UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Automatic News Analysis for Polarization Understanding

THÈSE

présentée et soutenue publiquement le 3 Septembre 2024

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Evan Dufraisse

Composition du jury

Président :	Slim Ouni	Professeur, Université de Lorraine
Rapporteurs :	Pascale Sébillot	Professeure, INSA de Rennes
	Symeon Papadopoulos	Professeur, Centre for Research and Technology Hellas
Examinateur :	Oana Goga	Chargée de Recherches CNRS, Ecole Polytechnique
Directrice :	Armelle Brun	Professeure, Université de Lorraine
Co-Directeur :	Julien Tourille	Docteur, Université Paris-Saclay, CEA List
Invité :	Adrian Popescu	Docteur, Université Paris-Saclay, CEA List

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Acknowledgments

I would like to express my sincerest gratitude to my advisors. Armelle Brun, for your guidance in navigating the complex realm of academia. Julien Tourille, for your guidance and technical support throughout this journey. And Adrian Popescu, for our rich exchange of ideas, which has fueled my research and driven it in innovative directions.

I am deeply grateful to my committee members. Special thanks go to my thesis reviewers, Pascale Sébillot and Symeon Papadopoulos, for their time, valuable feedback, and constructive criticism, which were instrumental in refining this final version. I am also thankful to the other committee members for their careful examination and insightful questions.

I am fortunate to have had the support of my colleagues and lab mates, particularly Michael, Aboubacar, and Tiphaine, whose stimulating discussions, collaboration, and camaraderie made both challenging and rewarding times memorable. I truly value the friendships and professional bonds we've built.

My heartfelt thanks go to my family and friends for their unwavering support, patience, and understanding. Their encouragement has been a constant source of strength throughout this journey. I am especially grateful to my parents, whose unconditional love and support have been a steadfast foundation.

This journey has equipped me with invaluable experiences and insights, which I look forward to carrying into future challenges and endeavors.

Résumé étendu

Contexte

Cette thèse s'inscrit dans le cadre de la transformation récente des modes d'accès à l'information permis par les avancées technologiques de la télécommunication, notamment l'avènement d'internet et la diffusion de l'utilisation des réseaux sociaux comme moyens de s'informer. Ces mutations nous ont menés dans ce que certains chercheurs qualifient de 'Crise de l'Information' [Beckett and Livingstone, 2018; Haider and Sundin, 2022], caractérisée entre autres par la fragmentation du paysage médiatique, le déclin du journalisme d'investigation au profit du sensationnalisme, et l'érosion de la confiance dans les institutions.

L'essor des médias numériques a entraîné une fragmentation marquée des audiences, conséquence de la diminution du coup de la diffusion permise par internet. Cette fragmentation s'est accrue avec la personnalisation des contenus, facilitée par des algorithmes de recommandation qui intensifient les préférences de l'utilisateur. L'apparition de modèles de monétisation principalement basés sur la publicité a également encouragé la production de contenus sensationnalistes et simplifiés au détriment du journalisme d'investigation, ce qui a aussi des répercussions sur la qualité de l'information disponible [Hargreaves, 2014; Und, 2016; Haider and Sundin, 2022].

Parallèlement, on observe la montée de la polarisation dans les pays occidentaux, à l'instar des États-Unis, où le phénomène est d'une importance croissante [1]. Ce phénomène s'exprime à la fois sur un plan politique et affectif, où les divisions idéologiques se renforcent et l'intolérance envers les opinions opposées s'accroît [Frimer et al., 2017; Martherus et al., 2021]. Bien que les mécanismes sous-jacents à cette polarisation ne soient pas encore complètement compris, les effets de la personnalisation de l'information et des algorithmes de recommandation des réseaux sociaux sont souvent suspectés. L'hypothèse la plus répandue suggère que l'exposition répétée à des informations homogènes tendrait à renforcer les convictions initiales des utilisateurs [Pariser, 2011]. D'autres explications existent également, mais indépendamment de celles-ci, la capacité à modéliser et caractériser précisément les opinions véhiculées par les contenus partagés et consultés en ligne pourrait permettre d'atteindre une compréhension plus approfondie des mécanismes de la polarisation et des moyens potentiels de la réduire. [Bakshy et al., 2015; Flaxman et al., 2016; Törnberg, 2022].

Face à ces défis, le projet ANR BOOM (ANR-20-CE23-0024), qui finance cette thèse, vise à proposer de nouveaux algorithmes de recommandation d'articles de presse visant à diversifier les opinions présentées aux lecteurs afin d'étudier l'impact de cette diversification sur la polarisation. Ce projet implique la capacité de caractériser finement les opinions contenues dans les articles de presse, en identifiant les biais et les positions idéologiques exprimées. C'est dans cette perspective que notre travail s'inscrit en s'interessant à ces trois questions de recherche :

- **RQ1** - Comment concevoir un système d'analyse d'opinion avec une granularité fine et préservant un haut niveau d'explicabilité ?
- **RQ2** - Comment permettre un transfert aisé de ce système d'analyse à d'autres langues ?
- **RQ3** - Comment agréger ces opinions pour permettre d'identifier les biais dans la presse ?

[1] https://www.pewresearch.org/politics/2014/06/12/appendix-a-the-ideological-consistency-scale/

À ces trois questions de recherche correspondent trois contributions principales :

1. **Le développement d'un système général pour l'analyse fine des opinions.** Ce système combine la reconnaissance d'entités nommées, la résolution d'entités par wikification, et une analyse de sentiment ciblée envers les entités, permettant ainsi une interprétation plus fine et une agrégation des sentiments basée sur les attributs démographiques ou politiques issus de bases comme Wikipédia et ParlGov. Cette approche offre un niveau d'explicabilité élevé, crucial pour comprendre les nuances des opinions exprimées dans les articles de presse.

2. **La conception d'un jeu de données multilingue pour la classification de sentiment ciblé.** Face à la rareté des jeux de données adaptés aux contextes multilingues et non-américains, nous avons créé MAD-TSC, un jeu de données aligné pour l'analyse de sentiments dans huit langues européennes. Ce corpus, réalisé à partir de traductions professionnelles d'articles VoxEurop, permet d'entraîner des modèles capables d'analyser des sentiments à l'égard d'entités spécifiques, facilitant des expériences multilingues et explorant le transfert de connaissances entre langues.

3. **L'application de l'analyse aux paysages médiatiques français et flamand.** En instanciant notre système pour analyser la presse française et flamande, nous démontrons sa capacité à identifier les biais politiques et démographiques des sources d'information. L'intégration de données telles que celles de ParlGov nous a permis de comparer la couverture médiatique des entités politiques selon leur orientation, révélant des tendances dans les traitements médiatiques nationaux et démontrant la flexibilité de ce système dans un contexte multilingue.

Ces contributions nous rapprochent d'une meilleure compréhension des mécanismes de polarisation dans les médias et permettent d'envisager des solutions pour diversifier les recommandations d'articles, avec un impact potentiel sur la réduction des bulles de filtres et la promotion d'une expérience informationnelle plus équilibrée.

Etat de l'art

Historiquement, l'analyse de la presse était effectuée dans les sciences humaines par la conception de guides d'annotation et une annotation manuelle des textes. L'augmentation de la puissance et de la disponibilité des moyens informatiques a progressivement permis de mener des analyses plus automatisées, d'abord avec des méthodes simples comme le comptage de mots. Malgré la progression des algorithmes, ces approches automatiques ont toujours représenté un compromis entre la quantité de données analysées et la qualité ou la complexité de l'annotation. Les avancées récentes en traitement automatique du langage naturel, dues principalement aux architectures basées sur les transformers et au pré-entraînement auto-supervisé, ont permis d'augmenter de manière significative les performances sur de nombreuses tâches en exploitant des représentations sémantiques riches issues du pré-entraînement sur de larges corpus.

Le domaine de l'analyse de presse et, plus particulièrement, celui de la détection de biais n'échappe pas à cette évolution. Dans la littérature, plusieurs approches existent pour détecter les biais médiatiques : on trouve des analyses basées sur la répartition thématique des contenus, la présence de mots ou expressions chargés émotionnellement ou idéologiquement, et la représentation de certaines catégories démographiques dans les contenus médiatiques. Par exemple, des méthodes comme la détection de cadrage (ou framing) examinent comment certains faits ou

thèmes sont mis en avant ou marginalisés dans les articles, influençant ainsi la perception du lecteur.

Dans ce contexte, les méthodes récentes de détection du cadrage utilisent des corpus annotés comme celui de Card et al. [2015] pour l'analyse de nouvelles politiques, tandis que des approches de détection de biais d'orientation politique visent à identifier des indices partisans ou idéologiques. D'autres études abordent les biais en se basant sur la représentation de catégories démographiques, par exemple l'analyse des représentations de genre ou d'ethnie dans les médias [Asr et al., 2021]. Toutefois, ces approches ont souvent une portée limitée en raison de la difficulté à généraliser les méthodes à d'autres langues ou cultures et du manque d'explicabilité pour les utilisateurs finaux.

Les travaux existants montrent également que certaines méthodes, comme l'analyse de sentiments à base de lexiques ou l'analyse fréquentielle, restent limitées pour des sujets complexes, car elles peinent à saisir des nuances plus subtiles comme le ton implicite envers une entité. Les systèmes pionniers, comme Lydia/Textmap [Lloyd et al., 2005], se sont concentrés sur des analyses dites 'objectives' en ne mesurant que des éléments quantifiables, comme les fréquences d'apparition et les relations de co-occurrence entre entités. Bien que d'autres, comme le système de Godbole et al. [2007], aient introduit une analyse de sentiments, ces approches demeurent simplistes et généralement limitées à l'anglais, ce qui limite leur applicabilité à des contextes multilingues.

Pour répondre à ces limitations, des approches plus récentes, comme celles de Hamborg [2023b], utilisent la classification de sentiment ciblée, c'est-à-dire l'identification d'opinions envers des entités spécifiques, pour comparer les biais de couverture entre sources médiatiques. Ces approches ajoutent de la granularité à l'analyse de presse en permettant d'étudier les positions implicites ou explicites d'un article ou d'un journal envers des entités politiques précises, mais nécessitent des corpus adaptés pour gérer des domaines différents. En outre, elles restent principalement adaptées à l'anglais, rendant difficile leur généralisation à d'autres langues et contextes culturels.

Dans le cadre de nos travaux, nous nous sommes concentrés principalement sur une caractérisation plus fine et plus généralisable du positionnement médiatique par l'analyse de sentiments ciblés envers des entités clés et leurs caractéristiques. L'utilisation de cette approche permet d'aller au-delà des méthodes basées uniquement sur la fréquence ou les mots connotés, en ajoutant un niveau d'explicabilité et de granularité qui tient compte des attributs des entités mentionnées. En intégrant des données comme celles de Wikidata et ParlGov pour enrichir les entités résolues, nous permettons une analyse multicritère et multilingue qui se distingue par sa flexibilité d'application à d'autres langues et potentiellement à d'autres domaines.

Système général pour l'analyse de la presse

Dans le chapitre 3, nous proposons un nouveau système pour l'analyse d'opinion dans la presse en se concentrant sur nos trois axes de recherche : la prise en compte de l'expression d'opinions, la généralisabilité vers d'autres langues et domaines, et une granularité élevée offrant une meilleure explicabilité. Ces trois aspects distinguent notre système des approches existantes et répondent aux principales limitations identifiées dans les travaux précédents, à savoir une couverture linguistique et culturelle restreinte, des analyses souvent superficielles, et un manque d'explicabilité [Hargreaves, 2014; Und, 2016].

Notre cadre d'analyse est composé de six étapes successives de traitement des données textuelles des articles : la collecte, le filtrage et déduplication, le découpage en phrases, la détection de mentions, la résolution d'entités nommées et enfin, la classification des sentiments. Ces étapes forment une chaîne de traitement modulaire conçue pour offrir une analyse fine et détaillée des opinions exprimées dans les articles, ce qui permet de mieux répondre à la question

de recherche **RQ1** : comment concevoir un système permettant une analyse fine des opinions tout en assurant une explicabilité accrue des résultats ? L'approche modulaire est également un atout pour atteindre la généralisation vers d'autres langues et domaines, explorée dans **RQ2**.

La première étape, la collecte des articles, repose sur différentes sources de données telles que le CommonCrawl, qui offre un vaste corpus multilingue et multimédia en accès libre. Dans notre étude, nous utilisons CCNews, un sous-ensemble de CommonCrawl, pour extraire des articles à partir d'un ensemble de domaines spécifiques aux pays ciblés, ce qui assure une couverture géographique et thématique adéquate. Le filtrage, deuxième étape, se compose d'une série de sous-étapes essentielles à la qualité et la pertinence de l'analyse : vérification de la langue des articles, déduplication pour éviter les biais introduits par les répétitions, et suppression des éléments non textuels ou des bannières publicitaires. Le filtrage repose sur un modèle d'identification de la langue (FastText)[Joulin et al., 2017], un outil de déduplication basé sur MinHashLSH[Leskovec et al., 2020], ainsi qu'une série de critères de pertinence basés sur des listes de mots-clés. Cette étape, en éliminant les contenus non pertinents et redondants, réduit les biais d'analyse et optimise le traitement des données.

Après filtrage, le découpage en phrases est effectué avec des outils comme le 'sentencizer' de Spacy, qui permet de segmenter les articles en unités de traitement cohérentes. Nous avons préféré Spacy à d'autres outils comme NLTK en raison de ses performances empiriques supérieures sur les tâches de segmentation dans de nombreuses langues [Honnibal and Montani, 2017].

Ensuite, nous procédons à la détection de mentions, une tâche essentielle à l'extraction des cibles, qui constituent la base de notre classification de sentiments ciblés. Nous utilisons des modèles de reconnaissance d'entités nommées (NER) multilingues, tels que GLiNER, capables d'extraire des entités spécifiques comme les noms de personnes, organisations et lieux dans plus de 100 langues [Zaratiana et al., 2023]. Pour les langues moins dotées, nous pourrions avoir recours à la traduction automatique avec des modèles récents tels que NLLB-200 [Team et al., 2022], offrant une couverture linguistique étendue qui permettrait de faire de la projection d'annotations d'entités nommées.

Il s'en suit d'une étape de résolution des entités nommées, aussi appelée dans notre cas 'wikification', qui associe les entités aux entrées de la base Wikidata, pour enrichir nos analyses ultérieures avec des attributs comme l'âge, le genre et l'affiliation politique des personnes citées [Cheng and Roth, 2013]. Pour la résolution multilingue, nous nous basons sur le modèle mGenre, un modèle autoregressif capable de lier des mentions dans le texte à des entités de Wikidata [De Cao et al., 2022]. Cette étape permet non seulement d'améliorer l'exactitude de l'analyse, mais aussi de la rendre plus explicable grâce à des informations externes enrichissantes sur les entités analysées.

Enfin, la classification de sentiments ciblée nous permet de déterminer si les mentions des entités identifiées sont associées à un sentiment positif, négatif ou neutre. Cette dernière étape constitue un des piliers de notre analyse de l'opinion, mais elle présente également le défi le plus important en termes de généralisabilité linguistique et contextuelle. La plupart des jeux de données de classification de sentiments sont disponibles uniquement en anglais, ce qui limite la robustesse de l'analyse pour les langues moins dotées. Bien que certains travaux aient explorés la traduction automatique pour appliquer les modèles de classification de sentiments en dehors de l'anglais [Bautin et al., 2021], cela reste un problème ouvert qui est traité en profondeur dans le chapitre 4 de la thèse.

Grâce une qualification des entités par une résolution utilisant Wikidata et Parlgov, nous réalisons une analyse multicritère des sentiments. Ce cadre d'analyse est particulièrement innovant dans la mesure où il combine des mesures objectives (fréquence des mentions) et subjectives (sentiments associés), offrant une approche hybride de la représentation médiatique dans la presse

écrite. Nous avons illustré l'efficacité de ce cadre au travers d'analyses qualitatives, notamment sur le paysage médiatique français, permettant une première validation empirique de sa pertinence au chapitre 5.

Analyse de sentiment ciblé dans un contexte politique

Dans le chapitre 4, nous présentons MAD-TSC, un jeu de données multilingues conçu pour la classification de sentiment ciblé pour la presse politique et disponible dans huit langues (anglais, espagnol, allemand, italien, français, portugais, néerlandais et roumain). MAD-TSC offre un cadre inédit d'expérimentation pour l'analyse de sentiment ciblé avec 5 110 mentions annotées alignées entre les langues. Nos données couvrent 286 sources d'information internationales, avec des traductions de haute qualité, au niveau des phrases, réalisées par des traducteurs professionnels. Le sentiment à l'égard des entités politiques est annoté du point de vue de l'auteur de l'article, ce qui ajoute une couche de complexité car l'expression des opinions des journalistes est souvent subtile et implicite [Dufraisse et al., 2023a].

La construction de ce jeu de données repose sur la sélection d'articles issus de VoxEurop, une source multilingue d'articles européens centrés sur des thématiques politiques et à fort impact sociétal. Après avoir collecté les articles, nous effectuons un traitement complet comprenant l'extraction des entités nommées, leur alignement entre les langues, et une résolution d'entités permettant d'obtenir des statistiques précises sur la diversité des figures politiques présentes dans notre jeu de données [De Cao et al., 2022]. Cela nous permet de représenter des personnalités et des partis politiques variés issus de contextes culturels et géographiques multiples, renforçant ainsi la diversité des perspectives couvertes.

Notre protocole d'annotation se fonde sur une échelle de cinq niveaux de sentiment, qui est ensuite consolidée en trois catégories (négatif, neutre, positif), comme cela est souvent le cas dans les jeux de données de classification de sentiment ciblé [Hamborg and Donnay, 2021]. Dans sa version anglaise, notre jeu de données se distingue par une complexité plus élevée par rapport aux jeux de données existants comme NewsMTSC, mesurée par des proxys comme la longueur des phrases et des métriques de lisibilité.

Nous avons exploré différentes architectures de modèles pour la classification de sentiment ciblé. Grâce à l'alignement multilingue de MAD-TSC, nous avons pu conduire des expériences de transfert cross-lingual qui explorent la capacité des modèles à généraliser d'une langue à une autre. En particulier, les modèles de traduction automatique ouverts tels que NLLB-200 [Team et al., 2022] montrent des performances très satisfaisantes pour le transfert de la tâche de classification de sentiment ciblé dans le domaine politique d'une langue à une autre. La traduction des phrases d'une langue cible vers l'anglais permet d'utiliser des modèles anglais préentraînés de plus haute qualité, atteignant des performances comparables à celles du modèle anglais sur les exemples anglais d'origine. Nos résultats soulignent également que les performances des modèles monolingues, souvent inférieures, sont dues à un inadéquation entre le pré-entraînement des modèles (qui n'est pas toujours orienté vers le domaine politique ou qui n'a pas bien convergé), ce qui limite leurs performances.

Nous avons également testé l'applicabilité de nos modèles, affinés sur notre jeu de données politique, dans d'autres domaines tels que la presse sportive ou les articles de santé. Cependant, nos résultats montrent que le transfert de modèle vers ces domaines se heurte à des obstacles, principalement en raison de différences sémantiques et contextuelles. Ce constat souligne l'importance d'un pré-entraînement et d'un ajustement spécifiques pour garantir la précision des modèles dans des domaines variés.

En conclusion, MAD-TSC permet d'effectuer une analyse de sentiment ciblée et automatique

dans huit langues pour le domaine politique, et montre une potentielle transférabilité à d'autres langues par traduction automatique. Cependant, l'extension vers d'autres domaines ou types d'entités nécessite des recherches complémentaires pour améliorer la transférabilité.

Analyse de la presse française

Dans le chapitre 5, nous présentons l'instanciation de notre système du chapitre 3 avec nos modèles du chapitre 4, et illustrons sa capacité d'utilisation dans le cadre d'une analyse de presse française, suivie d'une succincte comparaison avec le cas de la presse flamande.

Nous montrons les limitations inhérentes aux systèmes d'analyse actuels, qui se contentent généralement d'approches fréquentistes autour de métriques 'objectives', et mettons en évidence l'intérêt de la mesure du sentiment ciblé comme complémentaire aux analyses quantitatives basées sur la fréquence de mentions. À travers l'analyse d'une quarantaine de journaux français, nous explorons les distributions de mentions et de sentiments en fonction des attributs des entités résolues par Wikidata, ainsi que des classifications des partis politiques établies par la base de données ParlGov.

Ces attributs nous permettent d'analyser les sentiments et les fréquences de mentions selon des facteurs démographiques (comme l'âge et le genre) et politiques (notamment l'orientation sur l'axe gauche-droite). Ces analyses peuvent être agrégées à différents niveaux : au niveau des journaux individuels, des thématiques abordées, et de manière temporelle.

Nos résultats confirment certaines tendances identifiées par des spécialistes en sciences politiques concernant le paysage médiatique français [Cointet et al., 2021]. Elles nous permettent aussi d'en relever de nouvelles. Nous notons par exemple que les principaux journaux français en termes de tirage tendent à avoir une représentation équilibrée des partis politiques centraux. Notre analyse révèle aussi un biais généralisé des médias envers les politiciens affiliés aux partis de droite radicale, qui sont souvent perçus plus négativement, alors que les politiciens de la gauche radicale bénéficient d'une perception relativement plus positive. Nous constatons également une faible progression de la représentation des femmes en politique, malgré une quasi-parité en nombre au sein du parlement et du gouvernement. Ce phénomène de sous-représentation féminine est moins marqué dans les journaux locaux, suggérant une couverture plus inclusive à ce niveau.

Une distinction claire émerge également dans le sentiment moyen exprimé en fonction du type de journal. Les journaux locaux et économiques présentent un sentiment moyen plus positif envers les entités politiques, tandis que les journaux d'investigation et ceux à orientations plus radicales tendent à véhiculer des sentiments plus critiques. Cette différence dans les tonalités éditoriales pourrait indiquer une distinction dans les lignes éditoriales et les publics visés par ces différents types de publications.

Ces analyses sont particulièrement pertinentes pour de multiples parties prenantes, y compris les journaux eux-mêmes, les régulateurs, et les citoyens. Par exemple, les éditeurs et journalistes peuvent s'appuyer sur ces analyses pour ajuster la diversité de leur couverture médiatique. Les citoyens, de leurs côté, pourraient bénéficier d'une transparence accrue leur permettant d'apprécier plus objectivement les biais médiatiques potentiels des sources qu'ils consultent. Les régulateurs peuvent aussi y voir un outil précieux pour évaluer les déséquilibres dans la couverture des thématiques politiques et démographiques.

L'application de notre analyse au contexte de la presse flamande révèle des phénomènes similaires concernant les biais médiatiques, avec une perception générale plus positive dans les médias régionaux et une distribution de sentiments en fonction des orientations politiques des entités mentionnées. Comme dans le cas français, nous observons un biais négatif envers les partis de droite radicale, et des différences de représentations en termes de genre et d'âge. La

représentation des femmes politiques est similaire à celle observée en France, marquée par une sous-représentation malgré des avancées dans la parité politique.

Enfin, nous discutons de l'extension potentielle de ce cadre d'analyse à un contexte européen plus large. Cela ouvrirait des perspectives pour une analyse comparative des paysages médiatiques entre les pays, permettant de mieux comprendre les spécificités et les points communs des représentations politiques et démographiques dans les médias à l'échelle européenne.

Conclusion

Les moyens d'accès à l'information ont considérablement évolué au cours des dernières décennies, mettant fin à l'ère des médias de masse et favorisant la fragmentation et l'individualisation de la consommation des nouvelles, notamment via les réseaux sociaux. Cette évolution s'accompagne d'une polarisation accrue, bien que ses mécanismes exacts restent incertains. Elle soulève de nouveaux défis pour la stabilité des sociétés occidentales. Or, pour étudier et atténuer cette polarisation, il est crucial de pouvoir analyser automatiquement le contenu des informations que consomment les individus. Cette thèse répond à cet enjeu en développant un cadre d'analyse des opinions qui repose sur la classification de sentiments ciblés pour permettre une meilleure compréhension et diversification des opinions médiatiques accessibles.

Les méthodes existantes, centrées sur le contexte américain et anglophone, sont souvent limitées dans leur applicabilité à d'autres cultures et langues, et reposent fréquemment sur des fréquences de mentions ou des mesures de subjectivité très globales. Nos contributions se concentrent sur l'élaboration d'un cadre d'analyse fine des opinions (chapitre 3), la création d'un jeu de données multilingue aligné pour la classification de sentiments ciblés dans le contexte des articles de presse politique (chapitre 4), et la mise en application de cette analyse pour explorer les biais politiques et démographiques dans la presse française et flamande (chapitre 5).

Plusieurs directions se dessinent concernant de futures perspectives de recherche. Il serait d'abord intéressant d'étendre la classification de sentiments ciblés à d'autres langues, évaluer notamment l'applicabilité de l'extension automatique aux langues non indo-européennes, pour une éventuelle analyse globale. Ensuite, l'application de ce cadre à d'autres domaines comme la santé ou l'économie permettrait de tester sa généralisation. L'amélioration des modèles d'analyse de sentiment ciblé en termes de performances, notamment par l'intégration de dispositifs de détection d'incertitude, constituerait un autre axe important. Enfin, au-delà de l'aspect technique, cette recherche pourrait être intégrée dans des outils numériques (extensions web, agrégateurs de nouvelles) pour une utilisation concrète, encourageant une consommation d'information diversifiée et contribuant potentiellement à réduire la polarisation.

En conclusion, ces travaux posent les bases d'une compréhension fine de la couverture médiatique et offrent des pistes prometteuses pour une meilleure information des citoyens et une société plus résiliente face aux défis de la polarisation.

Contents

Chapter 1
Introduction **1**

1.1 General Context 2
1.2 Problem Definition 4
1.3 Contributions 6
1.4 Scientific Publications 7

Chapter 2
State of The Art and Related Literature

2.1 The Internet and the Mutations of the Informational Landscape 10
2.1.1 Transformations in Media Dynamics 10
2.1.2 Mutations in Individual Media Consumptions 11
2.2 Polarization: Origins, Consequences and Mitigation 12
2.2.1 Definition and Origins of Polarization 12
2.2.2 Internet and Social Media's Roles in Fostering Polarization 12
2.2.3 Consequences of Polarization 13
2.2.4 Mitigation of Polarization 14
2.3 Analysis of News 16
2.3.1 From Manual to Automated News Analysis 16
2.3.2 Recent Evolution of Natural Language Processing 17
2.3.3 Natural Language Processing for News Analysis 17
2.4 Characterizing Stances through Media Biases 24
2.4.1 Root Causes of Media Biases 24
2.4.2 Types of Media Biases 25
2.5 Automatic Detection of Media Biases 29
2.5.1 Framing Bias Detection 29
2.5.2 Stance Bias Detection 32
2.5.3 Phrasing Bias Detection 37

2.5.4 Social Category Biases . . . 39
2.6 Aspect-Based Sentiment Classification . . . 40
2.6.1 Definition . . . 40
2.6.2 Aspect-Term Extraction (ATE) . . . 40
2.6.3 Aspect-Based Sentiment Classification (ABSC) . . . 41
2.7 Proposed Approaches . . . 43
2.7.1 Distant Supervision using Twitter as Signal . . . 43
2.7.2 Stance Detection through Targeted-Sentiment Analysis . . . 44

Chapter 3
General News Analysis Framework

3.1 Introduction . . . 48
3.2 Existing News Analysis Frameworks . . . 48
3.3 A General Framework for News Analysis . . . 49
3.3.1 Source of News . . . 50
3.3.2 Filtering Step . . . 50
3.3.3 Sentencizer Step . . . 50
3.3.4 Mention Detection . . . 50
3.3.5 Entity-Linking . . . 51
3.3.6 Enrichment . . . 51
3.3.7 Sentiment Classification . . . 51
3.4 Instantiation for Political News Analysis . . . 51
3.4.1 Source of Articles . . . 52
3.4.2 Filtering Steps . . . 52
3.4.3 Entity-Linking . . . 52
3.4.4 Enrichment . . . 53
3.4.5 Sentiment Classification . . . 54
3.4.6 Analyses . . . 54
3.5 Conclusion . . . 54

Chapter 4
Targeted Sentiment Analysis for Political News

4.1 Introduction . . . 56
4.2 Related work . . . 57
4.3 Dataset Construction . . . 58
4.3.1 Data Sources . . . 58
4.3.2 Sample Selection . . . 59

4.3.3 Sample Annotation . . . 60
4.3.4 Label Aggregation . . . 60
4.3.5 MAD-TSC Dataset Characteristics . . . 62
4.4 Experiments . . . 63
4.4.1 Compared TSC Methods . . . 64
4.4.2 Dataset Splits and Metrics . . . 65
4.4.3 Experiments with MAD-TSC and NewsMTSC . . . 66
4.4.4 Experiments with Individual Languages . . . 66
4.4.5 Cross-lingual Experiments and Machine Translation . . . 67
4.4.6 Analysis of Example Complexity . . . 68
4.4.7 Ablation of Training Set . . . 68
4.4.8 Cross-Domain Experiments . . . 69
4.5 Limitations . . . 71
4.6 Conclusion . . . 71
4.6.1 Perspectives . . . 72

Chapter 5
French Press Analysis

5.1 Introduction . . . 74
5.2 Related Work . . . 75
5.3 Instantiation Details of the News Processing Framework . . . 76
5.4 Political News Analysis . . . 78
5.4.1 Outlet-oriented Analysis . . . 79
5.4.2 Topic Oriented Analysis . . . 82
5.4.3 Politician Oriented Analysis . . . 83
5.4.4 Gender Representation . . . 86
5.4.5 Age Representation . . . 87
5.4.6 Limitations . . . 87
5.4.7 Conlusion . . . 89
5.5 Analysis of Other Countries . . . 90
5.5.1 Multilingual Political Pre-Filtering Step . . . 90
5.5.2 Sentiment Inference . . . 92
5.5.3 Flemish Political News Analysis . . . 93
5.6 Conclusion . . . 97

Chapter 6
Conclusion

6.1 Summary and Contributions 100
6.2 Perspectives 101
6.2.1 Perspectives for Targeted Sentiment Classification in the News 101
6.2.2 Perspectives for the News Characterization Pipeline 103
6.2.3 Perspectives on Opinion and Stance Detection in the News 104

Appendix A
Social Media and Analysis of Polarization in the News

A.1 Introduction . . . 107
A.2 Automatic Analysis of Political News Based on Social Media . . . 107
A.3 Preliminary Results and Future Work . . . 108

Appendixs

Appendix B
Targeted Sentiment Analysis for Political News

B.1 Supplementary results . . . 111
B.2 Annotator Demographics . . . 111

Appendix C
French Press Analysis

C.1 Description of news sources . . . 115

Bibliography

List of Figures

1.1 Distribution of Democrats and Republicans along a 10-point liberal-conservative axis. The graph shows that both parties are becoming more ideologically consistent, with the median Democrat and Republican consistently drifting apart from 1994 to 2017. 4

2.1 Media biases can occur from the gathering of facts to the presentation of the news to the end user. Illustration from Hamborg [2023b] 24
2.2 The Asian Disease Experiment [Kahneman and Tversky, 1984] 27

3.1 Different steps of our General framework for news analysis 49
3.2 Instantiated framework for political news analysis 52
3.3 SPARQL query for Wikidata newspaper query 53

4.1 Page for an example of annotation. 61
4.2 Distribution of linkable entities per original language of source article. "other" merges rare source languages. Linking is done with Blink [Wu et al., 2020c]. . . . 62
4.3 Distribution of linkable entities from MAD-TSC (top) and NewsMTSC (bottom) per country. The country associated to each entry is the most frequent one from the Wikipedia article of the entity. 63
4.4 Frequent word clouds for MAD-TSC and NewsMTSC corpora (minimum frequency 20 occurrences). 63
4.5 $F1_m$ scores with training sets of variable size for NewsMTSC (NM) and MAD-TSC (MAD). Results are reported with monolingual pretraining and SPC. The rightmost points represent scores for the full dataset. 69

5.1 Number of articles per news outlet over the 20 most frequent outlets in the corpus 77
5.2 Number of articles per year in the corpus 77
5.3 Average sentiment (blue dots) for the most frequent 40 news outlets in the corpus. The dashed line represents the average sentiment for the entire corpus. Outlet frequency decreases from left to right. 78
5.4 Distribution of mentions of political orientations in news outlets. 80
5.5 Deviation of the sentiment associated with major political orientations from the average sentiment of each source. Average sentiment of linked politicians is indicated in parentheses for each source. 80
5.6 Mentions of political orientations across time. 81
5.8 Distribution of mentions of political orientations for ten impactful political topics. 82
5.7 Sentiment associated with political orientations in time. 82

5.9 Deviation of the sentiment for major political orientations from the average sentiment of each topic. Average topic sentiment is given in parentheses. 83
5.10 Mentions of most frequently mentioned politicians in time. 85
5.11 Sentiment associated with the most frequently mentioned politicians in time. . . 85
5.12 Percentage of gender mentions in the top-10 news outlets and on average. 87
5.13 Percentage of gender mentions per year. 87
5.14 Mean sentiment classification scores by gender in the top-10 news outlets and on average. 88
5.15 Average age of the politicians mentioned at the time of publication per news outlet. 88
5.16 Number of articles per class and country in CCNews 91
5.17 Number of articles per language and class after balancing in the classification dataset 92
5.18 Average sentiment (blue dots) for the 24 most frequent news outlets in the corpus sorted by decreasing number of articles; the dashed line represents the average sentiment for the entire corpus. 94
5.19 Deviation of the sentiment associated with major political orientations from the average sentiment of each source. Average sentiment of linked politicians is indicated in parentheses for each source. 95
5.20 Gender mention proportions across various media outlets. The bars indicate the percentage of mentions per gender for each media outlet. 96
5.21 Gender representation across various media outlets from 2019 to 2023. The bars show the percentage of mentions per gender each year, including an average (AVG) column. 96
5.22 Average age of politicians mentioned per source. 97

A.1 Newspapers Distribution over Political Space 109

Chapter 1

Introduction

1.1 General Context

Over recent decades, technological advances in telecommunications have significantly changed the mechanisms by which information is disseminated in society and how the public interacts with it, switching from a mass media communication controlled by a few established actors, to a more competitive field with diverse sources of information [Hargreaves, 2014; Und, 2016; Van Aelst and Strömbäck, 2017]. In Western societies, where the freedom of the press and freedom of speech have historically been upheld, some of the changes brought about by this mutation may be detrimental to the well-functioning and stability of democracy. Several scholars refer to this situation as a time of "information crisis" [Beckett and Livingstone, 2018; Haider and Sundin, 2022].

This information crisis is characterized by several features. One is the high number of available media, which has fragmented a historically cohesive audience. The dramatic increase in this number is due to the advent of the Internet, which, due to lower entry costs, has moved us from a low-choice to a high-choice media landscape, marking the end of the concentration of mass media communication [Haider and Sundin, 2022].

This audience fragmentation has been accompanied by a notable increase in the personalization of news consumption. This trend has led to a more individualized experience of current events, as consumers are able to select news sources that closely match their personal preferences and viewpoints [Haider and Sundin, 2022].

As competition intensified with this fragmentation, news media increased their preference for cheaper sensational and emotional content over more traditional forms of in-depth or investigative journalism. This shift toward "infotainment" has been noted by several scholars and is seen as undermining the quality and depth of public discourse [Berry and Sobieraj, 2013; Jones, 2012; Boukes, 2019a; Jenks, 2022; Haider and Sundin, 2022]. This fierce competition has also fostered editorial differentiation among news outlets, with many choosing to specialize in niche or partisan perspectives to appeal to specific audiences [Mancini, 2013]. The period is also characterized by a shift to partial or complete advertising-based economic models for news outlets, which can directly or indirectly influence their editorial decisions [Gilens and Hertzman, 2000; Besley and Prat, 2006; Brighton and Foy, 2007]. We're also seeing an increasingly blurred line between professional journalists and bloggers, with many people not distinguishing between the two. This blurring of lines raises questions about the mutation of the field, questions about accountability and the role of expertise [Pintak and Fouda, 2009; Fenton, 2010; Witschge and Nygren, 2009; Lasorsa et al., 2012]. Finally, the erosion of public trust in official institutions, including the news media, is also a challenge of our period. This loss of trust complicates the efforts to address this information crisis, and leads many individuals to doubt and struggle to find truthful information [Haider and Sundin, 2022; Park and Lee, 2023].

An examination of the history of the dissemination of news information reveals that these mutations occurred in a very short period of time compared to previous revolutions [Hargreaves, 2014; Und, 2016]. While Gutenberg's invention of the printing press in the 15th century marked a revolutionary leap in the dissemination and permanence of information, printed newspapers really developed in the 17th and 18th centuries. They were thoroughly regulated by most governments until the 19th century [Und, 2016], before becoming a cornerstone of society, ensuring institutional transparency, informing the public, and constituting the first form of mass communication [Chaffee and Metzger, 2001]. The spread of the radio in the early 20th century, and its free and rapid dissemination of information, forced the newspapers to differentiate their content beyond simple news coverage [Hargreaves, 2014; Und, 2016]. The spread of television in the mid-20th

century made information more appealing through its visual elements and increased this pressure. As the number of channels increased, notably through satellite television, we witnessed the beginnings of the fragmentation of the media landscape and the individualization of information consumption [Katz, 1996]. The development of the 24-hour news channels, marked a new leap in the news reporting process by prioritizing immediacy and continuous updates, often at the expense of in-depth analysis [Lewis et al., 2005].

Although the advent of the Internet revolutionized the access to information by providing instant and on-demand access to news, it fragmented in unprecedented ways the media landscape by multiplying the number of accessible news sources. This fragmentation enabled individuals to be more selective in their information consumption leading to the individualization of the media consumption experiences. But this "selective exposure" - the natural tendency of people to seek out information that is consistent with their own beliefs [Freedman and Sears, 1965; Aronson, 1969] - got notably amplified by the invention of smartphones and the rise of social media. Social media transformed the relationship to information from curated and editorialized forms to automated and incidental receptions, notably through the concept of recommendation feeds [Park and Lee, 2023]. Confronted with a vast flow of content, social media platforms resort to personalization algorithms to offer pertinent recommendations to the end-user based on the similarity of new content with their historical consumption. These recommendation strategies have led to the phenomenon of "filter bubble" - a situation in which an Internet user is only exposed to information and opinions that are consistent with and reinforce their own beliefs - coined by Pariser [2011]. This rise in social media use, facilitated by the penetration of smartphones, has been accompanied by a rise in misinformation, polarization, and an erosion of trust in the institutions. A study conducted by the Pew Research Center shows a significant increase in ideological polarization from 1994 to 2017 in the U.S[2]. Using a consistent series of 10 questions, they assess individuals' tendencies toward liberal or conservative viewpoints. The results, as seen in Figure 1.1, indicate that people's opinions on various issues are becoming more closely aligned with their party's positions, leading to a growing distance between the leanings of the median Democrat and the median Republican.

While the polarization of society - the growing ideological divide, with the radicalization of ideas and the reduction of tolerance for opposing views - has been linked to the emergence of "filter bubbles", the scientific consensus on the real underlying causal mechanisms between these bubbles and polarization has not been reached yet [Bakshy et al., 2015; Flaxman et al., 2016; Törnberg, 2022]. The solution to this problem is threefold: regulatory bodies could force platforms to increase the transparency of their algorithms and impose guardrails on their functioning; media literacy initiatives could be enforced to sensitize the public to the stakes of information and develop skills to critically analyze media content; finally, automatic mitigation by modifying recommendation algorithms to increase diversity could help reduce polarization. On the mitigation side, the seemingly sensible solution of adding randomness to recommendations to increase diversity could actually be counterproductive, as presenting strongly opposed viewpoints has been shown to actually exacerbate polarization [Bail et al., 2018], and this would also assume that randomly selected articles represent a well-balanced mix of opinions.

In this context, it seems that the ability to understand news in a fine-grained way could allow researchers to better study polarization. According to the current understanding of its candidate mechanisms, it could even help mitigate it through intelligent diversification schemes in recommendations that would allow users to get a more complete picture of events and opinions [Helberger et al., 2018]. Developing this ability is even more important as the tendency

[2]https://www.pewresearch.org/politics/2014/06/12/appendix-a-the-ideological-consistency-scale/

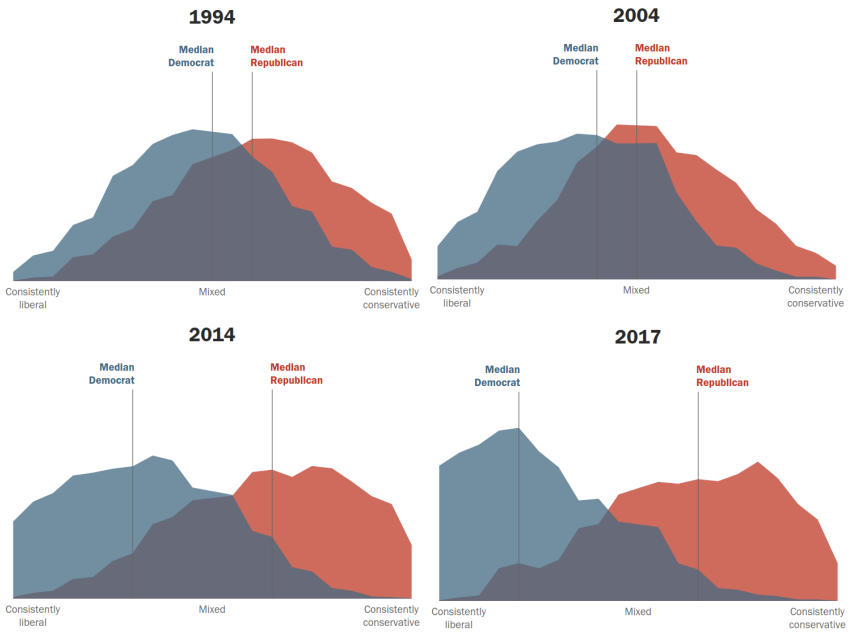


Figure 1.1: Distribution of Democrats and Republicans along a 10-point liberal-conservative axis. The graph shows that both parties are becoming more ideologically consistent, with the median Democrat and Republican consistently drifting apart from 1994 to 2017.

to use social media as a gateway to access information is growing worldwide, especially among younger demographics, and is already the primary access to news in some countries [Newman et al., 2023]. It is also important because of the negative impact that polarization can have on the proper functioning of democracy, from legislative gridlocks [Jones, 2001] to the dehumanization of political opponents [Martherus et al., 2021].

1.2 Problem Definition

The analysis and characterization of news is a research task that predates the emergence of computational analysis in social sciences [Entman, 1989; Iyengar and Simon, 1993]. The development of automatic methods for news characterization is an important research objective, in order to be able to process the large amount of news published daily and to try to quantitatively understand the characteristics of the media landscape, without resorting to the prohibitive cost and work that manual annotation by experts would require. Although the gap of complexity between the phenomena that can be identified manually and those that can be identified automatically is still wide, recent advances in Natural Language Processing (NLP), especially since the advent of the transformer architecture [Vaswani et al., 2017], allow the derivation of increasingly complex automated analysis methods for the characterization of news.

These recent advances in NLP motivate the research topic of this thesis, which is funded in the context of the ANR BOOM[3] research project; which aims to develop new algorithms for news recommendation systems to diversify the opinionated content presented to the end user and open filter bubbles. The ability to perform this diversification implies the ability to characterize opinions contained in articles. In the literature, opinion characterization of news has evolved

[3]https://anr.fr/Projet-ANR-20-CE23-0024

around the detection of proxy measures, specifically the detection of biases that are indicative of leanings and opinions. The terminology and definitions of bias may vary, but we can identify two primary interests in the detection of content-based bias:

1. **What is reported** – This focuses on the selection of events and details within those events
2. **How it is reported** – This examines the portrayal of events, including the identification of tone and bias, and the analysis of the narrative stance.

In the first approach, some researchers focus on analyzing the distribution of topics through topic modeling, while others delve into the concept of "framing detection". Although the concept of framing is broad and encompasses more types of biases in social sciences, in NLP research, it primarily refers to the identification of specific subtopics within an article. By comparing the distribution of these topics and subtopics across different sources, one can infer potential biases in content selection or framing by the media outlet. While these biases may be indicative of underlying opinions, articles discussing the same topics or subtopics may actually present opposing views. Extending automated framing analysis to identify competing narratives around an event could provide more detailed insight, but remains an open research question.

The second approach focuses on identifying the stance of the articles using classifiers or lexicons that indicate bias at the level of the whole text. While sometimes effective, these methods often lack explanability or rely on proxy measures that are too imprecise to reliably determine stance. This leads us to a primary question that this work seeks to answer:

> **RQ1**: How to design a framework that allows for fine-grained analysis of opinions expressed in text with greater explanability?

A second limitation of existing methods is their lack of generalizability across languages and domains. While the advent of the transformer architecture has shifted natural language processing from lexicon, rule-based, and probabilistic modeling to a "pretraining/finetuning" paradigm, models still require substantial finetuning data to perform downstream tasks. To address this challenge, research is increasingly focusing on zero-shot, few-shot, in-context, and transfer learning scenarios. In the context of stance detection, most techniques and datasets cover the English language, focus on the U.S. news landscape, and are bound to a specific time frame. These methods often aim to categorize articles based on a left-right or pro/anti stance. While this approach fits well with binary issues, it struggles to capture the complexity of positions across the political spectrum, or issues with more diverse sets of positions, and it does not account for shifts in cultural and temporal understandings of political orientations. Moreover, these strategies often do not extend to non-political domains, and adapting datasets or models to different languages is a significant challenge. Thus the second question we wish to address in this thesis is:

> **RQ2**: How generalizable can we make this opinion detection pipeline to facilitate transfer to other languages and other domains?

Once we have collected the opinions expressed in the articles, the next question is how these opinions are aggregated across articles to determine the stance of the outlet. Since our ultimate goal is to characterize political attitudes in news outlets, the final research question we wish to answer is:

RQ3: How can we use this framework to characterize the political leanings of news outlets?

1.3 Contributions

This work has resulted in three main contributions that correspond to Chapters 3, 4 and 5 respectively:

- **First contribution: Development of a framework for fine-grained opinion analysis in text.** Our approach addresses the challenge of identifying opinions expressed towards particular entities within a text, framing it as a sentiment analysis task. Advancing beyond current methodologies, our framework integrates Named Entity Recognition (NER) with wikification (linking entities to Wikipedia pages). Wikification allows us to aggregate the results while being robust to the different name variations of entity mentions. This framework not only enhances the explainability and detail of the analyzed opinions, but also allows to leverage Wikipedia's entity attributes, enriching the interpretation of the obtained sentiment scores with characteristics such as demographics.

- **Second contribution: Creation of a multilingual dataset for targeted sentiment classification in the news.** Prior to this contribution, the sole dataset available for targeted sentiment classification in news, created by Hamborg and Donnay [2021], had multiple drawbacks. It was exclusively in English, focused on the U.S. political news, and exhibited a significant bias in entity distribution. These challenges, alongside the absence of datasets for assessing language and domain transferability, hindered our ability to evaluate the dataset's adequacy for French political news analysis. Therefore, we created a new dataset specifically designed for targeted sentiment classification in the context of European political news. Using the manually translated news from the VoxEurop initiative, we compiled a dataset of aligned, annotated sentences in eight European languages. This dataset demonstrated that performance in individual languages is highly dependent on the quality of available pre-trained language models. It also confirmed the feasibility of using machine translation from low-resource languages into English as a suitable strategy for targeted sentiment classification. Our results indicate that manual and automatic translations achieve comparable results in this domain. Furthermore, our dataset presents a wider variety of entities and more complex sentence structures compared to those found in Hamborg and Donnay [2021]'s dataset.

- **Third contribution: Analysis of the French news landscape and extension to under-resourced languages.** Most news analytical tools focus on objective metrics or coarse subjective metrics, notably through general sentiment analysis or political leaning classification, leading to limited granularity and explanability. Our approach overcomes these challenges by utilizing the opinion analysis framework we defined as first contribution and training models on our targeted-sentiment classification dataset. We enhance Wikification attributes by incorporating data from the "ParlGov" database, which provides political leaning scores for political parties for most OECD countries. This allow us to aggregate results of sentiment analysis in a bottom-up fashion to analyze biases of news sources, while maintaining explainability. Through an extensive analysis of the French press,

we illustrate the valuable insights gained from our automated analysis. We also present the application of this framework to Flemmish news analysis, and show how this framework, unlike leaning-based ones, could be extended to other low-resource languages.

