser, tantôt à mobiliser.

L'approche par traitement automatique (TAL) a permis de quantifier ces tendances sur un volume de texte difficile à traiter manuellement. Elle a révélé la persistance de thèmes spécifiques (comme la victimisation) et a mis en lumière des connexions inattendues (pilote/submersible).

Cependant, l'analyse souligne la nécessité impérative d'une critique des sources : l'algorithme ne "comprend" pas l'ironie du Faux Soir de 1943 [6] ni la douleur réelle derrière les statistiques de bombardements. L'outil informatique reste un puissant auxiliaire pour l'historien, capable de cartographier le discours, mais c'est l'interprétation humaine qui lui donne son sens final.

Bibliographie

1. De Wilde, M., *TP4 : Traitement automatique de corpus*, ULB, 2025.
2. Le Soir volé — Wikipédia, consulté le 17 décembre, 2025, https://fr.wikipedia.org/wiki/Le_Soir_vol%C3%A9
3. Le Soir - Belgium WWII, consulté le 17 décembre, 2025, <https://www.belgiumwwii.be/belgique-en-guerre/articles/le-soir.html>
4. Bombardement Mortsel - Lint, consulté le 17 décembre, 2025, <https://www.lint.be/bombardement-mortsel>
5. Bombardement op Mortsel - Wikipedia, consulté le 17 décembre, 2025, https://nl.wikipedia.org/wiki/Bombardement_op_Mortsel
6. Le Faux Soir » : un éclat de rire dans la Deuxième Guerre mondiale, consulté le 17 décembre, 2025, <https://www.bdzoom.com/172477/interviews/%C2%AB-le-faux-soir-%C2%BB-un-eclat-de-rire-dans-la-deuxieme-guerre-mondiale/>
7. Faux Soir — Wikipédia, consulté le 17 décembre, 2025, https://fr.wikipedia.org/wiki/Faux_Soir
8. La Libre Belgique | The Belgian War Press - Cegesoma |, consulté le 17 décembre, 2025, <https://warpres.cgesoma.be/en/node/8946>
9. La Libre Belgique clandestine (1940-1944) - Wikipédia, consulté le 17 décembre, 2025, [https://fr.wikipedia.org/wiki/La_Libre_Belgique_clandestine_\(1940-1944\)](https://fr.wikipedia.org/wiki/La_Libre_Belgique_clandestine_(1940-1944))
10. Céline Rase, Les ondes en uniforme. La propagande de Radio Bruxelles en Belgique occupée (1940-1944), Louvain-la-Neuve, Faculté d'histoire, UCL, 2008-2009. Consulté le 17 décembre, 2025, https://www.journalbelgianhistory.be/en/system/files/article_pdf/chpt23_005_Rase.pdf
11. La Libre Belgique (Peter Pan) - Belgium WWII, consulté le 17 décembre, 2025, <https://www.belgiumwwii.be/belgique-en-guerre/articles/la-libre-belgique-peter-pan.html>
12. File:La Libre Belgique journal clandestin 02.jpg - Wikimedia Commons, consulté le 17 décembre, 2025, https://commons.wikimedia.org/wiki/File:La_Libre_Belgique_journal_clandestin_02.jpg
13. François Hirsch, Les « soldats de la plume » : La presse clandestine en Belgique occupée pendant la Première Guerre mondiale. Consulté le 17 décembre, 2025, <https://warpres.cgesoma.be/sites/warpres.cgesoma.be/files/M%C3%A9moire%20Presse%20clandestine%20PGM.pdf>

