

Master 2 ECAP
NLP

ETUDE DES DISCOURS DU TRÔNE AU ROYAUME UNI PAR NLP SUR UNE PÉRIODE ALLANT DE 1911 À 2012

**UNE ANALYSE SUR LES THÉMATIQUES, LES PARTIS
POLITIQUES, ET LES SENTIMENTS DES DISCOURS**

Cassandre ORHAN, Isabel PALACIO BARCO,
Lauriane BELANE TANGA & Marie KERHOAS

Mai 2025

Sommaire

I -	Introduction	3
II -	Présentation de la base de données	3
III -	Préparation des données et analyse descriptive	4
IV -	Modélisation thématique avec LDA	7
V -	Classification selon les partis politiques	11
VI -	Analyse de sentiments	12
VII -	Conclusion et Discussion	13
	Annexes	16

I - Introduction

Le Discours du Trône, également appelé Discours du Roi ou de la Reine, constitue un moment institutionnel central de la vie politique britannique. Prononcé à l'ouverture de chaque session parlementaire par le monarque, mais rédigé par le gouvernement en place, il énonce les priorités politiques et législatives pour l'année à venir. Ces discours, riches en indications programmatiques, reflètent les grandes orientations des exécutifs successifs et offrent un aperçu précieux des préoccupations politiques dominantes au fil du temps.

Dans le cadre de ce projet de traitement automatique du langage naturel (NLP), nous nous sommes intéressés à l'évolution du contenu de 105 de ces discours sur une période allant de 1911 à 2012, couvrant ainsi les règnes de George V, George VI et Elizabeth II.

Nos objectifs dans cette étude sont d'étudier les thématiques des priorités politiques britanniques en utilisant la méthode LDA, de proposer une classification de l'appartenance politique des discours selon leur contenu vectorisé via la méthode TF-IDF en mettant en place des modèles de machine learning, mais également de réaliser une analyse de sentiments en utilisant le lexique VADER.

II - Présentation de la base de données

Notre base de données est composée de plusieurs éléments qui ont été assemblés.

Nous avons en premier lieu choisi des données provenant du site du UK Policy Agendas Project, un projet académique qui met à disposition des jeux de données liés à l'agenda politique britannique. Nous utilisons ici le jeu de données intitulé « Speech from the Throne »¹, qui comprend 105 discours du Trônes prononcés au Royaume-Uni entre 1911 et 2012, lors de l'ouverture officielle des sessions parlementaires. Ces discours, écrits par le gouvernement en place mais lus par le monarque ou son représentant devant la Chambre des Lords et la Chambre des Communes, exposent les priorités législatives du gouvernement pour l'année à venir. Il est à noter que ces discours avaient été recodés en quasi-phrase auxquelles des thèmes avaient été attribués. Mais certaines quasi-phrase étaient une répétition des phrases d'origines afin de prendre en compte les différents thèmes lui correspondant. Aussi, nous avons choisi de reconstituer intégralement chaque discours et avons ainsi obtenu un fichier dans lequel une ligne correspond à un discours. Nous disposons également des dates auxquelles ils ont été prononcés. Il est important de préciser que ces discours n'ont pas eu lieu à une date fixe chaque année, en effet ils sont prononcés selon le contexte politique du pays (élection d'un

¹Source des données : <https://policyagendasuk.wordpress.com/datasets/>

nouveau premier ministre, ...) ou les événements majeurs (périodes de guerre, ...) auquel il fait face.

Afin de mieux comprendre le contexte politique associé à ces discours, nous avons également choisi d'indiquer les noms des premiers ministres en poste au moment où ils ont été prononcés. Cela avait pour objectif de nous permettre d'associer un parti politique à chacun des discours.

III - Préparation des données et analyse descriptive

III.1 - L'évolution du nombre de mots par discours

Dans un premier temps, il est intéressant de se rendre compte de l'évolution de la longueur des discours, qui constitue un indicateur révélateur des priorités politiques et des circonstances historiques de chaque période.