1.4 Scientific Publications

This work resulted in the following publications:

- **Evan Dufraisse**, Célina Treuillier, Armelle Brun, Julien Tourille, Sylvain Castagnos, and Adrian Popescu. (2022). *Don't Burst Blindly: For a Better Use of Natural Language Processing to Fight Opinion Bubbles in News Recommendations.* In Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences, pages 79–85, Marseille, France. European Language Resources Association. [Dufraisse et al., 2022]

- **Evan Dufraisse**, Adrian Popescu, Julien Tourille, Armelle Brun and Jerome Deshayes. (2023). *MAD-TSC: A Multilingual Aligned News Dataset for Target-dependent Sentiment Classification.* In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8286–8305 [Dufraisse et al., 2023a]

- Célina Treuillier, Sylvain Castagnos, **Evan Dufraisse** and Armelle Brun. (2022). *Being diverse is not enough: Rethinking diversity evaluation to meet challenges of news recommender systems.* In UMAP '22 Adjunct: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, pages 222–233 [Treuillier et al., 2022]

- **Evan Dufraisse**, Julien Tourille, Armelle Brun and Adrian Popescu. (2023). *Analysis of Polarization in the News.* In Workshop Infox-sur-Seine 2023 [Dufraisse et al., 2023b]

Chapter 2

State of The Art and Related Literature

Contents

2.1 The Internet and the Mutations of the Informational Landscape . . 10
2.1.1 Transformations in Media Dynamics 10
2.1.2 Mutations in Individual Media Consumptions 11
2.2 Polarization: Origins, Consequences and Mitigation 12
2.2.1 Definition and Origins of Polarization 12
2.2.2 Internet and Social Media's Roles in Fostering Polarization 12
2.2.3 Consequences of Polarization 13
2.2.4 Mitigation of Polarization 14
2.3 Analysis of News 16
2.3.1 From Manual to Automated News Analysis 16
2.3.2 Recent Evolution of Natural Language Processing 17
2.3.3 Natural Language Processing for News Analysis 17
2.4 Characterizing Stances through Media Biases 24
2.4.1 Root Causes of Media Biases 24
2.4.2 Types of Media Biases 25
2.5 Automatic Detection of Media Biases 29
2.5.1 Framing Bias Detection 29
2.5.2 Stance Bias Detection 32
2.5.3 Phrasing Bias Detection 37
2.5.4 Social Category Biases 39
2.6 Aspect-Based Sentiment Classification 40
2.6.1 Definition 40
2.6.2 Aspect-Term Extraction (ATE) 40
2.6.3 Aspect-Based Sentiment Classification (ABSC) 41
2.7 Proposed Approaches 43
2.7.1 Distant Supervision using Twitter as Signal 43
2.7.2 Stance Detection through Targeted-Sentiment Analysis 44

This chapter is organized as follows: Section 2.1 explains the changes in the news landscape driven by information technology innovations and the challenges they pose. Section 2.2 defines polarization, explores its causes, and suggests strategies for understanding or mitigating it. Section 2.3 describes the evolution of news analysis from manual methods to sophisticated automated analysis made possible by increasingly refined algorithms. Section 2.4 presents media biases, their emergence, and their use as proxy indicators for detecting attitudes in the news. Section 2.5 reviews the literature on automatic media bias detection, its various methods, and their limitations. Section 2.6 discusses Aspect-Based Sentiment Classification, one of the approaches to identifying stance bias in news articles. Finally, section 2.7 presents the research challenges and questions addressed in this thesis.

2.1 The Internet and the Mutations of the Informational Landscape

As seen in Chapter 1, the technological advances that paved the 20th century have significantly disrupted the flow of news information and the way the public interacts with it. But in many ways, the Internet, and the era it has brought with it, cannot be compared to the earlier transitions that historically took place from newspapers to radio, radio to television, or television to cable, in that it has drastically changed the environment of news production as well as the news consumption habits of citizens [Beckett and Livingstone, 2018; Haider and Sundin, 2022]. At the time of writing, the increasing public availability of Large Language Models (LLM) chatbots might announce another revolutionary prospect in the future of humans interactions with information through the use of AI tools. It also comes with new concerns, such as over-reliance on technology [Nashwan and Jaradat, 2023], the correctness and fairness of model knowledge [Pan et al., 2023; Si et al., 2023; Goyal et al., 2023], its impact on recipients' perceptions, and the potential re-concentration of information dissemination in the hands of a few private actors [Perrigo, 2024].

2.1.1 Transformations in Media Dynamics

Anticipated Disruptions and Audience Fragmentation

The disruptions brought by the deployment of the Internet were already anticipated by scholars in the 1990s, who extrapolated the potential consequences of the spread of this new information infrastructure based on their knowledge of their respective fields and their observations of the effects of the previous introduction of cable television [Katz, 1996; Hallin, 1992; Chaffee and Metzger, 2001; Sunstein, 2001]. Most of their questions at that time happen to be more relevant than ever. Katz [1996], Chaffee and Metzger [2001] foresaw the end of mass media communication, that is, the end of the concentration of the power of news dissemination in the hands of a few historical actors, and the beginning of the era we know, one of audience fragmentation among an increasing number of media [Webster and Ksiazek, 2012; Mancini, 2013]. In this environment of abundant choice and easily accessible free news, combined with a high proportion of citizens unwilling to pay for news [Newman et al., 2023], media outlets are being forced to differentiate themselves by focusing on niche markets and adopting strongly opinionated content [Mancini, 2013].

Emergence of New Media Genres and Impact of Social Media

This shift has led to the emergence of extreme new genres, such as outrage journalism [Berry and Sobieraj, 2013], and a preference for more profitable formats, such as breaking news or infotainment [Jones, 2012; Boukes, 2019a; Jenks, 2022], often at the expense of in-depth analysis or investigative reporting, and thus good information of citizens [Haider and Sundin, 2022]. This has also led media organizations to become increasingly dependent on advertising revenue, which affected media independence, with advertisers able to directly or indirectly influence editorial decisions [Gilens and Hertzman, 2000; Besley and Prat, 2006; Brighton and Foy, 2007]. This fragmentation has been pushed even further with the emergence of social media platforms, that enabled the creation of centralized virtual spaces where everyone could express their opinions. This has notably led to the rise of influencers and virality [Mills, 2012], diverging from conventional celebrity by emerging organically from user interactions [Mic; Susarla et al., 2016]. It has also weakened the consumption of editorialized information by enabling further personalization, changing the way individuals access information.

2.1.2 Mutations in Individual Media Consumptions

Evolution of Journalism and its Perception

At the level of the individuals, the development of new actors has blurred the line between what constitutes journalism, with an increasing numbers of actors, who are not trained journalists, producing news articles and challenging this distinction [Pintak and Fouda, 2009; Fenton, 2010; Witschge and Nygren, 2009; Lasorsa et al., 2012]. As Sunstein envisioned, the high-choice media environment that resulted from this fragmentation allowed people to choose to consume only what they were interested in, leading to extreme phenomena such as "news avoidance" altogether [Newman et al., 2023]. Recommendation systems intensify this filtering effect by crudely endorsing the content selection roles, once held by editors, on online platforms like social media. Given the vast amount of content flowing into these platforms, recommendation systems are necessary to help users filter out content of interest [Raza and Ding, 2021]. Far from the ideals that some might have expected to emerge from this new information communication support that is the Internet, the availability of information does not imply its accessibility, nor the interest of users in accessing it. Overall, the use of the Internet has translated into an unchanging knowledge of current events by Americans when compared to the pre-Internet era [Kohut et al., 2007], and a potential widening of the knowledge gap between those interested in politics and those who are not [Boukes, 2019b].

Recommendation Systems and their Impact on News Consumption

The increased reliance on recommendation systems as gateways to information rather than intermediary editorial institutions has led to a strong individualization of the news consumption experience, sometimes leading people into a narrowed view of current events and a narrower understanding of the diversity of existing opinions in society [Pariser, 2011; Beam, 2014; Heitz et al., 2022]. As early as 1996, Katz [1996] envisioned that a high degree of personalization of news content could undermine the cohesion of nations by suppressing the common and cohesive meeting ground historically provided by mass media. This personalization culminated with the advent of social media, with the explicit term "filter bubble" coined by Pariser [2011], refined to "opinion bubble" when referring to opinion-rich content, and characterizing a state in which the user encounters only information and opinions that match their beliefs and reinforce them.

These opinion bubbles are suspected to be involved in the political fracturing of societies as we experience it today through a phenomenon known as "polarization"; the U.S. illustrates this with the steady increase in ideological consistency among its population that divides its society on every issue along the liberal-conservative axis [Dimock et al., p18-24].

2.2 Polarization: Origins, Consequences and Mitigation

2.2.1 Definition and Origins of Polarization

In political and social science, polarization often refers to the growing ideological divide in society, with the radicalization of ideas and the reduction of tolerance for opposing views. The underlying mechanisms behind polarization can be understood through considerations of social and cognitive psychology. Its characterization predates the spread of the Internet. In 1999, Sunstein [1999] described how "group polarization" occurs in various real-world social contexts, that is, how like-minded individuals tend to reinforce each other's pre-existing beliefs through their interactions. In this regard, he noted in particular how the expansion of the Internet, by facilitating digital encounters between like-minded individuals in online forums, could actually promote the radicalization of those individuals. Polarization is also driven by cognitive biases such as "selective exposure" - the tendency to seek out information that is consistent with one's beliefs while avoiding contradictory information to avoid cognitive dissonance [Freedman and Sears, 1965; Aronson, 1969]; and "confirmation bias" - the human tendency to interpret information as confirming one's existing beliefs [Klayman, 1995]. Polarization can also exacerbate other cognitive biases, such as the "false consensus" effect, where individuals overestimate how widespread their views are in society. [Ross et al., 1977; Luzsa and Mayr, 2021].

2.2.2 Internet and Social Media's Roles in Fostering Polarization

The emergence of the Internet and social media platforms has played a pivotal role in shaping the contemporary crisis surrounding the dissemination and management of information. While the Internet has markedly enhanced the accessibility of information, it has simultaneously led to an explosion in the number of information sources. This surge has placed a significant responsibility on individual users to discerningly select and assess the information they encounter. Such a high-choice environment fosters the manifestation of psychological and cognitive biases we described in subsection 2.2.1, with users favoring the selection of content they like and agree with.

The Necessity and Limitations of Recommendation Algorithms

This amplification of content selection is two-sided: the overwhelming volume of content has made it difficult for users to find relevant information and has led to the rise of search engines and recommendation algorithms as key navigational aids on the Internet. These algorithms, notably those used by social media platforms, are fully automated and tailored to an individual's past consumption patterns and their similarity to the consumption patterns of other users, following the principle of homophily [Flaxman et al., 2016; Hamborg et al., 2019b], and reinforcing the phenomenon of "selective exposure". This recommendation strategy, called "collaborative filtering" [Schafer et al., 2007], addresses the problem of tailoring the content offfered to the user while remaining agnostic about the nature of the content itself. The consumption of news on social media is thus automatically selected and incidental, as users do not actively seek out the

news themselves [Park and Lee, 2023]. This similarity-based recommendation has been shown to homogenize content consumption, which in the context of political news means that confrontation with opposing viewpoints is reduced, leading to the formation of opinion bubbles [Pariser, 2011], sometimes called "echo chambers", a metaphor used to portray how users only interact with their own ideas.

Citizen Autonomy and the Emergence of Social Media as Information Gateways

These mutations have increased the autonomy of citizens and further individualized their experience in the consumption of information [Mossberger et al., 2007]. The fading of traditional editorial roles and the growing ambiguity in defining authentic journalism [Pintak and Fouda, 2009; Fenton, 2010; Witschge and Nygren, 2009; Lasorsa et al., 2012] have also led to the rise of misinformation, disinformation (commonly called fake news), and even the use of information warfare among geopolitical competing entities [Lin, 2019]. In this evolving landscape, the reliance on social media as the primary gateway for accessing news has been progressively intensifying. This shift is evidenced by a rise in the proportion of individuals utilizing social media for news, increasing from 23% in 2018 to 30% in 2023 across a wide collective of representative countries [Newman et al., 2023]. Concurrently, Newman et al. [2023] shows that there has been a decline in the direct usage of news platforms, from 32% to 22%. This trend is especially notable among the younger demographic, with 43% of individuals aged 18-24 turning to social media as their principal news source. The phenomenon is even more pronounced in emerging countries, where for a significant segment of the population (exceeding 50% in nations like Thailand, Chile, and the Philippines), social media serves as the sole access to news. While traditional news media articles remain the foremost source of information for the majority, the primary access point has transitioned from media outlet front pages to social media platforms [Wilding, 2018].

Assessing Social Media's Role in Polarization

The increasing popularity of social media has led researchers to try to understand their impact on polarization, with the most common explanation being the creation of opinion bubbles [Bakshy et al., 2015; Flaxman et al., 2016]. More recently, other causal mechanisms come to challenge this assumption. Törnberg [2022] argue that the intensification of polarization on social media is, on the contrary, due to users facing strong opposite opinions to theirs, without the phenomenon of "local heterogeneity" that emerge from social interactions in real life. In real life interactions, people might not agree on every subject with their peers, but through their sociabilization they will also acknowledge points of agreement, making them more tolerant of their differences. It's this lack of local heterogeneity, this lack of perception of some degree of agreement in others' opinions, that could lead to stronger stances and ultimately polarization. This study is partly on par with Bail et al. [2018], that shows that the exposure to opposed views on social media can lead to an increase in polarization. On the opposite end, other research seem to suggest social media would have low to no impact on the actual polarization of its users, which seem to vary depending on the political context and the characteristics of the user (such as being politicized or not) [Kubin and Von Sikorski, 2021].

2.2.3 Consequences of Polarization

The impacts of polarization are complex. While it can promote political engagement and higher electoral participation [Argyle and Pope, 2022; Wagner, 2021], its effects on the functioning of

democracy are largely negative. Polarization promotes partisan cohesion, which has led to the centralization of power in the U.S. institutions [Lee, 2015]; contributes to legislative gridlock [Jones, 2001]; provokes the active avoidance of interaction with political opponents [Frimer et al., 2017]; and can go as far as dehumanizing those opponents [Martherus et al., 2021]. In this context, regardless of the real underlying causal mechanisms of polarization, be it opinion bubbles, lack of local heterogeneity or external factors, its mitigation is desirable as a way to improve the stability of our societies and ensure the well-functioning of democracy through healthy debate.

2.2.4 Mitigation of Polarization

The previous subsection clarified that the root causes of polarization are still not fully understood, but considering the current hypotheses on its underlying mechanisms, three avenues can be identified to tackle the phenomenon: Increasing the transparency of platforms through regulation, increasing the public awareness on cognitive biases and the role of media, and designing tools to help users develop healthy media consumption.

Regulation of Platforms

Government institutions can help mitigate polarization and allow researchers to understand its mechanisms through specific regulatory measures. Such measures include:

- The regulation of targeted political advertising on digital platforms, to foster transparency and prevent misinformation.
  - ◇ The “EU Code of Practice on Disinformation” (EU, 2018), encourages platform to auto-regulate their political advertising.
  - ◇ The “For The People Act” (USA, 2021), requires platforms to maintain public database of their political ads purchases, and to clearly disclose the source of funding on those ads.
  - ◇ The “Digital Service Act” (EU, 2022) requires platforms to clearly indicate advertising content, its sponsor, as well as the targeting criteria.
- Demanding oversight and transparency in the process by which platforms recommend content.
  - ◇ The “Digital Service Act” (EU, 2022) requires platforms to add at least one recommendation option that doesn’t rely on profiling, to be more transparent about their recommendation system process, and to be subject to an independent external audit.
- Holding platforms accountable for their content and requiring them to moderate misinformation.
  - ◇ The “NetzDG” law (Germany, 2017), requires platforms to remove "clearly illegal" content within 24 hours and all illegal content within 7 days of it being posted.
  - ◇ The “Loi infox” (France, 2018), requires platforms to cooperate and deploy a dispositive allowing users to report fake news.
  - ◇ In the US the Section 230 of the “Section 230 of the Communications Decency Act” has been subjected to several amendments to increasingly hold platforms more accountable.
  - ◇ Finally, the “Digital Service Act” (EU, 2022) also forces platforms to have effective processes for reporting and suppressing misinformation.

Media Literacy

Media literacy encompasses the skills necessary for critical interaction with media content, enabling people to analyze and evaluate the sources of information and the information itself. It includes: understanding the media production processes, understanding the role and impact of media on society, evaluating media messages, recognizing their biases and purposes. Overall, it's the development of the skills necessary to develop a critical and healthy interaction with informational content. It has been found to be an effective way to help people identify misinformation and disinformation, and several researchers recommend teaching it to children [Potter, 2013; Jones-Jang et al., 2021].

Several state-led initiatives can be cited:

- In Europe, the "The Media For All" and the "European Media Literacy Week" initiatives
- In the US, the "Digital Citizenship and Media Literacy Act" encourages the development of media literacy programs in education curricula.

While media literacy initiatives primarily aim to help people identify disinformation and misinformation, they can also encourage the use of diverse media sources for a more balanced perspective [Van Der Meer and Hameleers, 2021].

Digital Tools

In addition to regulatory measures and media literacy initiatives, there are digital tools that can help users develop or maintain healthy news consumption habits. These tools can take the form of passive support, such as "news recommender systems", or more active approaches, such as "news aggregators", where users can manually refine what they want to see.

- **Web Extensions**: Web extensions help users throughout their navigation by indicating the reliability of news sites, the trustworthiness of the articles, their neutrality, or their factuality [Vargas et al., 2023]. They can use external databases [NewsGuard, 2018; GroundNews, 2020], machine learning models, or heuristics to provide guidance. In addition to semantic and stylistic cues, they can rely on other information such as traffic data [Chalkiadakis et al., 2021], crowd-sourced initiatives [AllSides, 2012], social networks [Zhao et al., 2020], trusted information sources [Van Zandt, 2015; Wang and Yu, 2021], or fact-checking sites [Botnevik et al., 2020]. Recent initiatives also address the problem of deep fake detection [Teyssou et al., 2017].
- **News Aggregators**: News aggregators, such as AllSides [AllSides, 2012] through crowdsourcing; or Verity [Foundation, 2020] through machine learning and knowledge bases, seek to aggregate articles from different outlets to form a comprehensive view of the press, empowering users to choose what they want to read and making it easier for them to break out of their bubble if they choose to.
- **News Recommender Systems**: News Recommender Systems (NRS) are designed to help readers find relevant information among the large amount of news available [Raza and Ding, 2021]. Diversifying the content offered to the user tends to degrade the classical accuracy metric used to evaluate these systems, and is referred to as the accuracy-diversity dilemma [Zhou et al., 2010]. New metrics and algorithms for diversification in terms of general sentiment and topics have been elaborated [Park et al., 2009; Tintarev et al., 2018;

Wu et al., 2020a; Raza and Ding, 2020], though these diversification metrics stay insufficient to effectively promote opinion diversification.

Digital tools complement media literacy efforts to improve healthy news consumption by end users. However, the precision of their recommendations and insights relies on the development of robust Natural Language Processing (NLP) techniques. In the following section, we explore the various areas of NLP research that can allow to build these news representations.

2.3 Analysis of News

As we look at news analysis, it is important to recognize the rapid evolution that has taken place from historical traditional manual methods to advanced automated processes enabled by digital technology. This transition has increased the efficiency and scale of news analysis at the expense of precision and nuance. However, the development of natural language processing techniques continues to improve, providing increasingly sophisticated tools that gradually bridge the gap between what requires manual annotation and what can be done automatically.

2.3.1 From Manual to Automated News Analysis

Understanding the impact of news on society and its portrayal of events predates digital technology and involves many academic disciplines. For instance, sociology can assess how news shapes social perceptions [Bennett, 2005; Gamson et al., 2024], political science can examine the interplay between media coverage and political behavior [Snyder Jr and Strömberg, 2010; Djourelova and Durante, 2022], economics can study market insights from news [Birz, 2017] or mutations of the news industry [Athey et al., 2013], and psychology can study the impact of news wording and framing on audience perception [Kahneman and Tversky, 1984].

Although qualitative and quantitative analyses coexist in the social sciences, this thesis only addresses quantitative analysis. Qualitative analysis focuses on understanding the depth of human experience and social phenomena through interpretation and detailed description by the researcher, while quantitative analysis quantifies phenomena using statistical methods to identify patterns and relationships.

Before the advent of computers, quantitative analyses were performed exclusively manually, while different workflows can be found, researchers generally defined a "codebook" that would guide annotators (also called coders) in their annotation process. With the rise of computer availability, new software emerged to help coders in their annotation process by targeting elements of interest, pre-annotating, and facilitating the annotation process [Lowe, 2002].

Computers also allowed researchers to perform the first automated analyses, using word counts from developed lexicons and generating statistics from them. Those automatically detected patterns are less fitted to the object of study, but the automation allows for a greater volume of annotations at a fraction of the cost. This reflects a trend that is still ongoing in text annotation, where there is a constant trade-off between the quality, the complexity of the annotation task, the cost of the annotation and the size of the annotated dataset. Depending on the phenomenon under study, a proxy annotation task can be designed; its complexity may require one or more annotations by experts or may be amenable to crowdsourcing or automated algorithms, the latter choice affecting both cost and quality [Spinde et al., 2021; Caselli et al., 2016].

In this context of news analysis, natural language processing stands as a transversal discipline that seeks to derive algorithms able of automatically extracting increasingly refined patterns in order to reconcile the complexity of annotation tasks and their cost.

2.3.2 Recent Evolution of Natural Language Processing

To better understand the technological advances in natural language processing that have moved news analysis from simple word count approaches to more sophisticated pattern recognition algorithms, a brief review of recent paradigm shifts is provided.

Prior to 1990, the majority of NLP research focused on logic, symbolic rules, and lexicons. Nevertheless, lexicon- and rule-based methods remain relevant today, as evidenced by the volume of publications using tools such as LIWC [Tausczik and Pennebaker, 2010] and VADER [Hutto and Gilbert, 2014].

The 90s and early 00s were marked by the emergence of probabilistic models, ranging from Latent Semantic Analysis (LSA) [Landauer et al., 1998] to the notable Latent Dirichlet Allocation (LDA) [Blei et al., 2003] and its further developments, enabling the derivation of word and document representations.

A pivotal shift occurred in 2013 with the advent of Word2Vec [Mikolov et al., 2013], which enabled the creation of static word embeddings using a shallow 2-layer neural network, sowing the seeds of the neural network era. The foundational paper on the transformer architecture was published in 2017, initially applied to translation tasks and without prior pre-training [Vaswani et al., 2017]. In 2018, the ELMo paper [Peters et al., 2018] showcased the use of a BiLSTM sequence model to achieve contextualized embeddings via unsupervised pre-training, enhancing performance on subsequent tasks. However, its sequential approach to language modeling resulted in slow convergence and embeddings that were disproportionately influenced by adjacent tokens.

That same year, OpenAI introduced its first model, GPT [Radford et al., 2018], a generative model for next-token prediction built on the decoding component of the transformer architecture. This model went beyond ELMo by leveraging the transformer architecture's high degree of parallelism, which is critical for pre-training and inference. In 2019, Google released BERT [Devlin et al., 2019], a model that uses a pre-training procedure based on masked language modeling and next-sentence prediction. By leveraging the encoder part of the transformer architecture, BERT is better suited for downstream tasks that require bidirectional information flow such as sequence labeling.

These technological breakthroughs led to a paradigm shift in the NLP community, where most problems would revolve around adapting and fine-tuning pre-trained large language models (LLMs) to specific tasks. While emerging architectures such as Mamba [Gu and Dao, 2023] may redirect future research efforts, it's undeniable that recent NLP research has mostly focused on the pre-training, adaptation, enrichment, fine-tuning, inference, optimization, alignment, and analysis of transformer-based LLMs for a wide range of applications.

As with many domains of application, news analysis methods to this day revolve around the use of all of the aforementioned technologies. The extensive knowledge gained during pre-training improves the generalization capabilities of large language models and their ability to extract more complex patterns from text. However, due to large language models requiring significantly more computational resources than earlier NLP methods, and the need for news analysis to sometimes handle vast amounts of text, authors occasionally opt for the less resource-intensive, older methods.

2.3.3 Natural Language Processing for News Analysis

News analysis encompasses a wide range of levels of analysis, from examining a single article [Baly et al., 2020a], to evaluating a group of articles [Fan et al., 2019], to analyzing an entire news

outlet [Kwak et al., 2020], to exploring the media landscape as a whole [Wevers, 2019]. Higher-level analysis often mitigates errors by employing aggregation strategies that combine measurements from more detailed levels. To these different depths of news analysis answers different granularities of NLP methods, which can also be applied at different levels; word, sentence, paragraph, document, or even multi-document levels, and can also be aggregated from detailed to coarser levels. Another peculiarity of news analysis compared to general text analysis is the high turnover of its topics, which poses challenges for the generalization of analysis methods across time [Li and Goldwasser, 2019].

In this subsection, we present an overview of the diverse set of news analyses that are commonly performed using NLP approaches at different levels of aggregation. Most advanced analyses performed today use a flavor of the transformer architecture [Vaswani et al., 2017], most of the time the encoder BERT [Devlin et al., 2019], but also graph-neural networks (GNN) [Wu et al., 2020a] and combinations of older methods and models.

We only present areas of work on the analysis of news, not on their generation or reformulation, although the boundaries are sometimes blurred in the literature, both because analyses can condition generations and generative models can be used to perform analyses. In the following presentation, we categorize typical NLP research tasks into three groups for the sake of clarity: *Information Extraction – Content Understanding and Modeling – Veracity Assessment and Fact Management.*

Information Extraction

Information Extraction focuses on automatically extracting structured information from unstructured text. In the context of news articles, this can help develop a broader understanding of a chain of events or extract relevant measurements for subsequent analysis. Key tasks within this domain include: Named Entity Recognition, Event Extraction and Entity Linking.

Named Entity Recognition (NER)

Definition

Named Entity Recognition (NER) identifies and classifies entities in text, such as names, places, and dates [Li et al., 2022]. Historically focused on a pre-defined, finite set of entities, NER has evolved to incorporate increasingly complex entities across multiple domains [Malmasi et al., 2022], even evolving to open entity types with zero-shot scenarios [Zhou et al., 2023; Zaratiana et al., 2023]. These recent advances are made possible by the broad knowledge acquired by large language models during pre-training.

Challenges

Nevertheless, smaller language models trained on domain-specific NER datasets still show superior performance with greater resource efficiency [Zhou et al., 2023]. Other challenges remain, such as improving cross-domain NER transfer [Liu et al., 2021], as well as developing datasets for languages other than English [Malmasi et al., 2022] or improving cross-lingual transfer [Kramov and Pogorilyy, 2022].

In News Analysis

Named Entity Recognition often constitutes a precursor step for further processing such as Entity Linking or Event Extraction. As quasi-direct use of this technology we can mention: the location of events such as crimes [Arulanandam et al., 2014] and floods [Flo, 2022]. The importance of these models as precursor step of more complex news analysis is driving initiatives in many languages to develop NER datasets [Cucerzan and Yarowsky, 1999; Pradhan et al., 2013; Szarvas et al., 2006; Ruokolainen et al., 2019; Alsaaran and Alrabiah, 2021; Adelani et al., 2021].

Entity Linking (EL):

Definition

Entity linking connects text mentions to relevant entities in a knowledge graph, enabling the use of the graph data for subsequent tasks [Sevgili et al., 2022]. The most common application is called Wikification and consists in linking text mentions to their Wikipedia entries. Models must rely on the textual context of the mention to perform entity disambiguation and match it with a node of the knowledge-graph. Various methods exist, including encoder-based models with indexed embeddings that perform entity ranking [Wu et al., 2020c] and encoder-decoder models with generation constraints that generate matching candidates ranked by perplexity [De Cao et al., 2022]. Some methods further consider the context by taking into account the other entities present in the text and their relationship in the reference knowledge graph [El Vaigh et al., 2019].

Challenges

Challenges in this task include enhancing multilingual performance, creating common entity representations across languages, improving the context-based disambiguation, and addressing knowledge graph incompleteness and new entity indexing [Zaporojets et al., 2022; Kassner et al., 2022].

In News Analysis

In news analysis, entity linking primarily precedes secondary applications such as sentiment analysis of linked entities [Sinha et al., 2022], triplet extraction for knowledge graph completion [Gohourou and Kuwabara, 2024], or event extraction [Rollo et al.; Zhang et al., 2021]. Its benefits include name disambiguation and linking different spellings of an entity. Its direct application allows entity frequency analysis [Nguyen et al., 2022], even in pathological cases of spelling errors, for example in OCR versions of historical news [Linhares Pontes et al., 2020].

Event Extraction (EE):

Definition

Event Extraction is a multi-step process that focuses on identifying events in text, classifying their type, extracting their arguments and labeling their roles depending on the identified event type [Xiang and Wang, 2019]. It aims to address the classic 5Ws and H framework (what, who, when, where, why, and how) associated with an event [Hamborg et al., 2019a]. Approaches often rely on NER, EL, and dependency parsing methods.

Challenges

One of the main challenges associated with this task is its extension to new domains, as it requires significant amounts of training data, leading to research focusing on few-shot learning, (e.g. through meta-learning [Tuo et al., 2022]), and zero-shot learning (e.g. by reframing the event extraction task as a question-answering problem [Lyu et al., 2021]).

In News Analysis

The most emblematic dataset for this task is ACE 2005 [Walker and Linguistic Data Consortium, 2006], which allows event extraction for the news domain. In news, event extraction is primarily used to identify new events [Lee et al., 2022], track their evolution [Nguyen, 2011] and reconstruct narratives, either as maps or linear representations, to help dealing with the overload of the informational landscape [Keith Norambuena et al., 2023].

Content Understanding and Modeling

This subsection contains tasks that aim to understand and model the abstract aspects of texts, such as their topics, opinions and arguments. These tasks help to understand the underlying semantics of the content and provide deeper insight of the textual information.

Topic Modeling (TM)

Definition

Topic Modeling, a long-standing area of research, focuses on the automatic detection of topics and the categorization of large collections of text [Churchill and Singh, 2022; Abdelrazek et al., 2023]. The concept of a "topic" is loosely defined as a pattern of word distribution. Its granularity can vary greatly depending on the subject under study, in particular in terms of the unit processed (from sentences to entire documents) and the level of detail of the topics selected (e.g. which level of the hierarchical IPTV news taxonomy used to categorize articles). Historically, topic models have played an important role in semantic extraction from text, which changed with the advent of more versatile large language models and the pre-training/fine-tuning paradigm.

Challenges

The fact that the term "topic" is loosely defined, and that the meaning one attaches to it depends on the task at hand, makes the selection of an algorithm a task-specific decision in terms of accuracy, which may also depend on the amount of text to be processed and the amount of computing power available. The selection of the best performing algorithm for a target task is thus unclear [Churchill and Singh, 2022]. Other ongoing challenges include the management of streaming documents in real time and their evolving topics, notably the emergence of new topics, and how to model noisy text such as the ones found on the Web.

In News Analysis

In news analysis, topic modeling can be used at different levels of granularity, but allows to track the counts and frequencies of occurence of those topics. From very specific examples; Ertek and Kailas [2021] tracks frequencies of the different types of wind turbine accidents reported in the news; to broader categorization such as the firsts levels of the IPTC news taxonomy[a]. Two important tasks, intermediate in their granularity, are particularly related to topic modeling: Agenda Setting Detection and News Frames Detection, both considered to be types of Media Biases. Agenda Setting consists in determining what are the issues prioritized by the news media in their coverage [Edgar, 1980; Field et al., 2018]. Frame Detection is loosely defined, but referring to the definition of Entman [1993], it involves selecting and making some aspects of reality more salient to modify the perception and understanding of an event by the reader. This task is further described in 2.5.

[a] https://iptc.org/standards/media-topics/

Opinion Mining

Definition

Opinion Mining, a term used interchangeably with Sentiment Analysis in the literature, is interested in inferring the sentiment expressed in a piece of text. This can be done at different levels of granularity, from the general sentiment assessed at the document level to the specific sentiment expressed towards an entity in the text [Hemmatian and Sohrabi, 2019]. Historically performed through the use of lexicon and rules [Hutto and Gilbert, 2014], sentiment analysis has also been the object of classification approaches using large language models [Cui et al., 2023].

Challenges

Key challenges in opinion mining and sentiment analysis include incorporating common sense knowledge into models to infer implicit sentiment from context, and improving cross-domain transfer learning.

In News Analysis

Sentiment analysis in the news has traditionally focused on the whole body or headline level to infer the stance of the article. Given the limitations of this approach in accurately representing an article's stance, the current focus within the community is shifting to more detailed sentiment analysis at the entity level [Bestvater and Monroe, 2023].

Argumentation Mining

Definition

Argumentation Mining is interested in extending opinion mining from "what" opinion is expressed to "why" the holder has this opinion and "how" it is expressed. This area of research mostly evolve around theories of discourse structuration and identification of argumentative units and their relationships. The process generally include two-steps: argument extraction, with the detection of discourse unit type (claim, premise), followed by the classification of the relations between the arguments (e.g. support, attack). Other dimensions of arguments can be evaluated such as their sentiment, their factuality or a quantification of their persuasiveness among others [Cabrio and Villata, 2018; Lawrence and Reed, 2020].

Challenges

Current challenges in argument mining include: the need to harmonize argumentation models, the expansion of methods to languages other than English, and the improvement of performances of current approaches, notably the integration of implicit prior domain knowledge [Cabrio and Villata, 2018; Lawrence and Reed, 2020]

In News Analysis

In news analysis, argumentation mining theory has been used to analyze news editorials [Khatib et al.], analyze stance of comments towards italian news blog [Basile et al., 2016], but also in the perspective of detecting fake news [Karimi and Tang, 2019; Delobelle et al., 2020a; Kammoun et al., 2022].

Veracity Assessment and Fact Management

This subsection presents tasks aimed at analyzing the veracity of information through stylometric analysis and fake news detection. It describes the common methods for evaluating the reliability or truthfulness of news content.

Stylometry Analysis

Definition

Stylometry is most often associated with attribution of authorship through analysis of lexical use, punctuation, and phrasing [Lagutina et al., 2019]. However, it is more broadly concerned with the writing style of a text. It can notably include: analysis of tone (subjectivity [Recasens et al., 2013], irony [Ravi and Ravi, 2017]), and readability [Štajner et al., 2012].

Challenges

Challenges in stylometry detection primarily stem from the challenge of creating metrics, as these concepts lack formal definitions. In the context of authorship attribution, the critical issues involve the creation, selection, and combination of different textual features.

In News Analysis

In news analysis, stylometric features are mostly used to qualify the subjectivity and neutrality of texts to assess their reliability and credibility in the context of misinformation detection [Schuster et al., 2020].

Fake News Detection

Definition

Fake news detection is designed to identify false or misleading information. Methods include analyzing proxy metadata like web traffic [Bhatt et al., 2018] and social network interactions [Castillo et al., 2011; Shu et al., 2017], or focusing directly on the textual content of articles [Mridha et al., 2021]. Content-based techniques focus on either the style of the content or its semantic meaning. Style-based detection uses factuality and subjectivity metrics to evaluate the credibility of sources [Potthast et al., 2017], while semantic-based approaches either use a trained text classifier based on specific features or check claims against externally maintained databases [Guo et al., 2022].

Challenges

Challenges in fake-news detection depends on the chosen approach. For style-based detection, it is hard if even possible to detect well written fake news. For fact-checking approaches, the challenge is to build those curated truthful databases of debunked and fact-checked assertions, it also leads to make a choice between labeling as true or false what is not the database.

While each of these natural language processing tasks used in news article analysis has its own unique challenges, some common areas for improvement among them include: limited language

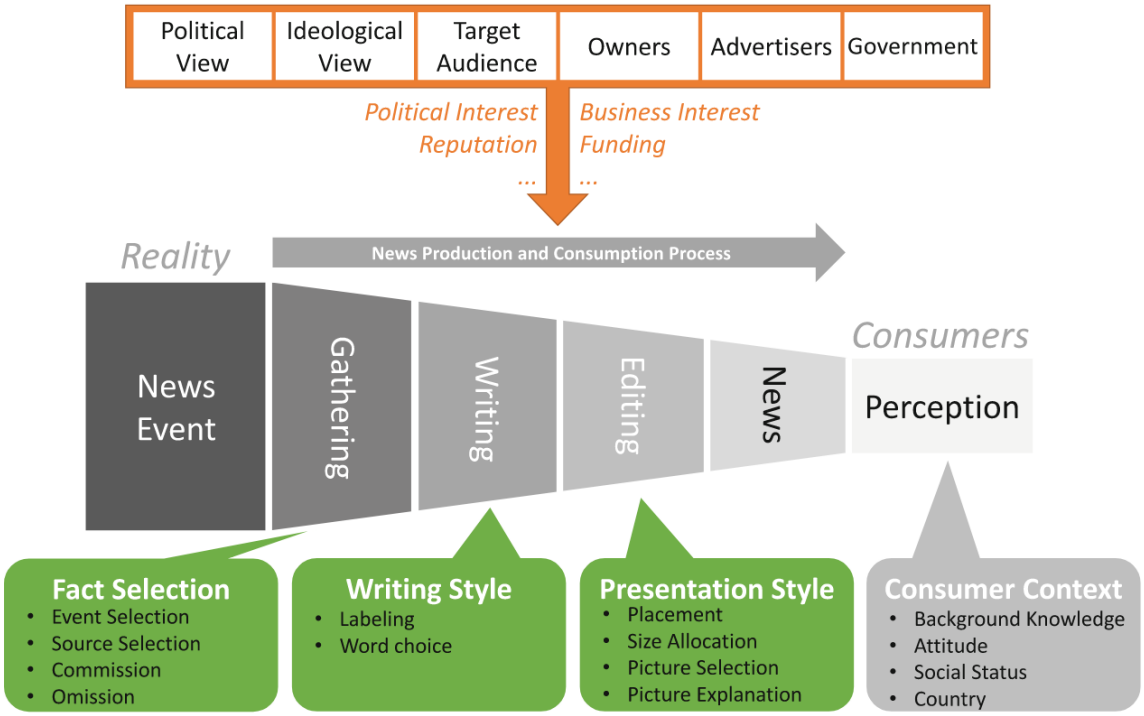


Figure 2.1: Media biases can occur from the gathering of facts to the presentation of the news to the end user. Illustration from Hamborg [2023b]

options (often only English), a lack of datasets for some domains, and the difficulties associated with cross-language and cross-domain transfers. Although these approaches do not aim to identify biases in the news directly, disparities in pattern distribution between "Information Extraction" and "Content Understanding and Modeling" methods could indicate the presence of biases. In the next subsection, we explore the causes of media biases and present specialized automated modeling tasks that have been developed to detect them.

2.4 Characterizing Stances through Media Biases

The opinions, most of the time implicitly expressed, of a news article or news outlets can be measured by examining and identifying the various biases that can manifest themselves throughout the information curation process of the media. This includes the treatment of information from its initial gathering to its final presentation, including the selection, prioritization, framing, and interpretation of news. All of these biases that alter perceptions of reality are grouped under the umbrella term of "Media Biases".

2.4.1 Root Causes of Media Biases

Among the various causes of media biases, the most prominent are the following:

- Media ownership [Gilens and Hertzman, 2000]: It is a well-documented phenomenon that media ownership has a significant impact on editorial content, often resulting in a bias in favor of the owners' viewpoints. This concept is exemplified through instances in France,

for example with the Bouygues group in 2008[4] and 2020[5], where the influence of media ownership extends to include a form of self-censorship.

- Carreer progression: Journalists sometimes skew their content to reflect the views of their superiors in order to increase the likelihood of being published and to advance their careers [Baron, 2006].

- Social Class and Ethnic Representation: Journalists often exhibit a disparity in their socioeconomic and ethnic backgrounds compared to the general population, which can unintentionally bias how news gets selected and reported. These discrepancies in representation can result in journalistic coverage that does not adequately represent the diverse perspectives and experiences present in society [Katsuo A. Nishikawa and Waltenburg, 2009; Spilsbury, 2017; Donsbach, 2004; Phillips, 2014].

- Media Audience: Newspapers often exhibit a tendency to bias their news coverage in ways that align with the preferences and expectations of their audiences. This practice reflects an editorial strategy aimed at maintaining reader engagement and loyalty by tailoring content to the perceived interests and viewpoints of their consumer base [Gentzkow and Shapiro, 2010; Mullainathan and Shleifer, 2005].

- Government: Even in the absence of overt censorship, media organizations often find it imperative to maintain positive relationships with government agencies to ensure access to timely and relevant information. This relationship dynamic can sometimes lead to biased reporting of news critical of the government [Besley and Prat, 2006; Brighton and Foy, 2007].

- Business Model: The business model of media outlets can contribute to biases in their reporting. A notable example is the need to attract advertisers, which can encourage a tendency towards sensationalism and the use of clickbait rather than in-depth and comprehensive analysis. This tendency is particularly pronounced in the context of the free press, which is more dependent on advertising revenue [Caple and Bednarek, 2016; Harcup and O'Neill, 2017].

Media biases, as illustrated by the listed causes, can be both intentional and unintentional, and can stem from a variety of underlying causes. Precisely defining these biases is challenging because it would require establishing what constitutes neutral reporting, which is unattainable. Most of the time, it is not even desirable because journalists are expected to contextualize the news they report to help their audiences understand it, which inevitably conveys their interpretation of the facts. Despite efforts to maintain objectivity, their reporting will inherently have a degree of subjectivity.

2.4.2 Types of Media Biases

Subject to these influences, the biases of a media outlet can be identified at different levels of the information processing pipeline as shown in Figure 2.1. At **the gathering phase**, with the selection of what information to report and not to report on; during **the writing phase**, with the choice of words and the framing of the information; and finally, at **the presentation phase**,

[4] `https://www.acrimed.org/EPR-de-Flamanville-la-condamnation-de-Bouygues`
[5] `https://www.acrimed.org/Cyberattaque-a-Bouygues-Construction-TF1-et-LCI`

with the choice of the placement, the size allocation and the images accompanying the article. A last important phase that is not directly related to the editorial process is also **the perception bias of the information** by the final user, which we don't tackle in this work but is important to take into consideration in digital tools.

Biases at the Gathering Phase

At this phase biases can occur as results of the news selection process, and are types of generally called **"Selection Bias"** or **"Gatekeeping Bias"** [White, 1950]. As news outlets can't report on every events or subjects, a choice needs to be made. These choices are the results of the influence of external factors as mentioned in subsection 2.4.1. For example, Saez-Trumper et al. [2013] show, among 80 international outlets, that cultural and geographic proximity are significant factors in press event selection.

Biases at the Writing Phase

At the writing stage, several biases can occur that are related to the way the information is semantically conveyed. The term "framing" is mostly used to conceptualize this idea of **selecting aspects** of a piece of information and making them more or less **salient** [Entman, 1993]. However, this definition remains quite vague and has led researchers across and within fields to adopt different definitions of what constitutes "framing" [Boydstun et al., 2014]. Overall, biases in the writing phase take the form of the recognition of semantically loaded text units in relation to objects. The objects can be: identified subgroups of the population [Dacon and Liu, 2021], a particular issue [Card et al., 2015; Kwak et al., 2020], or a side of an issue [Gentzkow and Shapiro, 2010; Bayram et al., 2019; D'Alonzo and Tegmark, 2021]. These biases are conveyed through word choice [Gentzkow and Shapiro, 2010], epistemological manipulation with the introduction of assertions, vagueness or use rhetorical device in presentation [Piskorski et al., 2023], or the insistence on or diminution of the importance of certain aspects by modulating the assigned text length within the article.

Other type of biases

Other types of biases can occur in the reporting of the news, such as the placement and size allocation of a story [Hamborg, 2023b], or the choice of a picture that affect the perception of the reader [Waldman and Devitt, 1998; Dearden, 2015]. These types of biases that concern other modalities than text are not studied as part of this thesis.

The Bias Perception Problem

Beyond the editorial biases that can alter the structuring and delivery of information, there is the subjective perception of those biases by the end reader, which is more related to their psychological idiosyncrasies and cognitive biases. This perception problem is important to expose as it should be considered within the development of digital tools. We can point out two problems in particular from the literature:

1. Two different factual phrasing of the same information can lead to opposed interpretations by readers.

The Asian Disease Experiment

Imagine that the U.S. is preparing for the outbreak of an unusual Asian Disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Set 1:

- **If Program A is adopted, 200 people will be saved.** (72% selected)
- If Program B is adopted, there is a one-third probability that 600 people will be saved, and a two-thirds probability that no people will be saved. (28% selected)

Set 2:

- If Program C is adopted, 400 people will die. (22% selected)
- **If Program D is adopted, there is a one-third probability that nobody will die, and a two-thirds probability that 600 people will die.** (78% selected)

Figure 2.2: The Asian Disease Experiment [Kahneman and Tversky, 1984]

2. The same information presented in the same way can be interpreted by different readers as being biased against opposing viewpoints.

The first problem is illustrated by Daniel Kahneman and Tversky in [Kahneman and Tversky, 1984]. The way information is presented can have a significant impact on the way it is perceived, regardless of its factuality. Their famous example of the "Asian disease" experiment is a powerful illustration of this fact. In this experiment, Kahneman and Tversky created a health crisis scenario for which they presented two equivalent sets consisting of the same two options, but phrased differently, as can be seen in Figure 2.2.

When presented with the first set, the majority favored program A, whereas, while being strictly equivalent in term of outcome, most people chose program D while presented the second set. This kind of phenomenon is even more pronounced in the case of news where the reader has less incentive to make an involved and rational analysis of the information presented to them [Daniel, 2017].

The second problem is called the "Hostile Media Effect" and was first brought to light in the late 1980's by the analysis of perceptions of the 1982 Beirut massacre [Vallone and Ross, 1985]. After viewing the same coverage of the massacre, pro-Israeli and pro-Arab partisans each reported that it was biased against their camp. More recent studies show that the stronger the affective polarization and the weaker the perceived status of one's group in society, the more likely a member of this group is to perceive an adversarial bias in what they read [Hartmann and Tanis, 2013].

While these idiosyncrasies should ideally be taken into account when assisting users in their

news consumption habits, we don't address this aspect in this manuscript. Most automatic bias detection methods focus primarily on textual cues, or use hybrid approaches that allow for the detection of biases that occurred at the gathering or writing stage. The next section discusses the specifics of these techniques.

2.5 Automatic Detection of Media Biases

The automatic detection of media biases revolves around identifying manifestations of the previously exposed sources, which often involves identifying proxy measures as cues of bias. Overall, Media Biases can be defined as the ensemble of processes that distort the reader's perception of reality. The notion of bias is difficult to define in absolute terms, but rather on the basis of relative comparisons with other outlets' coverage

2.5.1 Framing Bias Detection

Summary

Definition:
Originally a political communication concept, the development of automatic framing detection methods and agreement on what constitutes framing in the computer science community is complex. The most widely accepted definition is that of Entman [1993], which describes it as the selection and modulation of the salience of aspects. Framing is used to promote specific problem definitions, causal interpretations, moral judgments, and treatment recommendations. In essence, framing shapes how information is presented to influence audience perception and interpretation.

Methods:
Most methods employ the 14 general frames established by Boydstun (2014) and operationalized through the creation of the Media Frame Corpus (MFC) by Card (2015). The granularity provided by these generic frames is limited, prompting researchers to explore alternatives like "micro-frames" or topic modeling techniques to accelerate and refine the frame definition process.

Limits:
In the field of automatic framing detection, as in many others in NLP, the main challenges are the scarcity of datasets and the difficulty of adapting models for use across domains and languages.

Framing is a slippery term in the media bias literature, especially in automated detection, because of the lack of a precise and procedural definition of this political communication concept [Boydstun et al., 2014]. The definition that has prevailed in the community is that of Entman [1993], which states:

> "*Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.*"
>
> – Robert M. Entman

Studying the framing of a news corpus involves two steps, framing definition – determining the existing perspectives around the subject of study – and framing analysis – quantifying the use

of these frames in the corpus. In the natural language processing community, automatic framing analysis has attempted to address both of these steps by facilitating the definition of frames and developing classifiers to automatically identify them in articles. It's important to note that the boundary between framing and stance detection is blurry in the literature; but in this subsection we focus on those methods that attempt to derive these so-called "frames" for news analysis.

Manual approaches to define the framing of a subject typically involve the development of extensive frame codebooks through inductive reasoning based on a collection of related news articles. For example, Kothari [2010] examine the framing of the Darfur conflict in the New York Times from 2003 to 2006 with examples of the "United States as savior of Sudanese people" frame or the "fatalist" frame. Rauch et al. [2007] examine the framing of the Movement for Democratic Globalization from 1999 to 2004 by identifying "legitimizing" and "deligitimizing" frames. The inductive approach ensures a precise and well-tailored analysis to the object of study, but it is costly and cannot be automated at this level of granularity.

With the ambition of easing the definition process Boydstun et al. [2014] introduce a "Policy Frames Codebook" in which they develop 14 general frames that are sufficiently general to be used in the study of any issue (e.g., economics, capacity and resources, morality, fairness and equality, constitutionality, etc.). They demonstrate the effectiveness of their approach by presenting a pilot dataset annotated with the primary frames of articles on three policy issues: smoking, immigration, and same-sex marriage from 13 U.S. newspapers between 1990 and 2021. They show the evolution of frames use over time and across issues, and also demonstrate the ability to learn effective classifiers using simple logistic regression on bag-of-words representations of documents.