14. OCR Challenges in Historical Newspapers, consulté le 17 décembre, 2025, <https://communityhistoryarchives.com/ocr-challenges-in-historical-newspapers-a-problem-300-years-in-the-making/>
15. Presse clandestine (la) - Belgium WWII, consulté le 17 décembre, 2025, <https://www.belgiumwwii.be/belgique-en-guerre/articles/presse-clandestine-la.html>
16. Coline Gérard et Louis Fortemps, La réception de la propagande allemande en Belgique occupée (1940-1944) à travers les journaux personnels. Consulté le 17, décembre 2025, https://www.journalbelgianhistory.be/en/system/files/article_pdf/05%20G%C3%A9rard%20%26%20Fortemps%20v2.pdf
17. L'Indépendance belge - Wikipédia, consulté le 17 décembre, 2025, https://fr.wikipedia.org/wiki/L%27Ind%C3%A9pendance_belge
18. L'Indépendance Belge - OPAC, consulté le 17 décembre, 2025, <https://opac.kbr.be/Library/doc/SYRACUSE/17229173/l-independance-belge?lg=nl-BE>
19. Citation: Laqua, Daniel (2016) Belgian Exiles, the British and the Great War: The Birtley Belgians of Elisabethville. Immigrants & Minorities, 34 (2). pp. 113-131. ISSN 0261-9288. Consulté le 17 décembre, 2025, <https://core.ac.uk/download/pdf/41075322.pdf>
20. Press/Journalism (Belgium) - 1914-1918 Online, consulté le 17 décembre, 2025, <https://encyclopedia.1914-1918-online.net/article/pressjournalism-belgium/>
21. World War, 1939-1945, Underground Resistance Collection, Belgium. Consulté le 17 décembre, 2025, https://library.mcmaster.ca/sites/default/files/uploaded_files/ww2belgium1.pdf
22. Front de l'Indépendance (Le) - Belgium WWII, consulté le 17 décembre, 2025, <https://www.belgiumwwii.be/belgique-en-guerre/articles/front-de-l-independance-le.html>
23. Escaping Nazi Europe: Understanding the Experiences of Belgian Soldiers and civilians in World War 2 (Routledge Studies in Second World War History). Consulté le 17 décembre, 2025, <https://dokumen.pub/escaping-nazi-europe-understanding-the-experiences-of-belgian-soldiers-and-civilians-in-world-war-ii-routledge-studies-in-second-world-war-history-0367136422-9780367136420-9781032621104-9780429027697.html>
24. Topic Modelling Discourse Dynamics in Historical Newspapers, consulté le 17 décembre, 2025, <https://ceur-ws.org/Vol-2865/paper6.pdf>
25. Danilo Najkov, Detecting political bias in online articles using NLP and classification models - Medium, consulté le 17 décembre, 2025, <https://medium.com/@danilo.najkov/detecting-political-bias-in-online-articles-using-nlp-and-classification-models-c1a40ec3989b>
26. Guerre aérienne en Belgique (La) - Belgium WWII, consulté le 17 décembre, 2025, <https://www.belgiumwwii.be/belgique-en-guerre/articles/guerre-aerienne-en-belgique-la.html>
27. Than Just A Nuisance: When Aerial Terror Bombing Works - DTIC, consulté le 17 décembre, 2025, <https://apps.dtic.mil/sti/tr/pdf/ADA338813.pdf>
28. Renaud Meltz, 24. Le collabo impopulaire : la lente agonie du régime de Vichy (1943 – 1944). Consulté le 17 décembre, 2025, <https://shs.cairn.info/pierre-laval-9782262040185-page-919?lang=fr>

29. Allied and German propaganda distributed by air drops and shelling, consulté le 17 décembre, 2025, <https://library.mcmaster.ca/archives/virtualmuseum/mod3-1.html>
30. The German V1 Rocket Leaflet Campaign - Psywarrior, consulté le 17 décembre, 2025, <https://www.psywarrior.com/V1RocketLeaf.html>
31. Belgian Resistance - Wikipedia, consulté le 17 décembre, 2025, https://en.wikipedia.org/wiki/Belgian_Resistance
32. La Province de Namur au cœur de la Seconde Guerre mondiale, consulté le 17 décembre, 2025, https://www.patrimoineculturel.org/documents/fichier/1/2/20200715_100111guide du visiteur la province de namur au coeur de la seconde guerre mondiale.pdf
33. (PDF) Topic modelling discourse dynamics in historical newspapers, consulté le 17 décembre, 2025, https://www.researchgate.net/publication/346089953_Topic_modelling_discourse_dynamics_in_historical_newspapers
34. Airborne Propaganda: The Battle for Hearts and Minds, consulté le 17 décembre, 2025, <https://britishonlinearchives.com/posts/category/articles/436/airborne-propaganda-the-battle-for-hearts-and-minds>
35. Molly Brandt Skelbye et Dana Dannélls, OCR Processing of Swedish Historical Newspapers Using Deep Hybrid CNN-LSTM Network, consulté le 17 décembre, 2025, <https://aclanthology.org/2021.ranlp-1.23.pdf>
36. La Belgique Indépendante. Journal périodique, consulté le 21 décembre 2025 <https://warpres.cgesoma.be/fr/node/40328>
37. Histoire de l'agence d'information Havas, consulté le 21 décembre 2025 https://fr.wikipedia.org/wiki/Histoire_de_l%27agence_d%27information_Havas#:~:text=En%201859%20Havas%20Reuters%20et,ao%C3%BBt%201927%20sur%20l'information.
38. Qu'est-ce que l'analyse discriminante linéaire (LDA) ? IBM, consulté le 21 décembre 2025. [https://www.ibm.com/fr-fr/think/topics/linear-discriminant-analysis#:~:text=L'analyse%20discriminante%20lin%C3%A9aire%20\(LDA\)%20est%20une%20approche%20utilis%C3%A9e,r%C3%A9duction%20de%20dimensionnalit%C3%A9%20des%20donn%C3%A9es.](https://www.ibm.com/fr-fr/think/topics/linear-discriminant-analysis#:~:text=L'analyse%20discriminante%20lin%C3%A9aire%20(LDA)%20est%20une%20approche%20utilis%C3%A9e,r%C3%A9duction%20de%20dimensionnalit%C3%A9%20des%20donn%C3%A9es.)
39. CAMille (Centre d'Archives sur les Médias et l'Information), Corpus de presse belge 1939-1945. Consulté le 2 décembre 2025 <https://camille.ulb.be/?query=aviation+AND+communication&sortcrit=dateasc&year from=1939&year to=1945&edition=01&paper=JB427&paper=JB555&paper=JB555A&paper=JB838>