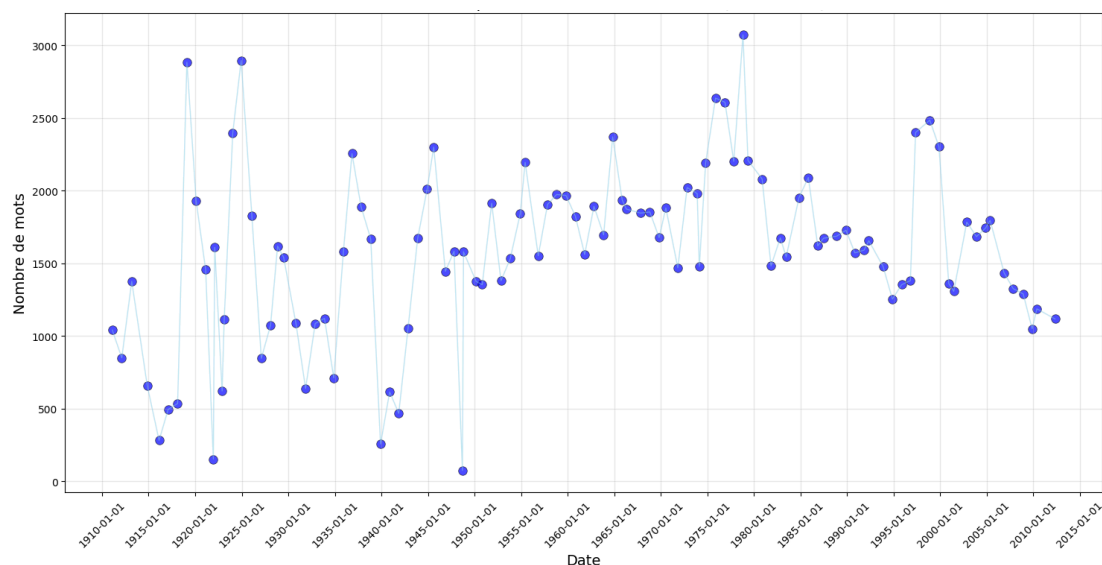


Fig. 1. – Evolution de la longueur des discours en nombre de tokens

Comme le montre le graphique, la longueur des discours du trône a globalement augmenté au fil des décennies. On observe des discours particulièrement courts à des moments clés de l'histoire :

- le 28 novembre 1939 : au début de la Seconde Guerre mondiale après que Varsovie se soit rendue, le discours se concentre exclusivement sur « la poursuite de la guerre » et « les efforts et sacrifices » demandés ;
- le 14 décembre 1921 : quelques jours après la signature du traité anglo-irlandais, dans un contexte de forte tension politique, le discours exprime « l'espoir sincère que par les Articles d'Accord maintenant soumis, les conflits séculaires puissent prendre fin » et que

« l'Irlande, en tant que partenaire libre du Commonwealth » puisse réaliser ses idéaux nationaux ;

- le 14 septembre 1948 : en pleine période de reconstruction économique et sociale après la guerre, le discours extraordinairement bref ne comporte que 82 mots et se concentre uniquement sur l'amendement de la loi parlementaire de 1911 ;

À l'inverse, les discours les plus longs ont été prononcés lors de sessions marquées par une densité législative importante, notamment :

- 1 novembre 1978 : précédant l'arrivée au pouvoir de Margaret Thatcher et marquant la fin d'une époque travailliste en crise, ce discours exceptionnellement long présente un vaste programme législatif couvrant la politique étrangère, les mesures économiques contre « les maux de l'inflation et du chômage », les droits des locataires, la santé, et la participation des employés à la gouvernance des entreprises ;
- 9 décembre 1924 : au début d'un nouveau gouvernement conservateur, ce discours détaillé aborde le meurtre de Sir Lee Stack au Caire, les relations avec la Société des Nations, les liens commerciaux avec l'Allemagne, le chômage, et la crise du logement ;
- 11 février 1919 : fin de la Première Guerre mondiale, ce discours met l'accent sur la reconstruction du pays, « les aspirations à un meilleur ordre social » et la création de nouveaux ministères chargés de la santé publique et des transports ;

Il est à noter que des fluctuations importantes ont eu lieu entre 1910 et 1950, probablement dues aux instabilités géopolitiques, avant une tendance à l'homogénéisation de la longueur des discours à partir des années 50.

III.2 - Nettoyage du jeu de données

Si nous avons pu réaliser une présentation de la longueur des discours selon les données d'origines dont nous disposions, il est à présent nécessaire de procéder à leur nettoyage afin d'obtenir une version des discours à laquelle nous pourrions appliquer des méthodes d'analyse de NLP.

Aussi, nous avons successivement converti les discours en minuscules, enlevé la ponctuation, les chiffres et les stop words, tokenisé puis lemmatisé le texte et conservé uniquement les mots alphabétiques de plus de 2 lettres.

Pour revenir plus en détails sur les stop word, ces mots fréquents mais peu informatifs sur le contenu, que nous avons éliminés, notre analyse révèle que les vingt les plus courants représentaient une part considérable du corpus, avec « the » (13 491 occurrences), « of » (9 064) et « to » (8 800) en tête. La forte présence de « will » (5 354) et « my » (3 111) reflétait la nature performative et personnalisée de ces allocutions royales annonçant les intentions gouvernementales. Au total, l'élimination de ces mots vides a réduit le volume textuel de

50,46% (de 166 085 à 83 808 mots), illustrant le caractère hautement formalisé de ces discours. Cette transformation, comme en témoigne l'exemple « My Lords, and Gentlemen, In opening... » devenant « lord gentleman open... », permet de concentrer l'analyse sur les termes véritablement porteurs de sens politique, optimisant ainsi l'efficacité des traitements ultérieurs.

