

Projet TP

Détection des doubles standards dans la couverture médiatique occidentale de la guerre à Gaza

1 Consignes générales

Évaluation

- Rapport de projet : 60%
- Organisation du code : 10%
- Démonstration de l'application : 30%

Dates importantes

- **Constitution des équipes** : avant le dimanche 9 novembre 2025
- **Remise finale** : samedi 20 décembre 2025 à minuit
- **Démonstration** : date à fixer ultérieurement

Composition et organisation des équipes

- Taille : 3 à 4 personnes
- Un leader d'équipe doit être désigné
- Email de constitution d'équipe à envoyer avant le 09/11, en précisant le leader d'équipe

Livrables

A Rapport (format PDF)

- Le rapport doit présenter la méthodologie, les résultats, les visualisations ainsi qu'une analyse critique des observations.
- Il doit être rédigé de manière claire, structurée et cohérente.
- Il doit inclure une déclaration précisant la contribution de chaque membre du groupe.
- La longueur maximale du rapport est de 20 pages (annexes incluses).

B Dépôt GitHub contenant

- Le corpus, organisé dans des répertoires clairement identifiés.
- Le code source complet, comprenant :
 - ✓ Les scripts de collecte automatique permettant de reconstruire le corpus.
 - ✓ Code de nettoyage et pré-traitement.

- ✓ Les scripts d'analyse lexicale, sémantique et éventuellement de sentiment
- ✓ Le code de l'application ou de l'interface utilisateur

Règles et ressources

- L'utilisation de bibliothèques ou modules non vus en TP nécessite une autorisation préalable.
- Toutes les sources (articles, livres, sites, bibliothèques etc.) doivent être citées.
- Le leader d'équipe doit envoyer la remise finale par email, incluant :
 - ✓ Le rapport PDF du projet.
 - ✓ Le lien vers le dépôt GitHub.
- Chaque membre doit comprendre l'ensemble du projet, même les parties qu'il n'a pas personnellement codées.
- Tout retard après la date limite entraînera une pénalité.

2 Description du projet

Objectif principal

Ce projet vise à explorer l'utilisation des techniques de traitement automatique du langage naturel (NLP) pour révéler et quantifier les biais dans le traitement médiatique de la guerre à Gaza depuis octobre 2023. Vous analyserez comment les médias occidentaux majeurs (CNN, BBC, The New York Times, etc.) couvrent la guerre à Gaza, en recherchant des doubles standards à deux niveaux :

- **Analyse 1** : Biais internes au conflit → Comment les médias décrivent les Palestiniens et les Israéliens dans la même guerre.
- **Analyse 2** : Biais systémiques → Comment la couverture de Gaza diffère-t-elle de celle de la guerre en Ukraine ?

Composantes du projet

1. Collecte des données

- **Corpus Gaza** : 50 à 100 articles issus de médias occidentaux couvrant la guerre à Gaza (depuis octobre 2023).
- **Corpus Ukraine** : 30 à 50 articles provenant des mêmes médias, portant sur la guerre en Ukraine.

2. Prétraitement : Nettoyer et prétraiter les textes .

3. Analyse lexicale : L'objectif est d'identifier les patterns de vocabulaire révélateurs de biais.

- Identifier les termes et expressions récurrents dans chaque corpus (Gaza et Ukraine) et pour chaque acteur, afin d'extraire les mots-clés les plus fréquents.
hamass | idf | ukraina | pales
- Quantifier et comparer l'usage des termes pour identifier les asymétries de traitement médiatique.

- Identifier systématiquement les variations lexicales (adjectifs, verbes, substantifs) utilisées pour décrire des situations comparables, révélant ainsi les cadres interprétatifs différenciés.
- Analyser les cooccurrences et associations de mots afin d'identifier les expressions fréquemment associées.

4. Analyse sémantique :

- Étudier les contextes d'utilisation des mots-clés (concordance) afin de repérer les différences de sens ou de ton selon le conflit.
- Comparer les champs sémantiques associés à chaque acteur.

5. Analyse de sentiment (optionnelle mais valorisée) :

- Comparer la tonalité utilisée pour décrire les victimes civiles.
- Comparer le ton selon l'acteur.

6. Développement de l'interface utilisateur : Créer une interface permettant :

- De consulter les corpus collectés.
- De visualiser les résultats.

Résultats attendus

Analyse 1 (biais internes)

- **Hypothèse principale** : Les médias occidentaux appliquent des standards différents aux acteurs palestiniens et israéliens.
- **Observations attendues** :
 - ✓ Les Palestiniens sont décrits avec des termes déshumanisants ("militants", "Hamas", "terrorist"); les Israéliens avec empathie ("civilians", "victims").
 - ✓ Attribution de responsabilité asymétrique.
 - ✓ Contextualisation différenciée.

Analyse 2 (biais systémiques)

- **Hypothèse principale** : La couverture de la guerre en Ukraine montre plus d'empathie et de soutien moral que celle de Gaza.
- **Observations attendues** :
 - ✓ Pour l'Ukraine, les médias emploient souvent un ton héroïque et empathique : ils parlent de "heroic resistance", "fight for freedom", montrent de nombreuses histoires humaines et condamnent clairement l'agresseur avec un vocabulaire direct comme ("invasion", "war crimes").
 - ✓ Pour Gaza, la couverture adopte un ton neutre et technique ("conflict", "military operation", "escalation") avec un vocabulaire euphémisé qui évite l'attribution de responsabilité. L'humanisation est limitée, privilégiant les statistiques aux récits personnels, et une tendance à présenter le conflit comme moralement ambigu avec une fausse équivalence entre "both sides".

Contraintes et critères d'évaluation

- Les visualisations (graphiques et statistiques) sont obligatoires dans le rapport et l'application.
- Les projets démontrant une analyse linguistique pertinente et argumentée seront valorisés.
- L'originalité méthodologique ou visuelle sera récompensée.

Bon courage !