Card et al. [2015] iterate on the approach of Boydstun et al. [2014] by creating the "Media Frame Corpus" (MFC), keeping the same three topics (immigration, smoking, and same-sex marriage) and the 14 general frames. They annotate each article both with its primary (i.e. majority) frame and, at the text level, with the different spans that constitute frames. This approach turns the problem of frame analysis into a supervised classification of text or spans of texts to 14 general categories. Kwak et al. [2020] use the MFC dataset to perform a frame analysis on 1.5 million articles from the New York Times between 2000 and 2017. They train a Bert-base model to predict the main frame at the article level, and enrich their analysis through the use of the VADER Sentiment Analyzer [Hutto and Gilbert, 2014]. This information allows them to perform several temporal analyses at the level of specific issues or the overall outlet. In particular, they show that elections, wars, recessions, and other events significantly change the framing proportion in news reporting. They also show that the framing of specific events tends to evolve and converge on a single majority frame. Their analysis of sentiment concludes to an overall neutral framing of events by the New York Times, even during disasters such as "Hurricane Katrina" and "Hurricane Sandy.

One limit of this supervised approach to frame classification is the application to other languages. Field et al. [2018] tackle this problem for Russian. They derive a lexicon from the annotated sentences of the MFC using pointwise mutual information [Church and Hanks, 1990]. They then automatically translate the resulting lexicon into Russian using Google Translate, before expanding the Russian lexicon using a Word2Vec CBOW model that measures similarities of translated content with new words [Mikolov et al., 2013]. They determine that a document contains a given frame from the MFC corpus if it contains at least three frame-associated words. They use this new word-based frame detection to analyze the Russian news agency "Izvestia". They show that during Russian economic downturns, not only are the articles about the U.S. significantly more frequent, but also the two frames "Security and Defense" and "Crime and Punishment" are more frequent in them, depicting the U.S. as villainous and contrasting it with Russia. Niven and Kao [2020], similar to Field et al. [2018], use normalized pointwise mutual

information to determine which word categories among the 79 categories of the LIWC [Tausczik and Pennebaker, 2010] are most related to the 14 frames of the "Media Frame Corpus" [Card et al., 2015]. They find high concordance of LIWC categories with four frames, and use their Mandarin Chinese version of those categories to analyze the alignment of various Taiwanese press outlets in their coverage of protests with that of the Chinese Communist Party (CCP) state media. In particular, they find a significantly higher occurrence of the LIWC's "Death" and "Religion" categories in CCP coverage of protests, and show that the prior knowledge of Taiwanese media alignment with the CCP correlates with this framing. The "Media Frame Corpus" approach to frame analysis has been recently expanded to other languages by Piskorski et al. [2023], in the context of the 17th edition of the International Workshop on Semantic Evaluation (SemEval 2023). They developed a framing dataset in 9 languages, building on the work and taxonomy developed by Card et al. [2015], but requiring to predict all frames present in the article and without annotating them at the level of spans.

The problem with these approaches is the a priori definition of a small set of 14 generic frames. To remedy to this coarse granularity Kwak et al. [2021] offer a new characterization of framing through the use of what they call "micro-frames." They construct semantic axes based on differences in Glove embeddings [Pennington et al., 2014] of 1,621 antonyms pairs selected from WordNet [Miller, 1995]. Using the Glove embeddings of words within a text, they compute their cosine distances to the semantic axes and derive bias indicators from the score distribution over these micro-frames. By analyzing 50,073 U.S. news headlines from AllSides, divided into several categories, they show that some "micro-frames" allow to differentiate between conservative and liberal articles. Mokhberian et al. [2020] extend the idea of "micro-frames" by Kwak et al. [2021] to establish a set of micro-frames characteristic of "Moral Foundation Theory" (MFT), they show that using the 5-axes of MFT can help detect partisanship as a complementary feature to textual ones. They train a classifier to determine the MFT dimensions of tweets using an annotated dataset, and apply it to a dataset of news headlines between 2016 and July 2017 from the AllSides dataset.

Other sets of methods try to stay closer to the manual process of the custom and fitted definition of frames to a subject. Tsur et al. [2015] propose a method that requires limited human annotation. Based on the LDA algorithm [Blei et al., 2003], they identify frames as topics in a corpus of US congressional speeches. Modeling the corpus with 100 topics, they cluster the transcribed speeches into four clusters that form macro-topics, which they manually label as health, energy, security, and the economy. They then automatically find the subtopics that most discriminate between Republican and Democratic speeches, which can later be used to differentiate partisan framing in news articles. More recently, Czymara and Klingeren [2022] used Structural Topic Modeling (STM) [Roberts et al., 2014] to identify 14 frames to proceed to a comparative analysis between print and online versions of three German news outlets of the coverage of the "migration crisis" from 2015 to 2017. They only consider articles that differ between online and print versions. They conclude that while some frames are favored by some media at certain times, print and online news frames tend to follow the same trends. They also conclude that there is greater frame diversity in print.

2.5.2 Stance Bias Detection

Summary

Definition:
We define "Stance Bias Detection" to include all methods that aim to identify the stance, partisanship, or ideology conveyed in a text. In the literature, we identify three primary approaches: detection based on lexical cues, supervised stance detection, and entity-based stance detection.

Methods:

- Lexical cues aim to determine the stance of a text by identifying specific sets of words that are characteristic of a particular position.
- Supervised stance detection includes all methods that use machine learning models trained on annotated datasets to assign stance labels to text.
- Entity-based stance detection methods focus on assessing the sentiments directed toward specific entities, which, along with other characteristics, can provide insight into the article's stance on a topic.

Limits:

- Most lexical approaches are predominantly US-centric, focusing on creating word sets that distinguish between speeches by Democratic and Republican members of Congress. While effective at identifying partisan text, these methods struggle to generalize effectively to new topics, over time, and across countries.
- Supervised methods are limited by their poor explainability and the possibility of spurious correlations. Utilizing the Twitter retweeting graph could have been a promising way to enhance and broaden these techniques. As it stands, these methods face similar generalization challenges as lexical approaches.
- Entity-based sentiment detection seems more promising in terms of generalization capabilities to other domains and languages. However, its limitations stem from the performance of the models, especially because sentiment is often implicitly expressed in news, making it more difficult to accurately determine sentiment towards entities.

Political Stance through Lexical Cues

Early efforts at automated partisanship detection used word choice in news articles, based on the premise that different ideological positions are reflected in specific word usage. Most of these methods are based on word frequency analysis of specific phrase cues. These cues are typically identified by comparing different partisan discourses to identify their distinctive lexical signals. A pioneering study in this context is the work of Groseclose and Milyo [2005], which, although through manual analysis, is the first attempt to categorize news media by comparing their alignment with congressional speeches whose political leanings are known. This method involved identifying and counting think tank citations and comparing their frequency in both news media and congressional discourse. Years later, Gentzkow and Shapiro [2010] also decide to

compare the wording of the news, this time automatically, with congressional speeches from 2005. They extract common two- and three-word phrases from congressional speeches and determine the 500 most discriminating between Republicans and Democrats using a Pearson's chi-squared test. By correlating the occurrence of the phrases with the ideology of the congressmen, they're able to determine a slant index for news outlets based on their own frequency of use of these discriminatory phrases. They show that this slant index is positively correlated with a manually established score "the Mondo Times conservativeness rating". Gentzkow et al. [2016] analyze the transcriptions of the United States Congressional Record from 1873 to 2016 to understand the evolution of political partisanship in congressional speech and interpret it as an indicator of polarization. They use a multinomial logit probabilistic model with a Poisson approximation to assess the frequency of phrases used by members of Congress as a function of their party. Their results show a significant increase in partisanship on several issues, especially immigration, crime, and religion, in recent years. Bayram et al. [2019] iterate on this work but through the use of a simple bag-of-word modeling approach, keeping only words that occurred at least 50 times and removing some common words, then they use the log frequency of the words and normalize the vector representing each speech to unit norm. They show that classifying these simple word-based feature vectors with a linear classifier can effectively predict stance with an accuracy close to that of more complex methods. One limitation of methods based on lexical analysis is their inability to discern how distinct political groups use different terminology to describe the same concepts, reflecting their attitudes toward specific groups (i.e. "migrants" vs "illegal aliens"). Webson et al. [2020] aim to solve this problem. Using the Congressional Records and the Partisan News Corpus, they retrieve relevant partisan phrases in a manner similar to previous methods, and then train a network to generate both denotation and connotation representations of these phrases. The denotation representation is intended to identify the entity to which a phrase refers, regardless of party affiliation. Conversely, the connotation representation serves as an indicator of the author's political stance. A related method used by Bianchi et al. [2021] consists in using static embedding models trained on different corpora to determine the difference in perception of a concept between these corpora. Finally, D'Alonzo and Tegmark [2021] use common phrases that appear among 100 sources in the U.S. from 2019 to 2020. They then use a generalized version of SVD, truncated to its two primary dimensions, which roughly correspond to the representation of phrases along the traditional left-right and pro-anti-establishment axes.

Supervised Stance Detection

Supervised stance detection are machine learning models that attempt to infer a stance or ideological label at the article or outlet level from textual and/or metadata inputs. Most labeled datasets exist in the context of U.S. newspapers, notably from initiatives such as "AllSides" – a crowdsourced platform of article bias annotations – and "Media Bias/Fact Check" (MB/FC) – that annotate left-right leanings and factuality at the outlet level. Limitations of such approaches lie in their poor generalizability across language, culture, and temporality; as well as their poor explainability. Other approaches attempt to infer labels from the use of social network communities, but often rely on significant manual curation and stance simplification, hindering their broad applicability.

Baly et al. [2018] are the first to use the annotated supervised approach to derive the left-right leaning of news outlets using a MB/FC dataset. They employ a variety of features, in particular the 141 mostly lexicon-based features introduced by Horne et al. [2018], which they then enrich using the Wikipedia pages, Twitter accounts, and AlexaRank traffic-related data of each outlet. They represent each outlet by the features of 10 to 100 articles, and train an SVM with an RBF

kernel to learn the political bias classification task. They show that the textual content of the articles is the best performing feature for prediction, although not very high, and performing on par with Wikipedia page content and article titles. Kulkarni et al. [2018] on the contrary, try to derive the left-right leaning at the article level. They use 30,000 articles from AllSides, annotated with position (left/center/right). The articles come from 59 different US outlets. They enrich textual representations by exploiting the graph of hyperlinks from the articles to other domain names to build representations of the media outlets using Node2Vec [Grover and Leskovec, 2016]. They then learn a multi-view classifier that must determine the label by fusing information from the title, content, and network information. Their ablation study shows that the "title" feature doesn't add information to the "content" feature, but that the addition of the network feature leads to further improvement. An important limitation of this work is the lack of split across outlets, even more when considering that the graph representation is computed at the outlet level. A limitation also present in later work such as the one of Li and Goldwasser [2019] and pointed-out among other limitations by Zhou et al. [2021].

The main problem with these supervised approaches is to not overfit on the style of the training outlets to predict the leaning. Zhou et al. [2021] notably shows that randomizing the labels of the outlets allow to conserve around the same prediction performance when not splitting according to them. Baly et al. [2020a] propose an adversarial loss to enforce independence of representations on outlet style. They build a dataset of 34,737 news articles using AllSides, across 109 topics from 73 different media sources to achieve a balanced ideological representation. The goal is to identify media bias at the individual article level, but as expected, they show the dataset possesses only 3% of the articles are not consistent with the general bias of their outlet. They demosntrate that traditional classification methods yield accuracy only slightly above chance when articles are splitted by outlet. To address this, the authors propose a novel use of triplet loss, which aims to closely group articles that share ideological perspectives across outlets, while distancing articles from the same outlet but with different learnings. This method shows improvement over a previous adversarial loss approach, but the modest performance underscores the inherent difficulties in accurately determining article-level bias. The problem of the caducity of annotations has been highlighted by several papers that have shown a strong decrease in performance when splitting over time [Li and Goldwasser, 2019; Zhou et al., 2021].

The issue of the validity of these annotations for other cultural contexts is also complex. Other initiatives have attempted to use social networks such as Twitter to derive pseudo-labels or to perform remote monitoring to obtain up-to-date annotations at low cost. Wong et al. [2013] are among the first to have the idea to use the information contained in microblogs such as Twitter to automatically label the stance of news outlets on the left-right axis. They select popular, most retweeted user accounts. Using a lexicon-based approach, they determine the sentiment of these users' tweets and whether the tweet is favorable or unfavorable to a candidate. Candidate detection is based on manual keyword detection. By averaging the identified tweet sentiments, they determine a sentiment score of Twitter users. From this average stance of users and their retweeting pattern of media outlets, they derive a stance score of those outlets. The obtained scores are significantly correlated with the manually derived liberal-conservative scores of the "Mondo Times". Several other papers aiming at categorizing the leanings of outlets have used Twitter to derive classification features, to perform pseudo-labeling, or as signal for distant supervision. In the realm of classification features, Baly et al. [2018] make use of outlet account descriptions and Baly et al. [2020b] augment it by using also features from followers' descriptions. In the same fashion as Kulkarni et al. [2018], we could also imagine using article representations derived from Node2Vec algorithm on the retweeting graph. Zhang et al. [2022a] notably perform such a strategy of enrichment but using other knowledge bases than Twitter.

In the context of pseudo-labeling, Darwish et al. [2020] make use of the Twitter retweet graph for classifying pro or against stance of users on issues. After selecting a topic and using keywords and hashtags to collect relevant tweets, the researchers examine the retweet behavior of users. They then apply UMAP [McInnes et al., 2018] (Uniform Manifold Approximation and Projection) for dimensionality reduction and Mean-Shift for clustering, keeping only the two largest clusters. These clusters are empirically shown to represent the pro and anti stances on the given topics. Stefanov et al. [2020] build upon this methodology of by using the clusters with automatically assigned labels as ground truth for classifying tweets. They identify the positions of users not included in the two primary clusters by using a FastText model trained on tweet data. They then assess the media's stance on eight different topics by calculating a valence score between $[-1, 1]$. This score measures the difference in how often an outlet's domain name appears in the tweets of left-leaning versus right-leaning communities. By correlating their findings with Media Bias/Fact Check (MBFC) labels, they show that valence serves as an effective metric for inferring the political leanings of outlets when averaged across topics. The two main limitations of these pseudo-labeling approaches are; 1- the need to manually identify topics and retrieve their tweets; 2- the granularity and complexity of the approach, which only allows modeling binary stance issues. Using the Twitter graph by considering users who have shared articles to perform distant supervision of text classifiers could solve this granularity problem. This seems to be even more accurate in the light of studies such as the one by [Morales et al., 2021], which shows that performing dimensionality reduction by correspondence analysis on the follower matrix of French MPs allows us to retrieve primary axes that correspond to the ideological positions of the parties, as they correlate with the party opinions of the Chapel Hill Expert Survey. Demonstrate the presence of ideological cues in the Twitter graph of interactions. In this respect Li and Goldwasser [2019] are the first to use the Twitter graph as a political signal to train text extractors. They retrieve 10,385 stance-annotated US articles from AllSides with their URLs. They then query the Twitter API to obtain active users who have reshared a portion of these articles, and construct a heterogeneous graph with articles, users, and politicians as nodes, and follows and shares as edges. The recent acquisition of Twitter by Elon Musk on October 27, 2022, marked the end of the availability of the academic API. Due to the prohibitive cost of the new API, these promising approaches using Twitter are no longer being explored, and it results difficult to find similar alternative platforms with sufficient data. Using crowdsourcing platforms to reduce annotation costs typically requires non-expert annotators to provide more annotations for reliable results compared to expert annotators [Lazaridou et al., 2020]. Alternative methods approximate the leaning of an article by the general leaning of its outlet [Liu et al., 2022]. However, the foundations of leaning annotation approaches have been criticized. Ganguly et al. [2020] show that the three common assumptions in the construction of media bias detection datasets do not hold in practice, they are: raters' leanings do not affect annotations, an article's leaning matches that of its outlet, and an outlet maintains consistent leaning across issues. Additionally, Zhou et al. [2021] point out that such methods lead to poor prediction explainability, with several bias detection datasets containing hidden biases that drive spurious correlations with their labels.

Stance Detection through Entity-based Sentiment Analysis

In addition to these supervised methods, which suffer from low explicability and generalizability, techniques have been developed to measure the opinions expressed in article texts. A common proxy measure for this task is to use sentiment analysis to determine the sentiments directed toward specific entities. By characterizing these entities, patterns of sentiment can be identified that can reveal consistent stances.

A simple proxy measure, used until recently in the literature, is to determine the attitude toward an entity by determining the overall sentiment of the text mentioning that entity. Dacon and Liu [2021] assign paragraphs to a gender based on the predominant gendered pronoun, and use VADER [Hutto and Gilbert, 2014] to identify differences in sentiment representation between genders. In a similar fashion, Zahid et al. [2020] analyze the statement biases of six media outlets by analyzing their official tweets on the Pakistani 2018 elections. They also assess the sentiment of tweets mentioning the three presidential candidates using VADER [Hutto and Gilbert, 2014]. Some approaches try to build more specific models by annotating tweets' stance towards specific entities, but generalization to other contexts and other entities is difficult Mohammad et al. [2016a].

Using text-level sentiment analysis methods to perform targeted sentiment analysis is rather imprecise in practice, in the sentence "I'm sad that John lost", although eminently negative, it actually shows a positive attitude towards John. Rashkin et al. [2016] tried to tackle this problem. They created a data set of what they call "Connotation Frames". They annotated the connotation associated with the most frequent 900 transitive verbs from the New-York Times. Each of these verbs articulate an agent and an object, and the annotation is made to establish the writer's sentiment toward the entities, the entities' sentiment toward themselves, the effect on the object, the value of the object, and the mental state of the object. They use these annotations to establish subtle sentiment expressed by writer towards entities in the news. Sap et al. [2017] iterate over this "Connotation Frames" dataset and extend its annotations to include "Power" and "Agency" dimensions for the arguments (agent and object) of its 900 transitive verbs, resulting in the "Connotation Frames of Power and Agency". They show that the use of these new dimensions is an effective way of demonstrating the persistence of gender stereotypes in film plots, with men characters on average being given more agency and power through their speech than their women counterparts. Finally, Field et al. [2019] use the "Connotation Frames of Power and Agency" to conduct a framing analysis of the MeToo movement in the global press landscape. They extend the limitations of the static representation and limited number of verbs in the original corpus by training logistic regression to predict the agency and power dimension from ELMo's contextualized embeddings of any verb [Peters et al., 2018].

Transitive verbs are not the only means of expressing sentiment towards entities, and these approaches also fail to account for how context can change the interpretation of verb connotations. Alternative methods address these limitations by using word embeddings to measure the overall sentiment towards entities across a corpus (multiple documents). Field and Tsvetkov [2019] use the NRC Valence, Arousal, and Dominance (VAD) lexicon Mohammad [2018] to train regression models that predict the valence scores of words. They infer the valence of entities in long texts by averaging the scores of their contextualized embedding representations. Additionally, by training static embeddings from Word2Vec on different corpora, Bianchi et al. [2021] analyze how sentiment valence varies across corpora for specific terms like 'capitalism', 'parenting', or 'Berlusconi'. However, the accuracy of the first method is limited as we'll see in chapter 4, and the second method's use of uncontextualized embeddings and a short window size results in reduced precision, rendering it ineffective at the article level.

Finally, several researchers have adapted aspect-based sentiment classification, typically used for products, to entities in news articles. Tran [2020] pioneered the application of BERT models, trained on the Restaurant dataset [Pontiki et al., 2016a], to the All-the-news corpus, which includes 14 different American sources. They ranked these sources according to left-leaning scores obtained in their articles, and demonstrated through Pearson correlations that their rankings were consistent with the ratings of major American bias assessors: AllSides, Media Bias/Fact Check, and AdFontes. However, as indicated by Rietzler et al. [2020], cross-domain transfers generally

perform poorly, as also discussed in Chapter 4, underscoring the need for domain-specific datasets. In addition, the method's reliance on the Democrat-Republican dichotomy for bias scoring and string-based matching for entity identification in text limits its effectiveness and may miss many entities. To address the issue of domain-adapted datasets for targeted sentiment classification, Hamborg and Donnay [2021] introduced a new annotated dataset focused on U.S. newspapers. While this enhances transferability, the US-centric nature of the dataset, with its significant bias in entity distribution, limits its applicability in European contexts, as detailed in chapter 4. Being one of the research directions we explored, the next subsection 2.6 presents the state of the art in aspect-based/targeted sentiment classification.

2.5.3 Phrasing Bias Detection

Summary

Definition:
By "Phrasing Bias Detection" we mean all methods that attempt to identify the use of subjective language or rhetorical devices that that may distort perceptions or misrepresent facts. We include factuality detection methods and information bias detection methods.

Methods:

- Methods for assessing factuality rely on features including lexicon, data from social networks, and traffic statistics. In supervised scenarios, models are trained using the factuality scores provided by the Media Bias/Fact Check institute.
- Informational bias detection primarily utilizes annotated datasets to identify its presence within specific text spans, up to determining the particular persuasion techniques employed.

Limits:
Both approaches are limited by their dependence on annotations for classification, which restricts their ability to generalize. Additionally, supervised models trained on the MB/FC dataset often lack explainability.

Factuality

A body of research focuses on determining the factuality of articles or news outlets, often using the results of the Media Bias/Fact Check initiative on U.S. newspapers to evaluate their methods. While these approaches do not determine the stance of news outlets, they do help identify potential presentation bias by analyzing the presence subjective phrasing.

Baly et al. [2018] developed a method for assessing the factuality of media outlets using a variety of features, particularly the 141 primarily lexicon-based features from Horne et al. [2018]. They augmented these features with data from Wikipedia pages, Twitter accounts, and AlexaRank traffic statistics for each outlet. Representing each outlet by features extracted from 10 to 100 articles, they trained an SVM with an RBF kernel to predict the 3-level factuality scores from MB/FC. The authors reiterate the experiment with a new set of features in Baly et al. [2020b]. Panayotov et al. [2022] show that by using graph representation features from Alexa website traffic data, their results outperform or are comparable to the feature-based approaches of Baly

et al. [2020b]. However, despite sometimes yielding satisfactory performances, these feature-based methods often lack explainability and appear unreliable beyond the scope of well-established news outlets. Other researchers have developed more fine-grained methods based on textual content and using annotated datasets to assess factuality at finer levels. Pryzant et al. [2020] identify what they call 'framing bias', which essentially involves the use of subjective language that may indicate a lack of factuality. They developed a new dataset, the 'Wiki Neutrality Corpus (WNC)', collected from Wikipedia's edit history of articles flagged for violating the 'Neutral Point Of View' (NPOV) policy. This dataset contains 180,000 sentence pairs, each consisting of a subjective sentence and its revised neutral version, facilitating the training of a classifier to detect problematic sentences and an encoder-decoder model for sentence reformulation. Sales et al. [2021] use the Portuguese subjectivity lexicon from Amorim et al. [2018], which assesses subjectivity across five dimensions (Argumentation, Presupposition, Modalization, Valuation, and Sentiment). They translate this lexicon into English, German, and Spanish using automatic translation. The subjectivity level of each article is determined by comparing the Word2Vec average sum of its word embeddings to the average sum of embeddings for each of the five lexicons (one per dimension). Scores are normalized across languages using the EuroParl parallel corpus. They validate their method and analyze coverage disparities in reporting on the Syrian War and the Venezuelan crisis across various countries (USA, Brazil, Germany, Venezuela, etc.). Finally, Spinde et al. [2021] introduce a manually annotated sentence-level dataset categorizing sentences as Opinionated, Factual, Mixed, or containing no agreement among annotators.

Informational Bias

The title of this subsection is derived from the definition of "informational bias" as described in [Fan et al., 2019], which is: "*sentences or clauses that provide information that is tangential, speculative, or background to the main event, with the goal of influencing readers' opinions about the entities involved in the news.*" Informational bias is therefore somewhat independent of factuality, focusing instead on persuasion techniques and rhetorical devices that shape readers' opinions within a specific context.

Fan et al. [2019] publishes the BASIL dataset consisting of 300 articles from three media: Fox News, New York Times, and Huffington Post, covering different events between 2010 and 2019. For each event, a set of three articles (one from each outlet) is assembled into a triplet. Among other annotations, annotators must specifically identify sentences that exhibit a lexical or informational bias, noting whether the bias directly targets a particular entity and whether it occurs within a quote. Van Den Berg and Markert [2020] leverage these informational bias span annotations to highlight the importance of considering a wider context beyond just the sentence level for accurately identifying informational bias. They propose the EvCIM model, which processes article triplets with different political leanings about the same event. This model uses a BiLSTM to generate representations of these articles, averages these representations, and concatenates them with the candidate sentence representation obtained by averaging the last four layers of the RoBERTa model. Finally, a linear layer is used for classification on the presence of informational bias. The simple but vague notion of "information bias" can be related to the detection of persuasion techniques, primarily in the context of propaganda detection [Dimitrov et al., 2021]. Recently, [Piskorski et al., 2023] presented a dataset for classifying 23 persuasion techniques in news articles as part of Task 3 of SemEval 2023.

2.5.4 Social Category Biases

Summary

Definition:
Social category bias is the study of the underrepresentation or misrepresentation of particular social groups (such as gender, age, race) in the media. This field examines the biased representation and coverage of these groups in various types of news content.

Methods:
Studies typically use content analysis techniques such as entity-centric sentiment analysis, topic modeling (LDA), and machine learning classifiers (SVM, Naive Bayes). They analyze text for semantic properties such as sentiment and topic association, often using datasets that have been automatically labeled with demographics.

Limits:
The limitations of research in Social Category Biases primarily stem from the inherent drawbacks of the diverse techniques employed, such as topic modeling, sentiment analysis, named-entity recognition, and entity-linking.

In Social Category Biases, we consolidate all research related to the under-representation or misrepresentation of certain social groups in the news, which may include demographics such as gender, age, or race.

A first type of studies try to extract characteristics associated with identified group in a similar manner as some early approaches of entity-centric sentiment analysis, that is after identifying the occurrence of the group under consideration, establishing semantic characteristics that distinguish it from others (sentiment, topic, occupations). An illustration of this approach is the work of Dacon and Liu [2021]. They use the MiND and NCD dataset articles to identify gender biases, assigning M and F labels to paragraphs using the predominant gendered pronoun. They then analyze career nouns, occupations, and sentiments for statistical insights. They notably use Centered Resonance Analysis on nouns and adjectives within paragraphs to show most associated words with male and female mentions. They put in evidence presence of biases and stereotypes in the depiction of men and women. In the same fashion, Leavy [2020] explore gender biases in how ministers are portrayed in Irish newspapers from 1997 to 2011. The analysis is carried out using bag-of-words representations of texts, excluding pronouns and names that could indicate gender. The study employs Support Vector Machines, Decision Trees, and Naive Bayes classifiers for this analysis. They find that although female ministers are mentioned 2.5 times more often than male ministers, their mentions also reveal gender stereotypical biases, particularly in the portrayal of family life. Finally, Rao and Taboada [2021] train LDA topic models on Canadian news sources and study the variation of frequency of male vs. female gendered mentions across time by topics. They show over recurrent topics such as politics, business, lifestyle, entertainment and healthcare that there are consistent gender disparities. These methods have the adavantage of being applicable at the scale of articles, but necessitate to clearly define the semantic concepts under study. Other methods using static embeddings learned on news corpora are also used to perform multiple association and bias analysis. Following the increasing use of static word embeddings after the release of Word2Vec by Mikolov et al. [2013], Bolukbasi et al. [2016] is the first to detail gender stereotypes and biases contained within those embeddings through similarity metrics and

analogy (notably in work occupations). They offer a method to find the gender subspace of the embedding space and neutralize words for which we're not expecting variation of similarity along that dimension. This led to the development of the WEAT test methodology by Caliskan et al. [2017], to determine whether embeddings and the corpora they're trained on are biased towards gender stereotypes along certain dimensions. As a direct application of this methodology, Wevers [2019] train Word2Vec models to analyze the gender biases of the Dutch Media landscape and its evolution from 1950 till 1990. They analyze six Dutch national newspapers that are known to have long-term ideological differences. They compare them and also compare their individual evolution over the four decades. In particular, they observe temporal tendencies of the subcategories "Money" and "Negative Emotions" towards the model representation of the concept "Male", and of the subcategories "TV", "Music", "Metaphysical Issues", and "Sexuality" towards the model representation of the concept "Female". Still in the analysis of the Dutch newspaper, Kroon et al. [2021] examine the progression of stereotypical biases between ingroup and outgroup ethnicities by comparing the evolution in the similarity of their embedding representations to those of words indicative of low-status and high-threats.

2.6 Aspect-Based Sentiment Classification

In computer science, "Opinion Mining" and "Sentiment Analysis" are often used interchangeably. Of the three levels of sentiment analysis usually distinguished (document, sentence, and entity/aspect), only the finer-grained level is suitable for stance characterization [Bestvater and Monroe, 2023] (although coarser evaluation can be good indicators in explicit formats such as product reviews). In a more general setting, a document or sentence could contain multiple entities about which sentiments are expressed.

2.6.1 Definition

Liu [2010] defines the task of ABSA as the extraction of quintuples $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, where e_i is the i^{th} entity considered, a_{ij} is the j^{th} aspect of that entity, and s_{ijkl} is the sentiment that the opinion holder h_k expresses towards the aspect a_{ij} at time t_l. Sentiment analysis at the aspect level is interested in extracting, sometimes partially, those quintuples. The two most studied extraction frames are "product aspect mining", with the extraction of (e_i, a_{ij}, s_{ij}) triplets, and "stance detection" with the extraction of (e_i, s_i) tuples. These framing of the task can classically be split into two steps: Aspect-Term Extraction (ATE), and Aspect-Term Sentiment Classification (ATSC). Aspect-Based Sentiment Analysis has mostly been applied to product reviews datasets [Pontiki et al., 2014, 2015, 2016b], but also to financial [Jangid et al., 2018; Gaillat et al., 2018] , and more recently, news datasets [Steinberger et al., 2017; Hamborg and Donnay, 2021]. For a more complete overview of all existing ABSA-related tasks, we refer the reader to the survey of Zhang et al. [2022b]. In the following, we focus on presenting the two main constitutive tasks of the ABSA literature, namely the detection of entities and the inference of the sentiment expressed towards them.

2.6.2 Aspect-Term Extraction (ATE)

Given an opinionated-text content as input, ATE aims at extracting the targets and aspects ($(a_{ij}$,$e_i)$ or e_i) towards which a sentiment might be expressed. Approaches can be grouped into four types. Frequency-based approaches extract most frequent nouns and noun phrases

after Part-of-Speech (PoS) tagging, notably missing less frequent aspects, and generating a large number of noisy targets. Other approaches such as the one described by Qiu et al. [2011] extract domain-specific opinion words and targets using syntactic relationships. They alternatively expand both an opinion word lexicon and a set of candidate targets in a bootstrapping fashion. Supervised methods, by formalizing ATE as a sequence-labeling task have also been implemented. Finally, non-lexicon-based unsupervised methods make use of topic-modeling approaches (e.g. LDA, PLSA). An example of such an approach is Titov and McDonald [2008], which models a document as a mixture of both local and global topics, local topics are derived at a window scale to capture target aspects. The above approaches assume explicitly mentioned aspects, however, aspects can also be implicit. We refer the reader to literature surveys [Hemmatian and Sohrabi, 2019; Nazir et al., 2020] for further information on the topic. In the context of our work, entities are characterized using named entity recognition systems. Thus, we don't get into an exhaustive presentation of the different sets of existing techniques used in ATE in settings such as product reviews.

2.6.3 Aspect-Based Sentiment Classification (ABSC)

Also called "Targeted-Sentiment Classification (TSC)", Aspect-Based Sentiment Classification consists in determining the sentiment expressed towards extracted entities. In this context, machine learning methods are often distinguished from historical rule-based ones. Rule-based methods mostly rely on PoS tagging and the adjective-noun proximity heuristic to associate adjectives to aspects. These methods have been largely supplanted by machine learning methods, more particularly through the use of pretrained language models.

Methods

Since its introduction by Devlin et al. [2019], resolving ABSC with a fine-tuned BERT has been a strong baseline that is hard to beat in practical scenarios. This approach frames the aspect sentiment prediction as a sentence pair classification task, using the format "[CLS] sentence [SEP] entity [SEP]"', such as "[CLS] This phone is fantastic! [SEP] phone [SEP]". Gao et al. [2019] proposes a different modeling approach that uses aspect final embedding of BERT model instead of the traditional sentence-pair classification. This method allows simultaneous sentiment prediction for multiple aspects.

Some approaches enrich the embedding representations by using external dependency parsers to identify useful patterns for sentiment classification. A problem with these approaches is their generalization to languages that don't have such external resources. Sun et al. [2019] use GloVe word embeddings. They are processed by a BiLSTM to create contextual representations. These are further enriched with the relationships from a dependency tree using a 2-layer Graph Convolutional Network (GCN). To predict polarity, the final aspect representation is obtained by averaging the embeddings of the words associated with the aspect under analysis.Wang et al. [2020] iterates on this by taking into account the different types of relationships from the dependency tree. By anchoring the dependency parser to the aspect words, their approach generates representations from neighboring words within the tree. These representations are generated using a tailored graph attention network (GAT). Zhang et al. [2019] employs an architecture similar to that of Sun et al. [2019], but introduce a common proximity heuristic by applying weights that force the model to consider the distance between words for polarity prediction. This modification is based on the observation that sentiment expressions tend to be

localized around the aspect. Similarly, Zeng et al. [2019] split the representation computation by distinguishing between global and local contexts. They introduce a Semantic Relevance Distance to limit the number of tokens considered around the aspect for local context representation. The global and local sequence representations are then concatenated and integrated using a multihead attention layer, followed by a pooling layer to determine the final sentiment prediction. More recently, Dai et al. [2021] question the need to include dependency structure in model inputs. They show that there's no clear advantage to adding dependency structure to Bert-based model inputs. First, they show that a form of dependency tree can be derived from these models by analyzing how the representation of the jth token changes when the ith token is masked. They find that incorporating these tree structures as additional input during fine-tuning yields better results than using dependency structures from classical parsers referenced in the literature. They conclude that transformers inherently have some sense of dependency structure from their pre-training. Furthermore, they show that using the simple sentence pair classification model without dependency structures often achieves state-of-the-art or near state-of-the-art results compared to methods that incorporate structured extensions.

Pretraining large language models is crucial for them to learn valuable representations for downstream tasks. Some approaches aim to improve these representations by continuing the pre-training process on specialized datasets (a process refered to as Domain Adaptation or Continual Pretraining). In this context, Rietzler et al. [2020] applies the standard sentence pair classification objective from Devlin et al. [2019], but improves model performance by further pretraining on related corpora (Yelp and Amazon reviews) in an unsupervised manner. Similarly, Xu et al. extends the pretraining on these corpora to include the Next Sentence Prediction (NSP) goal. It also fine-tunes BERT on the Squad dataset to improve his reading comprehension.

Recently, several methods have attempted to exploit the zero-shot capabilities of language models, especially generative models. A notable example is Seoh et al. [2021], which adapts the ABSC task to a gap-filling format, building on the initial concepts of Schick and Schütze [2021]. This approach effectively uses the token prediction skills learned in pretraining. They develop a verbalizer and evaluate the performance of GPT-2 and BERT under zero-shot, few-shot, and fully supervised conditions, demonstrating a modest improvement over traditional sentence-pair modeling. Scaria et al. [2023] employ an instruct-tuned language model derived from a T5 model, which was trained by AllenAI [Wang et al., 2022] using the Super-NaturalInstructions dataset. They adapt this model to perform ABSA by transforming the Restaurant and Laptops datasets into an instructional format that includes task definitions and examples as input. They report robust performances, particularly in cross-domain settings, while maintaining a constrained model size. Zhang et al. [2023] introduce a new benchmark for evaluating language models in sentiment analysis, assessing their performance at different levels of sentiment granularity. They find that large language models are effective and perform well in zero and few-shot scenarios. However, in more complex, supervised contexts, smaller language models significantly outperform their larger counterparts. This finding is also corroborated by Wang et al. [2023], who compare ChatGPT's performance in a zero-shot setting against a supervised BERT model.

Datasets

Most datasets used for aspect-based sentiment classification focus on online reviews of products, restaurants, hotels, cosmetics, or microblog messages, primarily available in English and Chinese [Pontiki et al., 2014, 2015, 2016b; Nakov et al., 2016; Jiang et al., 2019; Wang et al.,

2019; Kiritchenko et al., 2014; Severyn et al., 2016]. These types of reviews tend to be more explicit than news articles, where journalists often express themselves more implicitly due to the expectation of maintaining a certain degree of neutrality. Current datasets for aspect-based sentiment analysis in news are scarce; Steinberger et al. [2017] focuses on European English news with few samples, while Hamborg and Donnay [2021] introduced a more comprehensive American-centric English dataset with approximately 11,000 samples.

2.7 Proposed Approaches

In this media bias detection literature, we can observe several critical points among the developed methods. They are limited in their ability to generalize, especially across time, domains, languages, and cultures. Most of them are heavily dependent on English-language or US-centric resources, and the effectiveness in automatically translating them to other languages is questionable. In addition, many supervised approaches suffer from a lack of explicability, poor granularity, and seem prone to spurious correlations, making them less attractive from a news media analysis perspective. Given these challenges, and given our focus on the analysis of the French press landscape, we aim to develop and investigate methods that specifically address these critical issues: improving cross-domain and cross-lingual generalization, reducing the costs associated with annotation, and improving explainability through more granular analysis. These improvements are essential for creating more robust and adaptable media bias detection tools that are better suited to diverse and evolving media environments.

For the purposes of this thesis, two approaches have been retained, but only the second has been thoroughly explored:

2.7.1 Distant Supervision using Twitter as Signal

The idea is to develop a methodology similar to the one proposed by Li and Goldwasser [2019], which uses a heterogeneous graph constructed from Twitter data in addition to textual data. This graph consists of three types of nodes-politicians with explicitly known positions, users, and articles-and two types of relationships: follows and shares. For each article, two different representations are generated: one by a graph convolutional network and another from textual features. The approach explores two scenarios: a supervised one with labeled articles and another relying only on labels for politicians. In particular, a compelling aspect of this method is the incorporation of a loss term designed to align the two article representations. This alignment potentially enhances the ability of the textual encoder to incorporate social information present in the social network.

Building on this concept seems even more promising when considering the findings of [Morales et al., 2021]. Their study applied Correspondence Analysis to reduce the dimensions of the follower matrix of French MPs, revealing that the first three dimensions correlate with the partisan political stances on various issues from the Chapel Hill Expert Survey. This suggests that it may be possible to finely characterize the political stance of articles shared on Twitter by analyzing the profiles of the users who shared them.

This observation reinforces several criticisms of the limitations of Li and Goldwasser [2019]. First, the outdated textual feature encoding they used is a pragmatic but less compelling problem; updating their experiments with modern transformational models is advisable. Second, their approach simplifies the problem into a three-class prediction model (“Right”, “Center”, “Left”), which

is probably unsuited to capture the nuanced political information available in social networks. Third, this oversimplified model may even lead to a loss of detailed political stance information within the graph representations of articles.

An idea we wanted to explore is to use a contrastive learning approach to directly align article representations with graph-based representations without relying on any labels. This method could have addressed the range of challenges identified in our literature review:

- **Generalization across Time**: Social media such as Twitter are organically updated and thus comprise latest information.
- **Generalization across Domains**: We're restricted to the type of content shared on Twitter.
- **Generalization across Language**: Twitter is used in many different countries.
- **Generalization across Culture**: Twitter is used in many different countries.
- **Annotation Cost**: Probably close to zero.
- **Granularity**: Potentially high, depending on the effective finesse of political stance we manage to extract from the social graph.

Unfortunately, as noted in the previous section, the acquisition of Twitter by Elon Musk on October 27, 2022, resulted in the termination of the academic API, making this approach costly under the current API fees. This is due to the need for extensive social graph data and a maximum number of user shares per article to accurately represent their stance. An initial study was conducted to explore the feasibility of using the Twitter graph to determine the political stance of media outlets. However, the data from this experiment could not be used to build article-level representations due to the low number of shares per article in the dataset: 85% of articles were shared only once, and 95% were shared less than twice. This preliminary study is provided in the Appendix A and was presented as part of the Infox sur Seine Workshop in 2023.

2.7.2 Stance Detection through Targeted-Sentiment Analysis

This line of research builds on the methods discussed in the subsection 2.5.2 on "Stance Detection through Entity-based Sentiment Analysis". It focuses on using the sentiments expressed toward key entities to determine the stance of an article or outlet. The full design of this approach is outlined in chapter 3. Compared to other media bias detection methods, this strategy is interesting under the selection criteria we have established:

- **Generalization across Time**: Shifts in sentiment expression can be expected to occur at a pace comparable to that of semantic shifts, which are slow.
- **Generalization across Domains**: More research is needed in this area, but a great deal of information about sentiment expression seems to be domain independent. We notably explore this aspect in Chapter 4.
- **Generalization across Language**: The method can be adapted to different languages, although the effectiveness of automatic dataset translation or cross-lingual transfer remains to be examined. This topic is further explored in Chapter 4 and Chapter 5.

- **Generalization across Culture**: Sentiment expression primarily presents a translation challenge, though cultural differences may exist. This aspect appears more amenable to cross-cultural transfer than other concepts such as left-right partisanship.
- **Annotation Cost**: The question of the annotation cost is central. As explained in Subsection 2.3.1, there's plenty of factors that intervene in the viability of an annotation based method, among which the complexity of the annotation task, the required expertise, and generalizability of the model learned on this task. We further address this issue in Chapter 4.
- **Granularity**: Relies on the precision of the method developed. Outcomes from this approach appear to be explainable with a satisfactory level of detail.

The detailed presentation of the work related to this approach is outlined in Chapters 3, 4, and 5.

Chapter 3

General News Analysis Framework

Contents

3.1 Introduction 48
3.2 Existing News Analysis Frameworks 48
3.3 A General Framework for News Analysis 49
3.3.1 Source of News 50
3.3.2 Filtering Step 50
3.3.3 Sentencizer Step 50
3.3.4 Mention Detection 50
3.3.5 Entity-Linking 51
3.3.6 Enrichment 51
3.3.7 Sentiment Classification 51
3.4 Instantiation for Political News Analysis 51
3.4.1 Source of Articles 52
3.4.2 Filtering Steps 52
3.4.3 Entity-Linking 52
3.4.4 Enrichment 53
3.4.5 Sentiment Classification 54
3.4.6 Analyses 54
3.5 Conclusion 54

3.1 Introduction

This chapter outlines the general news analysis framework proposed in this thesis, which is later specifically applied to political news analysis. We begin by reviewing various existing news analysis frameworks and highlighting how our approach differs. As noted in Chapter 2, the primary limitations of current news analysis techniques include their limited generalizability across languages, cultures, and time, and often, their inadequate explainability due to a coarse-grained approach to news characterization. This chapter attempts to address these limitations and thus focuses on answering our first research question **RQ1**, that is, "How to design a framework that allows for fine-grained analysis of opinions expressed in text with greater explanability?".

3.2 Existing News Analysis Frameworks

We aim to develop a comprehensive framework for news analysis, as discussed in Chapter 2. This framework should facilitate the extraction of insights from the textual content of news articles and outlets, support multilingual transferability, cross-domain and temporal generalization, and provide a satisfying degree of transparency and explicability.

In section 2.5, we discussed several methods used in computer science for detecting media biases, most of which are specifically tailored to the political domain. For instance, automatic framing detection extends the political communication concept of framing, and most methods primarily use the annotated dataset of Card et al. [2015], which is designed for political news. Similarly, stance bias detection typically aims to identify partisan cues in articles. Although other methods such as phrasing and social category biases are likely more applicable to other domains such as health, sports, or finance, they fall short in identifying stances on divisive issues. Focusing on entity-centric stance detection, by concentrating on expressed sentiments, may provide a more generalizable approach that improves cross-time generalization, explainability and cross-lingual transferability. Chapter 4 addresses these different points.

Several general news analysis frameworks have been proposed in literature. Lloyd et al. [2005] develop a news analysis system called Lydia/Textmap that processes daily American newspaper publications. Their approach uses entity recognition, entity disambiguation, and newspaper geolocation to facilitate several types of analyses, notably what they call "juxtaposition analysis" which assesses the relatedness of entities from their co-occurence statistics. Their analysis doesn't involve any stance detection as they rely fully on objective metrics. Godbole et al. [2007] extend the news analysis system of Lloyd et al. [2005] by adding a sentiment dimension through the use of a lexicon. They bootstrap the sentiment lexicon through adjective co-occurrence heuristics, considering negation as a polarity reversal, and linking sentiment expressions to entities present within the same sentence. Despite these refinements, the approach has limitations: the polarity reversal is somewhat elementary, the association method remains imprecise, and the method is limited to the analysis of English content. Bautin et al. [2021] tackle this problem by iterating on the solution of Godbole et al. [2007] to adapt it for multiple languages. They use automatic translation to convert news articles from the target language to the source language before applying the English models. They acknowledge cultural variations in sentiment expression and adjust sentiment scores for each language to accommodate these differences. Ward et al. [2009] propose adding demographic analysis to investigate ethnic representation discrepancies, using a Wikipedia-based knowledge classifier that associates $(name, surname)$ tuples with ethnicities. This method is later adopted by Asr et al. [2021] for a general news gender analysis framework, which assigns the most likely gender based on names using a training set derived from U.S. census

data. However, this approach has notable shortcomings, including its binary treatment of gender and its primary suitability for the Anglo-Saxon context. More recently, Hamborg [2023b] presents a system for comparing coverage differences across a set of articles using targeted sentiment classification on key entities within the texts. They use cross-document coreference resolution to align entities across articles and aggregate sentiment scores, normalized by their frequency, to cluster articles into three perspectives. While this approach yields interesting results, it has limitations: it requires manual collection of articles on the same event, relies on English-only resources for co-reference resolution and sentiment classification, and doesn't incorporate entity linking or external knowledge for more informed analysis, which also limits aggregation and analysis at the outlet level.

3.3 A General Framework for News Analysis

Our aim is to create a news analysis framework to address **RQ1** and **RQ2**, whose perspectives are:

- **RQ1**: How to design a framework that allows for fine-grained analysis of opinions expressed in text with greater explanability?
- **RQ2**: How generalizable can we make this opinion detection pipeline to facilitate transfer to other languages and other domains?

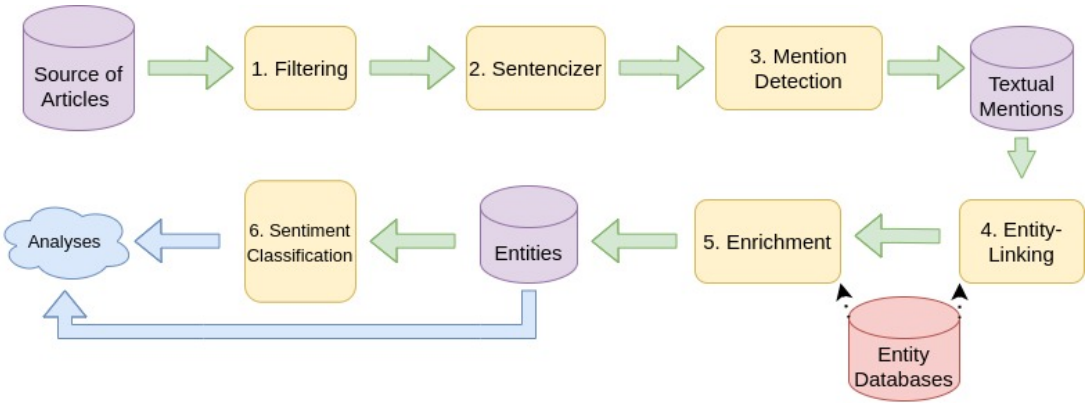


Figure 3.1: Different steps of our General framework for news analysis

We propose the general framework for news analysis shown in Figure 3.1. This framework consists of six processing steps designed for fine-grained analysis of news articles and outlets, incorporating both opinion and demographic perspectives. By measuring sentiment classification scores at the mention level, we ensure high granularity and explainability. Steps 1 through 5 are chosen because they are applicable to multiple languages, since many multilingual resources exist for them, and their cross-domain application should allow for adequate performance (each step is thoroughly presented later in this chapter); the generalizability of the sentiment classification component is explored in chapter 4. The effectiveness of such a framework is difficult to evaluate quantitatively, we offer a qualitative evaluation of it for the political domain analysis through an extensive analysis of the French news landscape in Chapter 5. We are the first to offer a multilingual news analysis framework that combines wikification and sentiment classification to uncover bias in news coverage at the outlet level.

3.3.1 Source of News

Articles can come from a variety of sources, including standard scraping methods, the CommonCrawl corpus, or its subset CCNews[6]. Below we outline the different processing steps, from raw article collection to detailed opinion analysis.

