

Étendre la caractérisation des chaînes de référence

Regards croisés sur la coréférence : linguistique, TAL et SIC

Jeudi 26 juin 2025

Maison de la Recherche SHS Annie Ernaux - CY Cergy Paris Université

Hélène Manuélian

helene.manuelian@cyu.fr

Marine Delaborde

marine.delaborde@cyu.fr



LT2D

lexiques, textes,
discours, dictionnaires

Plan

1. Contexte
2. La notion de chaîne et les études liées
3. Description/caractérisation linguistique d'une chaîne
4. Les mesures sur les chaînes
5. Les textes en réseau : application à un exemple de presse en ligne

Contexte

- **LT2D** = Équipe pluridisciplinaire
 - Sciences du langage (section CNU 7)
 - Sciences de l'information et de la communication (section CNU 71)
- **Objet d'étude commun** : le discours
- Projet Democis
- Nécessité de dialogue et de partage des outils
- Les travaux sur les chaînes de référence semblent un bon premier point d'entrée entre linguistique et étude des discours

La notion de chaîne

- Notion qui a 50 ans : Chastain (1975)
 - “Let an anaphoric chain be a sequence of singular expressions occurring in a context, such that if one of them refers to something the all of the others also refer to it.”
 - “So far we have anaphoric chains within contexts (...) and we also have referential chains between contexts”
- Au départ, donc, la notion n’est pas une notion limitée à l’unité “texte”

Etude des chaînes et de la coréférence

- Travaux sur la coréférence (notamment)
 - Premiers travaux : résumés de films, romans, articles de presse
 - Corpus DEMOCRAT : différents types et genres de textes
 - Corpus ANCOR : conversations à l'oral
 - Études des écrits sur réseaux / SMS ou les notes de blogs
- Notion de coréférence toujours liée à la notion de texte (chaînes anaphoriques de Chastain ?)
- Une notion proche : les chaînes topicales (Péry-Woodley *et al.* 2011)
- Des difficultés se posent déjà “en interne” aux textes, et la question de la structure textuelle et de l'intertextualité ont déjà été posées

Structure textuelle / intertextualité

« Alors, oui il faut affronter **la rage**, qui se libère et s'exprime. **C'est celle dont** parle un ancien élève du lycée Masséna,

“Oh mon cœur étonné, triste jusqu'à la mort,

J'ai promené **ma rage** en ces jours blancs et froids.”

Cet élève de Nice, c'était Guillaume APOLLINAIRE et **cette rage**, je le sais, beaucoup d'entre vous **la** portent encore au creux de l'estomac. »

Exemple relevé dans un devoir de “Constitution et exploitation de corpus numériques” par Alexandra FARASOGLU (M2 Master SDL)

Transcription du discours d'Emmanuel Macron lors de la cérémonie d'hommage aux victimes de l'attentat de Nice

Source : <https://www.elysee.fr/front/pdf/elysee-module-762-fr.pdf>

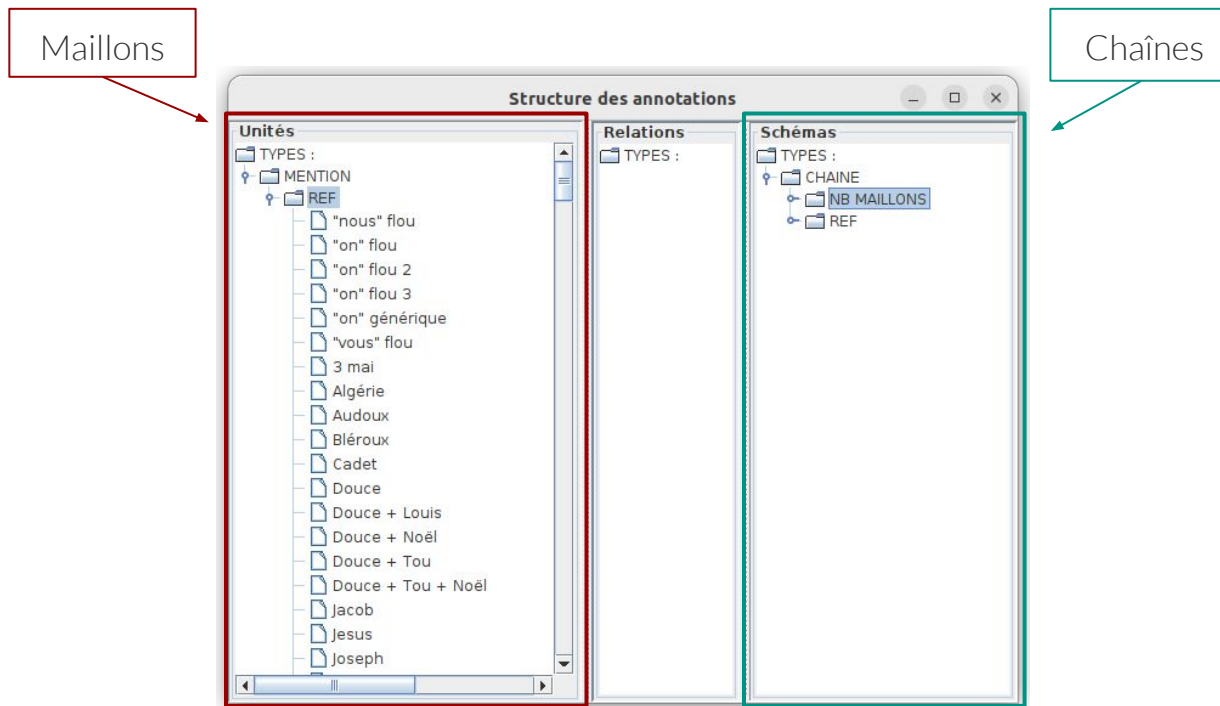
Description et caractérisation de chaîne(s)