III.3 - Analyse descriptive par parti politique

Dans l'optique d'améliorer notre compréhension du jeu de données, il est intéressant de se rendre compte de la distribution des discours selon le parti politique du gouvernement en place. Parmi les 105 discours : 56 sont issus d'un gouvernement conservateur, 37 d'un gouvernement travailliste et 12 d'un gouvernement libéral.

D'un point de vue chronologique, ils se répartissent de la façon suivante :

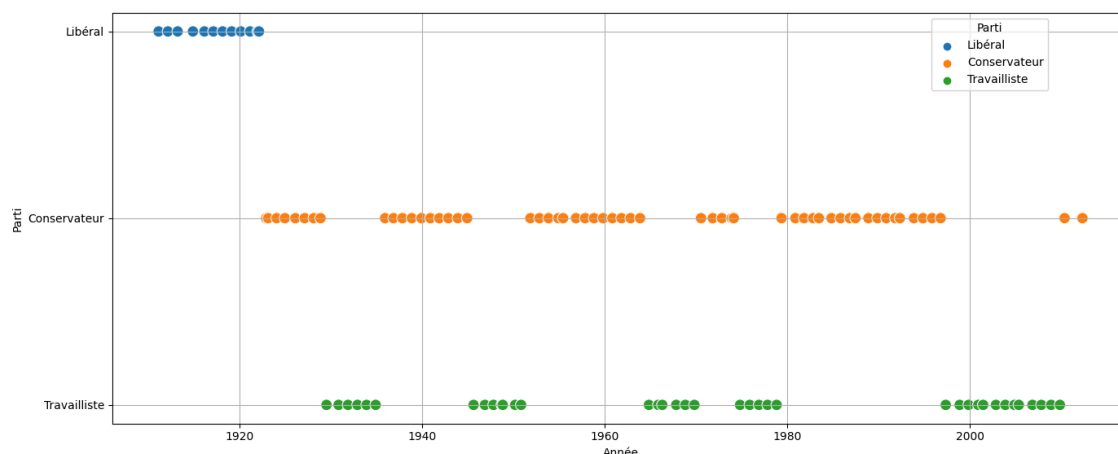


Fig. 2. – Répartition des discours selon les partis politiques et au cours du temps

Si l'on observe que le parti libéral n'a pas été celui du premier ministre depuis le milieu des années 1920, il y a en revanche eu une alternance au fil des années entre les chefs de gouvernement des partis travailliste et conservateur.

Afin d'avoir une idée des différences visibles entre les discours de ces différents partis, nous avons cherché à connaître les mots les plus fréquemment employés dans les discours associés aux 3 partis politiques.

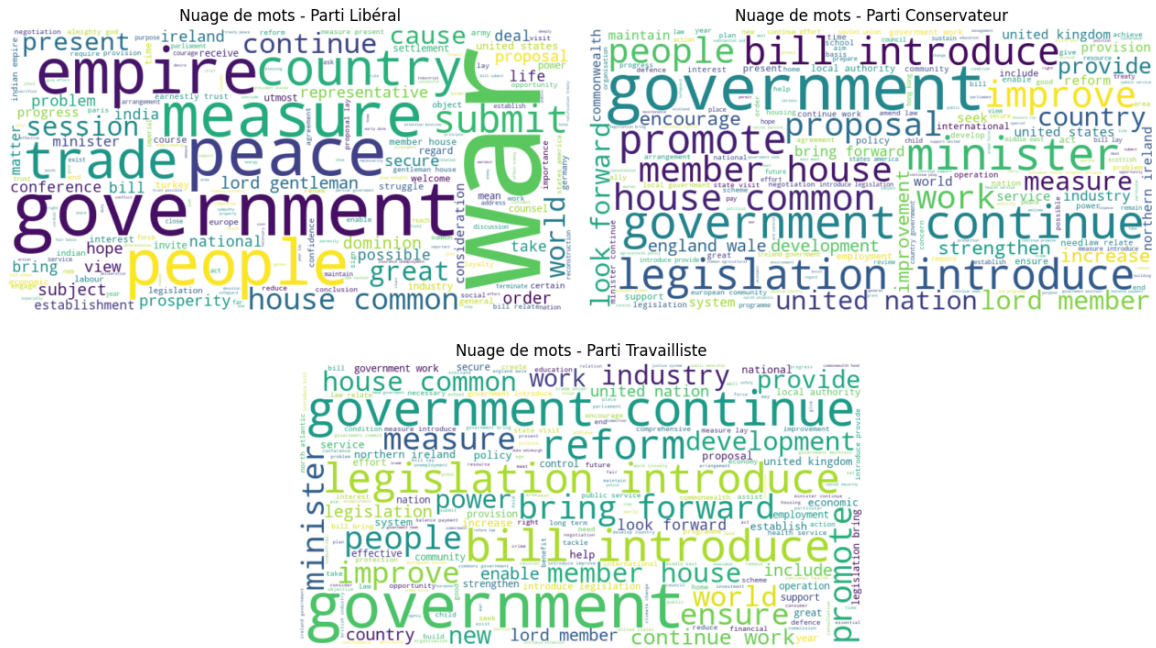


Fig. 3. – Nuage de mots des différents partis politiques

Étant donné le contexte historique (1ère guerre mondiale) des discours prononcés lorsque le gouvernement en charge était le parti libéral, on comprend la forte présence de termes liés à cette thématique.