3.3.2 Filtering Step

In the filtering step, we focus on excluding irrelevant sources, such as specific domain names, subdomains, or languages. This step includes verifying articles' languages with Fasttext [Joulin et al., 2017], verifying presence of textual content, and removing boilerplates such as cookie notifications or subscription banners. In addition, we perform near-deduplication per domain name to avoid redundant content using the MinHashLSH algorithm [Leskovec et al., 2020, p.73] with 256 permutations and set to a threshold that is likely to eliminate articles with more than 50% similarity in their word-based 5-grams. This deduplication is crucial to prevent redundant article contributions when aggregating and deriving bias metrics. Moreover, this deduplication approach has the advantage of being language agnostic, unlike semantic embedding-based methods [Abbas et al., 2023].

3.3.3 Sentencizer Step

In this step, we segment the text of the article into individual sentences for subsequent processing. There are several sentence splitting libraries, of which NLTK's "Punkt" [Bird et al., 2009] and Spacy's "sentencizer" [Honnibal and Montani, 2017] are the most popular. Combining their unique language supports allows us to handle 29 languages. For specific and under-resourced languages, we can use the closest available language which possess a sentencizer. In this operation, we favor Spacy to NLTK, because the former empirically outperforms the latter on the sentence splitting task.

3.3.4 Mention Detection

Mention detection is the Aspect Term Extraction (ATE) task of our sentiment analysis process (see Section 2.6), but also allows the derivation of our objective frequency-based metrics. In news analysis, we focus on extracting entity types commonly identified in Named Entity Recognition (NER) tasks, including categories such as person, organization, and location. While datasets exist for many languages, some languages don't have any public NER extractor and require training one. Even for languages without suitable NER datasets, recent advances in multiligual NER approaches and machine translation are increasingly helping. A notable development is the introduction of multilingual NER models such as GliNER [Zaratiana et al., 2023], which effectively generalizes to the pre-training languages of its base model, mDebertaV3 [He et al., 2023], which is trained on over 100 languages from the CC100 corpus [Conneau et al., 2020].

Advances in translation models such as NLLB [Team et al., 2022], which supports translation between 200 languages, and new strategies such as the EasyProject [Chen et al., 2023], which enables translation with span markers, provide two routes for extending support to under-resourced languages. These include translating from the target language to English to use English NER recognition models and then translating back, or translating English NER datasets into the target

[6] https://commoncrawl.org/blog/news-dataset-available

language to develop custom models. In our analyses, we either utilized existing NER models or retrained monolingual ones for under-resourced languages.

3.3.5 Entity-Linking

For entity linking, we perform the traditional task of Wikification [Cheng and Roth, 2013], since Wikipedia represents a comprehensive knowledge base of most entities mentioned in the news and is also available in 300+ languages. Two types of models coexist: Encoder-based models, that contextually embed each mention and match them with the most similar embedded entities contained in an index, this operation is sometimes followed by a re-ranking of candidate entities using a cross-encoder that acknowledges the entity and the context [Wu et al., 2020c]. More recently, autoregressive constrained models have also been employed. This approach involves generating the reference of the contextualized entity that is given as input to the model [De Cao et al., 2022]. Encoder-based linking offers the benefits of reduced computational requirements through a one pass inference process, unlike autoregressive models, some recent variants even allow inferring multiple entities in one pass [Plekhanov et al., 2023]. Autoregressive models, on the other hand, eliminate the need to consult an index to retrieve references for the most probable entities, but are inherently more costly at inference time due to the multi-step constrained generation process. At the time of the framework conception, the top multilingual entity-linking model was the autoregressive mGenre [De Cao et al., 2022]. More recently, the release of the Bela model [Plekhanov et al., 2023] introduced an encoder-based linking approach that allows single-pass inference for all entities within a sentence, which should results more being efficient.

3.3.6 Enrichment

Once the linking has been made, the possibility to enrich the analyses by retrieving entity metadata from the Wikidata knowledge base allows notably to perform demographics analyses. Other databases can also be used to enrich the representations of these entities based on their Wikidata attributes. This allows to derive results along different classes of attributes to get refined insights into the representation of subgroups in the news.

3.3.7 Sentiment Classification

In our pipeline, sentiment classification is performed through target-based sentiment classification models (see Section 2.6). The objective is to determine if a given sentence mentions negatively, neutrally or positevely a given entity. It's through significant differences of sentiment towards key entities that we compare the stance of sets of articles and outlets in the analysis part.

3.4 Instantiation for Political News Analysis

In Figure 3.2, we show how the general framework shown in Figure 3.1 is instantiated to analyze political news of a given country. The first step is to select relevant news sources for the country under study. Next, select and enrich the representations of identified political entities before proceeding to the sentiment analysis phase. We detail the adaptations to the filtering, enrichment, and sentiment classification steps accordingly.

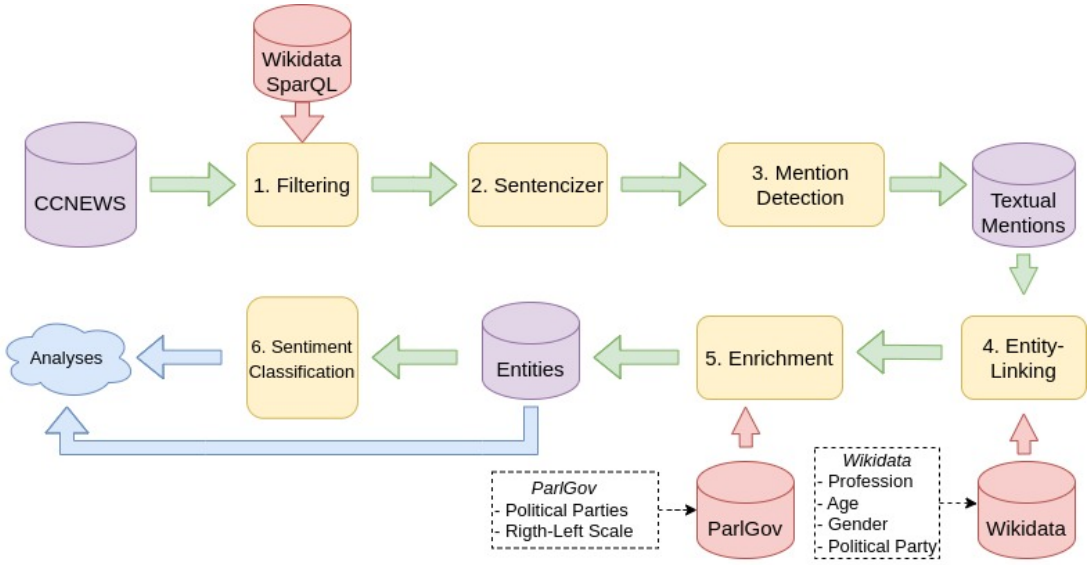


Figure 3.2: Instantiated framework for political news analysis

3.4.1 Source of Articles

This framework is widely applicable to many countries thanks to initiatives like CCNews, a subset of the CommonCrawl project. Its outlet selection is performed using URLs from the DMOZ[7] initiative, which includes approximately 50k domain names of outlets and covers over 50 languages. Although the framework can in practice be instantiated using any custom news source, here we illustrate the following processing steps with CCNews.

3.4.2 Filtering Steps

For country-specific analysis, we first filter the CCNews domain names for each country using a list of valid newspapers obtained using the wikidata SPARQL query of Figure 3.3. We supplement this first list with other domain names extracted from the w3newspapers site[8]. After obtaining the list of candidate domain names, we identify the most common website extensions for each country and use them to further select domain names present in CCNews that are not listed in Wikidata but are related to that country.
An additional pre-filtering step can be implemented to reduce the amount of text to be processed by training a classifier to quickly exclude non-political articles (see subsection 5.5.1).

3.4.3 Entity-Linking

Linking across the entire Wikipedia database is particularly slow. By focusing on a predefined subset of relevant entities, we can significantly reduce the number of potential links, thereby speeding up the process. As mentioned previously, we use the mGenre model [De Cao et al., 2022] for linking mentions to Wikidata entities, but optimize performance and memory in two ways. First, we restrict the set of entities to referenced politicians or party-affiliated entities in Wikidata, reducing the number of entities from 63M to 3M, which significantly reduces the size of

[7]https://web.archive.org/web/20180125073812/http://dmoztools.net/
[8]https://www.w3newspapers.com/

```
SELECT DISTINCT ?item ?itemLabel ?lang ?url
WHERE
{
    # Select items that are instances of or subclasses of "newspapers"
      ?item wdt:P31/wdt:P279* wd:Q11032.
    # that have an official website
      ?item wdt:P856 ?url.
    # that are based in the following country
      ?item wdt:P17 wd:Q28.
    # retrieve the language of the newspaper
      ?item wdt:P407 ?language.
    # retrieve the lang code of the newspaper
      ?language wdt:P220 ?lang.
      SERVICE wikibase:label { bd:serviceParam wikibase:language \
      "en,de,ru,fr,es,it,ja,zh,*" }
}
```

Figure 3.3: SPARQL query for Wikidata newspaper query

the tree and increases the generation speed by a factor of 5. Second, we adapt the functionality of the tree. The model, originally trained to generate the corresponding Wikipedia page title of the entity with the language of the page, has been biased into predicting the language associated with the language of the input text. When confronted with an entity that has a Wikipedia page but not in the language of the input text, the perplexity (i.e. the exponentiated inverse probability of the model's predictions, a measure of model's uncertainty) of the generated reference is greatly increased because of the non-matching languages, even though the linked entity is often correct. To address this, we decouple the input language from the entity linking by allowing arbitrary language codes to be generated during inference. In this setting, we experimentally verify that matching based on the perplexity of the generated name vs. generated name + language code leads to the same results. Accordingly, we revise the matching dictionary to depend only on the name of the entity without the language code of its original page. While these improvements can speed up the generation process, autoregressive inference linking remains resource-intensive.

3.4.4 Enrichment

To effectively analyze media bias, we can use entities' Wikidata attributes and augment them with information from other external databases. In particular, we want to explore subgroups that reflect different political leanings. Using data from the Wikidata dump, we extract the political parties of matched entities using the "member of a political party" from the "P102" Wikidata property. We semi-automatically align these party names with those in the ParlGov[9] database for the corresponding country. This alignment gives us a score for most political parties from OECD countries on a 10-scale left-right axis, which allows us to measure the frequency of political groups, and later, the overall sentiment expressed towards them in specific sets of articles.

[9]https://www.parlgov.org/

3.4.5 Sentiment Classification

We've mentioned in Subsection 2.6.3 the scarcity of datasets for targeted sentiment classification in the news domain, in a task where cross-domain and cross-lingual performance pose significant challenges. The elaboration of this step component and the exploration of its generalizability in these scenarios are discussed in chapter 4.

3.4.6 Analyses

Our analyses of political news articles allows a multifaceted understanding of the news media landscape. It identifies demographic characteristics of the mentioned entities, such as their age, geographic location, and socioeconomic status. It also allows us to derive their political parties. These objective frequency-based metrics are enriched by the use of sentiment analysis, which assesses the sentiment expressed toward the mentioned entities, allowing us to determine the biases of news outlets.

3.5 Conclusion

In this chapter, we introduced a novel general framework for news analysis that integrates both objective and subjective metrics to create a thorough representation of the news media landscape. By employing entity linking with Wikidata, we enhance the representations of the detected mentions with their attributes. While the targeted-sentiment classification module should allow for an explainable characterization of stance, several challenges remain for this framework to work effectively within the constraints of our study: ensuring good cross-lingual and cross-cultural generalizability, improving cross-domain transfer, and maintaining a high level of explainability. These constraints and their applicability are discussed in Chapter 4, and the framework is instantiated in Chapter 5 for the analysis of the French news landscape, allowing its qualitative evaluation.

Chapter 4

Targeted Sentiment Analysis for Political News

Contents

4.1 Introduction 56
4.2 Related work 57
4.3 Dataset Construction 58
4.3.1 Data Sources 58
4.3.2 Sample Selection 59
4.3.3 Sample Annotation 60
4.3.4 Label Aggregation 60
4.3.5 MAD-TSC Dataset Characteristics 62
4.4 Experiments 63
4.4.1 Compared TSC Methods 64
4.4.2 Dataset Splits and Metrics 65
4.4.3 Experiments with MAD-TSC and NewsMTSC 66
4.4.4 Experiments with Individual Languages 66
4.4.5 Cross-lingual Experiments and Machine Translation 67
4.4.6 Analysis of Example Complexity 68
4.4.7 Ablation of Training Set 68
4.4.8 Cross-Domain Experiments 69
4.5 Limitations 71
4.6 Conclusion 71
4.6.1 Perspectives 72

4.1 Introduction

In Chapter 3, we detailed the steps of our processing framework tailored for political news analysis, highlighting the problem of resource scarcity for performing targeted-sentiment classification in the news domain and in languages other than English. This scarcity is problematic because sentiment expression is known to vary across data sources, languages, and text domains, making cross-lingual and cross-domain transfer difficult. This scarcity is widely observed in other sentiment analysis tasks, but is even more challenging in target-dependent sentiment classification (TSC), where we rely on the understanding of the context relationship to the entity to identify the sentiment expressed towards it.

Currently, the bulk of TSC-related research efforts are focused on major languages. This focus is an effect of the availability of generic and of task-specific resources in these languages [Brauwers and Frasincar, 2022; Nazir et al., 2020]. A majority of datasets are monolingual and, when they are multilingual [Balahur and Turchi, 2014; Barriere and Balahur, 2020; Cortis and Davis, 2021; Pontiki et al., 2016a; Severyn et al., 2016], the examples are not aligned across languages, preventing a fair comparison of cross-lingual model performances. Equally, a majority of existing datasets and methods are devised for texts such as tweets, reviews or comments [Nakov et al., 2016; Pontiki et al., 2016a; Severyn et al., 2016] in which sentiment is most often expressed explicitly. Somewhat surprisingly, less attention is given to TSC in the news, despite the usefulness of an automatic analysis of its content for the understanding of societally impactful processes such as disinformation or polarization [Hamborg and Donnay, 2021].

The main contribution of this chapter has been the subject of one of our publications [Dufraisse et al., 2023a], it consists in the introduction of MAD-TSC, the first large multilingual aligned dataset for TSC in news articles. It includes 5,110 annotated entity mentions from 4,714 unique sentences. Each sentence has professionally-translated and aligned versions in eight languages (English, Spanish, German, Italian, French, Portuguese, Dutch, and Romanian). Sentences originate from 286 news sources published in over 30 countries. These characteristics differentiate the proposed dataset from existing resources, and in particular from NewsMTSC [Hamborg and Donnay, 2021], a monolingual dataset focused on American politics which is the closest to MAD-TSC. We first present the dataset creation process, and provide a qualitative analysis of its content. Then, we propose a comprehensive evaluation of state-of-the-art TSC methods on MAD-TSC to validate the cross-lingual transferability of TSC models for our news analysis framework in Chapter 3. Multilingual experiments are performed, with a particular focus on the usability of machine translation in TSC. Preliminary results in cross-domain scenarios are also presented. The main findings are the following:

- The proposed dataset is more challenging since it includes more complex examples compared to existing resources, as detailed in Subsection 4.3.5.
- Performance for individual languages varies due to the fact that the quality of available pretrained language models is itself variable, with the best scores being obtained for English.
- Results with machine translation of training sets from English toward target languages and with manual translations are on par.
- The performance level for other languages matches that of English by translating test examples to English in order to take advantage of the pretrained language models available in this language. The same is true when both the training and test sets are automatically translated to English.

- Preliminary results in cross-domain transfer show that transferability between domains is difficult, especially from specialized to general domains.

The last two findings are particularly interesting since they show that if a domain-specific TSC dataset is available, it can be effectively used for multiple languages. Overall, the introduction of MAD-TSC will facilitate progress in multilingual TSC. The dataset and the full evaluation protocol are available online[10] to encourage further research, and to ensure reproducibility.

4.2 Related work

Target-dependent sentiment classification (also named aspect-based sentiment analysis [Nazir et al., 2020] or classification [Brauwers and Frasincar, 2022]) is a complex task due to the numerous factors which influence the way sentiments are expressed in texts [Brauwers and Frasincar, 2022], such as the language, the domain or the personal biases of the author/reader.

TSC is often evaluated on short texts such as tweets [Nakov et al., 2016], reviews [Pontiki et al., 2016a] or comments [Severyn et al., 2016]. A common characteristic of these resources is that sentiment is often expressed in an explicit way. News are more challenging texts because sentiments are expressed implicitly or indirectly [Hamborg and Donnay, 2021], often include multiple targets in a single sentence [Brauwers and Frasincar, 2022], and both negative and positive arguments about a target entity are combined due to the fact that journalists are supposed to be objective [Balahur et al., 2010; Hamborg et al., 2019b; Liu, 2010]. Multilinguality is important in order to be able to analyze texts in different languages.

Multilingual datasets are proposed for tweets [Lampert and Lampert, 2021; Vilares et al., 2017], reviews [Jiménez Zafra et al., 2015; Pontiki et al., 2016a] and institutional texts [Cortis and Davis, 2021]. While interesting, these datasets differ from MAD-TSC because they do not focus on news. Equally important, they are only aligned at a domain level, but not at an example level. This limits their utility in terms of comparative evaluation in monolingual and multilingual settings. Multilingual approaches were also explored for news representation. For instance, bilingual word embeddings were used to compensate data scarcity in under-resourced languages [Akhtar et al., 2018] or to transfer models between source and target languages in zero-shot settings [Jebbara and Cimiano, 2019].

Classical TSC methods rely on engineered features based on lexicons and syntactic analysis [Biber and Finegan, 1989; Baccianella et al., 2010; Jiang et al., 2011; Mohammad et al., 2016a; Vo and Zhang, 2015]. Strong progress in TSC was made possible by the introduction of deep language models, such as BERT [Devlin et al., 2019; Zeng et al., 2019]. Improvements are obtained when pretraining includes a larger proportion of news [Hamborg and Donnay, 2021] – this is the case for English RoBERTa [Liu et al., 2019b] or XLNET [Yang et al., 2019] – or when including an intermediate tuning on domain-related data [Du et al., 2020]. Naturally, further improvements are obtained by introducing TSC-specific architectures [Brauwers and Frasincar, 2022; Nazir et al., 2020; Zhou et al., 2019]. We follow this trend and use pretrained language models in our experiments.

To our knowledge, there are only four datasets dedicated to TSC in the news. The first one [Balahur et al., 2010] has 1,592 examples, and includes only quotes from political news. Quotes are interesting because they include a lot of sentiment expressions, but they are also easier since sentiment is often expressed explicitly [Hamborg and Donnay, 2021]. The second one [Steinberger et al., 2017] has 1,274 examples. The size of these datasets makes them difficult

[10] `https://github.com/EvanDufraisse/MAD_TSC`

to use with deep-learning-based TSC methods. The third dataset [Hamborg et al., 2019b] includes 3,002 examples. However, as noted later by its authors [Hamborg and Donnay, 2021], the dataset is imbalanced and its sentiment expressions are predominantly explicit. The dataset which is closest to ours is NewsMTSC [Hamborg and Donnay, 2021]. Their common characteristics include: a focus on political news, an identical definition of the task, and a similar annotation process. Importantly, we follow Hamborg and Donnay [2021] and instruct annotators to think from the author's perspective in a holistic way. They are instructed to consider the "what" of the sentence (events, facts) but also the "how" (choice of words, author's attitude). This choice contrasts with previous works [Balahur et al., 2010; Steinberger et al., 2017], which distinguish author- and reader-levels in TSC, and is important in order to minimize the influence of personal biases. The main differences between MAD-TSC and NewsMTSC relate to multilingualism, complexity of examples, political topics and geography of examples, while maintaining the same order of magnitude in the number of samples, with 5,110 for MAD-TSC , and 11,361 for NewsMTSC. These differences are detailed in Subsection 4.3.5, and they make MAD-TSC appropriate for a thorough evaluation of TSC in multilingual settings.

4.3 Dataset Construction

We build on previous works to develop sentiment classification datasets [Nakov et al., 2016; Pontiki et al., 2016a], particularly those focused on TSC datasets [Balahur et al., 2010; Hamborg and Donnay, 2021; Steinberger et al., 2017]. This includes the typical stages of dataset creation: collecting data, identifying candidate samples for annotation, performing the annotation process, and consolidating the results. We describe the main steps of this dataset creation process and analyze the resulting dataset.

4.3.1 Data Sources

Our objective is to create a dataset which:

1. is multilingual and aligned at a sentence level across languages to enable a comprehensive evaluation of TSC in cross-lingual setting.

2. includes content from a large number of high-quality journalistic sources, which offer a diversified view of the included topics.

3. covers societally-impactful topics in different countries.

In this perspective, Voxeurop[11], as a multilingual news website dedicated to providing interesting and high quality news to European audiences, is a valuable resource.

The project inherits from Presseurop, which ran from 2009 to 2013 and aimed to make European news from over 200 sources available in as many European languages as possible. Voxeurop and Presseurop articles can be available in up to ten languages[12].

The articles are translated by professional translators to ensure high quality texts in all available languages. In addition, the content is published under a Creative Commons BY-NC-ND open license, which facilitates its redistribution and reuse for non-commercial purposes, and which is also used to distribute MAD-TSC.

[11] `https://voxeurop.eu`

[12] English, Spanish, German, French, Italian, Dutch, Portuguese, Polish, Czech, and Romanian

In order to create a multilingual and aligned dataset, we limited our data collection to the following eight languages: English (EN), Spanish (ES), German (DE), Italian (IT), French (FR), Portuguese (PT), Dutch (NL), and Romanian (RO). We then collected 7,370 news articles with translations in these languages, resulting in a total of 122,263 English sentences, with similar amounts in the other languages.

The vast majority of the examples are related to politics, with the others pertaining to business, culture and society. Most of the entities mentioned in articles refer to prominent public figures from different European countries at the time of publication (2009–2013). Well-represented political sub-topics include: the economic crisis which started in 2008, European Union evolution process, election-related news, political crisis at a national level, and corruption scandals in different countries.

4.3.2 Sample Selection

We use English as the pivot language for selecting candidate sentences because of the extensive availability of its resources for Named Entity Recognition and Entity Linking. Our goal is to identify mentions of political entities in sentences for the annotation process.

Named entity recognition was performed using the Flair model for English [Akbik et al., 2018], resulting in an initial pool of 30,303 candidate sentences containing at least one mention of a person. We use the encoder-based Blink [Wu et al., 2020c] model for entity linking and combine it with coreference resolution using neuralcoref[13] to obtain reliable entity counts in articles. This count helps to eliminate entities that are only mentioned once in an article, as they are not considered central and therefore unlikely to express a sentiment that is not neutral towards them. This filtering led to 19,223 candidate sentences. The alignment of candidate sentences in the eight languages is inspired by the multilingual book alignment tool lingtrain[14]. It is based on sentence embeddings from a multilingual-sentence BERT model [Yang et al., 2020], with English as source and the other languages as targets. We use two matching criteria for two sentences to be considered aligned: (1) the need for a reciprocal best match (inter-match) between source and target sentences, and (2) a cosine similarity threshold of 0.5. Both criteria need to be met for all eight language pairs in order to select a sentence in the final dataset. This automatic alignment method was manually checked for three languages (EN, FR, RO), with a sample of 1,000 examples. It was correct in 98.1% of cases. The remaining imperfections usually relate to additional context being provided by the translator in one of the languages. This does not affect the sentiment expressed about the target entity, and the obtained alignments of sentences are usable.

Following sentence alignment, it is also necessary to align entity mentions across languages. We used a rule-based approach for this task. We first applied NFKD unicode normalization to examples in all languages. We then computed a normalized Levenshtein distance between the English mention of the entity and the words from the target sentence. A similarity threshold of 80% between the English and the target mention in any of the other languages was used. To add flexibility to the matching process, we also considered nearly contiguous sequences as valid matches. This matching strategy implies that only proper nouns are matched, since Levenshtein distances generally cannot account for the semantic proximity of common nouns, especially in a multilingual setting. We have checked this matching and it is correct in over 99% of cases on a subset of 1,000 mentions.

[13] `https://github.com/huggingface/neuralcoref`
[14] `https://github.com/averkij/lingtrain-aligner`

Another problem is the unbalanced distribution of sentiment classes in the news, with the neutral class dominating by a large margin, so we followed an initial selection procedure inspired by the one introduced in Hamborg and Donnay [2021]. It involves an undersampling of potentially neutral mentions as predicted by a simple binary classifier. This led to a pool of 11,000 examples which were selected and proposed to annotators.

4.3.3 Sample Annotation

Annotations were crowdsourced using a custom web interface. Aware of the challenges of news annotations, which often lead to low inter-annotator agreement, Balahur et al. [2010]; Hamborg and Donnay [2021] devised a process in which participants are asked to annotate following the news author's perspective. While some news articles express sentiments in a manner which is easy to recognize, others convey them in an intricate and/or implicit manner. Equally, the sentiment often depends on text parts distant from the target entity, which are not included in the text presented for annotations. The annotation guide made the annotators aware of the complexity of the task. They were presented with examples of sentences which include intricate and/or implicit sentiments expressions, as well as irony. Annotators also had the possibility to label examples as unknown whenever they could not determine the label of an example.

TSC annotations are usually collected using 3-, 5- or 7-points Likert scales [Balahur et al., 2010; Hamborg and Donnay, 2021; Nakov et al., 2016; Pontiki et al., 2016a]. Following an initial experiment which involved 50 sentences, we opted for a 5-points scale which offers a good balance between annotation simplicity and expressiveness. Possible labels were: negative, weakly negative, neutral, weakly positive and positive. The annotation was supported by a Web interface which is illustrated in Figure 4.1. Examples were provided in English, French and Romanian to facilitate the annotation.

Annotations were provided by a total of 21 volunteer participants, whose demographics are presented in Appendix B.2. They were recruited via a call for participation which was circulated via group and personal e-mails. Participants provided explicit consent to use their annotations and demographic data at the beginning of the experiment. The choice to work with volunteers is motivated by the fact that crowdsourcing performance is similar for paid and volunteered participation [Mao et al., 2013]. Samples were presented randomly in order to avoid any ordering effect, and users were free to stop at any point. Each sample was labeled by three annotators in order to allow annotation consolidation. All users speak at least two of the three languages used in the annotation interface.

4.3.4 Label Aggregation

As noted in previous studies [Hamborg and Donnay, 2021; Steinberger et al., 2017], the annotation task is prone to disagreement and consolidation of annotations is essential. Therefore, we first excluded all samples containing at least one "unknown" label. Following Hamborg and Donnay [2021], we reduced the five initial labels to three classes (negative, neutral and positive) by aggregating the two possible labels for the negative and positive sentiments. Finally, we kept only samples for which there was an unanimous voting or majority agreement with a third vote in a neighboring class (for instance two positive, one neutral). The inter-rater reliability, measured using Fleiss' kappa [Fleiss, 1971], reaches $\mathcal{K}_F = 0.58$ and $\mathcal{K}_F = 0.67$, before and after consolidation, respectively. The two values indicate that the task is challenging, but that the final reliability score corresponds to a good agreement [Hallgren, 2012]. This consolidation strategy leads to a coherent labeling of MAD-TSC, and is used in experiments.



Figure 4.1: Page for an example of annotation.

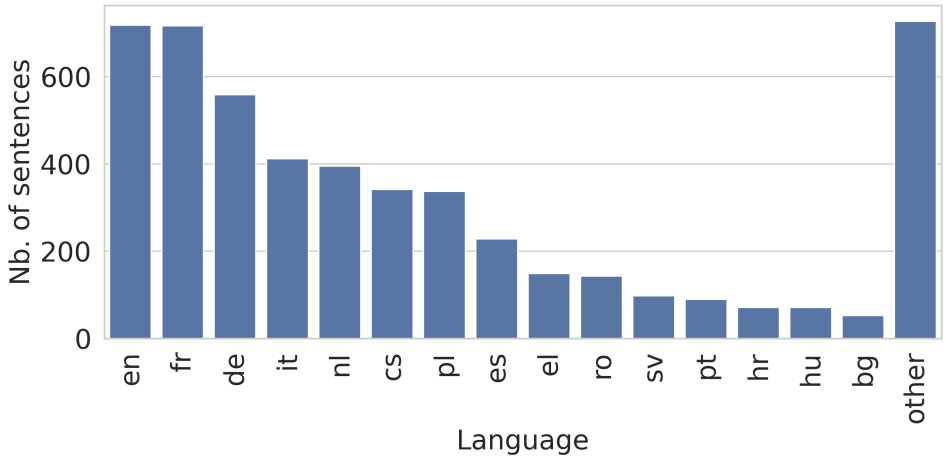


Figure 4.2: Distribution of linkable entities per original language of source article. "other" merges rare source languages. Linking is done with Blink [Wu et al., 2020c].

4.3.5 MAD-TSC Dataset Characteristics

MAD-TSC includes a total of 5,110 labeled target entity mentions for all eight languages, with 1,839, 2,011, 1,260 of them labeled as negative, neutral and positive, respectively. There can be more than one target entity per example, and there are 4,714 unique examples in MAD-TSC.

Figure 4.2 shows the total number of unique linkable entities, grouped by the language of the original article as indicated by the VoxEurop website. These sources are generally correlated to journals from different countries, indicating that the proposed dataset comes from a variety of sources. The total number of unique linkable entities aggregated across all source languages is lower, with 1,007 entities. This is in contrast with NewsMTSC, which is sampled only from American newspapers.

We illustrate the content of the dataset by a number of quantitative and qualitative characteristics. We first present example-related statistics for MAD-TSC, and compare them to NewsMTSC [Hamborg and Donnay, 2021]. Hamborg and Donnay [2021] notably use sentence length as a proxy for the complexity of the dataset, and showed that texts in NewsMTSC are longer than those from previous datasets. We follow their approach and report the statistics regarding the number of characters for English examples included in MAD-TSC. The mean is 192.3 characters (72.1 stdev), while the corresponding value for NewsMTSC is 152.2 (109.1 stdev). The number of words per example is a related measure, but more informative from a semantic perspective. MAD-TSC examples contain an average of 31.1 words (11.6 stdev), while the corresponding values for NewsMTSC are 25.2 (15.5 stdev). MAD-TSC can thus be considered more complex than NewsMTSC according to this proxy.

A second analysis focuses on entities which can be linked to English Wikipedia articles using Blink [Wu et al., 2020c]. MAD-TSC contains 1,007 distinct linkable entities, with a mean of 6.3 mentions per entity (28.9 stdev). NewsMTSC includes 975 linkable entities, 5.5 mentions per entity (43.9 stdev). Donald Trump, Hillary Clinton, and Barack Obama, the top-3 entities from NewsMTSC cover 20.1%, 11.7% and 7.7% of mentions, respectively. In MAD-TSC, Angela Merkel, Silvio Berlusconi, and Nicolas Sarkozy cover 10.2%, 4.9%, and 3.5% of mentions, respectively. We conclude that MAD-TSC has a distribution of entity mentions which is less skewed.

Third, we examine the geographic distribution of linkable entities in both datasets. The

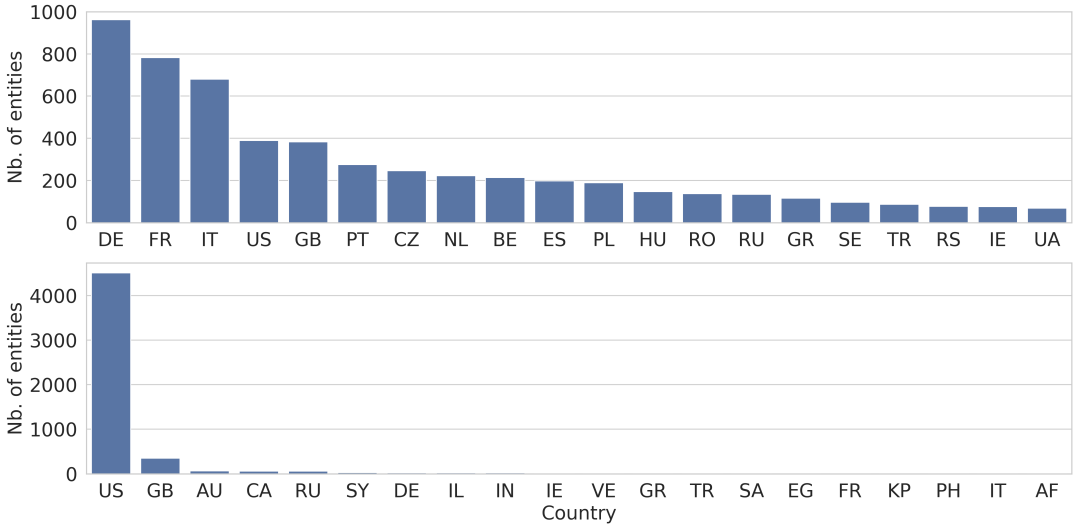
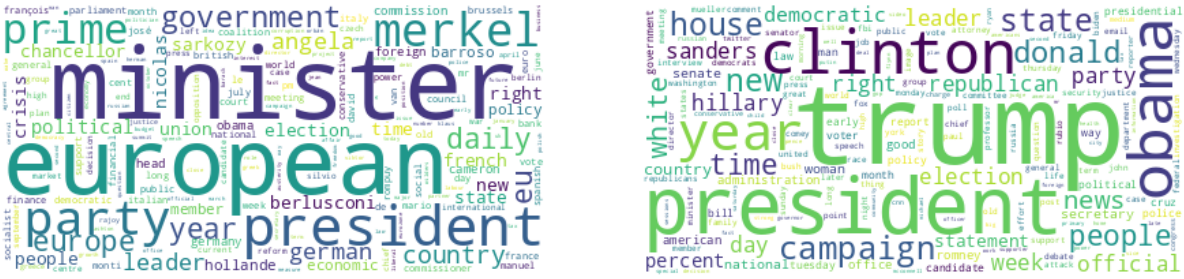


Figure 4.3: Distribution of linkable entities from MAD-TSC (top) and NewsMTSC (bottom) per country. The country associated to each entry is the most frequent one from the Wikipedia article of the entity.



(a) MAD-TSC dataset

(b) NewsMTSC dataset

Figure 4.4: Frequent word clouds for MAD-TSC and NewsMTSC corpora (minimum frequency 20 occurrences).

obtained distributions are presented in Figure 4.3, and they confirm that MAD-TSC is much more diversified than NewsMTSC from a geographical point of view.

Finally, we present clouds of frequent words in the two datasets in Figure 4.4. This qualitative representation of the two datasets confirms that the main topics are different, with focus on European and on American topics, respectively.

4.4 Experiments

With this new dataset, we perform several experiments. In Subsection 4.4.1, we evaluate the performance of different TSC model architectures and assess the domain transferability and complementarity between NewsMTSC and MAD-TSC. The aligned nature of our dataset also allows us to perform extensive multilingual experiments, which will also help us to validate the transferability of the sentiment analysis component of our news analysis framework approach, presented in Chapter 3, to other languages. In Subsection 4.4.4, we compare the results of

monolingual models and a multilingual model, observing significant disparities between the monolingual ones and mitigated interest in using the multilingual one. In Subsection 4.4.5, we show that the differences arise from undertraining or domain gaps in the pre-training of some monolingual models. We also discuss the effectiveness of automatic translation in addressing these issues. In Subsection 4.4.6, we further analyze the impact of two quantitative factors on sample complexity: sentence length and the number of entities per sentence. In Subsection 4.4.7, we perform an ablation study on the training sets of NewsMTSC and MAD-TSC. This analysis shows that the smaller number of samples in MAD-TSC compared to NewsMTSC has minimal impact on the final performance, as adding more training than 3000 samples results in only marginal F1m improvements for both datasets, making the increased dataset size seem not cost-effective. Finally, in Subsection 4.4.8, we conduct a preliminary study on the transferability of models trained on MAD-TSC and NewsMTSC to other news domains, specifically the health domain. The results highlight the challenges and limitations of applying targeted sentiment classification models across domains.

4.4.1 Compared TSC Methods

We evaluate the effectiveness of four modeling approaches of the TSC task using both monolingual and multilingual pre-trained encoder models. In all experiments, we opt for the most proficient Bert-like architecture available in each language. RoBERTa-training inspired backbones [Liu et al., 2019b] are preferred due to their superior performance compared to BERT models in several tasks, especially in TSC for news [Hamborg and Donnay, 2021], due to Bert's relatively undertrained state. When RoBERTa-training inspired models are not available, we select the best available BERT-training inspired backbone [Devlin et al., 2019]. The specific models used for each language are described in Table 4.1.

Lang	architecture	link	reference
EN	roberta-base	RoBERTa-base	Liu et al. [2019b]
ES	roberta-base	Bertin v2	De la Rosa et al. [2022]
DE	bert-base	bert-base german	-
IT	bert-base	bert-base italian	-
FR	roberta-base	Camembert	Martin et al. [2020]
PT	bert-base	BERTimbau	Souza et al. [2020]
NL	roberta-base	RobBERT	Delobelle et al. [2020b]
RO	bert-base	Romanian Bert	Dumitrescu et al. [2020]
MULTI	roberta-base	XLM-RoBERTa	Conneau et al. [2020]

Table 4.1: List of pretrained models, with associated language, base architecture, URL and reference when available.

Below is a brief description of the four different approaches to modeling TSC:

- SPC (Sentence Pair Classification) [Song et al., 2019] is based on the classical Sentence Pair Classification task of BERT models. The input is designed as "[CLS] <sentence> [SEP] <entity> [SEP]".

- TD (Target Dependent) [Gao et al., 2019] only considers the last hidden states of the entities' tokens and merges their representation through a max-pooling layer.

- PM (Prompt Model) [Seoh et al., 2021] employs a prompt model for TSC. Our implementation of this method uses the simple prompt "<entity> is [mask]" with ["good", "ok", "bad"] proposed in the original paper as verbalizer. This prompt is translated in all MAD-TSC languages.

- BASE (Baseline) - version of SPC without access to the entity mention in its input. It can be deployed for any general sentiment classification strategy since annotations of the entity are not used.

We apply an efficient fine-tuning process inspired from the findings and optimization method of Mosbach et al. [2021] which shows that models are typically under-trained during fine-tuning, leading to poor and unstable performance between runs. This type of optimization was successfully used in another TSC context [Seoh et al., 2021]. We fix the learning rate to 2e-5 and train with early-stopping conditions up to 40 epochs using AdamW optimizer, and batches of size 32. Full optimization details are visible in Table 4.2.

Hyperparam	Fine-tuning
Learning Rate	2e-5
Batch Size	32
Weight Decay	0.01
Warmup	0.06
Max Epochs	40
Adam ϵ	1e-6
β_1	0.9
β_2	0.98
Optimizer	AdamW
Seeds	[42, 302, 668, 745, 343]

Table 4.2: Values of hyperparameters used for training TSC models.

Initial experiments confirmed that this approach outperforms the traditional hyperparameter search used in BERT [Devlin et al., 2019], demonstrating improved and more stable performance. In Table 4.3, we notably show significantly improved results reported for fine-tuning on NewsMTSC compared to those in [Hamborg and Donnay, 2021] for both SPC and TD.

All results are reported by averaging scores of five runs launched with different seeds. Pytorch [Paszke et al., 2019] is used for all implementations.

4.4.2 Dataset Splits and Metrics

We run experiments with MAD-TSC in monolingual and multilingual settings, and also use NewsMTSC for experiments in English. The training/validation/test subsets are sampled randomly and include 3,810/300/1,000 labeled mentions, respectively. Results for NewsMTSC are reported with the official splits from [Hamborg and Donnay, 2021]. TSC evaluation can be performed with different metrics [Hamborg and Donnay, 2021; Nakov et al., 2016; Pontiki et al., 2016a], we use macro F1 ($F1_m$) on all classes as primary metric. Performance with other metrics, such as F1 only on positive and negative classes ($F1_{pn}$), accuracy (acc), and average recall (rec) follow similar trends. A selection of such results, along with standard deviations are reported in Appendix B.1.

Train	NM		MAD		MIX	
Test	NM	MAD	NM	MAD	NM	MAD
SPC_{EN}	83.2	67.1	76.7	72.3	**83.4**	72.7
SPC_{ML}	81.0	61.5	70.5	67.8	80.3	67.9
TD_{EN}	**83.6**	69.7	75.2	**73.2**	83.0	**74.1**
TD_{ML}	81.6	63.3	71.9	68.5	80.4	68.4
PM_{EN}	**83.6**	**67.3**	**77.3**	72.8	82.6	72.2
PM_{ML}	81.7	60.9	69.3	66.6	82.0	67.0
$BASE_{EN}$	76.7	64.0	68.8	69.9	76.7	71.0
$BASE_{ML}$	74.1	59.2	66.5	64.8	74.1	66.7

Table 4.3: $F1_m$ results for individual train and test sets of the English subset of MAD-TSC (MAD) and NewsMTSC (NM), and for their combination (MIX). TSC methods are applied on top of language models pretrained specifically for English (EN) or with a multilingual corpus (ML).

4.4.3 Experiments with MAD-TSC and NewsMTSC

We compare the TSC methods from Subsection 4.4.1 on both the English subset of MAD-TSC and on NewsMTSC. Models are trained and evaluated on each dataset and on their combination. Results with different train and test set combinations are reported in Table 4.3.

A first observation is that the $F1_m$ results are similar for the different TSC models - SPC, TD, and PM - regardless of the specific dataset combination used. Complementary, the fact that the BASE model has significantly lower performance (about 6 and 3 points for NewsMTSC and MAD-TSC, respectively) is reassuring and confirms the need to provide the target entity to perform the TSC task. We also note the performance gains from the fine-tuning strategy we employ, with the performance of SPC and TD obtained when training and testing on NewsMTSC being 3 to 4 points better than originally reported in [Hamborg and Donnay, 2021]. SPC and TD scores are even on a par with those of the more complex GRU method [Hamborg and Donnay, 2021], which needs an external knowledge source that is not available for languages other than English.

In addition, the results confirm that MAD-TSC is more challenging than NewsMTSC with lower performance of the models, also confirmed for iso-number of training instances later in Figure 4.5. We note that, although the domains are expected to be semantically similar, transferring models trained on one dataset to another test set yields suboptimal results. However, the combination of the two training sets still has a marginal positive effect on each test set, demonstrating the complementarity of the two datasets.

Finally, the language model pretrained specifically for English is clearly better than its multilingual counterpart, probably due to the curse of multilinguality which affects multilingual models [Conneau et al., 2020].

4.4.4 Experiments with Individual Languages

Results for the eight MAD-TSC languages are presented in Table 4.4. Regarding the similar performances of SPC, TD and PM models in Table 4.3, they are only reported for SPC, a commonly used TSC baseline method [Cao et al., 2022; Hamborg and Donnay, 2021; Seoh et al., 2021; Song et al., 2019]. The best $F1_m$ scores are obtained for English and French, and the lowest scores are reported for Dutch and Spanish. When using monolingual pretraining (TG), the

Pretrain	EN	ES	DE	IT	FR	PT	NL	RO
TG	**72.3**	63.9	**66.1**	**65.8**	**70.8**	**68.2**	62.1	66.9
ML	67.8	**67.2**	64.8	65.1	67.2	66.2	**66.4**	**68.5**

Table 4.4: $F1_m$ results for the eight languages included in MAD-TSC. SPC is applied on top of models pretrained specifically for each target language (TG) or with a multilingual corpus (ML) using SPC.

Train	Test	ES	DE	IT	FR	PT	NL	RO
EN	EN_{M2M}	73.3	70.8	71.4	70.6	71.9	71.1	73.0
EN	EN_{DL}	**73.9**	**73.2**	**72.5**	**73.5**	**73.1**	**72.1**	**73.8**
TG	TG	*63.9*	*66.1*	*65.8*	*70.8*	*68.2*	*62.1*	*66.9*
TG_{M2M}	TG	**64.7**	65.0	**66.0**	70.6	66.9	**64.2**	65.7
TG_{DL}	TG	63.7	**65.2**	65.8	**71.3**	**68.3**	62.0	**67.5**

Table 4.5: $F1_m$ results for machine translation languages included in MAD-TSC, compared to the results obtained when without machine translation for English-only (72.3 in Table 4.3) and monolingual models (fourth row copied from Table 4.4). Notations: EN - English, TG - target language. The original train/test sets were used if no subscripts are present. DL (DeepL) and $M2M$ [Fan et al., 2020] subscripts give the machine translation model used. All results are reported with language-specific pretrained models. TSC models are trained with SPC.

difference between the best and worst scores is over 10 points. In contrast, the results obtained with multilingual pretraining (ML) are much more similar across all languages. The variability in the observed performance can be attributed to the quality of the pre-trained models, in particular the size of the pretraining datasets and their proportion relevant to the analysis of political news. Interestingly, the multilingual pretraining is much better than the language-specific one for Dutch and Spanish, and is also better for Romanian. Inversely, monolingual models are clearly better for English and French, that are the two languages which have the best monolingual pretraining. The results of Table 4.4 show that effective monolingual pretraining is preferred for training a TSC model for a target language, but multilingual pretraining can serve as a viable alternative when the dataset for a specific language is limited.

4.4.5 Cross-lingual Experiments and Machine Translation

Machine translation (MT) has strongly progressed in recent years, notably thanks to the introduction of neural architectures [Stahlberg, 2020]. A successful deployment of MT for sentiment classification would greatly facilitate the transfer of TSC task in the multilingual setting because it would reduce, or even remove, the need for specific annotations in each language. Building on previous works which apply MT to TSC [Balahur and Turchi, 2014; Mohammad et al., 2016b], we report results with English as pivot language. Test and/or train subsets of the other languages are translated to English. The translation is performed with two methods: (1) M2M100 [Fan et al., 2020], a recent massively multilingual translation model, by using the largest available model (12B parameters); (2) the API of DeepL[15], a well-known commercial machine translation service.

[15] `https://www.deepl.com/en/docs-api`

The $F1_m$ scores obtained with different MT configurations are reported in Table 4.5. The results are very interesting, particularly when translating the test set to English with DeepL (row with EN train and EN_{DL} test). $F1_m$ scores are globally better than 72.3, the performance of SPC obtained with manual translations for English in Table 4.3. The maximum gain is 1.6 points for Spanish, and Dutch is the only language for which DeepL translations are slightly worse (-0.2 points). $F1_m$ is also interesting with M2M100, albeit lower than that of EN_{DL}. These results surprised us initially, but we reach the same conclusions after running the experiments a second time in an independent manner. When translating the test set, all languages benefit from the strong pretraining available for English, and strong performance can be obtained for them if a good translation model is available from the target language toward English. This condition is met for all languages included in MAD-TSC . Qualitatively, the findings reported here could be explained by the fact that, while the polarity of sentiments is preserved by both machine and human translators, professional translators are more creative and sometimes add context in sentences for their international public. While useful for human readers, these changes can have a slight negative influence on sentiment classification.

Results are also interesting when the English training set is translated toward the target languages using DeepL and M2M100 (rows with TG_{DL} / TG_{M2M} train and TG test). The associated $F1_m$ scores are on par with those obtained with the manual translations. However, the translation of training sets is less effective than that of test sets. This happens because TSC training is done in languages other than English, and is based on weaker pretrained language models.

We also translated both the training and test sets from Dutch and Romanian to English, two of the languages which have low performance in Table 4.4, using DeepL. The $F1_m$ scores for the two languages are 72.3 and 73.6, respectively. We conclude that interesting performance can be obtained by automatically translating TSC datasets from other languages to English.

4.4.6 Analysis of Example Complexity

We complement the quantitative results by a qualitative analysis of factors which influence TSC performance. The analysis is done for English, and findings are similar for the other languages.

We first test the hypothesis that the complexity of examples is correlated to their length [Hamborg and Donnay, 2021]. We split sentences in five subsets depending on the number of words per example and report $F1_m$ per subset: 75.0 for up to 20 words per sentence; 72.9 for 21 to 30 words; 70.9 for 31 to 40 words; 70.0 for 41 to 50 words; 70.0 for 50 words and more. These scores confirm that TSC difficulty increases with example length, but differences become smaller above 30 words.

The number of entities detected in each example is an interesting proxy for complexity since the expressed sentiments can vary for multiple entities. We compute $F1_m$ separately for examples which include 1, 2, 3 or more detected entities. The scores obtained for the three subsets are 73.6, 70.1, and 67.4, respectively. They confirm that the presence of more entities makes TSC more difficult.

4.4.7 Ablation of Training Set

The annotation of TSC datasets is a cumbersome task, and it is important to minimize this effort, while preserving the final performance. In Figure 4.5, we present the results obtained by sampling 1,000, 2,000, 3,000 and 4,000 training examples from NewsMTSC and MAD-TSC, and with the full datasets. $F1_m$ scores increase up to 3,000 samples, but the added benefit of more samples is

reduced beyond this dataset size. The trend is similar for MAD-TSC , and the results reported in Figure 4.5 indicate that its size is sufficient to tackle the TSC task effectively.