Selon quels critères ?

- Identification du référent : qui ou quoi ?
- Type du référent : selon l'objectif de l'analyse et le cadre théorique
- Longueur de la chaîne (nombre de maillons)
- Nature et fonction des maillons (nombre, ordre dans la chaîne)
- Distance entre les maillons
- Ordre d'apparition des chaînes
- Portée des chaînes
- Certaines mesures sur les expressions référentielles ne permettent pas de décrire / caractériser des chaînes mais des textes
 - Ex : densité référentielle

Les mesures sur les chaînes

Democrat : URS et TXM



Les mesures sur les chaînes

Democrat : URS et TXM

- Exploitation :
 - **Schema Lengths** : affiche la longueur moyenne (en unités) des schémas sélectionnés et l'index hiérarchique des longueurs
 - **Schemas Index** : index hiérarchique d'une propriété des schémas sélectionnés
 - **Schemas List** : afficher les schémas sélectionnés avec leur nombre, leurs propriétés et la liste de leurs unités
 - **Schemas Progression** : affiche le graphique de progression des schémas en mettant en valeur les unités sélectionnées (la sélection d'unités ne participe pas à la sélection des schémas)
 - **Schemas Ratio** : affiche pour chaque schéma sélectionné, le ratio *unitsslectionnees/unitsdemarge*
 - **Schemas Summary** : dénombre les unités sélectionnées ainsi que les valeurs de leurs propriétés
 - **Units Index** : index hiérarchique d'une propriété des unités sélectionnées
 - **Units Interdistance** : affiche les moyennes, médianes et quartiles des distances et des cadences pour les unités sélectionnées
 - **Units List** : affiche la liste des unités sélectionnées avec différents modes d'affichage
 - **Units Referential Density** : calcule le rapport entre le nombre d'unités d'un certain type et le nombre de mots du corpus (en %)
 - **Units Referential Stability** : calcule le rapport entre le nombre d'unités sélectionnées et le nombre de formes différentes utilisées par ces unités (taille du vocabulaire des unités) dans le corpus
 - **Units Summary** : dénombre les unités sélectionnées ainsi que les valeurs de leurs propriétés

Des réseaux de chaînes ?

- Avec les médias numériques, l'organisation des textes (et des discours) est bouleversée
- On consulte un article en ligne et l'hypernavigation / l'algorithme nous renvoie à des articles antérieurs / postérieurs, complémentaires, etc...
- Sortir de l'unité de temps et de lieu
 - “contextes multiples” : énoncés isolés
 - Conserver une approche “syntagmatisante”, relier des chaînes en réseau

Un exemple éphémère

Récits sur la Flottille pour la Liberté sur le site de CNews

<https://www.cnews.fr/monde/2025-06-09/flottille-pour-gaza-apres-larraisonnage-le-ministre-israelien-de-la-defense>

<https://www.cnews.fr/insolite/2025-06-09/greta-thunberg-photographiee-tout-sourire-alors-quun-soldat-israelien-lui>


<https://www.cnews.fr/monde/2025-06-09/ils-vont-retourner-dans-leurs-pays-le-bateau-humanitaire-de-greta-thunberg-et-rima>

<https://www.cnews.fr/monde/2025-06-10/bateau-pour-gaza-nous-avons-ete-kidnapes-dans-les-eaux-internationales-par-israel>