Concernant les termes qui ressortent des discours associés aux 2 autres partis, on retrouve principalement des notions communes liées au gouvernement, à l'imposition, au développement, à l'amélioration, ... Mais il est également à noter que des mots semblent plus propres à un parti que l'autre tels réforme (« reform ») pour le parti travailliste ou renforcer (« strengthen ») pour le parti conservateur.

IV - Modélisation thématique avec LDA

Afin d'identifier les thématiques majeures abordées dans les 105 discours du Trône, nous avons choisi d'appliquer la méthode LDA (Latent Dirichlet Allocation). Cette approche non supervisée est bien adaptée à notre corpus, puisqu'elle permet de faire émerger automatiquement des sujets latents à partir de textes, sans nécessité d'étiquettes préalables. LDA est particulièrement pertinente dans notre cas, car elle facilite une lecture diachronique des priorités gouvernementales, ce qui nous permet de comparer les thématiques dominantes au fil des décennies.

IV.1 - Vectorisation

Dans un premier temps, nous avons procédé à la vectorisation des discours à l'aide de la méthode CountVectorizer, qui permet de transformer les textes en une matrice document-terme (DTM). Cette matrice a pour structure : une ligne par discours ; une colonne par mot retenu ; et une valeur correspondant au nombre d'occurrences de ce mot dans le discours. Afin d'épurer le vocabulaire et de ne conserver que les mots pertinents, nous avons exclu :

- les mots trop fréquents : apparaissant dans plus de 90% des discours
- les mots trop rares : présents dans moins de 10 discours
- les mots vides usuels en anglais

Ce filtrage nous a permis de travailler sur un vocabulaire à la fois représentatif et discriminant du contenu des discours.

IV.2 - Entraînement du modèle LDA et exploration du vocabulaire

Nous avons ensuite appliqué le modèle LDA avec un nombre de sujets fixés arbitrairement à 3. Ce choix, bien qu'empirique, vise à proposer une première exploration thématique du corpus. Le modèle permet de représenter chaque discours comme une combinaison pondérée de plusieurs thématiques, et chaque thématique comme une distribution probabiliste de mots.

A l'issue de l'entraînement, nous avons vérifié le lexique exploité par le modèle en observant le nombre total de mots retenus dans la matrice document-terme (788 termes), et en examinant un échantillon aléatoire de mots présents dans le corpus filtré. Cette étape de contrôle visait à s'assurer de la pertinence et de la qualité du vocabulaire mobilisé par le modèle.

IV.3 - Détermination optimale des thématiques

Ce premier choix de 3 sujets, bien qu'exploratoire, soulève une question dans l'utilisation de LDA : combien de thématiques convient-il d'extraire pour restituer au mieux la diversité des discours ? En l'absence de connaissance préalable sur le nombre réel de sujets abordés, il n'est pas possible de fixer arbitrairement ce paramètre sans risquer de simplifier ou de fragmenter à l'excès les contenus. Pour affiner notre modélisation, nous avons donc testé plusieurs configurations et évalué la perplexité du modèle pour un nombre de topics variant entre 2 et 10. Cette mesure, utilisée pour les modèles de langages probabilistes, permet d'apprécier leur capacité à représenter les données: plus la perplexité est faible, meilleur est l'ajustement.

Le modèle à 5 topics s'est avéré le plus performant selon ce critère. (*Annexe 1*) Si ce modèle obtient la meilleure perplexité, cela ne garantit pas que les topics soient facilement

interprétables. Il est donc essentiel de les analyser manuellement, en commençant par observer les mots les plus représentatifs des différents topics.

Pour chaque topic, nous avons extrait les 20 termes ayant les poids les plus élevés selon LDA. Ces mots peuvent être visualisés sous forme de graphiques à barres, où leur poids estimé permet de mieux cerner la thématique dominante. Cependant, pour affiner cette analyse, il est pertinent de comparer les poids attribués par LDA à la fréquence réelle d'apparition de ces mots dans les discours attribués à chaque topic. Cette comparaison permet de vérifier si les mots les plus importants selon le modèle sont également fréquents dans les discours réellement classés dans le topic concerné. En d'autres termes, cela nous aide à évaluer la cohérence sémantique du modèle.