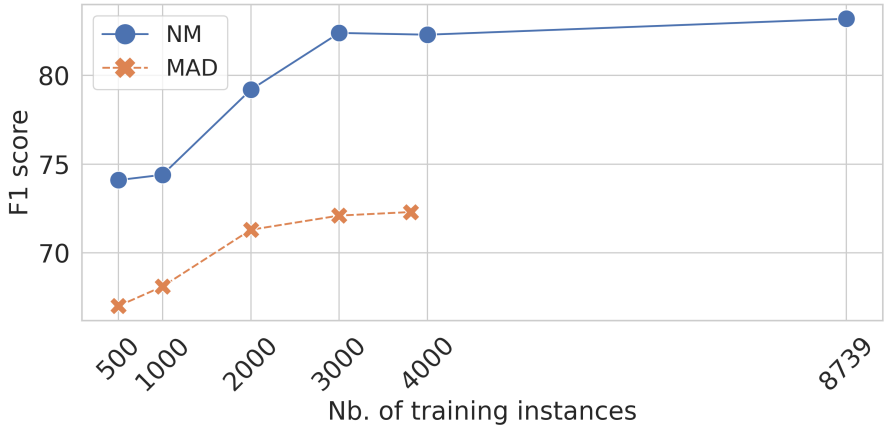


Figure 4.5: $F1_m$ scores with training sets of variable size for NewsMTSC (NM) and MAD-TSC (MAD). Results are reported with monolingual pretraining and SPC. The rightmost points represent scores for the full dataset.

4.4.8 Cross-Domain Experiments

In accordance with the our second research question **RQ2**, we aim to study and improve the cross-domain transferability of our TSC models, focusing on targeted sentiment analysis as discussed in the literature review in Section 2.6. We tested this by starting the annotation for datasets in three specialized news domains: finance, health, and sports.

We collected the articles from the MiND dataset [Wu et al., 2020b], a US-based news dataset designed for experiments in news recommendation systems, tagged according to a system that partially implements the IPTC taxonomy [16]. We extracted text from articles categorized under the headings of finance, health, and sports. Since we find that not all sentences of a given article are necessarily related to the target category; as they may be too general, or the article may cover multiple topics; we decide to further filter our sentences by learning a custom classifier.

For each selected MiND article, we extract only one unique sentence to ensure diversity of sources. We then extract and annotate 1,000 sentences per domain to determine whether they are representative; those that are not are left without label. In addition, we supplement the dataset with 1,000 random sentences from Polusa, another U.S.-based political news dataset, for which we assign the "politics" label. The final composition of the dataset is as follows: 533 finance, 701 health, 1,000 politics, 959 sports, and 807 unlabeled sentences. We partition the dataset, ensuring class balance with a 70%/15%/15% distribution. Next, we fine-tuned a RoBerta-based model for multi-label classification using a strategy similar to the training of our TSC models, with the hyperparameters listed in Table 4.6. This approach excludes the modeling of an "other" class, which increase robustness against future out-of-domain sentences. We used the final layer representation of the $[CLS]$ token for the classification head, the logits for each class are then processed through sigmoids. Commonly interpreted as the probabilities of class membership, we use the common threshold of 0.5 to assess the sentence's membership to a given class. This provides satisfactory precision and recall for each class as seen in Table 4.7:

[16] https://iptc.org/standards/media-topics/

Hyperparam	Fine-tuning
Learning Rate	1e-5
Batch Size	32
Weight Decay	0.01
Warmup	0.06
Max Epochs	5
Adam ϵ	1e-6
β_1	0.9
β_2	0.98
Optimizer	AdamW
Seeds	42

Table 4.6: Values of hyperparameters used for finetuning RoBerta classifier.

Domain	Finance	Health	Sports	Politics
Precision	70.4	77.8	92.5	91.9
Recall	62.5	69.8	94.4	84.0

Table 4.7: Classification performance on the test set with 0.5 threshold for the classes Finance, Health, Sports and Politics

We proceed by selecting candidate sentences for targeted sentiment classification annotation. We collect articles from the CCNEWS corpus from 2019 to 2023, filtering to include only those from the United Kingdom, which amounts to 16M articles and around 154M sentences. This selection, as described in Section 3.4, mitigates the domain-shift effect caused by performance differences between American and European cultural contexts, as observed in our comparison of transfer performance between MAD-TSC and NewsMTSC in Table 4.3. For each of our three domains of study - finance, health, and sports - we sample candidate sentences for annotation. We randomly select 5% of the articles from CCNEWS and use our classifier to annotate each sentence. For each article, we count the sentences that meet the 0.5 threshold for a given class. If a class is dominant and appears at least three times, we randomly select one sentence from those that exceed the 0.5 threshold.

Currently, annotations for finance and sports are incomplete. We offer first results for cross-domain transfer between Health and Politics in Table 4.8. The health dataset is annotated with 3,314 examples, balanced in sentiment, and results are reported for the SPC model.

Train\Test	Health $F1_m$	MAD $F1_m$	NM $F1_m$
Health	$\mathbf{74.6 \pm 1.7}$	45.2 ± 1.3	55.4 ± 2.7
MAD	70.4 ± 1.3	$\mathbf{73.1 \pm 0.8}$	77.3 ± 0.8
NM	71.7 ± 2.1	67.3 ± 1.0	$\mathbf{83.3 \pm 0.7}$

Table 4.8: Cross-domain TSC performance (Macro-$F1$) with SPC model between "Health" and "Politics" domains

As we can see in Table 4.8, the performance decreases slightly when transferring from "Politics" to "Health", and more significantly when transferring from "Health" to "Politics", possibly due to

a higher diversity of topics in political news, which may facilitate better model transferability to other domains. This study should be continued in the future by consolidating a dataset for sports and finance. This consolidation will allow the development of new cross-domain techniques for TSC in the news domain.

4.5 Limitations

The analysis in Subsection 4.4.6 shows that the number of entities per example significantly affects the results, a finding consistent with previous research [Jiang et al., 2019; Hamborg and Donnay, 2021]. MAD-TSC reflects the real-world distribution of entities in sentences, with predominantly single-entity examples (62%) versus multi-entity examples (38%). Jiang et al. [2019] further attempts to sample examples with contrasting sentiments for different entities within the same sentence, but this approach seems less viable in news analysis, where sentiments are expressed more subtly and such examples are rarer than in reviews. Although our baseline model "BASE", which doesn't access the considered entity, shows worse performance, enriching the dataset with examples containing multiple entities could further limit a potential bias of the model to rely on the overall polarity of the sentence instead of the targeted sentiment expression.

Another limitation of our work is that our dataset focuses primarily on the political domain. Efforts are underway to expand it to other domains to explore the potential for cross-domain transfer, but more work is needed. In addition to cross-domain generalization, cross-entity generalization also poses challenges.

Currently, our dataset has also two major limitations with respect to entities. First, it only includes sentiment annotations on entities categorized as "Person", but other types such as "Organization", "Location", and "Event" involve different ways of expressing sentiment and might be interesting and relevant to include. Second, our multilingual mention matching uses an edit distance metric, which limits the dataset to proper nouns and potentially reduces its applicability to common nouns. Finally, as discussed in Subsection 4.3.4, the task of sentence annotation presents inherent challenges for humans, often requiring additional context to accurately identify sentiments, especially in cases of irony. As a result, our dataset may not fully capture the complexity of the task, as sentences with ambiguous valence are typically excluded during the annotation process.

4.6 Conclusion

We introduced MAD-TSC, a dataset for multilingual target-dependent sentiment classification. Compared to existing resources, the proposed dataset is aligned across languages, includes examples about geographically diversified entities. Examples are longer and more complex because sentiment is often expressed in an implicit way. Given its aligned character, MAD-TSC dataset enables a comparison of sentiment classification between languages. Performance varies significantly, and this variation is to a large extent explained by the quality of pretrained models available for each language.

Importantly, the MT experiments show that human translations can be replaced by automatic ones. The automatic translation of test sets from target languages to English is particularly interesting since it brings target-dependent sentiment classification in different languages to the same quality level as that of English. This allows TSC to be scaled for the languages included in this study without the need to develop language-specific training sets. The only condition is to

have a labeled domain-related dataset in one language, English or other. In this context, we only partially addressed **RQ2**, that is the cross-lingual and cross-domain transferability of the targeted sentiment analysis component of the news analysis framework presented in Chapter 3. While we explored the feasibility of language transferability, we only briefly touched on cross-domain applicability. Our multilingual experiments confirmed that current automatic machine translation models are effective for applying TSC models across languages. It can be achieved either by translating into a high-resource language and using its models, or by translating the dataset to train a model in the target language. Cross-domain generalization will be explored in future research.

4.6.1 Perspectives

We can point out several perspectives of work on targeted sentiment classification in news.

First, while encoder-based models are widely used, study on the structuring of their input (e.g., prompts, entity markers, separators), output (e.g., verbalizers with different vocabulary configurations, CLS tokens, token aggregation) or loss (e.g. hierarchical, contrastive) remain underexplored. In addition, the potential of generative models, especially large language models in few-shot and zero-shot settings (e.g., prompting, in-context learning), deserves further investigation, although their inference costs may be prohibitive outside the creation of synthetic annotations. Current trends in the field have largely focused on augmenting input with syntactic trees, proximity heuristics, or sentiment lexicons, which are not available for all languages and have recently come under scrutiny for their effectiveness [Dai et al., 2021].

Second, regarding the use of annotations, we might reconsider the typical 3-class (negative, neutral, positive) framework for TSC classification and its predominant aggregation method. It might be beneficial to predict annotations on an individual annotator basis, which would allow the model to make better use of the information they provide. In addition, prioritizing samples with high inter-annotator agreement could be beneficial, as several studies suggest that removing more difficult/noisy examples in scenarios with limited data can lead to improved generalization [Forouzesh and Thiran, 2023]. This action could also be taken in the context of the curriculum learning approach [Soviany et al., 2022].

Third, when multiple domains become available, it will be possible to explore strategies for cross-domain transfer of TSC skills in the news domain. Historically, the main goal in transfer-learning has been to learn separate representations for domain-specific knowledge and transferable cross-domain knowledge [Cai and Wan, 2019; Luo et al., 2022]. However, the question of effectively leveraging the extensive common knowledge acquired by LLMs during pre-training to adapt a task to new domains remains a challenge.

In the next chapter 5, we use the model learned using MAD-TSC to perform an analysis of the French news media landscape.

Chapter 5

French Press Analysis

Contents

5.1 Introduction 74
5.2 Related Work 75
5.3 Instantiation Details of the News Processing Framework 76
5.4 Political News Analysis 78
5.4.1 Outlet-oriented Analysis 79
5.4.2 Topic Oriented Analysis 82
5.4.3 Politician Oriented Analysis 83
5.4.4 Gender Representation 86
5.4.5 Age Representation 87
5.4.6 Limitations 87
5.4.7 Conlusion 89
5.5 Analysis of Other Countries 90
5.5.1 Multilingual Political Pre-Filtering Step 90
5.5.2 Sentiment Inference 92
5.5.3 Flemish Political News Analysis 93
5.6 Conclusion 97

5.1 Introduction

In Chapter 3 we presented a general framework and its instantiation for political news analysis. In Chapter 4 we presented MAD-TSC, an aligned multilingual dataset specialized for sentiment classification of political news. We also demonstrated the suitability of using current machine translation (MT) models to translate TSC datasets "from" or "to" English in order to build models for a target language. In this chapter, we answer our third research question (**RQ3**) and demonstrate, by instantiating the framework for an analysis of the French news landscape, the comprehensive analytical capabilities it allows. As discussed in chapter 3, most news analysis tools focus on "objective metrics", which we define as the measurement of a pattern that, when considered atomically, cannot be indicative of a stance or opinion. This means that only relative distributional differences in the occurrence of these patterns can potentially be indicative of a stance or opinion. For instance, a citation graph analysis allows the clustering of news outlets based on their political affinity, as detailed in Cointet et al. [2021]. Similarly, an aggregation of TV and radio broadcast metadata can be used to estimate political biases, as explored in Cagé et al. [2022].

Conversely, we define "subjective metrics" as the measurement of a pattern that is indicative of a stance or opinion even when considered atomically, which generally includes sentiment classification or subjectivity classification methods. Based on this distinction, we note that most current news analysis tools focus on objective metrics or include only rudimentary subjective metrics. The original Textmap framework Lloyd et al. [2005] detects entities and their relationships, and grounds them in space and time. Godbole et al. [2007]; Ploch et al.; Bautin et al. [2021] extend Textmap's purely objective analysis using lexicon-based sentiment analysis at the article level. However, such an approach is still too coarse to accurately derive stance, as pointed out by Bestvater and Monroe [2023].

More recently, stance qualification has been approached with a focus on classifying the political leanings of entire news articles using large language models Baly et al. [2020a]; Ko et al. [2023]; Kamal et al. [2023]. These approaches have important shortcomings, notably in terms of granularity and explainability. Also they require to compile a new dataset when the political context changes. Furthermore, we argue that fine-grained subjective metrics are inherently better able to capture subtle and explainable stances. In this chapter we show through a comprehensive analysis of a French political news corpus that subjective metrics analyses are complementary to commonly used objective ones and allow to derive insights about news outlets, political topics, individual entities and demographic segments. To obtain these results, the framework integrates recent NLP components, information retrieval techniques, news outlet metadata, temporal metadata and external knowledge bases. The advantage of this framework is that it is designed to work seamlessly for multiple countries since all NLP components are multilingual and the knowledge bases are not country specific. Here we instantiate it for political news, as presented in Chapter 3, using a French political news corpus. We further show its inherent multilingual and multicultural nature by applying it to the analysis of the Flemish news landscape in Section 5.5.

The main findings of the comprehensive French political news analysis are:

- mainstream political orientations are presented in a rather balanced way in the major news outlets, but sentiment towards the radical left and right are respectively positively and negatively biased;
- the mentions and sentiments associated with political orientations vary across impactful political topics;
- sentiment scores are generally negative, with important variation between news outlets;
- the most positive and negative sentiment scores for politicians are correlated with the public

perception of their actions;
- mentions frequency is biased towards male politicians, but sentiment scores of female politicians are more positive;
- there is an age bias toward older politicians;
- the French semi-presidential system is reflected in the news, with a dominance of presidents or presidential candidates.

These findings can be useful for multiple stakeholders. News outlets can analyze their positioning and make it more transparent to the general public. Social scientists can gain new insights into the online representation of the political landscape. Political organizations can monitor the online reporting of their political actions. Citizens can be informed about potential political, thematic and demographic biases of news outlets.

5.2 Related Work

Media analysis studies are based on tools and insights from different disciplines, including political science [Kim et al., 2022] economics [Martin and Yurukoglu, 2017], and NLP [Agrawal et al., 2022; Gangula et al., 2019; Spinde et al., 2021]. Research efforts that reveal political, thematic or demographic biases are of particular interest.

Gender representation has been shown to be biased in multiple countries. Thesen and Yildirim [2022] show that male politicians are more present in news reporting for countries with proportional representation. Dacon and Liu [2021] analyze news abstracts and conclude to an underrepresentation of women in English news. Wevers [2019] and Bastin [2022] investigate gender bias in Dutch and French news, respectively. Online gender bias tracking tools were also developed, such as Gender Gap Tracker [Asr et al., 2021] for Canada and Gendered News [Richard et al., 2022] for France.

Aside from gender, topic-related bias in local newspapers is investigated in Rivas-de Roca et al. [2022], with focus on the European Union. Spinde et al. [2020] study the bias of words in German news. Cagé et al. [2022] show that political orientations are unevenly represented across broadcast outlets by aggregating mentions of individual political and non-political figures. Cointet et al. [2021] analyze the French media ecosystem by considering the online interaction graph between them. These studies are interesting but mostly rely on mentions and patterns frequency without addressing any measurement of sentiment or subjectivity.

Framing [Entman, 1993], the action of emphasizing certain facts over others, has also been studied. Liu et al. [2019a] introduce the Gun Violence Frame Corpus and implement an approach for news framing detection. Their results show that framing is dynamic and follows closely the gun-related violent crimes. Young et al. [2021] investigate how immigration is framed in US newspapers over the last decade, and show an increase in negative messages about immigrants. Hamborg [2020] studies framing through sentiment classification and word choice. The author implements a method for highlighting media outlet position over a specific topic. Alonso del Barrio and Gatica-Perez [2023] very recently tested large language models for frames detection in news headlines. Stance detection [Küçük and Can, 2020], the subjective positioning of the author towards a target, is also actively investigated. Conforti et al. [2020] build a corpus of news focused on company merging stance detection. Sepúlveda-Torres et al. [2023] address the detection of misleading news headlines through stance comparison of the body and the text. Other approaches treat stance detection as a text-level sentiment analysis task [Godbole et al., 2007; Ploch et al.; Wu et al., 2020a; Bautin et al., 2021], but as Bestvater and Monroe [2023] points out, this level is too coarse to bring satisfactory insights about stance.

More recently, some approaches [Baly et al., 2020a; Ko et al., 2023; Kamal et al., 2023] have used stance-labeled article databases such as AllSides to train language models for stance-classification in an end-to-end fashion. As presented in our literature review in Subsection 2.5.2 those approaches have several shortcomings: (1) stance is characterized at a coarse level of granularity, and it is impossible to provide detailed insights about politicians, political organizations, or demographic segments; (2) the explanatory power of the results obtained is limited, since only one class label per article is provided; and (3) the task is country, time-dependent and relies on supervised annotated datasets, thus requiring the creation of new datasets for each different political context.

Although they provide useful analytical tools for capturing media bias, most studies use lexical, mention-count based, or metadata approaches that fail to account for the sentiment expressed toward key entities in the text. This is mainly an effect of the relative lack of TSC resources for the political domain. Target-dependent Sentiment Classification (TSC) predicts the opinion towards a precise entity and can be aggregated to better understand the positioning of a news outlet toward that entity [Hamborg, 2023a]. We build upon our work in Chapter 4 to train and integrate our TSC models for political news analysis.

5.3 Instantiation Details of the News Processing Framework

This section expands on the description of the overall framework presented in Chapter 3 by presenting the instantiation details based on the pipeline illustrated in Figure 3.2.

1. Filtering - Contrary to the general pipeline presented in Chapter 3, our studied French corpus isn't based on CommonCrawl, but was instead built by collecting RSS streams from more than 280 politically diverse French news outlets between January 2016 to December 2022. Texts of the articles were extracted from their HTML using Trafilatura [Barbaresi, 2021]. As RSS streams were not completely mutually exclusive, and because some articles may be available from multiple links due to revision versions, a near deduplication using MinHashLsh [Broder, 1998] was performed to only keep one version per domain name as described in Subsection 3.3.2. Finally, only news including 200 characters or more were retained, totaling over 457,000 articles. Due to the different availability of content from news outlets, the distribution of the number of articles among them is uneven, as illustrated in Figure 5.1 with the distribution of articles over the 20 most frequent news outlets.

The collection of news articles was also limited by the availability of historical articles on the media platforms. Figure 5.2 shows how uneven is the distribution. One solution in the future could be to expand the number of articles for past years using CommonCrawl snapshots.

2. Sentencizer - Articles were split into sentences using the French "fr_web_core_sm" Spacy model.

3. Mention Detection - We then used the French version of Flair [Akbik et al., 2019] for named entity recognition. We detected 2.27M mentions of persons in the corpus.

4. Entity Linking - Entities were linked to Wikidata [Pellissier Tanon et al., 2016] using mGenre [De Cao et al., 2022], a recent multilingual linking model. Linking was considered

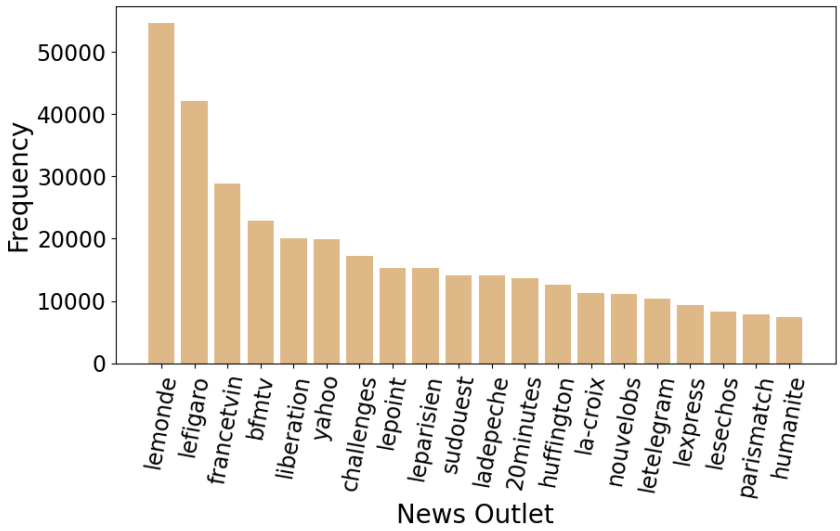


Figure 5.1: Number of articles per news outlet over the 20 most frequent outlets in the corpus

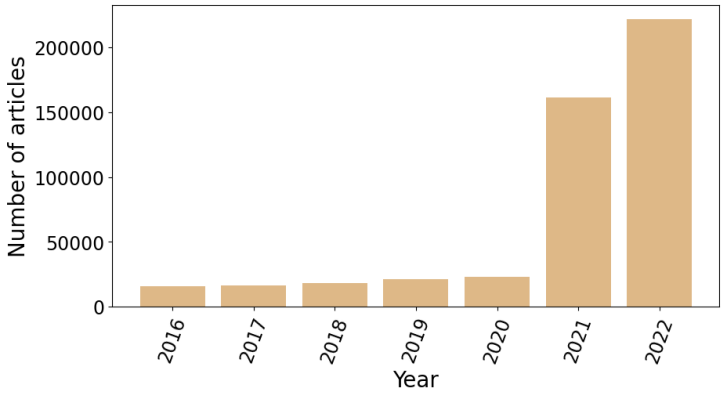


Figure 5.2: Number of articles per year in the corpus

successful if the log-likelihood of the best candidate was greater than -0.2. The threshold was determined by labeling 200 predictions per 0.1 range between 0 and 0.5, and it provides over 95% accuracy. There are 1.27M linked persons left.

5. Enrichment - Wikidata [Pellissier Tanon et al., 2016] entries of linked persons were used to select politicians and extract their gender, birth date, country, and political affiliation(s). ParlGov [Döring and Manow, 2012] includes information about political parties in different countries. It uses a scale from 0 to 10 (left to right) to encode political orientation. There are 520K mentions of politicians (89.9K female), out of which 345K are French and 333K can be linked to political orientations from ParlGov. To facilitate interpretability, we reduce this to a 5 points scale by merging pairs of neighboring orientations. The resulting scale used in experiments includes the following orientations: RL - radical left (30.5K mentions); CL - center left (56.4K); C - center (113.3K); CR - center right (87.3K); RR - radical right (45.8K). We finally match French political party names from Wikidata and ParlGov to attribute an orientation to politicians.

6. Sentiment Classification - In this step we use the TSC resource presented in Chapter 4.

We use the French version of the dataset to train a SPC-Bert model [Song et al., 2019], which is based on the classical Sentence Pair Classification task of Bert [Devlin et al., 2019] and provides strong TSC performances, as demonstrated in Chapter 4. The underlying transformer architecture we use is CamemBert [Martin et al., 2020], the French equivalent of RoBERTa [Liu et al., 2019b]. We follow the training procedure described in Chapter 4. The final model achieves an $F1_{macro}$ score of 70.8. It is important to assess the impact of errors made by the TSC component. We do a preliminary experiment with two sample sets. A first one where we keep all sentiment predictions, and a second one using a threshold where we keep only the 50% most confident for each class, assuming that less confident predictions are more error prone. For both sets, we compute the average sentiment scores and the number of mentions for the 1,000 most cited politicians from the corpus. We then compute the Pearson correlation between the sampled distributions, and obtain values of 0.993 and 0.966 for the number of mentions and the sentiments distributions, respectively. These results show that, while individual predictions are imperfect, their aggregation leads to very stable results, and all sentiment predictions are usable.

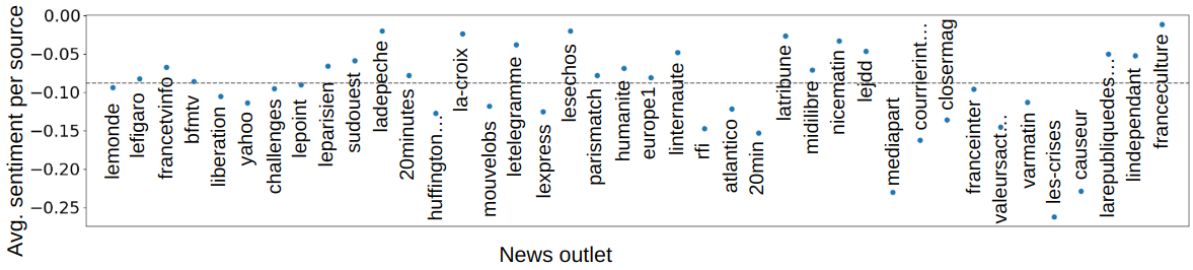


Figure 5.3: Average sentiment (blue dots) for the most frequent 40 news outlets in the corpus. The dashed line represents the average sentiment for the entire corpus. Outlet frequency decreases from left to right.

Topic Analysis - In addition to the targeted sentiment analysis we perform on the political entities of the articles, we also decide to perform a topic-oriented analysis. We use an information retrieval approach to select relevant subsets of documents for political topics. This type of approach is interesting because it allows a simple and flexible definition of topics. We index the news corpus using BM25 [Robertson et al., 2009], a probabilistic text representation method which provides strong performance when compared to more recent neural-based approaches [Deshmukh and Sethi, 2020], [Khloponin and Kosseim, 2021]. We chose a set of 10 impactful political topics associated with queries for retrieval. The set includes both diversified topics and similar ones. The list of topics includes: (1) climate change, (2) corruption in the political domain, (3) the economic consequences of the Covid-19 crisis, (4) the health-related issues of Covid-19, (5) the yellow vests protest which took place in France, (6) immigration, (7) purchasing power, (8) the war in Syria, (9) the war in Ukraine and (10) the economic consequences of the war in Ukraine. These topics are used in the analysis of the Subsection 5.4.2. Pertinence thresholds are needed to select relevant subsets of documents for each topic. These were determined empirically by examining the titles and snippets of BM25 returned results. There are between 2.1K and 8K mentions per topic in the corpus.

5.4 Political News Analysis

We propose a comprehensive analysis of political representation in the news by combining objective and subjective perspectives. We analyze news outlets, political topics, politicians, and

demographic segments (gender, age).

5.4.1 Outlet-oriented Analysis

Overview of the French Media Landscape To better understand the alignment of our findings with the known characteristics of French media outlets, we briefly introduce our corpus through the prism of the French news ecosystem as described in Cointet et al. [2021] and Lenoir [2019]. This news ecosystem is characterized by a diversity of media outlets that can be structured along left-right and pro-establishment axes. The set of newspapers analyzed in our corpus mainly includes the core of the French media landscape. Major dailies such as *Le Monde* and *Le Figaro* represent centrist and center-right views, respectively, and offer a wide range of general news. *Libération* and *L'Humanité* offer a left-wing perspective, the former being more contemporary while the latter has historical ties to the French Communist Party. *Mediapart*, a left-leaning investigative journal, is available only online but is also considered a core informational medium.

Regional publications such as *La Dépêche, Nice Matin*, and *Le Télégramme* play a crucial role in providing local news, each maintaining a centrist to center-right stance. This regional focus allows for a more localized perspective on news. Business news is prominently covered by *Les Echos* and *La Tribune*, both of which hold a centrist viewpoint and are important sources of financial and economic news.

On the fringes of mainstream media, outlets such as *Valeurs Actuelles*, *Causeur* and *Les-Crises* offer alternative viewpoints. *Valeurs Actuelles* and *Causeur* offer a right-wing perspective that focuses on conservative and traditional values. *Les-Crises* is considered a right-wing "counter-informational" media known for its anti-establishment stance, often criticizing and challenging the perspectives held by the government and the core media. A short description of each source in the corpus is further proposed in Appendix C.1.

We note that French TV and radio stations need to balance the speaking time between different political tendencies, and also have objectives regarding gender parity ARCOM [2021]. This is not the case for newspapers, and these sources choose the discussed topics and the political tendency coverage only based on their editorial policies. The same observation applies to the coverage of demographic segments.

Source Sentiment. Figure 5.3 presents the average sentiment for mentions of politicians for the most frequent news outlets in the corpus. The presented political leanings of sources are those established in Cointet et al. [2021]. The average sentiment across all outlets is negative with an average of -0.083, but with significant variability between sources, which can be explained by various factors such as their nature, geographic focus and/or political leanings. The lowest scores are obtained for *les-crises, causeur* and *mediapart*, which is consistent with their characterization by experts as right-wing anti-establishment; right-wing and left-wing outlets, respectively. High scores are obtained for regional or local news outlets such as *la depeche*, *nicematin* or *letelegramme*, as well as for economic media such as *lesechos* or *latribune*. As for the major outlets, their scores are close to the average, with *lefigaro* being the least negative, followed by *lemonde* and *liberation*.
Mapping of Political Orientations. We analyze the positioning of news outlets with respect to major political orientations by aggregating mentions and sentiment scores. The mention-oriented analysis is similar to existing ones [Cagé et al., 2022], while the sentiment-oriented analysis adds a subjective perspective. We use the five-points political orientation mapping described in Section 5.3.

Figures 5.4 and 5.5 summarize the mappings of mentions and of sentiment scores *vs.* political orientations. There is no correlation between mentions and sentiment since the Pearson corre-

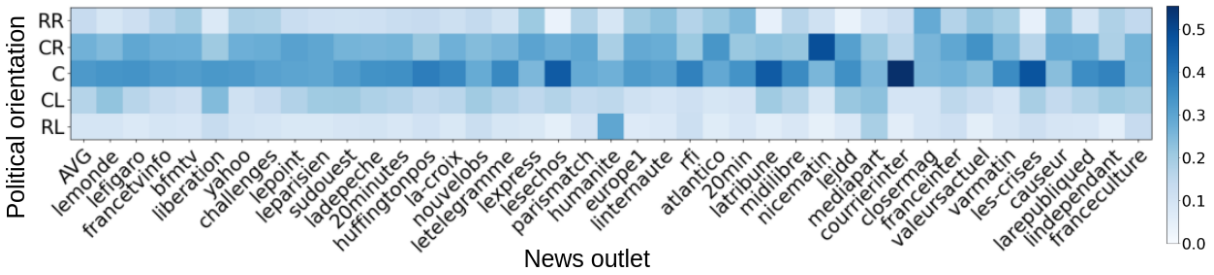


Figure 5.4: Distribution of mentions of political orientations in news outlets.

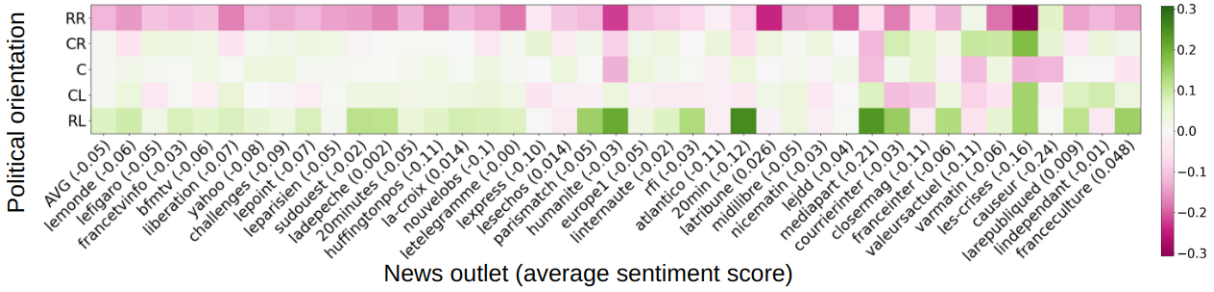


Figure 5.5: Deviation of the sentiment associated with major political orientations from the average sentiment of each source. Average sentiment of linked politicians is indicated in parentheses for each source.

lation coefficient between the AVG columns in the two figures is 0.06. This result shows that the two types of analyses are complementary and provide additional insights into political news. Right-leaning orientations are more cited than their left-leaning counterparts but the sentiment associated with right-leaning orientations is more negative. This contrast is even clearer for RR and RL, the two radical orientations that are represented in the two figures.

Figure 5.4 shows that the center, including the French governing party, is the most mentioned orientation. The right-wing tendencies are more represented than their left-wing counterparts. Outlet-wise, the center is overrepresented in economic newspapers (*latribune* and *lesechos*), but also in *les-crises*, a right-wing anti-establishment outlet. RL is strongly present in *humanite*, the newspaper of the French Communist Party, while the extreme right is often cited by sources such as *closermag*, a tabloid, or *causeur*, a right-wing outlet.

Figure 5.5 illustrates the sentiment-oriented positioning of news outlets. It is quantified by the difference between the sentiment score per orientation and the average sentiment score of the source. The positioning of sources towards political orientations varies, and this is a positive finding since diversified opinions are required for a healthy democratic debate. The representation of mainstream orientations (CL, C, CR) is rather balanced in the major sources (left of the figure) and more variable in other source (right of the figure). We note that, overall, there is a positive bias toward radical left (RL) politicians and a negative bias toward radical right (RR) ones. Notable exceptions for RL include *valeursactuelles*, *closermag* and *lindependant*. The only two sources which have a slightly positive positioning toward RR relative to their average positioning are *valeursactuelles* and *causeur*, but the average sentiment scores of RR remain overall negative even for these two sources. Intuitively, the center, which includes the current ruling party, is most criticized by news outlets which are left-wing (*humanite* and *mediapart*) and right-wing (*valeursactuelles* and *causeur*).We also note that the average sentiment of sources

differs between Figures 5.3 and 5.5. The scores obtained for French-only politicians are higher than when including all politicians. This may be explained by a more critical stance of French media on international issues than on national ones.

A temporal analysis can complement outlet-oriented ones. Figures 5.6 and 5.7 illustrate the temporal evolution of mentions and average sentiment associated with political orientations.

The distribution of mentions changes strongly between 2016 and 2017. The year 2016 is dominated by the center-left, which included the governing party at the time, and the center right. The central orientation gains momentum in 2017, when its party became majoritary, and is dominant between 2018 and 2020. The center right is well represented in 2021, when its primary elections for the 2022 presidential election took place and were frequently discussed in the media, with the center also being frequently mentioned. Then, the center becomes dominant again in 2022. The representation of the radical right became stronger in 2021 and 2022, when *Marine Le Pen* and *Eric Zemmour* played important roles in the campaign for the presidential election. A similar trend is observed for the radical left, whose main representative, *Jean-Luc Mélenchon* was also well placed in the presidential race.

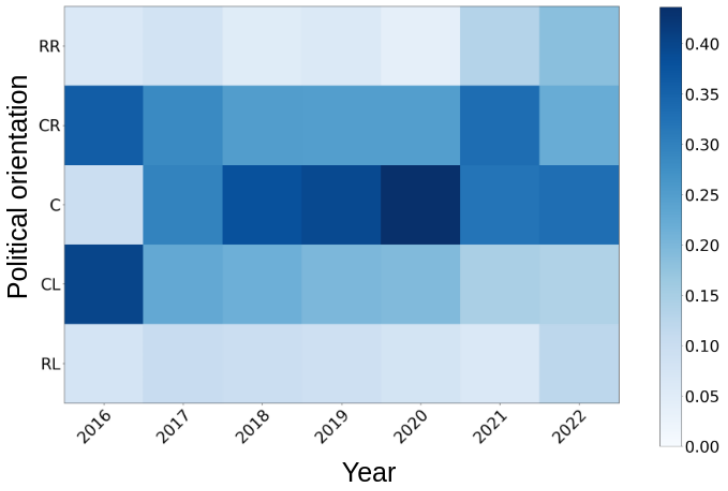


Figure 5.6: Mentions of political orientations across time.

The temporal evolution of sentiment is presented in Figure 5.7. It confirms that left-wing orientations have a media coverage which is overall more favorable compared to that of their right counterparts. The contrast is most pronounced for the radical orientations, with RL sentiment being positive in most years, especially in 2016 and 2022. Conversely, the sentiment toward the radical right is more negative in 2017, when *Marine Le Pen* lost the presidential election, and in 2021. The sentiment toward the center was most positive in 2017, the year when *Emmanuel Macron* won his first presidential run. However, this sentiment turned negative in 2018, the year that the *Yellow Vests* movement erupted.

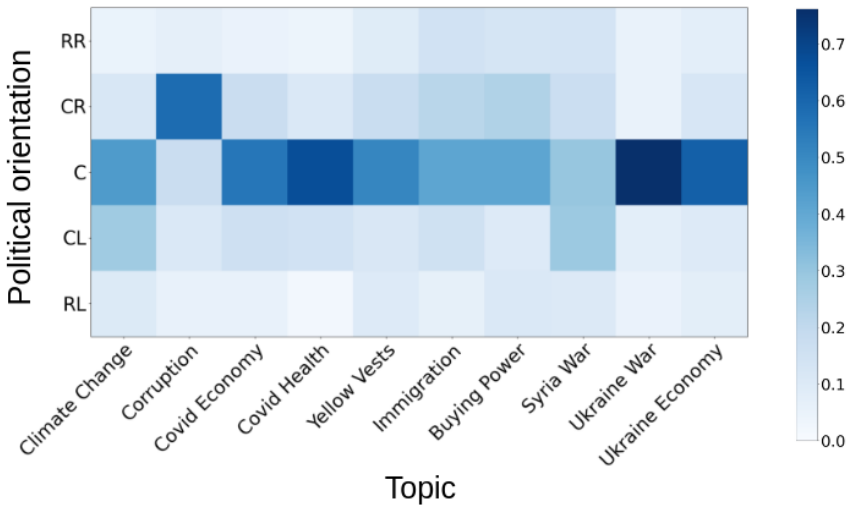


Figure 5.8: Distribution of mentions of political orientations for ten impactful political topics.

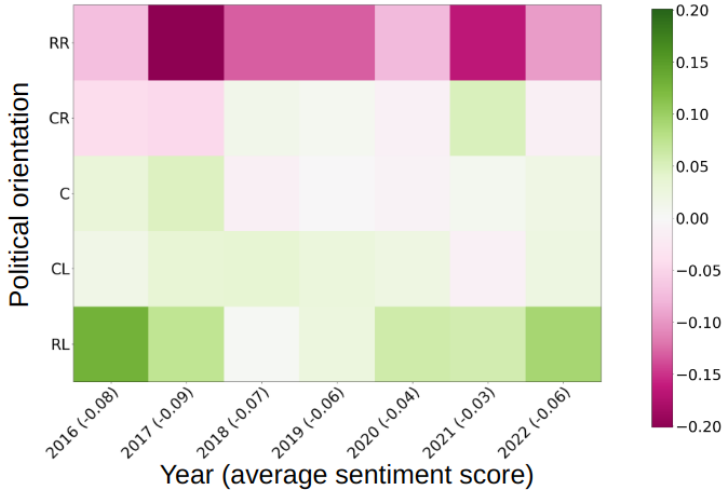


Figure 5.7: Sentiment associated with political orientations in time.

5.4.2 Topic Oriented Analysis

The results for mentions and sentiments associated with political topics are presented in Figures 5.8 and 5.9. The mentions of centrist orientation, which includes the governing party, dominate most topics. This is particularly the case for the health effects of Covid-19 and for the war in Ukraine, two topics for which the government's communication is prevalent. Corruption is one notable exception, with the center-right being the most cited because several of its politicians, such as the three with the lowest average scores from Table 5.2 were involved in major corruption scandals. There are some differences even for closely related topics, such as the pairs related to Covid-19 and Ukraine. There, the mentions of the center are more dominant for Covid Health and for Ukraine War.

The deviation of sentiment associated with political orientations from the average of the topic, illustrated in Figure 5.9, can be interpreted as a proxy for the credibility of political orientations for the respective topic. For instance, corruption has a particularly negative representation, with

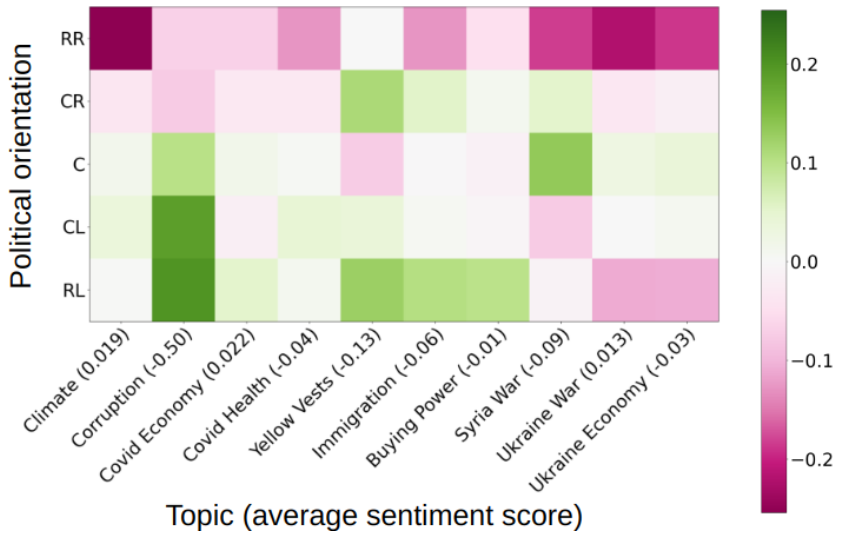


Figure 5.9: Deviation of the sentiment for major political orientations from the average sentiment of each topic. Average topic sentiment is given in parentheses.

an average sentiment of -0.5. While still in the negative range, RL, CL and C orientations are perceived better than CR and RR on this crucial political topic. The radical right has the lowest scores on average, with particularly negative positioning for climate change and the war in Ukraine. This finding is probably explained by the relatively small importance given to environment in RR platforms and by the favorable positioning with respect to Russia in the past for the war in Ukraine. The radical left is also perceived negatively on the two Ukraine related topics for a similar reason. However, the sentiment toward RL is more positive than that of other orientations for most other topics.

5.4.3 Politician Oriented Analysis

Frequently-mentioned Politicians. Table 5.1 provides insights about the top-10 most frequently mentioned politicians: overall mentions, average sentiment in the corpus, sentiment variation over time (i.e. the standard deviation of sentiment per year between 2016 and 2022). The table shows that the political discussion in France is focused on presidency since all top-10 names are of French or foreign presidents or presidential candidates. To verify the potential bias due to presidential elections, we ran the analysis separately for years without election and results are similar. *Emmanuel Macron* is by far the most mentioned politician, with Russian and American presidents also being mentioned frequently. The other French politicians from Table 5.1 are mainly 2022 presidential candidates. There is no correlation between their mentions in the corpus and their electoral results. *Eric Zemmour* and *Valérie Pécresse* scored lower in the elections than *Marine Le Pen* and *Jean-Luc Mélenchon*, but are more frequently mentioned.

The average sentiment scores from Table 5.1 vary substantially. *Donald Trump*, *Vladimir Putin* and *Nicolas Sarkozy* have the lowest scores. This is probably due to the critical perception of their political actions for the first two, and by ongoing legal actions against the third. Interestingly, *Valérie Pécresse* has the highest sentiment score in Table 5.1, despite her poor electoral performance in 2022. We verified her mentions and many of them correspond to her unexpected victory in the primary election of her party. We note that sentiment is more negative for far-right candidates (*Eric Zemmour* and *Marine Le Pen*) compared to *Jean-Luc Mélenchon*, the main

Name	Entity statistics		Sentiment variation
	Mentions	Sentiment	
Emmanuel Macron	49552	-0.074	0.036
Vladimir Putin	22925	-0.355	0.085
Éric Zemmour	21201	-0.176	0.106
Valérie Pécresse	15055	-0.006	0.09
Jean-Luc Mélenchon	12223	-0.052	0.097
Joe Biden	9907	-0.035	0.159
Donald Trump	8383	-0.430	0.071
Anne Hidalgo	7539	-0.090	0.032
Nicolas Sarkozy	6852	-0.319	0.042
Marine Le Pen	6639	-0.265	0.066

Table 5.1: Top-10 most frequently mentioned politicians, with: number of mentions; average sentiment scores computed using the entire corpus; standard deviation of average sentiment per year between 2016 and 2022.

left-wing candidate. This finding confirms the aggregated results from Figure 5.5. Sentiment varies to a different extent during the examined period, and this is explained by different factors: the political comeback for *Joe Biden*, the emergence of *Eric Zemmour* as presidential candidate in 2022, the invasion of Ukraine for *Vladimir Putin*, etc.

We also present the temporal evolution of mentions and sentiments of these ten most frequently mentioned politicians in Figures 5.10 and 5.11. The detailed view of mentions from Figure 5.10 illustrates the strong representation of *Emmanuel Macron* in the corpus. We also note the surge of mentions of *Vladimir Putin* in 2022, and that for *Eric Zemmour* in 2021. Sentiment changes in variable proportions in Figure 5.11. Stronger variations are explained by different factors. *Joe Biden* was initially absent from the corpus, then was presented negatively, but sentiment became more positive when his 2020 candidacy gained traction and he won the US presidential elections; *Eric Zemmour* was portrayed very negatively prior to 2021, but his image improved for some news outlets when his 2022 presidential run gained momentum; *Vladimir Putin* has an overall negative representation that has worsened since the invasion of Ukraine. Weaker variations are observed for *Emmanuel Macron* whose sentiment score is close to the average from Figure 5.3, except 2018, the year of the *Yellow Vests* movement, when his image was degraded. *Nicolas Sarkozy* has a constantly negative score which is probably explained by the array of legal actions opened against him.

Polarized Sentiment Scores. Table 5.2 presents the top-10 politicians with the highest and lowest sentiment scores. The first group includes local politicians whose action is well-appreciated (*Carole Delga* and *David Lisnard*), national politicians who are favorably viewed by the public (*Roselyne Bachelot* and *Jean Lassalle*), and international politicians with a centrist or center-left positioning who took office during the analyzed period (*Olaf Scholtz* and *Kamala Harris*). *Valéry Giscard* is an exception in that this former French president died in 2020 and was already retired from politics for a long time. These two factors could explain his favorable representation in the corpus. Strongly negative scores are mainly associated with politicians involved in prominent political scandals (*Claude Guéant, Penelope Fillon, Patrick Balkany*) and to authoritarian leaders (*Alexander Lukashenko*) or populist leaders (*Donald Trump, Jair Bolsonaro, Viktor Orban*). We

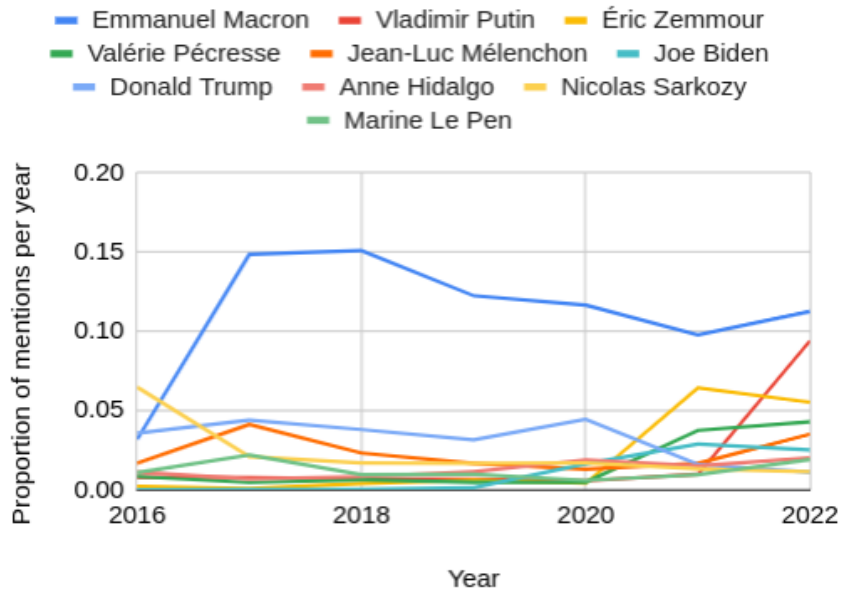


Figure 5.10: Mentions of most frequently mentioned politicians in time.