<https://www.cnews.fr/france/2025-06-11/bateau-pour-gaza-jean-luc-melenchon-demande-emmanuel-macron-detre-recu-lellysee>

RETRAITES ISRAËLIEN POLITIQUE FAITS DIVERS GUERRE EN UKRAÏNE FRANCE PARIS SPORT FOOTBALL

GRETA THUNBERG PHOTOGRAPHÉE TOUT SOURIRE ALORS QU'UN SOLDAT ISRAËLIEN LUI PROPOSE UN SANDWICH



Un soldat israélien propose un sandwich à Greta Thunberg, âgée de 16 ans, lors de son arrestation à bord du bateau pour Gaza. Elle sourit et accepte le sandwich.

Le 7/6/2025 à 14:24 - Mis à jour le 05/06/2025 à 14:24

La navire de militants en route pour Gaza avec Greta Thunberg et Rima Hassan intercepté par la marine israélienne. Nouvelle et eau ont été proposés par Tsoah.

Dans la nuit du dimanche 6 au lundi 7 juin, le navire transportant des militants en route pour Gaza a été intercepté par la marine israélienne.

A bord de ce bateau, port d'attache le far pour cibler le blocus israélien de la bande de Gaza, se trouvent douze militants, dont Rima Hassan, militante française de l'ONG Tsoah. Ils attendaient d'acheminer des vivres dans l'enclave palestinienne. L'offensive israélienne a été photographiée au moment de l'interception du navire.

Soudain et presque étonnée, elle s'est vu proposer un sandwich par des membres de l'armée israélienne. Selon les autorités israéliennes, tous les passagers ont refusé et seuls les deux militants ont été proposés aux vivres.


Le bateau a été escorté jusqu'à une zone israélienne.

Israel hater is a good thing. Good morning to all our followers. In case you missed it. The 'safe zone' is safety meeting us way to the shores of Israel. The passengers are safe and were provided with sandwiches and water, and are expected to return to their home countries.

LA LIBERTÉ

RETRAITES ISRAËLIEN POLITIQUE FAITS DIVERS GUERRE EN UKRAÏNE FRANCE PARIS SPORT FOOTBALL

FLOTTELLE POUR GAZA : APRÈS L'ARRAISONNAGE, LE MINISTRE ISRAËLIEN DE LA DÉFENSE ORDONNE À TSAHAL DE PROJETER AUX PARTICIPANTS LA VIDE DE MASSACRES DU 7-OCTOBRE



Après l'arraisonnement du bateau, le ministre de la Défense israélien, Benny Gantz, a ordonné à l'armée israélienne (Tsoah) de projeter aux participants la vidéo des massacres du 7 octobre.

Le 7/6/2025 à 14:22 - Mis à jour le 05/06/2025 à 14:22

Le Maïdan, le bateau qui tentait de relier Gaza avec le sud sans interrompre l'offensive israélienne. Greta Thunberg et l'européenne Rima Hassan, a été dérangé dans la nuit de dimanche à lundi par les autorités israéliennes. Après l'arraisonnement, le ministre de la Défense israeli Kati a ordonné à l'armée de montrer aux participants les vidéos d'horreur des massacres du 7 Octobre.

Après plusieurs mises en garde, la flottille pour Gaza a été interrompue par les autorités israéliennes dans la nuit du dimanche à lundi et les participants, dont Greta Thunberg et Rima Hassan, ont été pris en charge par l'armée.

Le sillon Maïdan est désormais dévié vers le port d'Ashdod, grand port commercial du sud d'Israël non loin de Gaza, où les passagers doivent des débriefings, a précisé le ministère israélien de la Défense. Il est attendu qu'ils soient libérés par la flottille, a indiqué le bureau de presse, cité par le média.

Mais avant cela, le ministre de la Défense israélien Benny Gantz a ordonné à l'armée de leur montrer des vidéos d'horreur de l'attaque du 7 Octobre, selon une information d'AFP. Ça, après qu'ils aient été escortés par les forces armées israéliennes.

Le 7/6/2025 à 14:22 - Mis à jour le 05/06/2025 à 14:22

Sur le même sujet

Un court reportage dans l'après-midi du dimanche 6 juin, le bateau humanitaire de Greta Thunberg et Rima Hassan, escorté par l'armée israélienne.