Afin de mieux interpréter les thématiques dégagées par le modèle, nous avons croisé deux éléments complémentaires : les poids LDA et les fréquences réelles. Et l'analyse conjointe de ces deux éléments permet de distinguer deux types de termes : les mots structurels et les mots anecdotiques. Cette double lecture renforce l'interprétation sémantique des 5 topics :

- **1. Gouvernance publique et coopération internationale :**

Ce topic fait émerger un lexique centré sur l'État, les politiques publiques et les relations institutionnelles. Des mots comme state, introduce, policy, community ou encore support reviennent fréquemment, aussi bien dans les poids LDA que dans les fréquences effectives. Ces éléments suggèrent une thématique structurée autour de la gestion publique, des réformes internes, mais aussi de la coopération entre entités politiques, notamment dans un cadre international comme le Commonwealth.

- **2. Diplomatie, commerce et législation internationale :**

Il se distingue par la présence de mots comme minister, trade, industry, peace, international ou proposal, qui évoquent une dimension économique-diplomatique forte. On y perçoit un langage technique tourné vers la régulation des échanges, les accords bilatéraux ou multilatéraux, ainsi que la législation internationale. Les discours associés à ce topic reflètent des enjeux de politique étrangère, mais aussi des dynamiques de négociation commerciale entre États.

- **3. Réformes économiques et développement public :**

Il présente un vocabulaire très ancré dans le champs du développement : reform, development, economic, policy, support, ou encore service. Les fréquences élevées de ces termes dans les discours renforcent l'idée que ce topic capte des interventions relatives à la modernisation de l'action publique, à la gestion économique, et à la mise en œuvre de programmes de développement.

- **4. Défense, conflits armés et sécurité nationale :**

Ce topic se distingue nettement par un vocabulaire guerrier : war, force, ally, nation, people, minister, army. Ces mots sont non seulement pondérés par LDA, mais aussi fortement présents dans les discours concernés, ce qui en fait un topic à forte cohérence sémantique. Il semble regrouper les discours traitant des conflits armés, de la défense nationale, ou encore de mobilisation militaire et stratégique.

- **5. Justice, sécurité intérieure et inclusion sociale :**

Le cinquième topic est porté par des mots tels que child, crime, secure, public, support, reform, include. Il articule deux dimensions souvent liées : d'un côté la justice sociale (protection de l'enfance, inclusion), et de l'autre la sécurité intérieure (criminalité, réformes de la justice ou de la police). On y retrouve une rhétorique axée sur la cohésion sociale et la lutte contre les inégalités, qui traverse les interventions portant sur la sécurité et les droits fondamentaux.

IV.4 - Évolution diachronique des thématiques dominantes

Nous avons ensuite voulu illustrer l'évolution au cours du temps de la répartition des discours entre les cinq thématiques majeures, en mettant en évidence les périodes durant lesquelles chaque sujet a été le plus fréquemment abordé dans les discours du Trône.

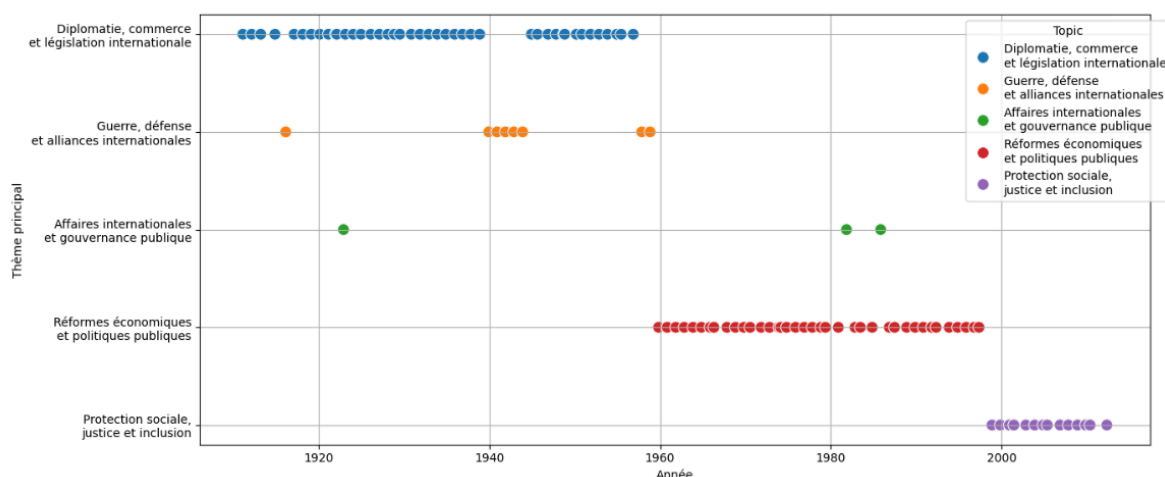


Fig. 4. – Thèmes dominants des discours

Cette visualisation révèle plusieurs dynamiques historiques intéressantes :

- Avant les années 1950, les discours sont très largement dominés par les thématiques de diplomatie, commerce et législation internationale, ce qui reflète les priorités d'un Empire britannique encore très tourné vers ses relations extérieures et commerciales.