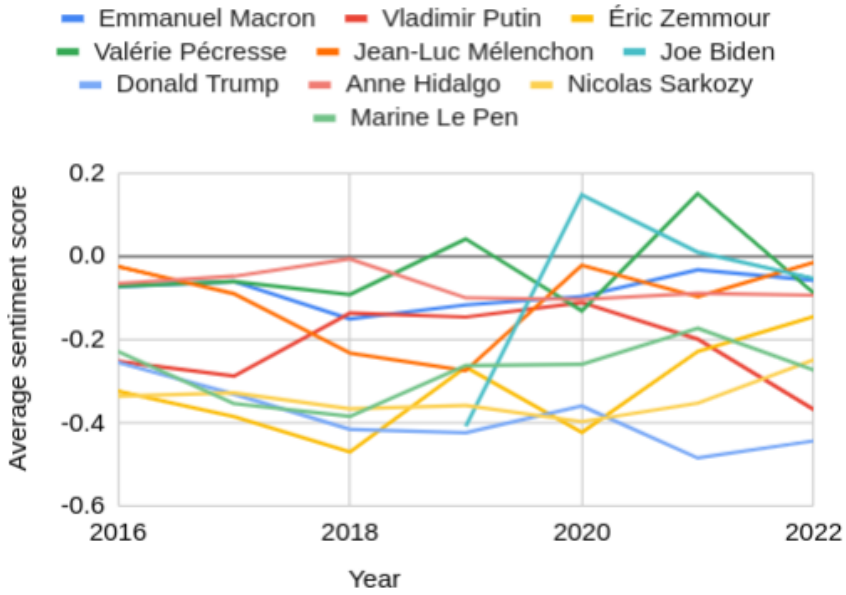


Figure 5.11: Sentiment associated with the most frequently mentioned politicians in time.

Scores	Politician name (average score)
Highest	Valéry Giscard (0.239); Carole Delga (0.193); Jean Lassalle (0.180); Roselyne Bachelot (0.163); Olaf Scholz (0.139); David Lisnard (0.135); Kamala Harris (0.125); Jerome Powell (0.104); Julien Denormandie (0.096); Michel Barnier (0.093)
Lowest	Claude Guéant (-0.70); Penelope Fillon (-0.61); Patrick Balkany (-0.52); Alexander Lukashenko (-0.50); Bernard Tapie (-0.48); Donald Trump (-0.43); Nicolas Hulot (-0.40); Jair Bolsonaro (-0.39); Vladimir Putin (-0.35); Viktor Orbán (-0.34)

Table 5.2: Politicians with the lowest and highest average sentiment scores (bottom and top of the figure, respectively). They were selected from the subset of the top-100 most frequently mentions politicians in our corpus.

note that *Penelope Fillon* has a much more negative score (-0.61) than *François Fillon* (-0.33), her husband, who provided her a no-show job for decades. This is explained by the fact that his representation in the corpus also includes the beginning of the 2017 presidential campaign, when *François Fillon* won his party's primary elections.

5.4.4 Gender Representation

Existing studies which quantify the representation of genders in politics [Asr et al., 2021; Doukhan et al., 2018] show that men are much more present than women. We deepen this analysis by adding a subjective component to understand how female and male politicians are represented in the news [17], with an illustration of mentions (Figure 5.12) and average sentiment (Figure 5.14). The results from Figure 5.12 show that the news from the analyzed corpus contain significantly more mentions of male than of female politicians. The percentage of female mentions varies from 16.7% (*lemonde*) to 22.3% (*leparisien*). Considering that the overall proportion of women in French politics is higher (38% in Parliament[18] and 50% in the Government), the results from Figure 5.12 show a clear under-representation of women. This bias is higher than the one observed for French audio-visual media, where female politicians amount for 30% of the total mentions [ARCOM, 2021]. The difference is probably explained by the stronger regulatory pressure on audio-visual media compared to written media. The temporal analysis of gender representation between 2016 and 2022 in Figure 5.13 shows that gender bias is relatively stable, with a slight decrease observed between 2017 and 2020, but which is reversed in 2021 and 2022. This finding indicates that the debate about the representation of gender in society have only a limited effect on the quantitative representation of women in politics.

The gender-focused analysis of sentiment from Figure 5.14 shows that the mentions of female politicians are more favorable than those of their male counterparts. This trend is respected for individual outlets, with some variation of the gap between genders. This positive coverage seems to be a positive signal for reducing gender bias in politics, but more efforts are needed to reduce the frequency gap.

[17] The corpus includes only 8 mentions of other genders.
[18] `https://data.ipu.org/content/france`

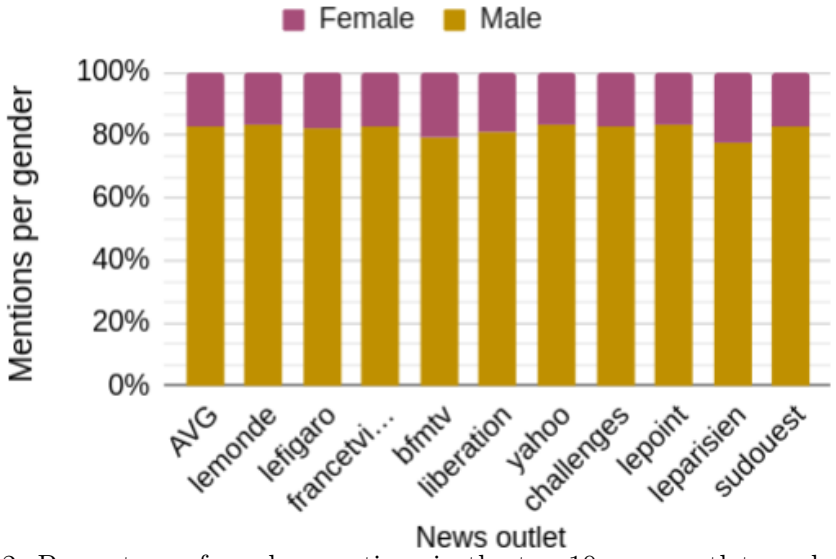


Figure 5.12: Percentage of gender mentions in the top-10 news outlets and on average.

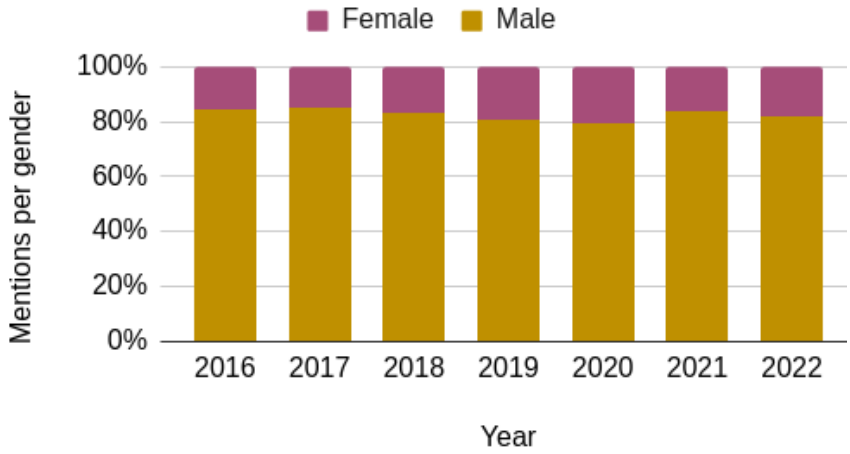


Figure 5.13: Percentage of gender mentions per year.

5.4.5 Age Representation

Figure 5.15 illustrates the average age of politicians mentioned in the corpus. The global average is 57.4 years, with individual averages varying between 54.6 for *leparisien* and 59.4 for *sudouest*. The average age of members of the French Parliament is 49.1, while the corresponding figure is even lower for the Government. This age-related bias confirms the qualitative results presented in Table 5.1, which shows that the majority of frequently mentioned politicians are older than the average. This is particularly the case for *Joe Biden, Donald Trump, Vladimir Putin* and *Nicolas Sarkozy*. The main exception is *Emmanuel Macron*, who is much younger than the average age reported in Figure 5.15.

5.4.6 Limitations

We discuss below a series of limitations of the proposed analyses, both in terms of data and processing.

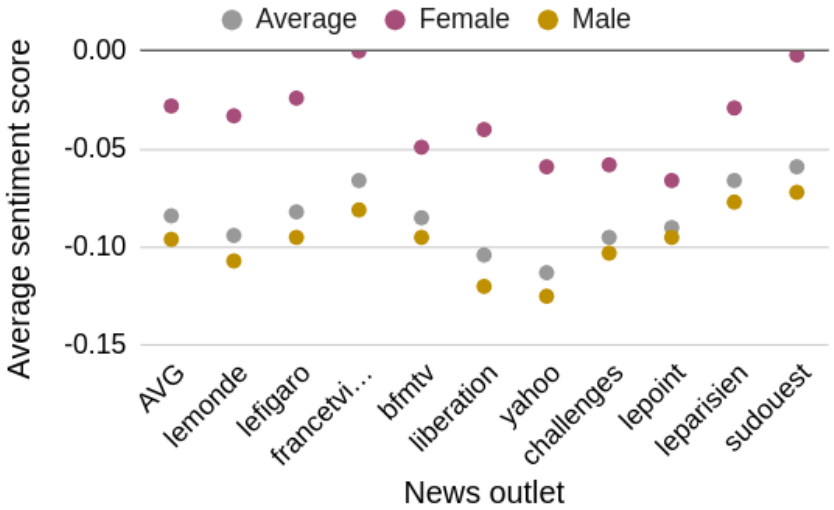


Figure 5.14: Mean sentiment classification scores by gender in the top-10 news outlets and on average.

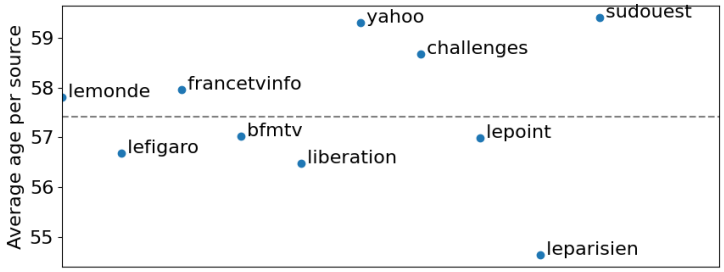


Figure 5.15: Average age of the politicians mentioned at the time of publication per news outlet.

Scope of the dataset. The collected dataset include a diversity of French media which are available online, but are focused on classical media. They do not include many articles published from alternative media, which could further increase diversity. The dataset content might induce biases regarding the results of the analysis. However, we carefully designed experiments so as to have sufficiently large samples in each case.

Dataset imbalance. The sources are represented by a variable number of news and of mentions of entities, as detailed in Figure 5.1. This is an effect of the collection strategies. The variable availibility of RSS streams and the existence of paywalls. However, this outlet-related imbalance is mitigated by the fact that the analyses are performed on aggregated samples of data.

Types of entities used for analyses. The entity detection and linking components handle persons, organizations, and locations. The first two types of entities are the most interesting for our work, but as pointed-out in Chapter 4, MAD-TSC only includes persons so far. The extension to organizations, political parties in particular, would be useful, although this absence of organizations is mitigated by the fact that politicians can be mapped to their political parties using Wikidata [Pellissier Tanon et al., 2016], which can be themselves assigned to political orientations via ParlGov [Döring and Manow, 2012].

Sentiment classification robustness. In section 5.3, we showed that sentiment classification leads to stable aggregated results despite imperfections for individual cases. However, two other potential limitations may affect the robustness of our analyses. First, although entity level aggregation is stable for popular (i.e. frequently mentioned) entities, we could imagine specific cases where the number of mentions isn't sufficient to obtain a robust aggregate. In this context, we should develop error estimation algorithms to estimate the confidence in our predictions. Although we know the error profile for the different classes of our model, we don't know the true underlying distribution of these classes in the articles of the outlets we study. This unknown distribution could be estimated using Bayesian simulations, which in turn could allow us to provide confidence bounds on our aggregate results. Another potential limitation is related to the robustness of results when inferring sentiment for a set of entities that are not in the training set. This is the case for the news corpus analyzed here. We perform a train/validation/test split of the target-dependent classification dataset, where the validation and test subsets contain 50% of the entities that appear in the training set and 50% that do not. After inference, we compute the $F1$ scores separately for the two test subsets, and the difference in performance is less than 1%. This indicates that the sentiment classification is transferable to new entities and can be used despite the temporal shift between MAD-TSC and the present news corpus.

5.4.7 Conlusion

We presented a set of analyses by instantiating our framework from Chapter 3 to comprehensively analyze a corpus of French news. These analyses can be of interest for different computational politics stakeholders. The NLP components are multilingual and the resources ensure international coverage. The approach could thus be easily transferred to other countries and other languages in order to understand the similarities and differences of political representations between them, as we'll briefly showcase with a concise Flemish news analysis in Subsection 5.5.3. We provide a comprehensive analysis focused on sources, politicians, and the representation of demographic segments. The obtained results are coherent with past findings [Cagé et al., 2022; Doukhan et al., 2018], but are significantly different, notably due to the important role given to the subjective perspective enabled by the use of target-dependent sentiment classification. At source level, we quantify the sentiment expressed of outlets and unveil their positioning with respect to the main political orientations. At topic level, we show that there are important differences between the distributions of mentions of political orientations and that the sentiment expressed about them varies significantly. At an individual level, we confirm that the French political representation in the media is dominated by the presidential function, and show that the automatic analysis is well-correlated with the public perception of politicians. Regarding demographic biases, we find that, unfortunately, the representation of politicians is still indicative of "old men's countries", as young and female politicians are still strongly underrepresented.

These findings are only examples of the insights which can be obtained with a rich and flexible representation of news, and can be used by multiple stakeholders. Newspaper editors and journalists can obtain feedback about their outlet's positioning and that of competitors, and push toward a more diversified coverage of political news. Social scientists can use this type of insights to deepen the analysis of the presentation of politicians and politics in the media. Political organizations can deploy the pipeline to monitor their political action and improve it. Regulatory bodies could use framework to quantify biases and push toward a more diversified representation of demographic segments. The proposed analyses are also a way to improve transparency about the political topics and positioning of news outlets. As such, they can improve the understanding of the political news landscape by the general public.

5.5 Analysis of Other Countries

In the following, we demonstrate the applicability of our pipeline to the analysis of the media landscape of other countries. We demonstrate it by analyzing the media landscape of Flemish Belgium which was done in collaboration with the Belgian media VRT to help them develop tools around the use of this framework. The whole work is also inscribed in our perspective to derive a multilingual and multicultural analysis at the European scale in the near future.

5.5.1 Multilingual Political Pre-Filtering Step

This pre-filtering step was carried out in the perspective of the future European Union news analysis that we plan to carry out, in this respect we present through the treatment of all the languages that were selected. Our goal is to characterize media outlets based on the sentiments they express toward key political entities. CCNews contains articles from a variety of sources, many of which have nothing to do with politics. Waiting till the entity linking step to filter out the non-political content would be very costly. To avoid processing the entire corpus, we propose the use of a coarse multilingual classifier that can exclude out-of-domain content. From observations of the URLs, we derive a list of keywords that, if present in an article's URL, quickly classify it as relevant or irrelevant for the political domain. Articles without any keyword are excluded from the classification dataset. We train a multilingual classifier to allow later the processing of other languages.

Data Set Construction

The list of keywords was derived empirically by looking at the different sets of urls. Since many news websites use the IPTC news categories in their URLs, it is useful to quickly generate classification datasets.

- Keywords for articles to keep: *"society", "international", "political", "economy"*
- Keywords for articles to discard: *"real estate", "foot", "sports", "video", "game", "scene", "lifestyle", "movie", "style", "announcement", "clip", "phone", "rugby", "drink", "mode", "automobile", "coffee", "culture", "espresso", "sport", "design", "interview", "history", "football", "beauty", "art", "journey", "music", "health", "television", "food", "pounds", "movable property", "kitchen", "weather", "gaming", "tv", "car", "financial"*

Using the Nllb-200-3.3B model from Team et al. [2022], we translate all these keywords into 10 target languages, which are the main languages of the countries from which we extracted and pre-processed the CCNews dataset: *Hungary (hu), Poland (pl), Romania (ro), Russia (ru), Ukraine (uk), Belgium (fr,nl), Denmark (da), France (fr), Germany (de), Netherlands (nl), Spain (es), United Kingdom (en).*

The resulting *relevant* (.i.e positive) and *irrelevant* (.i.e negative) articles are quite unbalanced across countries as seen in Figure 5.16. It is also imbalanced for languages and matched keywords.

We want to balance across classes, topics (i.e. keywords), countries, and languages. First, we impose a near 50% distribution between positive and negative classes. Then we subsample articles on a per-country basis, allowing only a doubling of relative quantity of articles between the most and least represented (country, language) pair. We also sample to balance as much as possible the keywords and source domains of the articles. The fact that some languages are spoken in several countries leads to this overrepresentation of French, Dutch and Russian compared

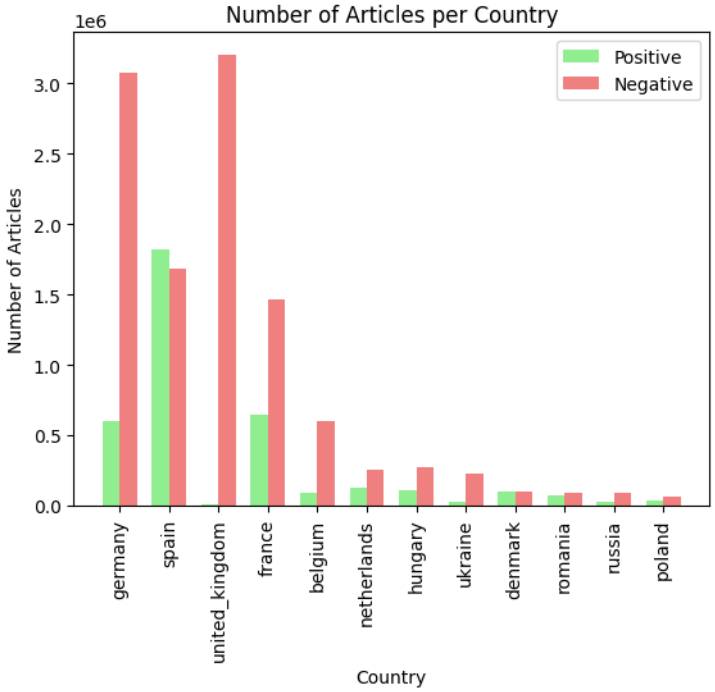


Figure 5.16: Number of articles per class and country in CCNews

to Ukrainian, as seen in Figure 5.17. The final dataset contains 359,823 instances, which we enrich with 14,740 positive instances from VoxEurop (selecting the language randomly among availables), as VoxEurop collects articles from a wide variety of European news outlets its an interesting source of positive articles. We then randomly split the dataset, leaving 7,500 entries in the validation and test sets. Without considering the domain names from VoxEurop articles, our dataset comprise 1,638 unique domain names.

Multilingual Classifier Training

We fine-tune a multilingual XLM-RoBERTa base in a single class classification setting, on 5 epochs using a linear scheduler with warmup, f16 precision, AdamW optimizer and 128 as effective batch size (See details in Table 5.3). We keep the best performing model on the validation set by checking the performance every third of an epoch. The final model has an accuracy of 92.8 on the test set.

Flemish News Corpus Filtering

We filter the whole corpus of Belgian news, preprocessing and filtering as described in subsection 3.3.2. We apply the classifier to the texts after deduplication and reduce the number of articles from 6,555,671 to 3,444,560 (French and Dutch together), of which 1,363,794 are Dutch. This results in 1,869,854 candidate politicians, which after applying the confidence threshold (See Section 5.3) leaves us with 427,811 politician mentions. We only keep the outlets for which we have more than 1,000 mentions, resulting in 24 outlets.

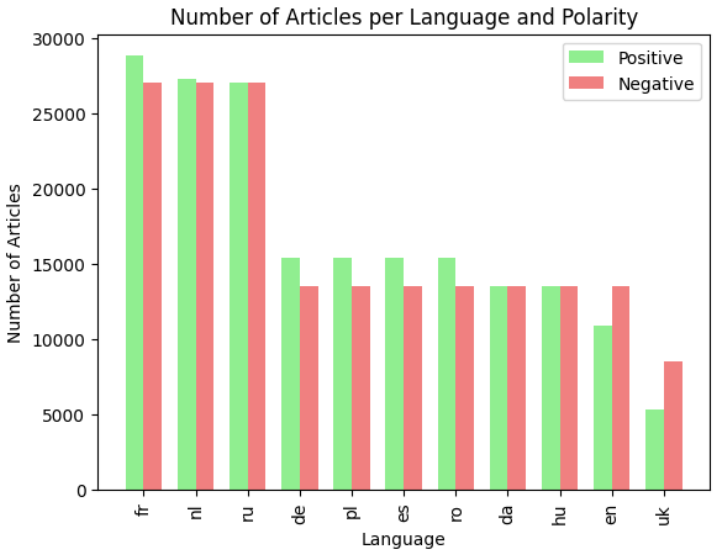


Figure 5.17: Number of articles per language and class after balancing in the classification dataset

Hyperparam	Fine-tuning
Learning Rate	1e-5
Batch Size	128
Weight Decay	0.01
Warmup	0.06
Max Epochs	5
Adam ϵ	1e-6
β_1	0.9
β_2	0.98
Optimizer	AdamW
Seeds	42

Table 5.3: Values of hyperparameters used for finetuning XLM-RoBerta binary classifier.

5.5.2 Sentiment Inference

As noted in subsection 4.4.4, the Dutch monolingual model has low performance compared to other monolingual models. This is due to the pre-training of the model, either symptomatic of a domain gap between the pretraining corpora and the news domain or an under-trained model. There are two possible solutions to this problem: adapting the monolingual model to improve its performance in the news domain, or translating Dutch sentences into English to apply the English model.

Model Adaptation

We follow the domain adaptation procedure of Rietzler et al. [2020] over a dataset of 733,431 Dutch political news articles, which after splitting into sentences and filtering by length leads us to 19,598,219 setnences. We train RobBERT [Delobelle et al., 2020b] using the hyperparameters in Table 5.4.

Hyperparam	Value
Learning Rate	5e-5
Batch Size	4096
Sequence Length	256
Precision	bf16
Weight Decay	0
Warmup	0.1
Epochs	4
Adam ϵ	1e-6
β_1	0.9
β_2	0.98
Optimizer	AdamW

Table 5.4: Values of hyperparameters used for domain adaptation of RobBERT [Delobelle et al., 2020b].

Fine-tuning the domain adapted model on MAD-TSC using the same procedure as for the original SPC_{NL} model (See Subsection 4.4.1) allows us to gain 4.6 points of $F1_m$ score, going from an average of 62.1 to 66.7 over 5 runs.

Automatic Translation with Markers

In our TSC machine translation experiments in subsection 3.2, we utilized the M2M100 translation 12B parameters model [Fan et al., 2020]) over the sentences. We then performed a match of mentions using a Levenshtein distance-based metric, thereby restricting our translation matching ability to proper nouns. A more recent approach by Chen et al. [2023], make use of the Nllb-200 3.3B parameters model [Team et al., 2022] by fine-tuning it to translate with markers. This approach allows the matching of arbitrary sequences between the source and the target language, notably common nouns. We translate the Dutch test set of MAD-TSC using this model to assess the applicability of this approach in the context of targeted-sentiment classification. It is to be noted that this translation with markers could also allow the low-cost creation of NER or TSC datasets for under-resourced languages.

Using the best English model trained on MAD-TSC over this translation leads to a score of 73.0 $F1_m$ compared to the translation using M2M100 that led to 72.1 $F1_m$. Although this result has been evaluated using only one model, it suggests the Nllb-200 model and its marker-based translation is an adequate strategy to use more performant English models.

Performances Comparison on MAD-TSC

Despite the increased performance allowed by the automatic translation for sentiment inference, we use the Dutch adapted model to make targeted sentiment classifications because processing the Flemish news corpora would involve high translation inference costs.

5.5.3 Flemish Political News Analysis

We compile the same set of figures as for the French political news analysis, although a stronger political knowledge of the Flemish political landscape would be necessary to fully interpret them.

Political Analysis

The analysis of the average sentiment in Figure 5.18 gives us interesting insights, among which tendencies that were already observed in the analysis of the French media landscape. Regional and local news outlets such as *kw, focus-wtv, ringtv, wtv*, and *tvl* show a significantly higher average sentiment compared to the general average, similar to what was observed with the French local outlets *la depeche*, *nicematin* and *letelegramme*. Likewise, business and economic news, represented by *tijd* and to a lesser extent *knack*, have a more positive sentiment, as observed with French business news *lesechos* and *latribune*. As in the French analysis, where *mediapart*, an investigative journalism outlet, is among the most negative outlets; *humo* and *apache*, also the only two involved in investigative journalism in our panel, are the most negative outlets in the Flemish landscape. With the exception of *zita*, which is an entertainment site that shares viral content and gossip, the others are general news providers that locally make up the rest of the sentiment distribution around the same level.

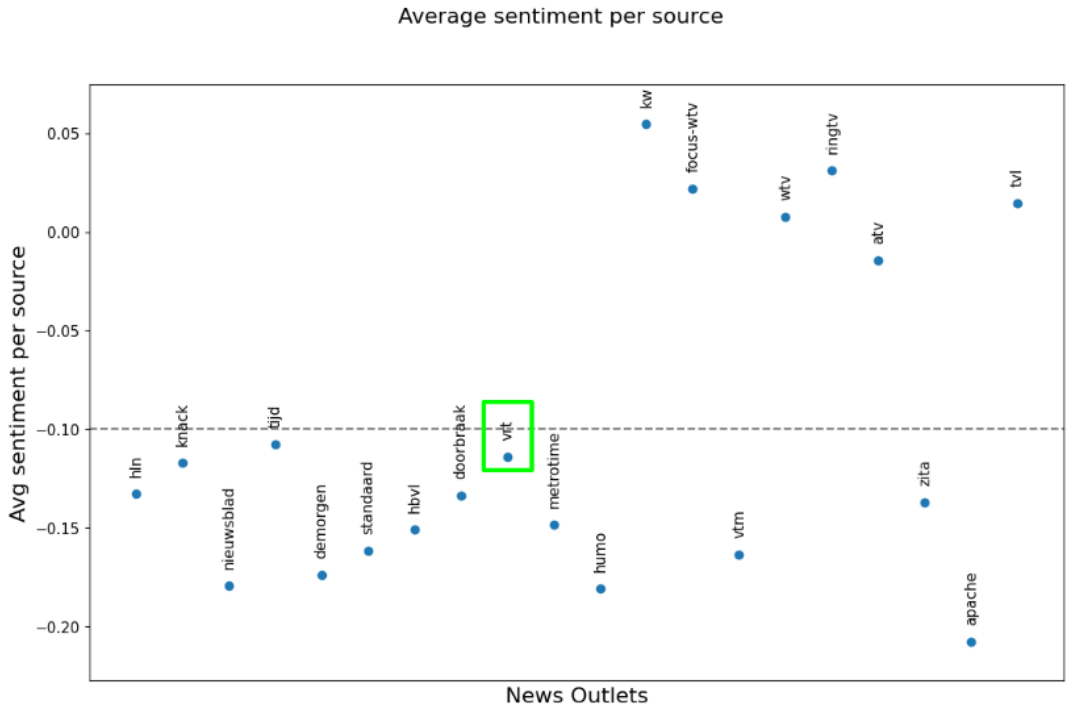


Figure 5.18: Average sentiment (blue dots) for the 24 most frequent news outlets in the corpus sorted by decreasing number of articles; the dashed line represents the average sentiment for the entire corpus.

The frequency and sentiment analysis of media outlets provides insights similar to those observed in the French landscape analysis. We again see a higher mention frequency of ruling center-right politicians. As in the French case, there is a noticeable negative sentiment bias towards the radical right (RR) and a positive sentiment bias towards the radical left (RL). A political expertise or reference on the Belgian media landscape would be necessary to further interpret the results.

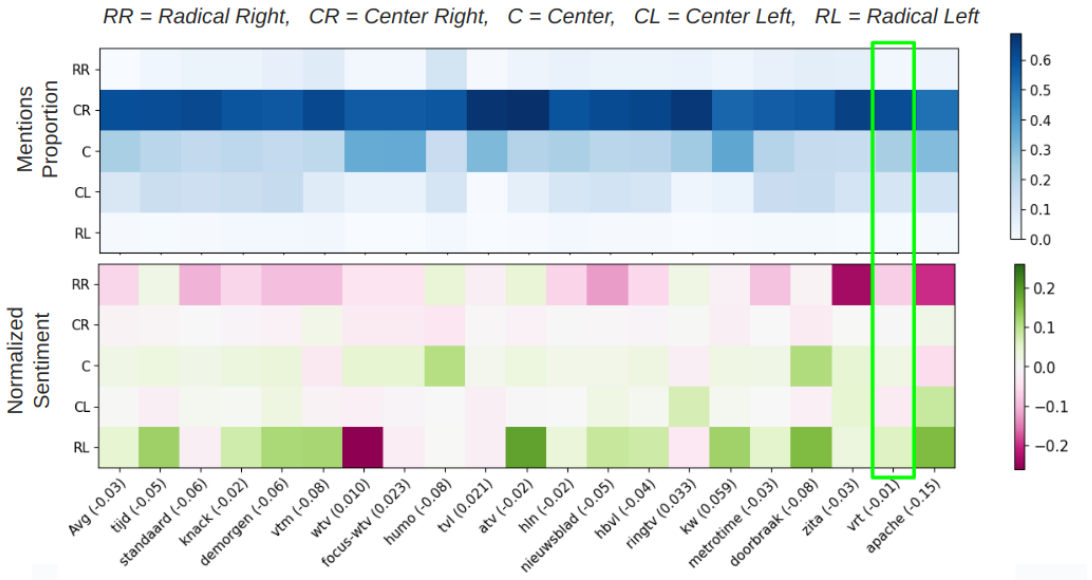


Figure 5.19: Deviation of the sentiment associated with major political orientations from the average sentiment of each source. Average sentiment of linked politicians is indicated in parentheses for each source.

Demographics Analysis

Gender Representation: Similar to the French analysis, the Belgian news landscape shows a significant gender disparity in political representation, as visible in Figure 5.20, despite the fact that women make up over 40% of parliamentary representatives[19]. Although the disparity in political representation is significantly greater at the local level, with only 15.8% of mayors being women, local news outlets tend to have better representation of women in their coverage on average. Similar to the trend in France, recent years have not shown a clear increase in the proportion of mentions of female politicians.

[19]Lukas Taylor, "Sharp rise in Belgian female politicians in 30 years," The Brussels Times, 2024. `https://www.brusselstimes.com/573261/sharp-rise-in-belgian-female-politicians-in-30-years`

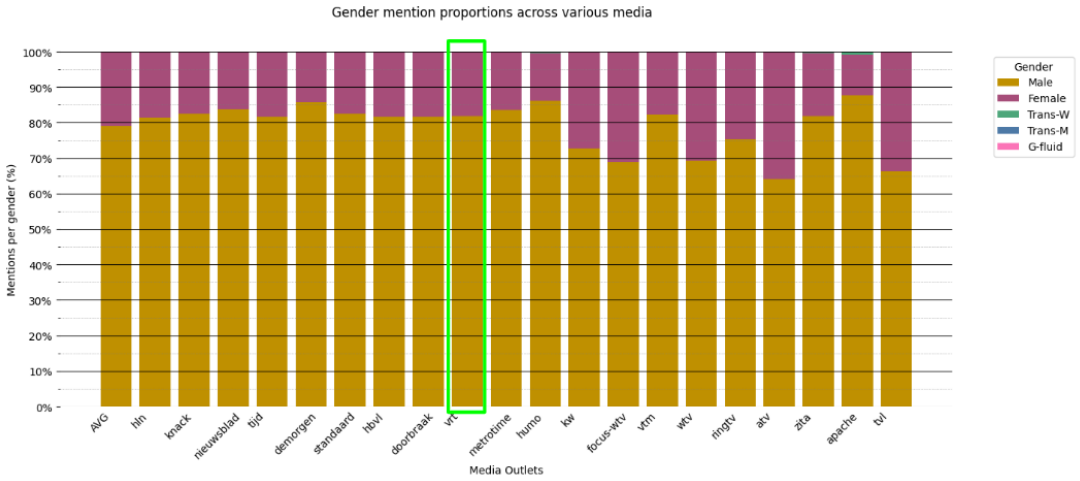


Figure 5.20: Gender mention proportions across various media outlets. The bars indicate the percentage of mentions per gender for each media outlet.

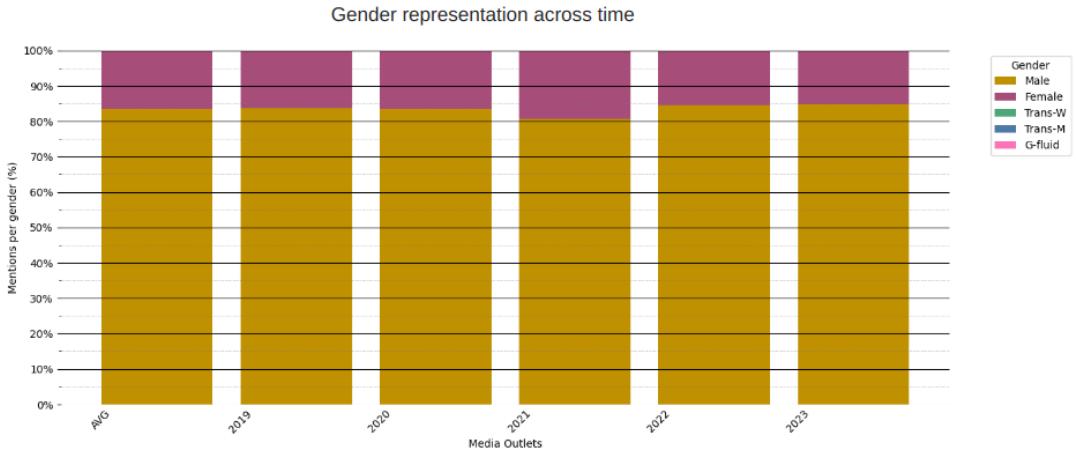


Figure 5.21: Gender representation across various media outlets from 2019 to 2023. The bars show the percentage of mentions per gender each year, including an average (AVG) column.

Age Representation: Similar to France, Figure 5.22 shows that while the average age of Belgian members of parliament is around 47 years[20] , the average age of mentioned Belgian politicians is approximately 57 years. Local news outlets tend to again differentiate themselves by mentioning younger politicians on average.

[20]IPU Parline, "Data on age: By country," available at `https://data.ipu.org/content/belgium?chamber_id=13343`

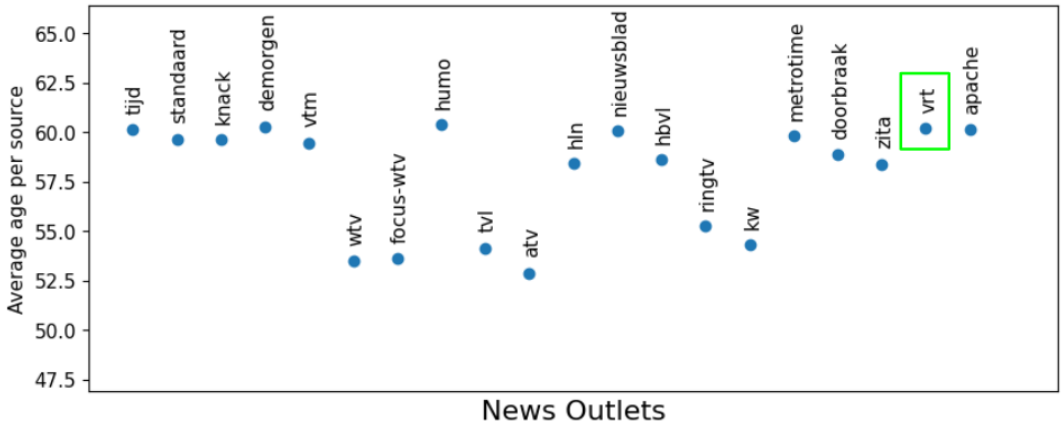


Figure 5.22: Average age of politicians mentioned per source.

5.6 Conclusion

The limitations of this Flemish news analysis are similar to those discussed in Subsection 5.4.6 for the French news. This additional study shows that our news analysis pipeline is easily adaptable to other countries, and it also shows that the Flemish and French media landscapes share certain trends, such as similar average sentiment levels across outlets of same nature (local, financial, investigative...) and similar trends in sentiment towards radical political groups, notably for outlet whose political leanings are known (progressive, conservative...). In addition, similar to France, the Flemish media landscape does not accurately reflect the distribution of politicians by gender and age. In the future, we aim to extend these analyses to other European countries, focusing on social issues and further exploring the similarities and differences in their news coverage.

Chapter 6

Conclusion

Contents

6.1 Summary and Contributions 100
6.2 Perspectives 101
6.2.1 Perspectives for Targeted Sentiment Classification in the News 101
6.2.2 Perspectives for the News Characterization Pipeline 103
6.2.3 Perspectives on Opinion and Stance Detection in the News 104

Over the last decades, our medium to access information have changed significantly. This has led to the end of the mass media era and rise of fragmentation and individualization of news consumption, especially through the use of social networks as main gateway to information. While the exact causal mechanism is unclear, polarization has risen concurrently, leading to instabilities in Western societies. The study of polarization and its mitigation is currently hindered by our inability to automatically understand the content of the news consumed by users. The development of algorithms capable of automatically characterizing the content shared is a crucial first step in developing solutions to improve the information available to the public through regulation, support tools, and improved recommendation systems. Most of the existing approaches in the literature have been developed from a US and English-centric perspective, which often limits their applicability to other cultures and languages. While there are some more general frameworks for news analysis, they are often based on frequency-based mentions metrics and do not use subjectivity metrics, or might measure very coarse ones. In this chapter, we first summarize the work done as part of this PhD thesis and its main contributions. Second, we present future research directions and perspectives that can be envisioned in the realm of automatic characterization of news media and articles.

6.1 Summary and Contributions

We began this manuscript with a comprehensive literature review in Chapter 2, discussing recent changes in information access and competing hypotheses about the causal mechanisms of polarization. We highlighted the need for automatic understanding and characterization of news content in order to develop solutions that can support users in their news consumption. We then presented natural language processing approaches that are commonly used in automatic news analysis, before shifting our focus to methods of media bias detection. Some of these methods are related to opinion and stance detection, which are of particular interest for the analysis of the political domain. We found that most methods for characterizing stances and political leanings in the literature are US-centric, tied to specific cultural, linguistic, and temporal contexts, and often lack modeling finesse and explanability. The only media bias detection techniques that are adapted to other languages in the literature are limited to the one relying on lexicons which are then automatically translated. It's in this context that we defined our three research questions:

- (**RQ1**) - How to design a framework that allows for fine-grained analysis of opinions expressed in text with greater explanability?
- (**RQ2**) - How generalizable can we make this opinion detection pipeline to facilitate transfer to other languages and other domains?
- (**RQ3**) - How can we use this framework to characterize the political leanings of news outlets?

We envisioned the exploration of two different tracks: the use of social networks sharing information as a remote monitoring signal for language models, and the use of targeted sentiment analysis towards key entities for the derivation of opinions. We briefly present preliminary work on the first option in Appendix A, although Elon Musk's acquisition of Twitter has hampered the ability to explore this avenue further. The rest of the manuscript therefore revolves around the second.

In Chapter 3, we explored **RQ1** by detailing our approach to fine-grained opinion analysis in the news. We introduced a versatile framework that should be applicable to all news domains,

and illustrated its instantiation for the political domain. The analysis revolves around the use of targeted sentiment analysis on key political entities linked to Wikidata and enriched with ParlGov. This pipelining allows us to derive trends and insights at different levels of granularity, responding to different levels of aggregation (outlet, entity, topic). This fine-grained approach ensures us to keep the explainability of the results and is also less susceptible to temporal deprecation due to the stability of sentiment expression over time. However, investigating the applicability of this approach and its ease of transfer across domains and languages has been impeded by the lack of targeted sentiment classification resources to perform these experiments.

To make up for this lack of resources, we developed MAD-TSC, the first large multilingual aligned dataset for Targeted Sentiment Classification (TSC) in political news, with 5,110 annotated entity mentions from 4,714 unique sentences, professionally translated and aligned in eight languages presented in Chapter 4. In addition to providing a unique resource for Targeted Sentiment Classification, this dataset also allowed us to answer **RQ2** by performing interesting experiments on the cross-lingual, cross-cultural and cross-domain transferability of TSC models. The aligned nature in 8 languages of the dataset allowed us to test the viability of transfer across languages using multilingual models or by performing automatic translation of the dataset. In particular, we showed that poor performance in less resourced languages was mostly due to weak pre-trained models in those languages. We also showed that automatic translation from less resourced languages to English could achieve similar performance to original English test set, facilitating the transfer of TSC-based approaches to other languages.

Finally, in chapter 5, we explored **RQ3** and showed that our targeted sentiment analysis pipeline towards political entities, complemented with objective frequency metrics enriched by the external databases ParlGov and Wikidata, allowed us to retrieve valuable insights into political and demographic characteristics of news outlets. We also presented the use of the pipeline in the context of Flemish news analysis, and observed recurrent patterns between both news landscapes characteristics.

6.2 Perspectives

There are several perspectives to automatic press characterization and its role in improving users' news consumption habits. We will first present direct iterations of our work, followed by a discussion of broader challenges in the field of automatic news analysis.

6.2.1 Perspectives for Targeted Sentiment Classification in the News

A first set of perspectives concern the improvements of Targeted Sentiment Classification (TSC) models in terms of languages and domain applicability, and performance.

Extension to other languages

It would be interesting to expand our work to perform an analysis at the scale of Europe and compare the depiction of political figures, the structuring of the press, and unveiling certain patterns or idiosyncracies of each country news media landscape.

While our cross-lingual experiments using MAD-TSC have demonstrated that expanding to other languages is feasible with recent automatic translation models, our experiments only included Germanic and Romance Indo-European languages. Since automatic translation performs better with closely related languages, further research is needed to explore the viability of the expansion to languages such as Hungarian and Finnish (both Uralic languages) or those from other distant families. More generally, while our experiments demonstrate strong machine translation performance on a selection of languages, they focus primarily on languages that are well-resourced relative to the majority of the world's languages. This machine translation paradigm also poses problem for the high cost of inference it is associated with. Current encoder-decoder translation models makes the solution expensive to deploy for massive text analyses. Potential improvements for encoder-decoder models could include research in size reduction (through distillation and pruning) and quantization, although most current efforts are focused on optimizing decoder-only models. Improving support for multilingual settings could also include enhancing cross-lingual transfer. While multilingual encoders show similarities in their representations across languages, achieving a unified semantic space across languages remains a challenge. Overcoming this could enable truly multilingual models through language-agnostic representations and broader cross-lingual applications.

Extension to other domains and entities

A current limitation of our work is the consideration of "Person" as the only entity type of interest for linking and sentiment analysis. If the sentiment expressed towards politicians seems to be the most indicative of political stance, the expression of sentiment towards "Organizations" and "Locations" might also be of interest in this perspective. The question of the transferability of TSC models across entity types in the news domain is also an open question that we want to explore. The literature highlights the challenges of transferring sentiment analysis methods across domains, and our preliminary result in transferring political news TSC models to the health news domain confirm this difficulty. The cross-domain transfer problem is difficult to define precisely; much like the ambiguous concept of "topic" in topic modeling, the notion of "domain" remains loosely defined. Most research in this area focuses on learning both domain-specific and domain-general representations. However, extending these approaches to zero-shot situations is complicated by the need for sufficient knowledge of the target domain and a clear conceptual understanding of the target task. In this context, instruction tuning of large language models shows promising results, even in TSC, although there's still room for improvement. In addition to the modeling challenge, the development of benchmarks to evaluate cross-domain transfer is essential. We also aim to address this issue by creating benchmarks for TSC in news across domains such as "Health", "Finance" and "Sports".

Modeling approaches to TSC

As pointed out in the perspectives Subsection 4.6.1 of Chapter 4, although encoder-based models are widely used in the literature, some modeling experiments should be conducted on the impact of input structuring (how to indicate the entity of interest), output structuring (how to retrieve the representation for classification), and loss functions (how to account for the hierarchical nature of the predicted labels). Another important avenue for improvement is the processing and aggregation of annotations. While most TSC datasets use majority voting for aggregation, this approach may be questionable for annotating political content, which often suffers from low inter-annotator agreement. Instead, predicting annotations for each annotator could pro-

vide additional signals for fine-tuning and allow for more precise control over predicted labels at inference time, although it poses the problem of the late aggregation strategy. Finally, if similarly sized decoder-based models haven't shown improvements in TSC tasks, large language models such as GPT-3 have shown comparable performance to Bert-finetuned models in certain sentiment analysis scenarios, particularly in zero-shot and few-shot settings. If fine-tuning some large decoder-based models can be costly, the question of synthetic data generation, automatic augmentation, or annotation is an interesting avenue to explore to compensate for the difficulty of annotating datasets [Mayhew et al., 2023]. However, it would also raise the question of the impact of political biases in LLMs, which could prevent their use in sentiment annotation of political news, ensuring fair and unbiased annotation being a challenge in itself.

6.2.2 Perspectives for the News Characterization Pipeline

Quantification of Uncertainties for Aggregated Predictions

While we discuss the relative robustness of our predictions in Chapter 3 on the average sentiment expressed towards the 1000 most frequent politicians, a clear measure of prediction uncertainty should be added, mostly to address the reliability of average sentiment prediction toward subgroups that might be underrepresented in a target outlet. There are three sources of uncertainty that we need to address. The first one comes from the TSC model itself, as the confusion matrices indicate that most errors made by the model are predictions in adjacent classes (i.e., neutral instead of positive), and that neutral class is majority in news articles . The second source of error arises from the underlying modeling assumption we make when averaging the sentiment scores (-1, 0, 1), where we need to establish the uncertainty of this average. The third type of error, although more difficult to address, can arise from differences in the distribution between the training set data and the target data, that could result in different model performances across news outlets.

The complexity around the error estimation of the average scores lies from the difference of classes distribution between the training set and the target outlet articles. Under these conditions, we could neglect the third source of error and try to estimate the true underlying distribution of positive, neutral, and negative sentiment in a given news outlet before assessing the error of the aggregated score.

With this simplification, conditioning on an outlet i and a subgroup j of interest, could be modeled as a 3-class multinomial distribution with two parameters p_{ij1}, p_{ij2}, such that we can define $p_{ij3} = 1 - (p_{ij1} + p_{ij2})$, with each parameter representing the probability that a given outlet i makes a negative, neutral or positive mention of the subgroup j. We should first derive the probability distribution of the true underlying polarities based on the model uncertainties, and use those distributions to establish the distributions of p_{ij1}, p_{ij2}. The score weighted distributions of p_{ij1}, p_{ij2} and p_{ij3} would allow us to derive the probability distribution of the average sentiment, and establish a confidence interval for it. The use of Bayesian simulation approach to automatically derive these distributions could allow to derive a credible interval around the mean of the distribution. However, this does not solve the problem of estimating the third type of error, which remains to be evaluated; we can expect cross-domain transfer to represent a higher bound, but it does not show good performance.

Automatic Segmentation of Analyses

In our analysis of the French press, we found interesting variations in the representation of political groups along manually defined topics using BM25. A better approach could be the automatic collection and classification around topics, perhaps following the IPTC news labeling, or by automatically matching topic-related articles. The choice of this automatic selection remains difficult because, as we described in subsection 2.3.3, topic models tend to have different definition and granularity depending on the studied object of interest. In addition, the news is subject to a high turnover of topics, and the automatic detection of the emergence of topics is a current challenge that would be interesting to address as part of automatic news analysis.

Similarly, our analyses of sentiment toward politicians have focused mostly on national politicians, but it may be interesting to include other subgroups in the study, such as international politicians, national and international institutions, countries, industries, companies. These analyses could, for example, complement the understanding of the peculiarities of media landscapes and cultural differences between countries. Another interesting avenue of research would be to automatically detect the distinguishing attributes of entities covered in the news on a given topic to automatically provide interesting subgroups to analyze, and provide more insights on the difference of perspectives between the different outlets.

6.2.3 Perspectives on Opinion and Stance Detection in the News

While the use of target-sentiment classification is the approach we chose to develop in this thesis, other different or complementary avenues can be envisioned in the near future to develop new methods that help better understand the news landscape using automatically.

Integrating Further Indications of Bias

There are other indicators of bias that are particularly understudied because of the difficulty to assess them automatically. A strong indicator of opinions can notably be found not in the text itself, but in the absence of text. In this respect, two avenues can be explored: (1) the identification of outlets that do not cover certain events which has already been explored in the past [Saez-Trumper et al., 2013]; (2) the other track consists in comparing how articles issued by different outlets and covering the same event differ by overemphasing, underemphasing or canceling alltogether certain informations; recent preliminary work in this are include [Jaradat et al., 2024].

While this indicator can be useful, three issues may hinder its effective use. First, determining the importance of missing information requires familiarity with current events and common knowledge. Second, it also supposes the ability to judge whether the absence of this information favors or disfavors a particular target, and which. Finally, it requires to ensure that the articles being compared were written at a time when journalists had the same level of information to avoid identifying nonexistent bias as the news develops.