- Durant les deux conflits mondiaux (1914–1918 et 1939–1945), on observe une montée des sujets liés à la défense, aux alliances militaires et à la sécurité nationale, ce qui est cohérent avec le contexte géopolitique.
- À partir des années 1960, un basculement s’opère avec l’émergence des réformes économiques et politiques publiques comme thématique centrale, marquant une réorientation vers les enjeux domestiques et la modernisation de l’État-providence.
- Les années 1980–1990 témoignent d’un intérêt croissant pour les questions de gouvernance publique et d’affaires internationales, notamment dans un contexte de globalisation et de redéfinition du rôle de l’État.
- Enfin, au tournant des années 2000, les préoccupations autour de la justice sociale, de la sécurité intérieure et de l’inclusion prennent de l’importance, probablement en lien avec l’évolution des politiques sociales et des enjeux sécuritaires post-11 septembre.

V - Classification selon les partis politiques

La partie suivante de notre étude a consisté à mettre en place des modèles afin de tenter de classer les discours selon leur parti politique et cela grâce à leur contenu. En effet, nous souhaitons comprendre si dans les discours de chaque parti il existe des dénominateurs communs permettant d’identifier leur appartenance politique.

V.1 - Méthodologie

Comme nous avons pu le voir précédemment, il existe 3 partis politiques auxquels appartenaient les premiers ministres en exercice au moment où les discours étudiés ont été prononcés. Mais étant donné que le parti Libéral a été au pouvoir uniquement lors des 12 premiers discours dont nous disposons et principalement dans un contexte de guerre, nous n’allons pas les prendre en compte pour réaliser cette classification. Aussi notre variable à expliquer comporte 2 modalités : le parti conservateur et le parti travailliste.

Afin d’être en mesure de classer les discours selon leur parti, nous avons besoin de variables explicatives. Elles ont été obtenues en vectorisant les textes nettoyés des discours selon la méthode TF-IDF et en se limitant aux 150 mots les plus significatifs. Nous avons ainsi obtenu une matrice indiquant l’importance de ces termes dans les différents discours.

L’étape suivante a alors consisté à séparer notre jeu de données entre un jeu d’entraînement comprenant 70 % de nos données, et un jeu de test composé des 30 % restant.

Nous avons alors mis en place 2 modèles de machine learning : une régression logistique ainsi qu’une forêt aléatoire, et nous les avons entraîné sur le jeu d’entraînement en testant différents hyper-paramètres.

V.2 - Résultats

Afin d'évaluer la pertinence de ces modèles, nous avons comparé leurs prédictions sur le jeu de test avec les valeurs réelles que nous connaissons.

		Prédiction			
		Régression Logistique		Forêt Aléatoire	
		Travailliste	Conservateur	Travailliste	Conservateur
Réel	Travailliste	3	9	4	8
	Conservateur	0	16	0	16

Tableau 1. – Matrices de confusion des modèles sur le jeu de test

On constate dans les 2 modèles que l'ensemble des discours rédigés lorsqu'un conservateur occupait le poste de premier ministre (16) ont été correctement classés. Toutefois, concernant les discours du parti travailliste, on observe que dans la régression logistique, seuls 3 d'entre eux parmi les 12, ont été correctement prédits. Dans le cas de la forêt aléatoire, ce nombre s'améliore légèrement en passant à 4. Cela signifie que les modèles sont bons pour prévoir les discours du parti conservateur, mais peinent à distinguer ceux du parti travailliste.

D'un point de vue plus global, la régression logistique et la forêt aléatoire ont respectivement correctement prédit 19 et 20 des 28 discours. Aussi, leur accuracy sont de 67.86 % et de 71.43 %. Ces mesures de performance sont plutôt satisfaisantes et témoignent de l'intérêt des modèles bien qu'ils sous-estiment le nombre de discours prononcés lorsque que le chef du gouvernement est travailliste.

VI - Analyse de sentiments

Nous avons également souhaité explorer la présence éventuelle de sentiments dans les discours, bien que ceux ci soient, par nature, marqués par une neutralité institutionnelle et un ton solennel. Cette démarche, avant tout exploratoire, visait à vérifier s'il était possible de détecter des émotions ou des expressions affectives, malgré le caractère formel. Pour cela, nous avons utilisé le lexique VADER (valence aware dictionary for sentiment reasoning).

VADER repose sur un dictionnaire lexical (en anglais) pondéré, associant à chaque mot une valeur de sentiment (positive, négative ou neutre) ainsi qu'un degré d'intensité. Le résultat est un score "compound" compris entre -1 (très négatif) et $+1$ (très positif). À partir de ce score, nous avons attribué une étiquette qualitative — positif, négatif ou neutre — en suivant les seuils recommandés par la documentation officielle (± 0.05).

La représentation graphique (*Annexe 2*) de l'évolution des scores confirme une tendance globale positive dans l'ensemble du corpus, avec des valeurs proches de +1 pour la quasi-totalité des années. Cette homogénéité reflète le registre solennel et affirmatif propre aux discours du trône auxquels on s'attendait, qui visent à afficher la stabilité de l'état et la continuité des priorités gouvernementales.