Another avenue of development might be the detection of rhetorical devices. Recently, Piskorski et al. [2023] provided comprehensive datasets for the detection of rhetorical devices in news as part of SemEval2023 subtask 3. However, this task does not include the identification of stance-related effects (e.g., how the presence of rhetorical devices influences the consideration of certain objects in the article). Although challenging because it also requires knowledge of current events and common knowledge, this direction could complement our approach by identifying the entities most affected by different rhetorical devices. Finally, the integration of multiple metrics

also raises the issue of aggregation to effectively compare news articles and outlets.

Exploiting Social Network Information

Social networks, while not necessarily representative of the population at large, contain valuable political information that could be extracted to remotely monitor textual models. This was the first approach we considered in our work. However, a clear limitation is the accessibility of the data, which has become prohibitively expensive since the disappearance of the Twitter Research API. Nevertheless, the political data within such networks could provide a constantly updated signal for informed analysis of the press. Currently, most methods use bipartisan user clustering to train models for pro-against classification of news articles on specific topics. Automatically identifying trending topics and the nuanced positions of different segments of the population could enable a deeper exploration of diverse perspectives on current events.

Integration Into Digital Tools

Finally, another line of research is related to the ultimate goal of this research: the integration of automated press analysis models into digital tools, such as web extensions, news aggregators or recommender systems, in order to improve users' news consumption experience. This requires multidisciplinary collaboration to evaluate the impact and identify the best ways to inform the public, potentially helping to reduce polarization and promote healthy democratic debate.

Appendix A

Social Media and Analysis of Polarization in the News

A.1 Introduction

Digital platforms have democratized news production and facilitated access to a wider range of sources. Many news providers are driven more by short-term financial goals and/or political interests, while readers are driven, at least in part, by the goal of making sense of the world through access to reliable information. This misalignment is exacerbated by the fact that social networks, which are also driven by profit, often act as intermediaries between sources and readers. The selection of news that social network users read is largely determined by recommendation algorithms that prioritize articles that are assumed to match users' preferences, regardless of quality and veracity. Existing work [Ludwig et al., 2023] shows that recommenders lead to polarization in the long run, a phenomenon that is detrimental to public debate and ultimately to democracy. We present work in progress from the BOOM project[21], which focuses on understanding and characterizing polarization, and then proposing recommendation strategies that reduce it. The focus of the presented work is the automatic and fine-grained characterization of the stance of news articles using Twitter as a source of political knowledge.

A.2 Automatic Analysis of Political News Based on Social Media

The French media landscape is vast, with around 400 different media outlets reported by the Montaigne Institute in 2019[22]. In theory, this wide range of options gives citizens access to a variety of viewpoints and helps them make better sense of the world in an unbiased way. In practice, the access to news is to a large extent mediated by recommendation algorithms, which suggest items that are similar to previous consumptions. This exposure to a limited range of opinions can lead to polarization, especially for political topics that are prone to controversy.

The development of algorithms that model the content of news articles is a first step towards news recommendation systems that account for user preferences, but also promote diversity among recommendations. Automatically determining the political stance of news typically requires the

[21] `http://boom.loria.fr/`

[22] `https://www.institutmontaigne.org/publications/media-polarization-la-francaise`

use of machine learning algorithms that are trained on labeled data. This approach has several drawbacks. First, annotating data is expensive and must be on a large scale. Second, these annotations quickly become outdated due to the dynamic nature of news, where topics discussed and opinions expressed change rapidly over time. Finally, most labeled datasets use a simple annotation scheme (3 or 5 labels on a political scale from left to right), which tends to oversimplify the stance expressed in the news.

The use of social media, and Twitter in particular, as a dynamic source of up-to-date political information is appealing, but also challenging since data need to be organized. Twitter was successfully used for different types of political analyses [Antonakaki et al., 2021], including community detection, controversy detection, information dissemination analysis, and polarization quantification, among others. It was equally used to automatically determine the political stance of news outlets. For instance, the authors of [Stefanov et al., 2020] infer the positioning of several US news outlets along eight different polarizing topics by using graphs of retweets from vocal users. While interesting, this work has several important limitations: (1) the eight topics were pre-determined by manually selecting characteristic hashtags, and are static, (2) they force the political stance on each topic to be represented on a five points scale between pro and con; and (3) the stance is only inferred at news outlet level, implying that individual articles necessarily have the same stance on a topic.

We propose to address the problem differently and attempt to qualify stance more dynamically, without resorting to the classical left/right predefined classification. Such an approach is motivated by recent events which show that classical political positioning becomes irrelevant for polarizing issues, such as vaccination, COVID-19, or pension reform. Our final objective is to position the political stance of news articles, and this entails two steps. First, topic detection will be performed by clustering article representations obtained by using the TF-IDF weighted sum of fasttext word embeddings. Second, the alignment between individual news articles and automatically detected Twitter political communities will be predicted using either clustering or graph neural networks. Such an approach is interesting since both detection and article mapping to communities can be updated to follow the pace of news. These tasks are ongoing and results will be detailed in future work.

A.3 Preliminary Results and Future Work

We describe below the mapping between community representations and news outlets, which are an intermediate step of our approach. Inspired by [Morales et al., 2021], we collect a list of Twitter accounts to obtain a comprehensive representation of the French political spectrum. This list includes: French MPs of the XVth legislature, economists, journalists, presidential candidates, NGOs, and accounts which are active on politically-relevant topics such as humanitarian aid, COVID-19, racism, immigration, feminism, etc. We collected data for accounts which have between 10,000 and 250,000 followers. Note that we did not apply this filter to MPs to reproduce [Morales et al., 2021]. The lower bound is applied to remove accounts which have low influence. The upper bound is needed due to limitation of the Twitter API, but also because of the risk of political misrepresentation for very popular accounts, as they have followers that do not represent support and that come from across the political spectrum. This is illustrated in Figure A.1 by the slight offset in the positioning of Jean-Luc Mélenchon and Marine Le Pen, both with over 2 million followers. The final dataset includes lists of followers from 1,672 seed accounts. We reduced the binary matrix of accounts and followers to a 2D representation using correspondence analysis. The obtained communities, illustrated in Figure A.1, are in line with

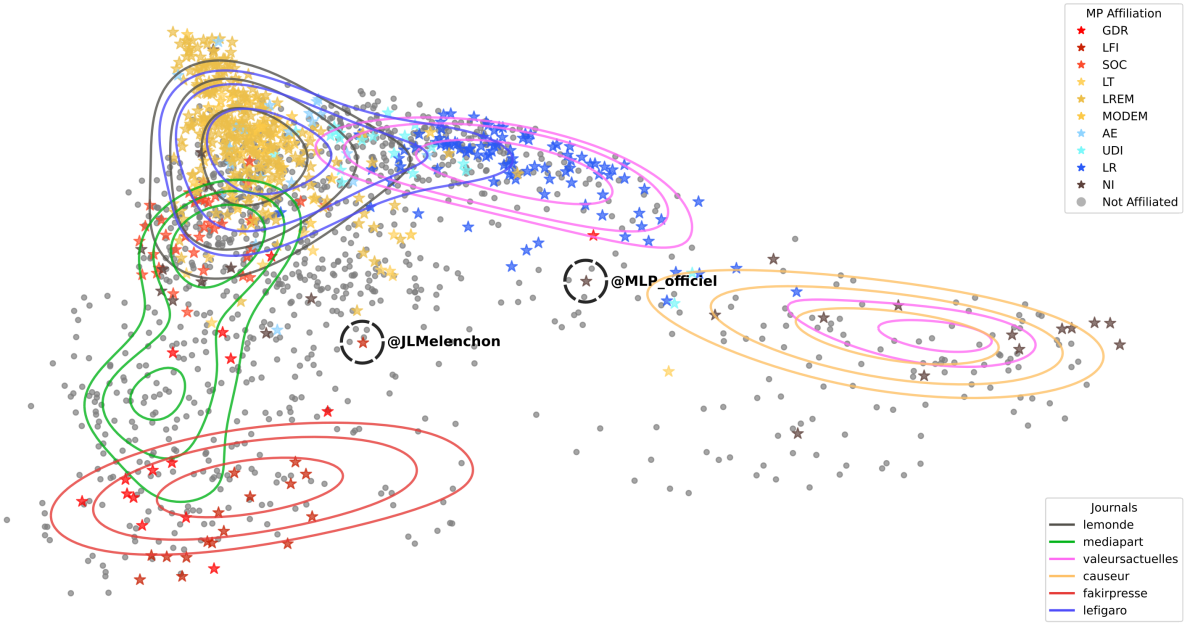


Figure A.1: Newspapers Distribution over Political Space

the results reported in [Morales et al., 2021] and confirm the usefulness of the approach.

We then collected, for all users, their tweets and retweets with URLs from 2016 until the results of the French presidential election of 2022. This raw dataset covers the entire XVth legislature, and include approximately two million tweets. We then selected only URLs with domain names of news outlets from the Montaigne Institute's mapping of the French media landscape. This subset includes 197 domain names, and over 500,000 URLs. We then used the distribution of domain names tweeted by different accounts to fit a mixture of Gaussians over the 2D projection of the political spectrum. The result for a subset of journals is illustrated in Figure A.1, with high densities being associated to the center of the contours.

These contour plots allow us to make several interesting observations. The alignment between newspapers and political communities follows intuition. Mainstream journals, such as Le Monde and Le Figaro, are associated to center-left and center-right groups, respectively. Causeur is clearly associated to far-right accounts, and Fakir to radical left. Interestingly, Mediapart and Valeurs Actuelles span over different political groups, toward the left and the right sides of the political spectrum, respectively. This might be indicative of agreement between political groups for a part of the topics, and of disagreement for others. Importantly, the results from Figure A.1 confirm the fragmentation of the news space. Readers who favor a limited number of sources are likely to be subject to polarization, it is thus important to propose recommender systems which propose diversified points of view, while being attentive not to antagonize readers.

A more detailed analysis, at topic and article levels, is needed in order to fully operationalize our approach. We have started collecting the articles linked in tweets. A preliminary analysis suggests that more than 85% of the URLs are still alive and that these articles are at least partially retrievable. Using the topic modeling information extracted from these articles, we will study the positioning of news outlets in the political space per topic. This will lead to a finer-grained characterization of the political communities, and a better mapping of individual articles in the political space. Then, given a user's current topic and her/his past consumption of news, we will be able to propose a diversification-driven recommendation strategy.

Appendix B

Targeted Sentiment Analysis for Political News

B.1 Supplementary results

Tables B.1, B.2, B.3 provide the standard deviation values for the experiments reported from Tables B.1, B.2, B.3 of the main text. Notations from the main tables are preserved.

B.2 Annotator Demographics

The main demographic characteristics of annotators are: (1) gender - 5 female/16 male; (2) age groups - 14 between 25 and 34, 5 between 35 and 44, 2 over 44, (3) countries of origin: France (16), Romania (4), Ivory Coast (1).

Model	Train	Test	$F1_m$	std_{F1_m}	$F1_{pn}$	$std_{F1_{pn}}$	acc	rec
SPC_{EN}	NM	NM	83.2	0.4	83.0	0.8	84.1	83.2
SPC_{EN}	NM	MAD	67.1	2.4	70.9	1.3	68.1	69.5
SPC_{ML}	NM	NM	81.0	0.3	80.3	0.5	81.6	81.1
SPC_{ML}	NM	MAD	61.5	1.3	64.7	1.3	61.8	63.2
TD_{EN}	NM	NM	83.7	0.7	83.4	0.8	83.8	83.0
TD_{EN}	NM	MAD	69.7	1.0	72.8	0.8	68.8	70.3
TD_{ML}	NM	NM	81.6	1.0	80.8	1.2	81.5	80.9
TD_{ML}	NM	MAD	63.3	1.5	65.5	1.4	62.9	64.3
PM_{EN}	NM	NM	83.6	0.6	83.1	1.0	84.1	83.5
PM_{EN}	NM	MAD	67.3	1.3	70.8	1.0	67.8	69.7
PM_{ML}	NM	NM	81.7	1.0	81.1	1.3	81.2	80.3
PM_{ML}	NM	MAD	60.9	1.9	63.7	1.3	61.7	62.5
BASE_{EN}	NM	NM	76.8	0.7	76.6	0.7	76.2	75.3
BASE_{EN}	NM	MAD	64.0	0.8	68.1	1.2	65.8	66.7
BASE_{ML}	NM	NM	74.1	0.7	74.0	1.2	74.6	74.3
BASE_{ML}	NM	MAD	59.2	1.5	62.8	0.8	60.5	62.3
SPC_{EN}	MAD	NM	76.7	1.0	76.6	1.7	77.3	75.7
SPC_{EN}	MAD	MAD	72.3	0.9	72.5	1.4	73.6	72.5
SPC_{ML}	MAD	NM	70.5	1.2	69.4	1.9	72.8	71.0
SPC_{ML}	MAD	MAD	67.8	1.6	66.9	1.7	68.6	68.0
TD_{EN}	MAD	NM	75.2	2.3	74.6	2.7	77.7	76.1
TD_{EN}	MAD	MAD	73.2	0.9	73.5	1.7	74.2	73.5
TD_{ML}	MAD	NM	71.9	1.2	70.0	2.8	75.0	72.6
TD_{ML}	MAD	MAD	68.5	1.1	66.6	2.0	69.8	68.5
PM_{EN}	MAD	NM	77.3	0.8	77.1	1.5	76.9	75.0
PM_{EN}	MAD	MAD	72.8	0.9	72.8	1.6	73.7	71.9
PM_{ML}	MAD	NM	69.3	4.8	67.3	6.5	75.0	73.5
PM_{ML}	MAD	MAD	66.6	3.1	64.7	4.6	69.3	68.8
BASE_{EN}	MAD	NM	68.8	0.6	68.6	1.4	69.4	68.2
BASE_{EN}	MAD	MAD	69.9	0.9	70.3	1.3	71.7	70.9
BASE_{ML}	MAD	NM	66.5	1.2	65.1	1.9	67.7	66.1
BASE_{ML}	MAD	MAD	64.8	1.1	62.9	2.2	66.9	65.9

Table B.1: Scores and standard deviations of $F1_m$, $F1_{pn}$, an accuracy (acc) and macro averaged recall (rec) scores for different configurations train/test configurations of MAD-TSC (MAD) and NewsMTSC (NM). These results complement those presented in Table 4.3 of the main text.

Pretrain	$F1_m$	std_{F1_m}	$F1_{pn}$	$std_{F1_{pn}}$	*acc*	*recall*
EN_{TG}	72.3	0.9	72.5	1.4	73.6	72.5
EN_{ML}	67.8	1.6	66.9	1.7	68.6	68.0
ES_{TG}	63.9	2.0	61.9	3.5	66.1	63.9
ES_{ML}	67.2	1.1	64.6	1.9	69.0	68.1
DE_{TG}	66.1	1.0	64.6	1.2	67.3	65.8
DE_{ML}	64.8	1.3	62.9	2.1	66.1	65.1
IT_{TG}	65.8	1.0	64.8	1.6	67.0	65.6
IT_{ML}	65.1	1.2	63.2	1.9	66.2	65.3
FR_{TG}	70.8	0.8	69.6	1.6	72.5	71.5
FR_{ML}	67.2	1.2	65.6	2.4	69.2	68.2
PT_{TG}	68.2	0.3	66.7	1.3	69.1	67.6
PT_{ML}	66.2	1.3	63.9	1.8	67.8	66.9
NL_{TG}	62.1	3.8	60.1	5.4	64.9	64.1
NL_{ML}	66.4	1.6	64.9	2.8	66.8	65.2
RO_{TG}	66.9	1.1	66.4	1.2	68.0	66.2
RO_{ML}	68.5	1.1	68.0	1.8	70.8	69.7

Table B.2: Scores and standard deviations of $F1_m$, $F1_{pn}$, an accuracy (*acc*) and macro averaged recall (*rec*) scores for monolingual experiments.These results complement those presented in Table 4.4 of the main paper.

Lang	Train	Test	$F1_m$	std_{F1_m}	$F1_{pn}$	$std_{F1_{pn}}$	acc	rec
ES	EN	EN_{M2M}	73.3	1.4	72.9	1.6	74.7	73.0
ES	EN	EN_{DL}	73.9	1.7	73.6	2.0	75.5	73.7
DE	EN	EN_{M2M}	70.8	1.5	70.1	1.6	72.2	70.6
DE	EN	EN_{DL}	73.2	1.0	72.6	1.4	74.7	72.9
IT	EN	EN_{M2M}	71.5	1.0	70.8	1.4	73.1	71.2
IT	EN	EN_{DL}	72.5	0.9	72.1	1.1	74.2	72.4
FR	EN	EN_{M2M}	70.6	1.3	70.0	1.6	72.6	70.4
FR	EN	EN_{DL}	73.5	1.3	73.3	1.5	75.2	73.3
PT	EN	EN_{M2M}	71.9	1.3	71.4	1.8	73.7	71.6
PT	EN	EN_{DL}	73.1	1.4	72.8	1.7	74.7	72.8
NL	EN	EN_{M2M}	71.1	1.3	70.7	1.6	72.7	70.7
NL	EN	EN_{DL}	72.1	0.9	71.8	1.2	74.0	71.8
RO	EN	EN_{M2M}	73.0	1.3	72.9	1.5	74.6	72.7
RO	EN	EN_{DL}	73.8	1.1	73.5	1.2	75.5	73.4
ES	TG	TG	63.9	2.0	61.9	3.5	66.1	63.9
ES	TG_{M2M}	TG	64.7	1.8	63.3	2.8	66.5	64.5
ES	TG_{DL}	TG	63.8	1.0	64.0	1.6	66.2	64.7
DE	TG	TG	66.1	1.0	64.6	1.2	67.3	65.8
DE	TG_{M2M}	TG	65.0	1.3	63.2	1.7	67.1	66.4
DE	TG_{DL}	TG	65.2	1.3	63.3	2.6	67.6	66.3
IT	TG	TG	65.8	1.0	64.8	1.6	67.0	65.6
IT	TG_{M2M}	TG	66.0	0.5	64.7	1.1	65.9	65.3
IT	TG_{DL}	TG	65.8	1.0	64.8	1.6	67.0	65.6
FR	TG	TG	70.8	0.8	69.6	1.6	72.5	71.5
FR	TG_{M2M}	TG	71.0	0.7	70.1	0.9	72.6	70.8
FR	TG_{DL}	TG	71.3	1.2	70.8	1.3	73.1	72.3
PT	TG	TG	68.2	0.3	66.7	1.3	69.1	67.6
PT	TG_{M2M}	TG	66.9	1.2	66.3	1.6	69.2	67.8
PT	TG_{DL}	TG	68.4	1.0	67.0	1.8	68.0	66.4
NL	TG	TG	62.1	3.8	60.1	5.4	64.9	64.1
NL	TG_{M2M}	TG	64.2	0.5	63.0	1.4	66.4	65.3
NL	TG_{DL}	TG	62.0	1.9	59.4	3.2	66.2	64.8
RO	TG	TG	66.9	1.1	66.4	1.2	68.0	66.2
RO	TG_{M2M}	TG	65.7	1.1	66.0	2.5	66.7	65.8
RO	TG_{DL}	TG	67.6	1.1	67.0	1.7	67.7	66.5

Table B.3: Scores and standard deviations of $F1_m$, $F1_{pn}$, an accuracy (acc) and macro averaged recall (rec) scores for machine translation experiments. These results complement those presented in Table 4.5 of the main paper.

Appendix C

French Press Analysis

C.1 Description of news sources

Table C.1 provides a brief description of top-10 outlets and the other outlets mentioned in the text. These descriptions are based on existing works such as Cointet et al. [2021] and Lenoir [2019]

Outlet	**Website**	**Description**
lemonde	`lemonde.fr`	Daily generalist newspaper with a centrist positioning
lefigaro	`lefigaro.fr`	Daily generalist newspaper with a center-right positioning
francetvinfo	`francetvinfo.fr`	News branch of the French national radio and TV broadcaster
bfmtv	`bfmtv.com`	News-focused TV channel
liberation	`liberation.fr`	Daily generalist newspaper with a left-wing positioning
yahoo	`yahoo.fr`	Generalist online news portal
challenges	`challenges.fr`	Weekly business magazine with centrist views
lepoint	`lepoint.fr`	Weekly political news magazine with centrist views
leparisien	`leparisien.fr`	Daily generalist newspaper with a centrist positioning
sudouest	`sudouest.fr`	Second largest regional daily newspaper in France, with a centrist positioning
les-crises	`les-crises.fr`	Business news website considered counter-information and anti-establishment right-leaning
causeur	`causeur.fr`	Monthly generalist newspaper with right-wing positioning
mediapart	`mediapart.fr`	Investigative French online newspaper with left-wing positioning
ladepeche	`ladepeche.fr`	Regional daily newspaper with a centrist positioning
nicematin	`nicematin.com`	Regional daily newspaper with a center-right positioning
letelegramme	`letelegramme.fr`	Regional daily newspaper with a centrist positioning
lesechos	`lesechos.fr`	Business news daily with a centrist positioning
latribune	`latribune.fr`	Business news weekly with a centrist positioning
humanite	`humanite.fr`	Daily generalist newspaper with a left positioning
closermag	`closer.fr`	Weekly press people magazine with a centrist positioning
valeursactuelles	`valeursactuelles.com`	Weekly news magazine with a right-wing positioning

Table C.1: The table presents the outlets used in the text, their websites, and a short description.

Bibliography

Micro-celebrities of information: Mapping calibrated expertise and knowledge influencers among social media veterinarians. ISSN 1369-118X. URL https://www.tandfonline.com/doi/epdf/10.1080/1369118X.2022.2109980?needAccess=true.

Understanding Media and Culture: An Introduction to Mass Communication. University of Minnesota Libraries Publishing, 01 edition, 2016. ISBN 978-1-946135-26-1. doi: 10.24926/8668.2601. URL http://open.lib.umn.edu/mediaandculture/.

Flood Mapping based on Online News using Named Entity Recognition. 2022. ISSN 18166075, 18180523. doi: 10.33168/JSMS.2022.0513. URL http://www.aasmr.org/jsms/Vol12/JSMS%20OCT%202022/Vol.12.No.05.13.pdf.

Amro Abbas, Kushal Tirumala, Dániel Simig, Surya Ganguli, and Ari S. Morcos. SemDeDup: Data-efficient learning at web-scale through semantic deduplication, March 2023.

Aly Abdelrazek, Yomna Eid, Eman Gawish, Walaa Medhat, and Ahmed Hassan. Topic modeling algorithms and applications: A survey. *Information Systems*, 112:102131, 2023.

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiu Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. MasakhaNER: Named Entity Recognition for African Languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131, 10 2021. ISSN 2307-387X. doi: 10.1162/tacl_a_00416. URL https://doi.org/10.1162/tacl_a_00416.

Samyak Agrawal, Kshitij Gupta, Devansh Gautam, and Radhika Mamidi. Towards detecting political bias in Hindi news articles. In Samuel Louvan, Andrea Madotto, and Brielen Madureira, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 239–244, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-srw.17. URL https://aclanthology.org/2022.acl-srw.17.

Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649. Association for Computational Linguistics, August 2018.

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. FLAIR: An easy-to-use framework for state-of-the-art NLP. In Waleed Ammar, Annie Louis, and Nasrin Mostafazadeh, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-4010. URL https://aclanthology.org/N19-4010.

Md Shad Akhtar, Palaash Sawant, Sukanta Sen, Asif Ekbal, and Pushpak Bhattacharyya. Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 572–582. Association for Computational Linguistics, June 2018.

AllSides. Allsides, 2012. URL https://www.allsides.com/.

David Alonso del Barrio and Daniel Gatica-Perez. Framing the news: From human perception to large language model inferences. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, pages 627–635, 2023.

Norah Alsaaran and Maha Alrabiah. Arabic named entity recognition: A bert-bgru approach. *Comput. Mater. Contin*, 68(1):471–485, 2021.

Evelin Amorim, Marcia Cançado, and Adriano Veloso. Automated Essay Scoring in the Presence of Biased Ratings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 229–237, New Orleans, Louisiana, 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1021.

Despoina Antonakaki, Paraskevi Fragopoulou, and Sotiris Ioannidis. A survey of Twitter research : Data model , graph structure , sentiment analysis. *Expert Syst. Appl.*, 164(July 2020):114006, 2021. ISSN 0957-4174. doi: 10.1016/j.eswa.2020.114006. URL https://doi.org/10.1016/j.eswa.2020.114006.

ARCOM. The representation of women on french tv and radio stations - 2021 report (in french). https://www.arcom.fr/sites/default/files/2022-03/La%20repr%C3%A9sentation%20des%20femmes%20%C3%A0%20la%20t%C3%A9l%C3%A9vision%20et%20%C3%A0%20la%20radio%20-%20Rapport%20sur%20l%27exercice%202021_0.pdf, 2021. Accessed: 2023-06-10.

Lisa P Argyle and Jeremy C Pope. Does Political Participation Contribute to Polarization in the United States? *Public Opinion Quarterly*, 86(3):697–707, 09 2022. ISSN 0033-362X. doi: 10.1093/poq/nfac036. URL https://doi.org/10.1093/poq/nfac036.

Elliot Aronson. The Theory of Cognitive Dissonance: A Current Perspective11This article and all the original work reported in it were supported by a grant from the National Science Foundation. volume 4 of *Advances in Experimental Social Psychology*, pages 1–34. Academic Press, 1969. doi: 10.1016/S0065-2601(08)60075-1. URL https://www.sciencedirect.com/science/article/pii/S0065260108600751.

Rexy Arulanandam, Bastin Tony Roy Savarimuthu, and Maryam A Purvis. Extracting crime information from online newspaper articles. In *Proceedings of the second australasian web conference-volume 155*, pages 31–38, 2014.

Fatemeh Torabi Asr, Mohammad Mazraeh, Alexandre Lopes, Vasundhara Gautam, Junette Gonzales, Prashanth Rao, and Maite Taboada. The gender gap tracker: Using natural language processing to measure gender bias in media. *PloS one*, 16(1):e0245533, 2021.

Susan Athey, Emilio Calvano, and Joshua Gans. The impact of the internet on advertising markets for news media, 2013. URL `http://www.nber.org/papers/w19419`.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), May 2010.

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. 115 (37):9216–9221, 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1804840115. URL `https://pnas.org/doi/full/10.1073/pnas.1804840115`.

Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. 348(6239):1130–1132, 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa1160. URL `https://www.science.org/doi/10.1126/science.aaa1160`.

Alexandra Balahur and Marco Turchi. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language*, 28(1): 56–75, 2014.

Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), May 2010.

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. Predicting Factuality of Reporting and Bias of News Media Sources, 2018. URL `http://arxiv.org/abs/1810.01765`.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991. Association for Computational Linguistics, 2020a.

Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3364–3374. Association for Computational Linguistics, 2020b. doi: 10.18653/v1/2020.acl-main.308. URL `https://www.aclweb.org/anthology/2020.acl-main.308`.

Adrien Barbaresi. Trafilatura: A web scraping library and command-line tool for text discovery and extraction. In Heng Ji, Jong C. Park, and Rui Xia, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 122–131, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-demo.15. URL https://aclanthology.org/2021.acl-demo.15.

David P. Baron. Persistent media bias. 90(1-2):1–36, 2006. ISSN 00472727. doi: 10.1016/j.jpubeco.2004.10.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0047272705000216.

Valentin Barriere and Alexandra Balahur. Improving sentiment analysis over non-English tweets using multilingual transformers and automatic translation for data-augmentation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 266–271. International Committee on Computational Linguistics, December 2020.

Pierpaolo Basile, Valerio Basile, Elena Cabrio, and Serena Villata. Argument Mining on Italian News Blogs. In Anna Corazza, Simonetta Montemagni, and Giovanni Semeraro, editors, *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016*. Accademia University Press, 2016. ISBN 978-88-99982-54-6. doi: 10.4000/books.aaccademia.1706.

Gilles Bastin. Gender imbalance in the media: Time lag or hysteresis?—french newspapers, gender parity shocks, and the long and winding road to the demasculinization of political reporting (1990–2020). *The International Journal of Press/Politics*, 0(0):19401612221143074, 2022. doi: 10.1177/19401612221143074. URL https://doi.org/10.1177/19401612221143074.

Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. International Sentiment Analysis for News and Blogs. 2(1):19–26, 2021. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v2i1.18606. URL https://ojs.aaai.org/index.php/ICWSM/article/view/18606.

Ulya Bayram, John Pestian, Daniel Santel, and Ali A. Minai. What's in a word? detecting partisan affiliation from word use in congressional speeches. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

Michael A. Beam. Automating the News: How Personalized News Recommender System Design Choices Impact News Reception. *Communication Research*, 41(8):1019–1041, December 2014. ISSN 0093-6502, 1552-3810. doi: 10.1177/0093650213497979.

Charlie Beckett and Sonia Livingstone. Tackling the information crisis: a policy framework for media system resilience-the report of the lse commission on truth trust and technology. 2018.

Tony Bennett. Media, 'reality', signification. In *Culture, Society and the Media*, pages 285–306. Routledge, 2005.

Jeffrey M Berry and Sarah Sobieraj. *The outrage industry: Political opinion media and the new incivility*. Oxford University Press, 2013.

Timothy Besley and Andrea Prat. Handcuffs for the grabbing hand? media capture and government accountability. *American Economic Review*, 96(3):720–736, June 2006. doi: 10.1257/aer.96.3.720. URL https://www.aeaweb.org/articles?id=10.1257/aer.96.3.720.

Samuel E. Bestvater and Burt L. Monroe. Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis*, 31(2):235–256, 2023. doi: 10.1017/pan.2022.10.

Shweta Bhatt, Sagar Joglekar, Shehar Bano, and Nishanth Sastry. Illuminating an Ecosystem of Partisan Websites. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, pages 545–554, Lyon, France, 2018. ACM Press. ISBN 978-1-4503-5640-4. doi: 10.1145/3184558.3188725.

Federico Bianchi, Marco Marelli, Paolo Nicoli, and Matteo Palmonari. SWEAT: Scoring Polarization of Topics across Different Corpora. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10065–10072, Online and Punta Cana, Dominican Republic, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.788.

Douglas Biber and Edward Finegan. Styles of stance in english: Lexical and grammatical marking of evidentiality and affect. *Text-interdisciplinary journal for the study of discourse*, 9(1):93–124, 1989.

Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

Gene Birz. Stale economic news, media and the stock market. 61:87–102, 2017. ISSN 01674870. doi: 10.1016/j.joep.2017.03.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S016748701730140X`.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Tolga Bolukbasi, Kai Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. pages 4356–4364, 2016. ISSN 10495258. URL `https://arxiv.org/abs/1607.06520v1`.

Bjarte Botnevik, Eirik Sakariassen, and Vinay Setty. BRENDA: Browser Extension for Fake News Detection. 2020.

Mark Boukes. Infotainment. In Tim P. Vos, Folker Hanusch, Dimitra Dimitrakopoulou, Margaretha Geertsema-Sligh, and Annika Sehl, editors, *The International Encyclopedia of Journalism Studies*, pages 1–9. Wiley, 1 edition, 2019a. ISBN 978-1-118-84167-9 978-1-118-84157-0. doi: 10.1002/9781118841570.iejs0132. URL `https://onlinelibrary.wiley.com/doi/10.1002/9781118841570.iejs0132`.

Mark Boukes. Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news. 16(1):36–51, 2019b. ISSN 1933-1681, 1933-169X. doi: 10.1080/19331681.2019.1572568. URL `https://www.tandfonline.com/doi/full/10.1080/19331681.2019.1572568`.

Amber E Boydstun, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. Tracking the development of media frames. *Work. Pap.*, pages 1–25, 2014. URL `http://www.cs.cmu.edu/$\sim$nasmith/papers/boydstun+card+gross+resnik+smith.apsa14.pdf`.

Gianni Brauwers and Flavius Frasincar. A survey on aspect-based sentiment classification. *ACM Comput. Surv.*, 55(4), nov 2022. ISSN 0360-0300.

Paul Brighton and Dennis Foy. *News values*. Sage, 2007.

A.Z. Broder. On the resemblance and containment of documents. In *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No.97TB100171)*, pages 21–29, Salerno, Italy, 1998. IEEE Comput. Soc. ISBN 978-0-8186-8132-5. doi: 10.1109/SEQUEN.1997.666900. URL http://ieeexplore.ieee.org/document/666900/.

Elena Cabrio and Serena Villata. Five Years of Argument Mining: A Data-driven Analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 5427–5433, Stockholm, Sweden, July 2018. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-2-7. doi: 10.24963/ijcai.2018/766.

Julia Cagé, Moritz Hengel, Nicolas Hervé, and Camille Urvoy. Hosting media bias: Evidence from the universe of french broadcasts, 2002-2020. *SSRN Electronic Journal*, 2022. doi: 10.2139/SSRN.4036211.

Yitao Cai and Xiaojun Wan. Multi-Domain Sentiment Classification Based on Domain-Aware Embedding and Attention. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pages 4904–4910, Macao, China, August 2019. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-4-1. doi: 10.24963/ijcai.2019/681.

Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. 356(6334):183–186, 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aal4230. URL https://www.science.org/doi/10.1126/science.aal4230.

Jiahao Cao, Rui Liu, Huailiang Peng, Lei Jiang, and Xu Bai. Aspect is not you need: No-aspect differential sentiment framework for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1599–1609. Association for Computational Linguistics, July 2022.

Helen Caple and Monika Bednarek. Rethinking news values: What a discursive approach can tell us about the construction of news discourse and news photography. *Journalism*, 17(4):435–455, 2016. doi: 10.1177/1464884914568078. URL https://doi.org/10.1177/1464884914568078.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. The media frames corpus: Annotations of frames across issues. volume 2, pages 438–444, 2015. ISBN 9781941643730. doi: 10.3115/v1/p15-2072. URL https://aclanthology.org/P15-2072.pdf.

Tommaso Caselli, Rachele Sprugnoli, and Oana Inel. Temporal Information Annotation: Crowd vs. Experts. 2016.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684, 2011.

Steven H. Chaffee and Miriam J. Metzger. The End of Mass Communication? 4(4): 365–379, 2001. ISSN 1520-5436, 1532-7825. doi: 10.1207/S15327825MCS0404_3. URL http://www.tandfonline.com/doi/abs/10.1207/S15327825MCS0404_3.

Manolis Chalkiadakis, Alexandros Kornilakis, Panagiotis Papadopoulos, Evangelos Markatos, and Nicolas Kourtellis. The Rise and Fall of Fake News sites: A Traffic Analysis. In *13th ACM Web Science Conference 2021*, pages 168–177. ACM, 2021. ISBN 978-1-4503-8330-1. doi: 10.1145/3447535.3462510. URL https://dl.acm.org/doi/10.1145/3447535.3462510.

Yang Chen, Chao Jiang, Alan Ritter, and Wei Xu. Frustratingly easy label projection for cross-lingual transfer. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5775–5796, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.357. URL https://aclanthology.org/2023.findings-acl.357.

Xiao Cheng and Dan Roth. Relational Inference for Wikification. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu, and Steven Bethard, editors, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1787–1796, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

Kenneth Church and Patrick Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

Rob Churchill and Lisa Singh. The evolution of topic modeling. *ACM Computing Surveys*, 54 (10s):1–35, 2022.

Jean-Philippe Cointet, Dominique Cardon, Andreï Mogoutov, Benjamin Ooghe-Tabanou, Guillaume Plique, and Pedro Morales. Uncovering the structure of the french media ecosystem. *arXiv preprint arXiv:2107.12073*, 2021.

Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. STANDER: An expert-annotated dataset for news stance detection and evidence retrieval. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4086–4101, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.365. URL https://aclanthology.org/2020.findings-emnlp.365.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July 2020. Association for Computational Linguistics.

Keith Cortis and Brian Davis. A dataset of multidimensional and multilingual social opinions for malta's annual government budget. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):971–981, May 2021.

Silviu Cucerzan and David Yarowsky. Language independent named entity recognition combining morphological and contextual evidence. In *1999 joint SIGDAT conference on empirical methods in natural language processing and very large corpora*, 1999.

Jingfeng Cui, Zhaoxia Wang, Seng-Beng Ho, and Erik Cambria. Survey on sentiment analysis: Evolution of research methods and topics. 56(8):8469–8510, 2023. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-022-10386-z. URL https://link.springer.com/10.1007/s10462-022-10386-z.

Christian S. Czymara and Marijn Van Klingeren. New perspective? Comparing frame occurrence in online and traditional news media reporting on Europeâ€™s â€œMigration Crisisâ€. 47(1):136–162, 2022. ISSN 1613-4087, 0341-2059. doi: 10.1515/commun-2019-0188. URL https://www.degruyter.com/document/doi/10.1515/commun-2019-0188/html.

Jamell Dacon and Haochen Liu. Does gender matter in the news? detecting and examining gender bias in news articles. In *Companion Proceedings of the Web Conference 2021*, WWW '21, pages 385–392, Ljubljana, Slovenia, 2021. Association for Computing Machinery. ISBN 9781450383134. doi: 10.1145/3442442.3452325. URL https://doi.org/10.1145/3442442.3452325.

Junqi Dai, Hang Yan, Tianxiang Sun, Pengfei Liu, and Xipeng Qiu. Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1816–1829, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.146.

Samantha D'Alonzo and Max Tegmark. Machine-Learning media bias. 2021. URL http://arxiv.org/abs/2109.00024.

Kahneman Daniel. *Thinking, fast and slow.* 2017.

Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. Unsupervised user stance detection on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):141–152, 2020.

Nicola De Cao, Ledell Wu, Kashyap Popat, Mikel Artetxe, Naman Goyal, Mikhail Plekhanov, Luke Zettlemoyer, Nicola Cancedda, Sebastian Riedel, and Fabio Petroni. Multilingual autoregressive entity linking. *Transactions of the Association for Computational Linguistics*, 10:274–290, 2022. doi: 10.1162/tacl_a_00460. URL https://aclanthology.org/2022.tacl-1.16.

Javier De la Rosa, Eduardo G. Ponferrada, Manu Romero, Paulo Villegas, Pablo González de Prado Salas, and María Grandury. Bertin: Efficient pre-training of a spanish language model using perplexity sampling. *Procesamiento del Lenguaje Natural*, 68(0):13–23, 2022. ISSN 1989-7553.

Lizzie Dearden. The fake refugee images that are being used to distort public opinion on asylum seekers. *The Independent*, 16(9):15, 2015.

Jérôme Delobelle, Amaury Delamaire, Elena Cabrio, Ramón Ruti, and Serena Villata. Sifting the Arguments in Fake News to Boost a Disinformation Analysis Tool. 2020a.

Pieter Delobelle, Thomas Winters, and Bettina Berendt. RobBERT: a Dutch RoBERTa-based Language Model. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3255–3265. Association for Computational Linguistics, November 2020b.

Anup Anand Deshmukh and Udhav Sethi. IR-BERT: Leveraging BERT for Semantic Search in Background Linking for News Articles. July 2020. URL http://arxiv.org/abs/2007.12603.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images, April 2021.

Michael Dimock, Carroll Doherty, Jocelyn Kiley, and Russ Oates. NUMBERS, FACTS AND TRENDS SHAPING THE WORLD.

Milena Djourelova and Ruben Durante. Media Attention and Strategic Timing in Politics: Evidence from U.S. Presidential Executive Orders. 66(4):813–834, 2022. ISSN 0092-5853, 1540-5907. doi: 10.1111/ajps.12683. URL `https://onlinelibrary.wiley.com/doi/10.1111/ajps.12683`.

Wolfgang Donsbach. Psychology of news decisions: Factors behind journalists' professional behavior. *Journalism*, 5(2):131–157, 2004.

Holger Döring and Philip Manow. Parliament and government composition database (parlgov). *An infrastructure for empirical information on parties, elections and governments in modern democracies. Version*, 12(10), 2012.

David Doukhan, Géraldine Poels, Zohra Rezgui, and Jean Carrive. Describing gender equality in french audiovisual streams with a deep learning approach. *VIEW Journal of European Television History and Culture*, 7(14):103–122, 2018.

Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, and Jianxin Liao. Adversarial and domain-aware BERT for cross-domain sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4019–4028, Online, July 2020. Association for Computational Linguistics.

Evan Dufraisse, Célina Treuillier, Armelle Brun, Julien Tourille, Sylvain Castagnos, and Adrian Popescu. Don't burst blindly: For a better use of natural language processing to fight opinion bubbles in news recommendations. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences*, pages 79–85. European Language Resources Association, June 2022.

Evan Dufraisse, Adrian Popescu, Julien Tourille, Armelle Brun, and Jerome Deshayes. MAD-TSC: A multilingual aligned news dataset for target-dependent sentiment classification. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8286–8305, Toronto, Canada, July 2023a. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.461. URL `https://aclanthology.org/2023.acl-long.461`.

Evan Dufraisse, Julien Tourille, Armelle Brun, and Adrian Popescu. Analysis of polarization in the news. 2023b.

Stefan Dumitrescu, Andrei-Marius Avram, and Sampo Pyysalo. The birth of Romanian BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4324–4328. Association for Computational Linguistics, November 2020.

Patricia Edgar. Agenda setting. 18(1):49–49, 1980. doi: 10.1177/1329878X8001800117. URL `https://doi.org/10.1177/1329878X8001800117`.

Cheikh Brahim El Vaigh, François Goasdoué, Guillaume Gravier, and Pascale Sébillot. Using Knowledge Base Semantics in Context-Aware Entity Linking. In *Proceedings of the ACM Symposium on Document Engineering 2019*, pages 1–10, Berlin Germany, September 2019. ACM. ISBN 978-1-4503-6887-2. doi: 10.1145/3342558.3345393.

Robert M. Entman. How the Media Affect What People Think: An Information Processing Approach. *The Journal of Politics*, 51(2):347–370, May 1989. ISSN 0022-3816, 1468-2508. doi: 10.2307/2131346.

Robert M. Entman. Framing: Toward Clarification of a Fractured Paradigm. 43(4):51–58, 1993. ISSN 0021-9916, 1460-2466. doi: 10.1111/j.1460-2466.1993.tb01304.x. URL https://academic.oup.com/joc/article/43/4/51-58/4160153.

Gürdal Ertek and Lakshmi Kailas. Analyzing a decade of wind turbine accident news with topic modeling. *Sustainability*, 13(22):12757, 2021.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. Beyond English-Centric Multilingual Machine Translation, October 2020.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. In Plain Sight: Media Bias Through the Lens of Factual Reporting, 2019. URL http://arxiv.org/abs/1909.02670.

Natalie Fenton. *Drowning or Waving? New Media, Journalism and Democracy*, pages 3–16. SAGE Publications Ltd, 2010. ISBN 978-1-84787-574-7 978-1-4462-8001-0. doi: 10.4135/9781446280010.n1. URL https://sk.sagepub.com/books/new-media-old-news/n1.xml.

Anjalie Field and Yulia Tsvetkov. Entity-Centric Contextual Affective Analysis. *arXiv:1906.01762 [cs]*, pages 2550–2560, April 2019. doi: 10.18653/v1/p19-1243.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. Framing and Agenda-setting in Russian News: A Computational Analysis of Intricate Political Strategies, 2018. URL http://arxiv.org/abs/1808.09386.

Anjalie Field, Gayatri Bhat, and Yulia Tsvetkov. Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories. 13:158–169, 2019. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v13i01.3358. URL https://ojs.aaai.org/index.php/ICWSM/article/view/3358.

Seth Flaxman, Sharad Goel, and Justin M. Rao. Filter Bubbles, Echo Chambers, and Online News Consumption. 80(S1):298–320, 2016. ISSN 0033-362X, 1537-5331. doi: 10.1093/poq/nfw006. URL https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfw006.

Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.

Mahsa Forouzesh and Patrick Thiran. Differences Between Hard and Noisy-labeled Samples: An Empirical Study, July 2023.

Improve The News Foundation. Verity, 2020. URL https://www.improvethenews.org/.

Jonathan L. Freedman and David O. Sears. Selective Exposure11The preparation of this paper was supported in part by NSF grants to the authors. volume 2 of *Advances in Experimental Social Psychology*, pages 57–97. Academic Press, 1965. doi: 10.1016/S0065-2601(08)60103-3. URL https://www.sciencedirect.com/science/article/pii/S0065260108601033.

Jeremy A. Frimer, Linda J. Skitka, and Matt Motyl. Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. 72:1–12, 2017. ISSN 00221031. doi: 10.1016/j.jesp.2017.04.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0022103116304024.

Thomas Gaillat, Bernardo Stearns, Gopal Sridhar, Ross McDermott, Manel Zarrouk, and Brian Davis. Implicit and explicit aspect extraction in financial microblogs. In *Proceedings of the First Workshop on Economics and Natural Language Processing*, pages 55–61. Association for Computational Linguistics, 2018.

William A Gamson, David Croteau, and William Hoynes. Media Images and the Social Construction of Reality. 2024.

Rama Rohit Reddy Gangula, Suma Reddy Duggenpudi, and Radhika Mamidi. Detecting political bias in news articles using headline attention. In Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov, and Dieuwke Hupkes, editors, *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 77–84, Florence, Italy, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4809. URL https://aclanthology.org/W19-4809.

Soumen Ganguly, Juhi Kulshrestha, Jisun An, and Haewoon Kwak. Empirical evaluation of three common assumptions in building political media bias datasets. pages 939–943, 2020. URL https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=7091&context=sis_research.

Zhengjie Gao, Ao Feng, Xinyu Song, and Xi Wu. Target-Dependent Sentiment Classification With BERT. *IEEE Access*, 7:154290–154299, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2946594.

Matthew Gentzkow and Jesse M Shapiro. What Drives Media Slant? Evidence From U.S. Daily Newspapers. 78(1):35–71, 2010. ISSN 0012-9682. doi: 10.3982/ECTA7195. URL http://doi.wiley.com/10.3982/ECTA7195.

Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. 2016.

Martin Gilens and Craig Hertzman. Corporate Ownership and News Bias: Newspaper Coverage of the 1996 Telecommunications Act. 62(2):369–386, 2000. ISSN 0022-3816, 1468-2508. doi: 10.1111/0022-3816.00017. URL https://www.journals.uchicago.edu/doi/10.1111/0022-3816.00017.

Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. Large-Scale Sentiment Analysis for News and Blogs (System Demonstration). 2007.

Didier Gohourou and Kazuhiro Kuwabara. Knowledge graph extraction of business interactions from news text for business networking analysis. 6(1):126–142, 2024. ISSN 2504-4990. doi: 10.3390/make6010007. URL https://www.mdpi.com/2504-4990/6/1/7.

Abhay Goyal, Muhammad Siddique, Nimay Parekh, Zach Schwitzky, Clara Broekaert, Connor Michelotti, Allie Wong, Lam Yin Cheung, Robin O. Hanlon, Lam Yin Cheung, Munmun De Choudhury, Roy Ka-Wei Lee, and Navin Kumar. ChatGPT and Bard Responses to Polarizing Questions, 2023. URL http://arxiv.org/abs/2307.12402.

Tim Groseclose and Jeffrey Milyo. A measure of media bias. 120(4):1191–1237, 2005. ISSN 00335533. doi: 10.1162/003355305775097542.

GroundNews. GroundNews browser extension, 2020. URL `https://ground.news/extension`.

Aditya Grover and Jure Leskovec. Node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864. ACM, 2016. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939754. URL `https://dl.acm.org/doi/10.1145/2939672.2939754`.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. A Survey on Automated Fact-Checking. 10:178–206, 2022. ISSN 2307-387X. doi: 10.1162/tacl_a_00454. URL `https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00454/109469/A-Survey-on-Automated-Fact-Checking`.

Jutta Haider and Olof Sundin. PARADOXES OF MEDIA AND INFORMATION LITERACY; The Crisis of Information. 2022.

Kevin A Hallgren. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23, 2012.

Daniel C. Hallin. The Passing of the â€œHigh Modernismâ€ of American Journalism. 42(3): 14–25, 1992. ISSN 0021-9916, 1460-2466. doi: 10.1111/j.1460-2466.1992.tb00794.x. URL `https://academic.oup.com/joc/article/42/3/14-25/4210083`.

Felix Hamborg. Media bias, the social sciences, and NLP: Automating frame analyses to identify bias by word choice and labeling. In Shruti Rijhwani, Jiangming Liu, Yizhong Wang, and Rotem Dror, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 79–87, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-srw.12. URL `https://aclanthology.org/2020.acl-srw.12`.

Felix Hamborg. *Person-Oriented Framing Analysis*, pages 55–70. Springer Nature Switzerland, Cham, 2023a. ISBN 978-3-031-17693-7. doi: 10.1007/978-3-031-17693-7_3. URL `https://doi.org/10.1007/978-3-031-17693-7_3`.

Felix Hamborg. *Revealing Media Bias in News Articles: NLP Techniques for Automated Frame Analysis*. Springer Nature Switzerland, 2023b. ISBN 978-3-031-17692-0 978-3-031-17693-7. doi: 10.1007/978-3-031-17693-7. URL `https://link.springer.com/10.1007/978-3-031-17693-7`.

Felix Hamborg and Karsten Donnay. NewsMTSC: A dataset for (multi-)target-dependent sentiment classification in political news articles. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, 2021.