Cependant, certaines exceptions ponctuelles apparaissent lors des années de guerre, notamment en 1940 ou en 1948. Le discours de 1940, prononcé en pleine seconde guerre mondiale, contient plusieurs segments faisant référence aux destructions, aux pertes humaines et aux contraintes économiques liées au conflit. Ces passages, bien que formulés dans un langage institutionnel, mobilisent un lexique anxiogène ("hostilities", "enemy attack", "destroyed", "damaged", "ordeal") que le modèle VADER interprète comme négatif. Ainsi, le passage :

"Measures will be submitted to you for compensating those whose home or business property has, at any time since the outbreak of hostilities, been destroyed or damaged by enemy attack..."

obtient un score de -0.97 . De même, d'autres segments du même discours évoquant les efforts financiers requis pour la guerre ou les conditions difficiles de la population sont également associés à des scores négatifs, par exemple:

"You will be asked to make further financial provision for the conduct of the war." (-0.60)

"Apart from these [...] My Government will take every possible step to sustain the health and wellbeing of My people in their ordeal." (-0.61)

Ces exemples montrent que même dans des allocutions globalement positives, certaines sections peuvent contenir un vocabulaire perçu comme négatif par le modèle. Et nous permet de voir les limites de VADER, qui applique une lecture lexicale directe sans forcément tenir compte du contexte. Les mots des « sentiments » détectées ne traduisent donc pas une tonalité globale pessimiste, mais bien l'évocation ponctuelle de circonstances graves, exprimées dans un cadre toujours solennel et volontariste.

VII - Conclusion et Discussion

L'analyse des discours du trône britannique sur la période 1911-2012 offre un éclairage précieux sur l'évolution des priorités politiques du Royaume-Uni au cours d'un siècle marqué par de profonds bouleversements.

Notre analyse descriptive a tout d'abord révélé des variations significatives dans la longueur des discours reflétant fidèlement les contextes historiques sous-jacents. Les allocutions parti-

culièrement brèves correspondent aux périodes de crise majeure, tandis que les discours plus développés signalent des moments de forte activité législative et de transformation politique. Ce premier constat confirme la pertinence d'étudier ces discours comme des indicateurs fiables des priorités gouvernementales.

La tokenization et l'élimination des mots vides (stop words) ont constitué une étape essentielle pour préparer notre corpus à une analyse thématique approfondie. Cette transformation a réduit le volume textuel de plus de 50%, témoignant du caractère hautement formalisé de ces allocutions royales.

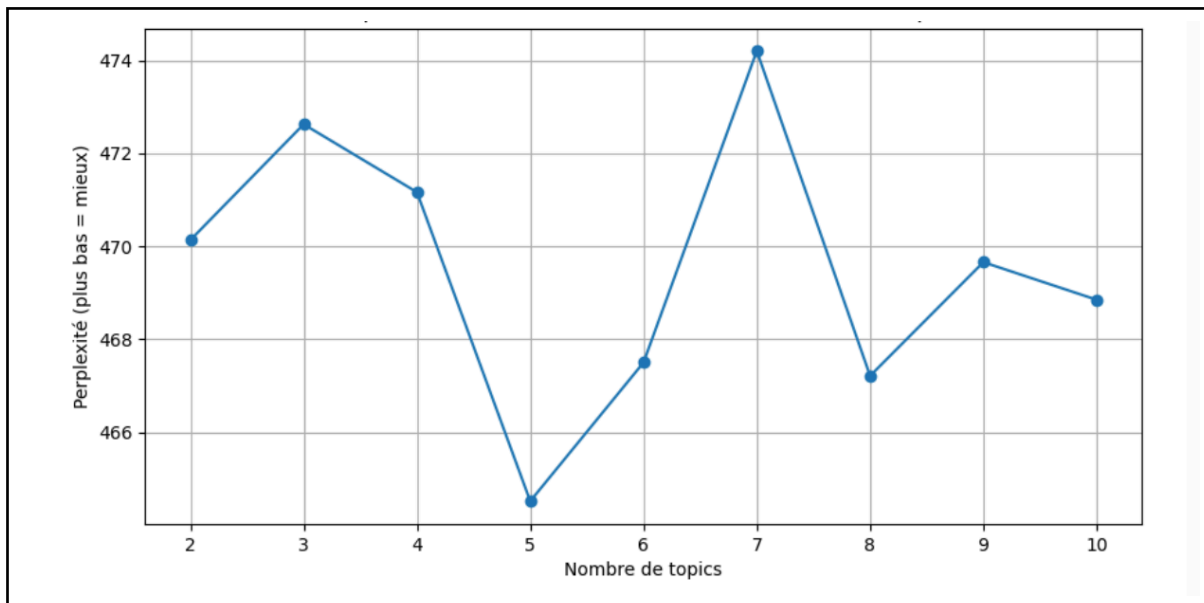
La modélisation thématique par LDA a fait émerger cinq grands domaines qui structurent ces discours : la gouvernance publique et la coopération internationale ; la diplomatie et le commerce ; les réformes économiques ; la défense et la sécurité nationale ; et enfin la justice et l'inclusion sociale. La prévalence du thème de la défense pendant les périodes de guerre, suivie d'une diversification progressive vers des préoccupations de gouvernance interne et de réformes sociales, illustre parfaitement l'adaptation du discours politique aux défis changeants de la société britannique.

Notre tentative de classification des discours selon l'appartenance politique du gouvernement en place a donné des résultats mitigés. Avec des taux de précision de 67,86% pour la régression logistique et 71,43% pour la forêt aléatoire, nos modèles parviennent à identifier correctement la majorité des discours, montrant l'existence de marqueurs lexicaux spécifiques à chaque parti. Toutefois, ils présentent une asymétrie notable : si les discours conservateurs sont systématiquement bien identifiés, les allocutions travaillistes sont souvent confondues avec celles des conservateurs. Cette difficulté suggère que les gouvernements travaillistes adoptent parfois une rhétorique plus proche de celle de leurs opposants, peut-être pour se positionner au centre de l'échiquier politique britannique.

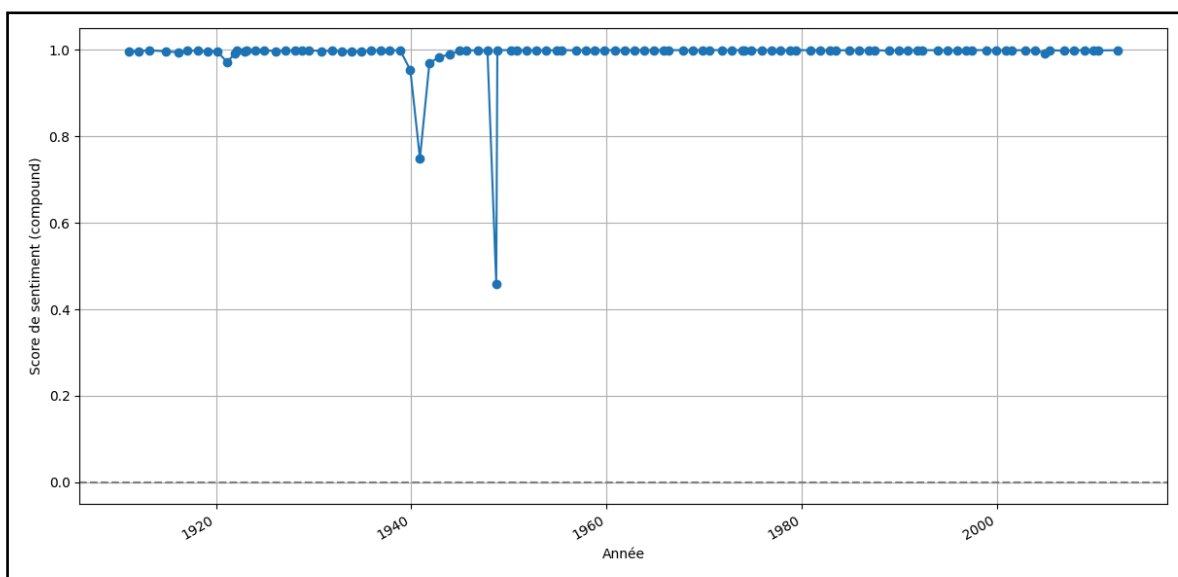
Cette recherche présente néanmoins certaines limites qu'il convient de souligner. Notre analyse de sentiment avec l'outil VADER, bien qu'ayant confirmé la prédominance d'une tonalité positive inhérente au format solennel de ces allocutions, a révélé ses limites face à un corpus institutionnel. Si quelques exceptions notables apparaissent durant les périodes de guerre avec un lexique plus anxiogène, l'outil reste peu adapté pour saisir les nuances d'un discours royal où même l'évocation de circonstances graves s'inscrit dans un cadre toujours solennel et volontariste. Plus généralement, nous avons constaté la difficulté d'analyser, par des méthodes automatiques, des discours institutionnels dont les subtilités politiques et économiques échappent souvent aux algorithmes. Notre approche, focalisée sur l'analyse lexicale et thématique, ne permet pas de capturer pleinement le sens contextuel et les implications de chaque phrase, ce qui constitue une limite inhérente aux méthodes de NLP appliquées à des corpus aussi spécifiques que les discours du trône.

Les techniques de NLP nous ont ainsi permis d'atteindre notre objectif initial : étudier le contenu des discours du trône, identifier les thématiques dominantes et observer leur évolution temporelle. Malgré la permanence du cadre cérémonial, le contenu de ces allocutions reflète les transformations profondes de la société britannique à travers le siècle.

Annexes



Annexe 1. – Perplexité du modèle LDA selon le nombre de topics



Annexe 2. – Évolution du sentiment des discours du Trône - VADER