Felix Hamborg, Corinna Breitinger, and Bela Gipp. Giveme5w1h: A universal system for extracting main events from news articles. *arXiv preprint arXiv:1909.02766*, 2019a.

Felix Hamborg, Karsten Donnay, and Bela Gipp. Automated identification of media bias in news articles: An interdisciplinary literature review. 20(4):391–415, 2019b. ISSN 1432-5012, 1432-1300. doi: 10.1007/s00799-018-0261-y. URL http://link.springer.com/10.1007/s00799-018-0261-y.

Tony Harcup and Deirdre O'Neill. What is News?: News values revisited (again). 18(12): 1470–1488, 2017. ISSN 1461-670X, 1469-9699. doi: 10.1080/1461670X.2016.1150193. URL https://www.tandfonline.com/doi/full/10.1080/1461670X.2016.1150193.

Ian Hargreaves. *Journalism: A Very Short Introduction*. Oxford University Press, 2014.

Tilo Hartmann and Martin Tanis. Examining the Hostile Media Effect as an Intergroup Phenomenon: The Role of Ingroup Identification and Status. 63(3):535–555, 2013. ISSN 1460-2466. doi: 10.1111/jcom.12031. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jcom.12031.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing, March 2023.

Lucien Heitz, Juliane A. Lischka, Alena Birrer, Bibek Paudel, Suzanne Tolmeijer, Laura Laugwitz, and Abraham Bernstein. Benefits of Diverse News Recommendations for Democracy: A User Study. *Digital Journalism*, 10(10):1710–1730, November 2022. ISSN 2167-0811, 2167-082X. doi: 10.1080/21670811.2021.2021804.

Natali Helberger, Kari Karppinen, and Lucia Dâ€™Acunto. Exposure diversity as a design principle for recommender systems. 21(2):191–207, 2018. ISSN 1369-118X, 1468-4462. doi: 10.1080/1369118X.2016.1271900. URL https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1271900.

Fatemeh Hemmatian and Mohammad Karim Sohrabi. A survey on classification techniques for opinion mining and sentiment analysis. 52(3):1495–1545, 2019. ISSN 0269-2821, 1573-7462. doi: 10.1007/s10462-017-9599-6. URL http://link.springer.com/10.1007/s10462-017-9599-6.

Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

Benjamin Horne, Sara Khedr, and Sibel Adali. Sampling the News Producers: A Large News and Feature Data Set for the Study of the Complex Media Landscape. 12(1), 2018. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v12i1.14982. URL https://ojs.aaai.org/index.php/ICWSM/article/view/14982.

C. Hutto and Eric Gilbert. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. 8(1):216–225, 2014. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v8i1.14550. URL https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

Shanto Iyengar and Adam Simon. News coverage of the gulf crisis and public opinion: A study of agenda-setting, priming, and framing. *Communication Research*, 20(3):365–383, 1993. doi: 10.1177/009365093020003002.

Hitkul Jangid, Shivangi Singhal, Rajiv Ratn Shah, and Roger Zimmermann. Aspect-based financial sentiment analysis using deep learning. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, pages 1961–1966. ACM Press, 2018.

Israa Jaradat, Haiqi Zhang, and Chengkai Li. On Detecting Cherry-picking in News Coverage Using Large Language Models, January 2024.

Soufian Jebbara and Philipp Cimiano. Zero-shot cross-lingual opinion target extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2486–2495. Association for Computational Linguistics, June 2019.

Christopher Jenks. Mock News: On the discourse of mocking in U.S. televised political discussions. 16(1):58–75, 2022. ISSN 1750-4813. doi: 10.1177/17504813211043719. URL `https://doi.org/10.1177/17504813211043719`.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 151–160. Association for Computational Linguistics, 2011.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. A challenge dataset and effective models for aspect-based sentiment analysis. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1654. URL `https://aclanthology.org/D19-1654`.

Salud M. Jiménez Zafra, Giacomo Berardi, Andrea Esuli, Diego Marcheggiani, María Teresa Martín-Valdivia, and Alejandro Moreo Fernández. A multi-lingual annotated dataset for aspect-oriented opinion mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2533–2538. Association for Computational Linguistics, September 2015.

David R Jones. Party Polarization and Legislative Gridlock. 2001.

Jeffrey P Jones. The 'new'news as no 'news': Us cable news channels as branded political entertainment television. *Media International Australia*, 144(1):146–155, 2012.

S. Mo Jones-Jang, Tara Mortensen, and Jingjing Liu. Does media literacy help identification of fake news? information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2):371–388, 2021. doi: 10.1177/0002764219869406. URL `https://doi.org/10.1177/0002764219869406`.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, volume 2, pages 427–431. Association for Computational Linguistics, March 2017. ISBN 978-1-5108-3860-4. doi: 10.18653/v1/e17-2068.

Daniel Kahneman and Amos Tversky. Choices, values, and frames. *American psychologist*, 39(4): 341, 1984.

Sadia Kamal, Jimmy Hartford, Jeremy Willis, and Arunkumar Bagavathi. Learning unbiased news article representations: A knowledge-infused approach. *arXiv preprint arXiv:2309.05981*, 2023.

Amina Kammoun, Rim Slama, Hedi Tabia, Tarek Ouni, and Mohmed Abid. Generative Adversarial Networks for face generation: A survey. page 1122445.1122456, 2022. ISSN 0360-0300, 1557-7341. doi: 10.1145/1122445.1122456. URL https://dl.acm.org/doi/10.1145/1122445.1122456.

Hamid Karimi and Jiliang Tang. Learning Hierarchical Discourse-level Structure for Fake News Detection. In *Proceedings of the 2019 Conference of the North*, pages 3432–3442. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1347. URL http://aclweb.org/anthology/N19-1347.

Nora Kassner, Fabio Petroni, Mikhail Plekhanov, Sebastian Riedel, and Nicola Cancedda. Edin: an end-to-end benchmark and pipeline for unknown entity discovery and indexing. *arXiv preprint arXiv:2205.12570*, 2022.

Rosalee A. Clawson Katsuo A. Nishikawa, Terri L. Towner and Eric N. Waltenburg. Interviewing the interviewers: Journalistic norms and racial diversity in the newsroom. *Howard Journal of Communications*, 20(3):242–259, 2009. doi: 10.1080/10646170903070175. URL https://doi.org/10.1080/10646170903070175.

Elihu Katz. And Deliver Us from Segmentation. 546(1):22–33, 1996. ISSN 0002-7162, 1552-3349. doi: 10.1177/0002716296546001003. URL http://journals.sagepub.com/doi/10.1177/0002716296546001003.

Brian Felipe Keith Norambuena, Tanushree Mitra, and Chris North. A Survey on Event-Based News Narrative Extraction. 55:1–39, 2023. ISSN 0360-0300, 1557-7341. doi: 10.1145/3584741. URL https://dl.acm.org/doi/10.1145/3584741.

Khalid Al Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. A News Editorial Corpus for Mining Argumentation Strategies.

Pavel Khloponin and Leila Kosseim. Using Document Embeddings for Background Linking of News Articles. 12801:317–329, 2021. doi: 10.1007/978-3-030-80599-9_28. URL https://link.springer.com/10.1007/978-3-030-80599-9_28.

Eunji Kim, Yphtach Lelkes, and Joshua McCrain. Measuring dynamic media bias. *Proceedings of the National Academy of Sciences of the United States of America*, 2022. doi: 10.1073/PNAS.2202197119.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Preslav Nakov and Torsten Zesch, editors, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland, August 2014. Association for Computational Linguistics. doi: 10.3115/v1/S14-2076. URL https://aclanthology.org/S14-2076.

Joshua Klayman. Varieties of confirmation bias. volume 32 of *Psychology of Learning and Motivation*, pages 385–418. Academic Press, 1995. doi: 10.1016/S0079-7421(08)60315-1. URL https://www.sciencedirect.com/science/article/pii/S0079742108603151.

Yunyong Ko, Seongeun Ryu, Soeun Han, Yeongseung Jeon, Jaehoon Kim, Sohyun Park, Kyungsik Han, Hanghang Tong, and Sang-Wook Kim. Khan: Knowledge-aware hierarchical attention networks for political stance prediction. *arXiv preprint arXiv:2302.12126*, 2023.

Andrew Kohut, Richard Morin, and Scott Keeter. PUBLIC KNOWLEDGE OF CURRENT AFFAIRS LITTLE CHANGED BY NEWS AND INFORMATION REVOLUTIONS. 2007.

Ammina Kothari. THE FRAMING OF THE DARFUR CONFLICT IN THE *NEW YORK TIMES* : 2003–2006. *Journalism Studies*, 11(2):209–224, April 2010. ISSN 1461-670X, 1469-9699. doi: 10.1080/14616700903481978.

Artem Kramov and Sergiy Pogorilyy. Cross-Lingual Named Entity Recognition for the Ukrainian Language Based on Word Alignment. In *2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT)*, pages 213–218. IEEE, 2022. ISBN 9798350332629. doi: 10.1109/ATIT58178.2022.10024219. URL `https://ieeexplore.ieee.org/document/10024219/`.

Anne C. Kroon, Damian Trilling, and Tamara Raats. Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage. *Journalism & Mass Communication Quarterly*, 98(2):451–477, June 2021. ISSN 1077-6990. doi: 10.1177/1077699020932304.

Emily Kubin and Christian Von Sikorski. The role of (social) media in political polarization: A systematic review. 45(3):188–206, 2021. ISSN 2380-8985, 2380-8977. doi: 10.1080/23808985.2021.1976070. URL `https://www.tandfonline.com/doi/full/10.1080/23808985.2021.1976070`.

Dilek Küçük and Fazli Can. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53 (1):1–37, 2020.

Vivek Kulkarni, Junting Ye, Steve Skiena, and William Yang Wang. Multi-view Models for Political Ideology Detection of News Articles. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3518–3527. Association for Computational Linguistics, 2018. doi: 10.18653/v1/D18-1388. URL `http://aclweb.org/anthology/D18-1388`.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. A Systematic Media Frame Analysis of 1.5 Million New York Times Articles from 2000 to 2017, 2020. URL `http://arxiv.org/abs/2005.01803`.

Haewoon Kwak, Jisun An, Elise Jing, and Yong-Yeol Ahn. FrameAxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7, 2021.

Ksenia Lagutina, Nadezhda Lagutina, Elena Boychuk, Inna Vorontsova, Elena Shliakhtina, Olga Belyaeva, Ilya Paramonov, and P.G. Demidov. A survey on stylometric text features. In *2019 25th Conference of Open Innovations Association (FRUCT)*, pages 184–195, 2019. doi: 10.23919/FRUCT48121.2019.8981504.

Jasmin Lampert and Christoph H Lampert. Overcoming rare-language discrimination in multilingual sentiment analysis. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5185–5192. IEEE, 2021.

Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.

Dominic L. Lasorsa, Seth C. Lewis, and Avery E. Holton. NORMALIZING TWITTER: Journalism practice in an emerging communication space. 13(1):19–36, 2012. ISSN 1461-670X, 1469-9699. doi: 10.1080/1461670X.2011.571825. URL `http://www.tandfonline.com/doi/abs/10.1080/1461670X.2011.571825`.

John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, 45(4): 765–818, 2020.

Konstantina Lazaridou, Alexander Löser, Maria Mestre, and Felix Naumann. Discovering Biased News Articles Leveraging Multiple Human Annotations. 2020.

Susan Leavy. Uncovering Gender Bias in Media Coverage of Politicians with Machine Learning, 2020. URL http://arxiv.org/abs/2005.07734.

Frances E. Lee. How Party Polarization Affects Governance. 18(1):261–282, 2015. ISSN 1094-2939, 1545-1577. doi: 10.1146/annurev-polisci-072012-113747. URL https://www.annualreviews.org/doi/10.1146/annurev-polisci-072012-113747.

Yen-Hsien Lee, Chih-Ping Wei, Paul Jen-Hwa Hu, Pao-Feng Wu, and How Jiang. A text summary-based method to detect new events from streams of online news articles. 59(6):103684, 2022. ISSN 03787206. doi: 10.1016/j.im.2022.103684. URL https://linkinghub.elsevier.com/retrieve/pii/S0378720622000933.

Théophile Lenoir. Media polarization "à la française" ? comparing the french and american ecosystems. Technical report, Institut Montaigne, 2019.

Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of Massive Datasets*. Cambridge University Press, 3 edition, 2020.

Justin Lewis, Stephen Cushion, and James Thomas. Immediacy, Convenience or Engagement? An analysis of 24-hour news channels in the UK. *Journalism Studies*, 6(4):461–477, November 2005. ISSN 1461-670X, 1469-9699. doi: 10.1080/14616700500250362.

Chang Li and Dan Goldwasser. Encoding Social Information with Graph Convolutional Networks forPolitical Perspective Detection in News Media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2594–2604. Association for Computational Linguistics, 2019. doi: 10.18653/v1/P19-1247. URL https://www.aclweb.org/anthology/P19-1247.

Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. A Survey on Deep Learning for Named Entity Recognition. 34(1):50–70, 2022. ISSN 1041-4347, 1558-2191, 2326-3865. doi: 10.1109/TKDE.2020.2981314. URL https://ieeexplore.ieee.org/document/9039685/.

Herbert Lin. The existential threat from cyber-enabled information warfare. *Bulletin of the Atomic Scientists*, 75(4):187–196, July 2019. ISSN 0096-3402, 1938-3282. doi: 10.1080/00963402.2019.1629574.

Elvys Linhares Pontes, Luis AdriÃ¡n Cabrera-Diego, Jose G. Moreno, Emanuela Boros, Ahmed Hamdi, Nicolas SidÃ¨re, MickaÃ«l Coustaty, and Antoine Doucet. Entity Linking for Historical Documents: Challenges and Solutions. In Emi Ishita, Natalie Lee San Pang, and Lihong Zhou, editors, *Digital Libraries at Times of Massive Societal Transition*, volume 12504, pages 215–231. Springer International Publishing, 2020. ISBN 978-3-030-64451-2 978-3-030-64452-9. doi: 10.1007/978-3-030-64452-9_19. URL https://link.springer.com/10.1007/978-3-030-64452-9_19.

Bing Liu. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, pages 627–666, 2010.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun

violence. In Mohit Bansal and Aline Villavicencio, editors, *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China, November 2019a. Association for Computational Linguistics. doi: 10.18653/v1/K19-1047. URL https://aclanthology.org/K19-1047.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019b.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nicholas Beauchamp, and Lu Wang. POLITICS: Pretraining with Same-story Article Comparison for Ideology Prediction and Stance Detection. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1354–1374, Seattle, United States, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-naacl.101.

Zihan Liu, Yan Xu, Tiezheng Yu, Wenliang Dai, Ziwei Ji, Samuel Cahyawijaya, Andrea Madotto, and Pascale Fung. Crossner: Evaluating cross-domain named entity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13452–13460, 2021.

Levon Lloyd, Dimitrios Kechagias, and Steven Skiena. Lydia: A system for large-scale news analysis. In *International Symposium on String Processing and Information Retrieval*, pages 161–166. Springer, 2005.

Will Lowe. Software for content analysis-A review. *Cambridge: Weatherhead Center for International Affairs and the Harvard Identity Project*, 2002.

Katharina Ludwig, Alexander Grote, Andreea Iana, Mehwish Alam, Heiko Paulheim, Harald Sack, Christof Weinhardt, and Philipp Müller. Divided by the algorithm? the (limited) effects of content-and sentiment-based news recommendation on affective, ideological, and perceived polarization. *Social Science Computer Review*, page 08944393221149290, 2023.

Yun Luo, Fang Guo, Zihan Liu, and Yue Zhang. Mere Contrastive Learning for Cross-Domain Sentiment Analysis, August 2022.

Robert Luzsa and Susanne Mayr. False consensus in the echo chamber: Exposure to favorably biased social media news feeds leads to increased perception of public support for own opinions. 15(1), 2021. ISSN 1802-7962. doi: 10.5817/CP2021-1-3. URL https://cyberpsychology.eu/article/view/12254.

Qing Lyu, Hongming Zhang, Elior Sulem, and Dan Roth. Zero-shot event extraction via transfer learning: Challenges and insights. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 322–332, 2021.

Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. Multiconer: A large-scale multilingual dataset for complex named entity recognition. *arXiv preprint arXiv:2208.14536*, 2022.

Paolo Mancini. Media Fragmentation, Party System, and Democracy. 18(1):43–60, 2013. ISSN 1940-1612, 1940-1620. doi: 10.1177/1940161212458200. URL http://journals.sagepub.com/doi/10.1177/1940161212458200.

Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E Schwamb, Chris J Lintott, and Arfon M Smith. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI conference on human computation and crowdsourcing*, 2013.

James L. Martherus, Andres G. Martinez, Paul K. Piff, and Alexander G. Theodoridis. Party Animals? Extreme Partisan Polarization and Dehumanization. 43(2):517–540, 2021. ISSN 0190-9320, 1573-6687. doi: 10.1007/s11109-019-09559-4. URL `https://link.springer.com/10.1007/s11109-019-09559-4`.

Gregory J. Martin and Ali Yurukoglu. Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99, September 2017. doi: 10.1257/aer.20160812. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20160812`.

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah, and Benoît Sagot. Camembert: a tasty french language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.

Stephen Mayhew, Terra Blevins, Shuheng Liu, Marek Šuppa, Hila Gonen, Joseph Marvin Imperial, Börje F. Karlsson, Peiqin Lin, Nikola Ljubešić, L. J. Miranda, Barbara Plank, Arij Riabi, and Yuval Pinter. Universal NER: A Gold-Standard Multilingual Named Entity Recognition Benchmark, November 2023.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

George A. Miller. WordNet: a lexical database for english. volume 38, pages 39–41, 1995.

Adam J. Mills. Virality in social media: The SPIN Framework. 12(2):162–169, 2012. ISSN 1472-3891, 1479-1854. doi: 10.1002/pa.1418. URL `https://onlinelibrary.wiley.com/doi/10.1002/pa.1418`.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41. Association for Computational Linguistics, 2016a. doi: 10.18653/v1/S16-1003. URL `http://aclweb.org/anthology/S16-1003`.

Saif M. Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, volume 1, pages 174–184. Association for Computational Linguistics, April 2018. ISBN 978-1-948087-32-2. doi: 10.18653/v1/p18-1017.

Saif M Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. How translation alters sentiment. *Journal of Artificial Intelligence Research*, 55:95–130, 2016b.

Negar Mokhberian, Andrés Abeliuk, Patrick Cummings, and Kristina Lerman. Moral Framing and Ideological Bias of News. volume 12467, pages 206–219. 2020. doi: 10.1007/978-3-030-60975-7_16.

Pedro Ramaciotti Morales, Jean-philippe Cointet, Gabriel Zolotoochin, Pedro Ramaciotti Morales, Jean-philippe Cointet, Gabriel Zolotoochin, and Pedro Ramaciotti Morales. Unfolding the dimensionality structure of social networks in ideological embeddings To cite this version : HAL Id : Hal-03315759 Unfolding the dimensionality structure of social networks in ideological embeddings. 2021.

Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow. On the stability of fine-tuning bert: Misconceptions, explanations, and strong baselines. In *International Conference on Learning Representations*, 2021.

Karen Mossberger, Caroline J. Tolbert, and Ramona S. McNeal. *Digital Citizenship: The Internet, Society, and Participation*. The MIT Press, 10 2007. ISBN 9780262280280. doi: 10.7551/mitpress/7428.001.0001. URL https://doi.org/10.7551/mitpress/7428.001.0001.

M. F. Mridha, Ashfia Jannat Keya, Md. Abdul Hamid, Muhammad Mostafa Monowar, and Md. Saifur Rahman. A Comprehensive Review on Fake News Detection With Deep Learning. *IEEE Access*, 9:156151–156170, 2021. ISSN 2169-3536. doi: 10.1109/ACCESS.2021.3129329.

Sendhil Mullainathan and Andrei Shleifer. The market for news. *The American Economic Review*, 95(4):1031–1053, 2005. ISSN 00028282. URL http://www.jstor.org/stable/4132704.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. SemEval-2016 task 4: Sentiment analysis in Twitter. In Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, San Diego, California, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/S16-1001. URL https://aclanthology.org/S16-1001.

Abdulqadir J Nashwan and Jaber H Jaradat. Streamlining Systematic Reviews: Harnessing Large Language Models for Quality Assessment and Risk-of-Bias Evaluation. 2023. ISSN 2168-8184. doi: 10.7759/cureus.43023. URL https://www.cureus.com/articles/178248-streamlining-systematic-reviews-harnessing-large-language-models-for-quality-assessmen

Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Transactions on Affective Computing*, 2020.

Nic Newman, Richard Fletcher, Kirsten Eddy, Craig T Robertson, and Rasmus Kleis Nielsen. Reuters Institute Digital News Report 2023. 2023.

NewsGuard. Newsguard, 2018. URL https://www.newsguardtech.com/.

Chi Mai Nguyen, Phat Trien Thai, Van Tuan Nguyen, and Duy Khang Lam. A Real-Time Open Public Sources Text Analysis System. 13(12), 2022. ISSN 21565570, 2158107X. doi: 10.14569/IJACSA.2022.01312105. URL http://thesai.org/Publications/ViewPaper?Volume=13&Issue=12&Code=IJACSA&SerialNo=105.

Ngoc Thanh Nguyen, editor. *Transactions on Computational Collective Intelligence V*, volume 6910 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-24015-7 978-3-642-24016-4. doi: 10.1007/978-3-642-24016-4. URL http://link.springer.com/10.1007/978-3-642-24016-4.

Timothy Niven and Hung-Yu Kao. Measuring Alignment to Authoritarian State Media as Framing Bias. In Giovanni Da San Martino, Chris Brew, Giovanni Luca Ciampaglia, Anna Feldman, Chris Leberknight, and Preslav Nakov, editors, *Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 11–21. International Committee on Computational Linguistics (ICCL), 2020. URL https://aclanthology.org/2020.nlp4if-1.2.

Yikang Pan, Liangming Pan, Wenhu Chen, Preslav Nakov, Min-Yen Kan, and William Yang Wang. On the Risk of Misinformation Pollution with Large Language Models, 2023. URL http://arxiv.org/abs/2305.13661.

Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel, and Preslav Nakov. GREENER: Graph Neural Networks for News Media Profiling. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7470–7480, Abu Dhabi, United Arab Emirates, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.506.

Eli Pariser. *The filter bubble: what the Internet is hiding from you.* Penguin Press, 2011.

Sora Park and Jee Young Lee. Incidental News Exposure on Facebook and Its Relation to Trust in News. 9(1):205630512311588, 2023. ISSN 2056-3051, 2056-3051. doi: 10.1177/20563051231158823. URL http://journals.sagepub.com/doi/10.1177/20563051231158823.

Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. NewsCube: Delivering multiple aspects of news to mitigate media bias. pages 443–452, 2009. doi: 10.1145/1518701.1518772.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. From freebase to wikidata: The great migration. In *Proceedings of the 25th international conference on world wide web*, pages 1419–1428, 2016.

Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162. URL https://aclanthology.org/D14-1162.

Billy Perrigo. Yann lecun on how an open source approach could shape ai, Feb 2024. URL https://time.com/6691705/time100-impact-awards-yann-lecun/.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1202. URL https://aclanthology.org/N18-1202.

Angela Phillips. *Journalism in context: Practice and theory for the digital age*. Routledge, 2014.

Lawrence Pintak and Yosri Fouda. Blogging in the middle east: Not necessarily journalistic. *Columbia Journalism Review*, 2009.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, Ritesh Kumar, and Elisa Sartori, editors, *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.semeval-1.317. URL https://aclanthology.org/2023.semeval-1.317.

Mikhail Plekhanov, Nora Kassner, Kashyap Popat, Louis Martin, Simone Merello, Borislav Kozlovskii, Frédéric A. Dreyer, and Nicola Cancedda. Multilingual End to End Entity Linking, June 2023.

Danuta Ploch, Andreas Lommatzsch, and Florian Schultze. An Advanced Press Review System Combining Deep News Analysis and Machine Learning Algorithms. In *Proceedings of ACL-2016 System Demonstrations*, pages 109–114. Association for Computational Linguistics. doi: 10.18653/v1/P16-4019. URL http://aclweb.org/anthology/P16-4019.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35. Association for Computational Linguistics, 2014.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495. Association for Computational Linguistics, 2015.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June 2016a. Association for Computational Linguistics. doi: 10.18653/v1/S16-1002. URL https://aclanthology.org/S16-1002.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics, 2016b.

W. James Potter. Review of Literature on Media Literacy. 7(6):417–435, 2013. ISSN 1751-9020, 1751-9020. doi: 10.1111/soc4.12041. URL https://compass.onlinelibrary.wiley.com/doi/10.1111/soc4.12041.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*, 2017.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. Towards robust linguistic analysis using ontonotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, 2013.

Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang. Automatically Neutralizing Subjective Bias in Text. 34(01):480–489, 2020. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v34i01.5385. URL `https://ojs.aaai.org/index.php/AAAI/article/view/5385`.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, 2011.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

Prashanth Rao and Maite Taboada. Gender Bias in the News: A Scalable Topic Modelling and Visualization Framework. *Frontiers in Artificial Intelligence*, 4:664737, June 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.664737.

Hannah Rashkin, Sameer Singh, and Yejin Choi. Connotation Frames: A Data-Driven Investigation, 2016. URL `http://arxiv.org/abs/1506.02739`.

Jennifer Rauch, Sunitha Chitrapu, Susan Tyler Eastman, John Christopher Evans, Christopher Paine, and Peter Mwesige. From Seattle 1999 to New York 2004: A Longitudinal Analysis of Journalistic Framing of the Movement for Democratic Globalization. *Social Movement Studies*, 6(2):131–145, September 2007. ISSN 1474-2837, 1474-2829. doi: 10.1080/14742830701497244.

Kumar Ravi and Vadlamani Ravi. A novel automatic satire and irony detection using ensembled feature selection and data mining. 120:15–33, 2017. ISSN 09507051. doi: 10.1016/j.knosys.2016.12.018. URL `https://linkinghub.elsevier.com/retrieve/pii/S0950705116305226`.

Shaina Raza and Chen Ding. A regularized model to trade-off between accuracy and diversity in a news recommender system. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 551–560. IEEE, 2020.

Shaina Raza and Chen Ding. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 55(1):749–800, 2021.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659. Association for Computational Linguistics, 2013.

Ange Richard, Gilles Bastin, and François Portet. Genderednews: Une approche computationnelle des écarts de représentation des genres dans la presse française, 2022.

Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, (May):4933–4941, 2020.

Rubén Rivas-de Roca, Francisco J. Caro-González, and Mar García-Gordillo. Covering the eu at local level: A multiple-case study in germany, the uk and spain. *International Communication Gazette*, page 17480485221134177, 2022. doi: 10.1177/17480485221134177. URL https://doi.org/10.1177/17480485221134177.

Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. Structural topic models for open-ended survey responses. *American journal of political science*, 58(4):1064–1082, 2014.

Stephen Robertson, Hugo Zaragoza, et al. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389, 2009.

Federica Rollo, Laura Po, and Giovanni Bonisoli. Online News Event Extraction for Crime Analysis.

Lee Ross, David Greene, and Pamela House. The â€œfalse consensus effectâ€: An egocentric bias in social perception and attribution processes. 13(3):279–301, 1977. ISSN 00221031. doi: 10.1016/0022-1031(77)90049-X. URL https://linkinghub.elsevier.com/retrieve/pii/002210317790049X.

Teemu Ruokolainen, Pekka Kauppinen, Miikka Silfverberg, and Krister Lindén. A finnish news corpus for named entity recognition. *Language Resources and Evaluation*, pages 1–26, 2019.

Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. Social media news communities: Gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management - CIKM '13*, pages 1679–1684. ACM Press, 2013. ISBN 978-1-4503-2263-8. doi: 10.1145/2505515.2505623. URL http://dl.acm.org/citation.cfm?doid=2505515.2505623.

Allan Sales, Albin Zehe, Leandro Balby Marinho, Adriano Veloso, Andreas Hotho, and Janna Omeliyanenko. Assessing Media Bias in Cross-Linguistic and Cross-National Populations. *Proceedings of the International AAAI Conference on Web and Social Media*, 15:561–572, May 2021. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v15i1.18084.

Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, and Yejin Choi. Connotation Frames of Power and Agency in Modern Films. 2017.

Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Arjun Sawant, Swaroop Mishra, and Chitta Baral. InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis, May 2023.

J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, *The Adaptive Web: Methods and Strategies of Web Personalization*, pages 291–324. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-72079-9. doi: 10.1007/978-3-540-72079-9_9. URL https://doi.org/10.1007/978-3-540-72079-9_9.

Timo Schick and Hinrich Schütze. Exploiting cloze questions for few shot text classification and natural language inference. *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pages 255–269, 2021. doi: 10.18653/v1/2021.eacl-main.20.

Tal Schuster, Roei Schuster, Darsh J Shah, and Regina Barzilay. The limitations of stylometry for detecting machine-generated fake news. *Computational Linguistics*, 46(2):499–510, 2020.

Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. Open aspect target sentiment classification with natural language prompts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6311–6322. Association for Computational Linguistics, November 2021.

Robiert Sepúlveda-Torres, Marta Vicente, Estela Saquete, Elena Lloret, and Manuel Palomar. Leveraging relevant summarized information and multi-layer classification to generalize the detection of misleading headlines. *Data & Knowledge Engineering*, 145:102176, 2023. ISSN 0169-023X. doi: https://doi.org/10.1016/j.datak.2023.102176. URL `https://www.sciencedirect.com/science/article/pii/S0169023X23000368`.

Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, and Katja Filippova. Multi-lingual opinion mining on YouTube. *Information Processing & Management*, 52(1):46–60, January 2016. ISSN 03064573. doi: 10.1016/j.ipm.2015.03.002.

Ã–zge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko, and Chris Biemann. Neural entity linking: A survey of models based on deep learning. 13(3):527–570, 2022. ISSN 22104968, 15700844. doi: 10.3233/SW-222986. URL `https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-222986`.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake News Detection on Social Media. 19(1), 2017.

Chenglei Si, Navita Goyal, Sherry Tongshuang Wu, Chen Zhao, Shi Feng, Hal DaumÃ© III, and Jordan Boyd-Graber. Large Language Models Help Humans Verify Truthfulness – Except When They Are Convincingly Wrong, 2023. URL `http://arxiv.org/abs/2310.12558`.

Ankur Sinha, Satishwar Kedas, Rishu Kumar, and Pekka Malo. SEntFiN 1.0: EntityâEaware sentiment analysis for financial news. 73(9):1314–1335, 2022. ISSN 2330-1635, 2330-1643. doi: 10.1002/asi.24634. URL `https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24634`.

James M Snyder Jr and David Strömberg. Press coverage and political accountability. *Journal of political Economy*, 118(2):355–408, 2010.

Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*, 2019.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*, 2020.

Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum Learning: A Survey, April 2022.

Mark Spilsbury. Diversity in journalism. *National Council for the Training of Journalist UK: Essex, UK*, pages 4–10, 2017.

Timo Spinde, Felix Hamborg, and Bela Gipp. An integrated approach to detect media bias in german news articles. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, JCDL '20, pages 505–506, Virtual Event, China, 2020. Association for Computing Machinery. ISBN 9781450375856. doi: 10.1145/3383583.3398585. URL https://doi.org/10.1145/3383583.3398585.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. Neural media bias detection using distant supervision with BABE - bias annotations by experts. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-emnlp.101. URL https://aclanthology.org/2021.findings-emnlp.101.

Felix Stahlberg. Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69:343–418, 2020.

Sanja Štajner, Richard Evans, Constantin Orasan, and Ruslan Mitkov. What can readability measures really tell us about text complexity. In *Proceedings of workshop on natural language processing for improving textual accessibility*, pages 14–22. Citeseer, 2012.

Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. Predicting the topical stance and political leaning of media using tweets. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 527–537, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.50. URL https://aclanthology.org/2020.acl-main.50.

Ralf Steinberger, Stefanie Hegele, Hristo Tanev, and Leonida Della Rocca. Large-scale news entity sentiment analysis. In Ruslan Mitkov and Galia Angelova, editors, *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 707–715, Varna, Bulgaria, September 2017. INCOMA Ltd. doi: 10.26615/978-954-452-049-6_091. URL https://doi.org/10.26615/978-954-452-049-6_091.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. Aspect-Level Sentiment Analysis Via Convolution over Dependency Tree. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5678–5687, Hong Kong, China, 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1569.

Cass R Sunstein. The Law of Group Polarization. 1999.

Cass R Sunstein. *Republic. com.* Princeton university press, 2001.

Anjana Susarla, Jeong-Ha Oh, and Yong Tan. Influentials, Imitables, or Susceptibles? Virality and Word-of-Mouth Conversations in Online Social Networks. 33(1):139–170, 2016. ISSN 0742-1222, 1557-928X. doi: 10.1080/07421222.2016.1172454. URL http://www.tandfonline.com/doi/full/10.1080/07421222.2016.1172454.

György Szarvas, Richárd Farkas, László Felföldi, András Kocsor, and János Csirik. A highly accurate named entity corpus for hungarian. *annotation*, 2:3–1, 2006.

Yla R. Tausczik and James W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. 29(1):24–54, 2010. ISSN 0261-927X, 1552-6526. doi: 10.1177/0261927X09351676. URL http://journals.sagepub.com/doi/10.1177/0261927X09351676.

NLLB Team, Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, Jeff Wang, and Meta Ai. No Language Left Behind: Scaling Human-Centered Machine Translation. 2022.

Denis Teyssou, Jean-Michel Leung, Evlampios Apostolidis, Konstantinos Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Olga Papadopoulou, and Vasileios Mezaris. The invid plug-in: Web video verification on the browser. In *Proceedings of the First International Workshop on Multimedia Verification*, MuVer '17, page 23–30, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450355100. doi: 10.1145/3132384.3132387. URL https://doi.org/10.1145/3132384.3132387.

Gunnar Thesen and T. M. Yildirim. Electoral systems and gender inequality in political news: Analyzing the news visibility of members of parliament in norway and the uk. *American Political Science Review*, 2022. doi: 10.1017/S0003055422000776.

Nava Tintarev, Emily Sullivan, Dror Guldin, Sihang Qiu, and Daan Odjik. Same, same, but different: Algorithmic diversification of viewpoints in news. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 7–13. ACM, 2018.

Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models. In *Proceeding of the 17th international conference on World Wide Web - WWW '08*. ACM Press, 2008.

Minh Tran. How biased are American media outlets? A framework for presentation bias regression. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 4359–4364, Atlanta, GA, USA, December 2020. IEEE. ISBN 978-1-72816-251-5. doi: 10.1109/BigData50022.2020.9377987.

Celina Treuillier, Sylvain Castagnos, Evan Dufraisse, and Armelle Brun. Being diverse is not enough: Rethinking diversity evaluation to meet challenges of news recommender systems. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 222–233, 2022.

Oren Tsur, Dan Calacci, and David Lazer. A Frame of Mind: Using Statistical Models for Detection of Framing and Agenda Setting Campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1629–1638. Association for Computational Linguistics, 2015. doi: 10.3115/v1/P15-1157. URL http://aclweb.org/anthology/P15-1157.

Aboubacar Tuo, Romaric Besançon, Olivier Ferret, and Julien Tourille. Better exploiting bert for few-shot event detection. In *International Conference on Applications of Natural Language to Information Systems*, pages 291–298. Springer, 2022.

Petter Törnberg. How digital media drive affective polarization through partisan sorting. 119 (42):e2207159119, 2022. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2207159119. URL https://pnas.org/doi/10.1073/pnas.2207159119.

Robert P Vallone and Lee Ross. The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre. 1985.

Peter Van Aelst and Strömbäck. Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1):3–27, 2017.

Esther Van Den Berg and Katja Markert. Context in Informational Bias Detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6315–6326. International Committee on Computational Linguistics, 2020. doi: 10.18653/v1/2020.coling-main.556. URL https://www.aclweb.org/anthology/2020.coling-main.556.

Toni G.L.A. Van Der Meer and Michael Hameleers. Fighting biased news diets: Using news media literacy interventions to stimulate online cross-cutting media exposure patterns. *New Media & Society*, 23(11):3156–3178, November 2021. ISSN 1461-4448, 1461-7315. doi: 10.1177/1461444820946455.

Dave M. Van Zandt. Media bias/fact check news, Jan 2015. URL https://mediabiasfactcheck.com/.

Francielle Vargas, Computer Science Department, Federal University of Minas Gerais, Brazil, Kokil Jaidka, Centre for Trusted Internet and Community, National University of Singapore, Singapore, Thiago A. S. Pardo, Institute of Mathematical and Computer Sciences, University of SÃ£o Paulo, Brazil, FabrÃcio Benevenuto, and Computer Science Department, Federal University of Minas Gerais, Brazil. Predicting Sentence-Level Factuality of News and Bias of Media Outlets. In *Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings*, pages 1197–1206. INCOMA Ltd., Shoumen, BULGARIA, 2023. ISBN 978-954-452-092-2. doi: 10.26615/978-954-452-092-2_127. URL https://acl-bg.org/proceedings/2023/RANLP%202023/pdf/2023.ranlp-1.127.pdf.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Åukasz Kaiser, and Illia Polosukhin. Attention is all you need. 2017-Decem: 5999–6009, 2017. ISSN 10495258. URL https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

David Vilares, Miguel A Alonso, and Carlos Gómez-Rodríguez. Supervised sentiment analysis in multilingual environments. *Information Processing & Management*, 53(3):595–607, 2017.

Duy-Tin Vo and Yue Zhang. Target-dependent twitter sentiment classification with rich automatic features. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

Markus Wagner. Affective polarization in multiparty systems. 69:102199, 2021. ISSN 02613794. doi: 10.1016/j.electstud.2020.102199. URL https://linkinghub.elsevier.com/retrieve/pii/S0261379420300822.

Paul Waldman and James Devitt. Newspaper photographs and the 1996 presidential election: The question of bias. *Journalism & Mass Communication Quarterly*, 75(2):302–311, 1998.

Christopher Walker and Linguistic Data Consortium. ACE 2005 multilingual training corpus, 2006.

Jingjing Wang, Changlong Sun, Shoushan Li, Xiaozhong Liu, Luo Si, Min Zhang, and Guodong Zhou. Aspect sentiment classification towards question-answering with reinforced bidirectional attention network. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3548–3557, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1345. URL `https://aclanthology.org/P19-1345`.

Jun Wang and Bei Yu. News2PubMed: A Browser Extension for Linking Health News to Medical Literature. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2605–2609. ACM, 2021. ISBN 978-1-4503-8037-9. doi: 10.1145/3404835.3462788. URL `https://dl.acm.org/doi/10.1145/3404835.3462788`.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. Relational Graph Attention Network for Aspect-based Sentiment Analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.295.

Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, A. Arunkumar, Arjun Ashok, Arut Selvan Dhanasekaran, Atharva Naik, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Gary Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Maitreya Patel, Kuntal Kumar Pal, M. Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Shailaja Keyur Sampat, Savan Doshi, Siddharth Deepak Mishra, Sujan C. Reddy, Sumanta Patro, Tanay Dixit, Xu dong Shen, Chitta Baral, Yejin Choi, Hannaneh Hajishirzi, Noah A. Smith, and Daniel Khashabi. Benchmarking generalization via in-context instructions on 1,600+ language tasks. 2022.

Zengzhi Wang, Qiming Xie, Zixiang Ding, Yi Feng, and Rui Xia. Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study, April 2023.

Charles B. Ward, Mikhail Bautin, and Steven Skiena. Identifying Differences in News Coverage between Cultural/Ethnic Groups. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pages 511–514, Milan, Italy, 2009. IEEE. ISBN 978-0-7695-3801-3. doi: 10.1109/WI-IAT.2009.337.

Albert Webson, Zhizhong Chen, Carsten Eickhoff, and Ellie Pavlick. Are "Undocumented Workers" the Same as "Illegal Aliens"? Disentangling Denotation and Connotation in Vector Spaces, 2020. URL `http://arxiv.org/abs/2010.02976`.

James G. Webster and Thomas B. Ksiazek. The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media. 62(1):39–56, 2012. ISSN 00219916. doi: 10.1111/j.1460-2466.2011.01616.x. URL `https://academic.oup.com/joc/article/62/1/39-56/4085792`.

Melvin Wevers. Using word embeddings to examine gender bias in Dutch newspapers, 1950-1990. In Nina Tahmasebi, Lars Borin, Adam Jatowt, and Yang Xu, editors, *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 92–97, Florence, Italy, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4712. URL `https://aclanthology.org/W19-4712`.

David Manning White. The "gate keeper": A case study in the selection of news. *Journalism quarterly*, 27(4):383–390, 1950.

Derek Wilding. The Impact of Digital Platforms on News and Journalistic Content. 2018.

Tamara Witschge and Gunnar Nygren. Journalism: A Profession Under Pressure? 2009.

Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. Quantifying Political Leaning from Tweets and Retweets. 7(1):640–649, 2013. doi: 10.1609/icwsm.v7i1.14422. URL `https://ojs.aaai.org/index.php/ICWSM/article/view/14422`.

Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. SentiRec: Sentiment diversity-aware neural news recommendation. In Kam-Fai Wong, Kevin Knight, and Hua Wu, editors, *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 44–53, Suzhou, China, December 2020a. Association for Computational Linguistics. URL `https://aclanthology.org/2020.aacl-main.6`.

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. MIND: A large-scale dataset for news recommendation. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3597–3606, Online, July 2020b. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.331. URL `https://aclanthology.org/2020.acl-main.331`.

Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. Scalable zero-shot entity linking with dense entity retrieval. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6397–6407. Association for Computational Linguistics, November 2020c.

Wei Xiang and Bang Wang. A survey of event extraction from text. *IEEE Access*, 7:173111–173137, 2019.

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis.

Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-hsuan Sung, Brian Strope, and Ray Kurzweil. Multilingual universal sentence encoder for semantic retrieval. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 87–94. Association for Computational Linguistics, July 2020.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

M. Young, Hannah Sarnoff, Dana Lang, and A. Ramírez. Coverage and framing of immigration policy in us newspapers. *The Milbank quarterly*, 2021. doi: 10.1111/1468-0009.12547.

Anam Zahid, Maham Nasir Khan, Ahmer Latif Khan, Faisal Kamiran, and Bilal Nasir. Modeling, Quantifying and Visualizing Media Bias on Twitter. 8:81812–81821, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2990800. URL `https://ieeexplore.ieee.org/document/9079528/`.

Klim Zaporojets, Lucie-Aimée Kaffee, Johannes Deleu, Thomas Demeester, Chris Develder, and Isabelle Augenstein. Tempel: Linking dynamically evolving and newly emerging entities. *Advances in Neural Information Processing Systems*, 35:1850–1866, 2022.

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer, November 2023.

Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification. *Applied Sciences*, 9(16):3389, August 2019. ISSN 2076-3417. doi: 10.3390/app9163389.

Chen Zhang, Qiuchi Li, and Dawei Song. Syntax-Aware Aspect-Level Sentiment Classification with Proximity-Weighted Convolution Network. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1145–1148, Paris France, July 2019. ACM. ISBN 978-1-4503-6172-9. doi: 10.1145/3331184.3331351.

Wenqian Zhang, Shangbin Feng, Zilong Chen, Zhenyu Lei, Jundong Li, and Minnan Luo. KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media, May 2022a.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges, November 2022b.

Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. Sentiment Analysis in the Era of Large Language Models: A Reality Check, May 2023.

Zixuan Zhang, Hongwei Wang, Han Zhao, Hanghang Tong, and Heng Ji. EventKE: Event-enhanced knowledge graph embedding. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1389–1400, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-emnlp.120. URL `https://aclanthology.org/2021.findings-emnlp.120`.

Zilong Zhao, Jichang Zhao, Yukie Sano, Orr Levy, Hideki Takayasu, Misako Takayasu, Daqing Li, Junjie Wu, and Shlomo Havlin. Fake news propagates differently from real news even at early stages of spreading. 9(1):7, 2020. ISSN 2193-1127. doi: 10.1140/epjds/s13688-020-00224-z. URL `https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-020-00224-z`.

Jie Zhou, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu, Tingting Wang, and Liang He. Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE access*, 7:78454–78483, 2019.

Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.

Wenxuan Zhou, Sheng Zhang, Yu Gu, Muhao Chen, and Hoifung Poon. Universalner: Targeted distillation from large language models for open named entity recognition. *arXiv preprint arXiv:2308.03279*, 2023.

Xiang Zhou, Heba Elfardy, Christos Christodoulopoulos, Thomas Butler, and Mohit Bansal. Hidden Biases in Unreliable News Detection Datasets, April 2021.

Résumé étendu

Les progrès technologiques dans le domaine des télécommunications ont radicalement changé la manière dont l'information est diffusée et dont le public interagit avec elle. Le paysage médiatique, autrefois dominé par une poignée d'acteurs, est devenu très concurrentiel avec l'avènement d'internet et a entraîné une forte augmentation du nombre de sources. Cette évolution est parfois qualifiée par les spécialistes de « crise de l'information », caractérisée par la fragmentation de l'audience et la personnalisation accrue de la consommation d'informations. Cette nouvelle structuration du paysage informationnel soulève des inquiétudes avec le développement d'autres genres médiatiques comme l'infodivertissement, et l'augmentation de la désinformation et de la polarisation qui posent des défis pour le bon fonctionnement démocratique. La mitigation et la compréhension de ces effets négatifs ne peuvent être envisagées que par une analyse automatisée de la presse, car le recours à une annotation manuelle serait d'un coût prohibitif considérant le flux d'articles journaliers. Les progrès récents dans le domaine du Traitement Automatique du Langage Naturel (TALN), en particulier avec l'avènement des architectures de type transformer, ont permis de développer de nouvelles méthodes pour l'analyse automatique de la presse. Cependant, il reste des défis à relever pour qualifier avec précision les biais et les opinions, pour généraliser ces méthodes à d'autres langues et domaines moins dotés, et pour maintenir l'explicabilité des résultats. Dans cette thèse, nous abordons ces limitations en nous penchant sur trois points clés : le développement d'un framework pour l'analyse fine des opinions, la création d'un jeu de données multilingues pour l'analyse de sentiment ciblée dans les news, et des analyses de paysages médiatiques en différentes langues. Notre framework intègre la reconnaissance d'entités nommées liées à Wikipedia pour fournir une analyse de sentiments enrichie et explicable. Nous présentons un nouveau jeu de données pour la classification de sentiment ciblée à destination de l'actualité politique européenne, qui comprend des phrases traduites alignées et annotées manuellement dans huit langues, ce nouveau jeu de données nous permets d'étudier l'applicabilité des modèles de traduction automatique pour le transfert de l'analyse des sentiments à d'autres langues. Enfin, nous appliquons notre framework pour conduire une analyse de la presse française et de la presse flamande, nous enrichissons cette analyse par l'incorporation de données issues de « ParlGov » qui nous permettent de caractériser les biais de représentation des différents journaux. Cette approche permet non seulement d'améliorer l'explicabilité des résultats, mais présente également un potentiel d'application dans les langues peu dotées et, dans l'ensemble, elle pourrait fournir des informations précieuses pour notre compréhension des problèmes modernes de consommation de l'information, tels que la polarisation et la désinformation.

Mots-clés: Traitement automatique des langues, Fouille d'opinions, Analyse de presse, Analyse de Biais

Abstract

Technological advances in telecommunications have drastically changed how information is disseminated and how the public interacts with it. The media landscape, once dominated by a handful of players, has become highly competitive with the advent of the Internet and has led to

a sharp increase in the number of sources. This shift is sometimes referred to by scholars as an "information crisis," characterized by fragmented audiences and increased personalization of news consumption. This new structuration of the news landscape raises concerns about sensationalism, misinformation, and polarization that pose challenges to democracy. The mitigation and understanding of these adverse effects can only be envisioned through the automated analysis of news, as traditional manual annotation is prohibitively expensive and the daily flow of news is colossal. Recent advances in natural language processing (NLP), particularly with transformer architectures, have provided new methods for complex automated news analysis. However, challenges remain in accurately detecting biases and opinions, generalizing these methods across languages and domains, and deriving explainable results. In this thesis, we address these limitations by investigating three key areas: the development of a framework for fine-grained opinion analysis in the news, the creation of a multilingual dataset for targeted sentiment classification, and the analysis of the news landscape in different languages. Our framework integrates Named Entity Recognition (NER) with wikification to provide detailed and explainable Targeted Sentiment Classification (TSC), using Wikipedia's entity attributes for enriched representations. We present a new dataset for Targeted Sentiment Classification in European political news, which includes manually translated and annotated sentences in eight languages, addressing the limitations of existing datasets and demonstrating the applicability of machine translation for sentiment analysis transfer. Finally, we apply our framework to analyze bias in the French and Flemish press, incorporating data from the "ParlGov" database to characterize the political leanings of news sources. This approach not only improves explainability, but also shows potential for application in low-resource languages, and overall could provide valuable insights into modern news consumption issues such as polarization and misinformation.

Keywords: Natural Language Processing, Opinion Mining, News Analysis, Bias Analysis