LA LIBERTÉ

Annotation de 4 chaînes

- La Flottille
- Les passagers du bateau
- Rima Hassan
- Greta Thunberg

=> Dans les 5 articles placés dans l'ordre chronologique

=> Dans chaque article indépendamment les uns des autres

ARTICLE 1 : 429 mots

ARTICLE 2 : 442 mots

ARTICLE 3 : 427 mots

ARTICLE 4 : 997 mots

ARTICLE 5 : 538 mots

TOTAL = 2833

Annotation de 4 chaînes

[REF:**Flottille**, NB MAILLONS:51, TYPE REFERENT:, GENRE:, NOMBRE:] (51/51): l'embarcation, Il, qui, du voilier parti d'Italie le 1er juin pour « briser le siège » de Gaza, FLOTTILLE POUR GAZA, le navire Madleen, BATEAU POUR GAZA, la « Flottille de la liberté », Le voilier Madleen, Le navire de militants en route pour Gaza avec Greta Thunberg et Rima Hassan, Ce bateau humanitaire, son, du Madleen, du « Madleen », « Flottille de la liberté », LE BATEAU HUMANITAIRE DE GRETA THUNBERG ET RIMA HASSAN, le navire transportant des militants en route pour Gaza, ses, Le bateau humanitaire Madleen, la flottille, du « Madleen », il, ce bateau , parti d'Italie le 1er juin pour « briser le blocus israélien » de la bande de Gaza, Le bateau humanitaire Madleen à destination de Gaza, Le Madleen, du navire, du bateau, du voilier Madleen, du voilier , le Madleen, le navire en route vers la bande de Gaza, le bateau, un bateau d'aide pour Gaza, le Madleen, la Flottille de la liberté, Le Madleen , le bateau qui tentait de rallier Gaza avec à son bord notamment l'activiste suédoise Greta Thunberg et l'eurodéputée Rima Hassan, l', le bateau humanitaire, Le bateau, celui -ci, la flottille, LE NAVIRE

[REF:**Equipage_complet**, NB MAILLONS:40, TYPE REFERENT:, GENRE:, NOMBRE:] (40/40): ses passagers, Les passagers du bateau, qui, notre, les militants, Rima Hassan et les autres équipiers, douze militants, l'équipage, Ils, ILS, se, des militants, Nous, 12 militants, aux participants, aux militants, CÉLÉBRITÉS, les volontaires internationaux, 12 de plusieurs nationalités, les célébrités, leur, célébrités, ils, les passagers, tous les passagers, Greta Thunberg et les autres, des militants français , allemand , brésilien , turc , suédois , espagnol et néerlandais, NOUS, AUX PARTICIPANTS, leurs, les 12 personnes

[REF:**GT**, NB MAILLONS:29, TYPE REFERENT:, GENRE:, NOMBRE:] (29/29): LUI, l'activiste suédoise Greta Thunberg, son, -elle, L'activiste suédoise, GRETA THUNBERG, l'activiste Greta Thunberg, la Suédoise Greta Thunberg, Greta Thunberg, elle

[REF:**RH**, NB MAILLONS:15, TYPE REFERENT:, GENRE:, NOMBRE:] (15/15): la députée européenne Rima Hassan, une parlementaire, RIMA HASSAN, l'eurodéputée Rima Hassan, Rima Hassan , eurodéputée française LFI, Rima Hassan, l'eurodéputée franco-palestinienne de gauche Rima Hassan, l'eurodéputée insoumise Rima Hassan

Résultats 1 : longueur des chaînes / nb ER

Longueur des chaînes

	Ensemble	Article 1	Article 2	Article 3	Article 4	Article 5
Flottille	51	9	7	14	14	6
Equipage	40	9	10	7	12	1
Greta Thunberg	29	3	10	3	10	1
Rima Hassan	15	3	2	2	2	5

Résultat 2 : Distance et cadence

Article 1 :

distance moyenne : $1241 / 20 = 62.05$

distance médiane : 48

cadence moyenne : $1144 / 24 = 57.2$

cadence médiane : 29

Article 2 :

distance moyenne : $818 / 25 = 32.72$

distance médiane : 24

cadence moyenne : $696 / 29 = 27.84$

cadence médiane : 11

Article 3 :

distance moyenne : $951 / 22 = 43.23$

distance médiane : 22

cadence moyenne : $760 / 26 = 34.54$

cadence médiane : 17

Article 4 :

distance moyenne : $2662 / 34 = 78.29$

distance médiane : 45

cadence moyenne : $2463 / 38 = 72.44$

cadence médiane : 18

Article 5 :

distance moyenne : $811 / 9 = 90.11$

distance médiane : 105

cadence moyenne : $773 / 11 = 85.89$

cadence médiane : 39

Ensemble des articles :

distance moyenne : $9986 / 131 = 76.23$

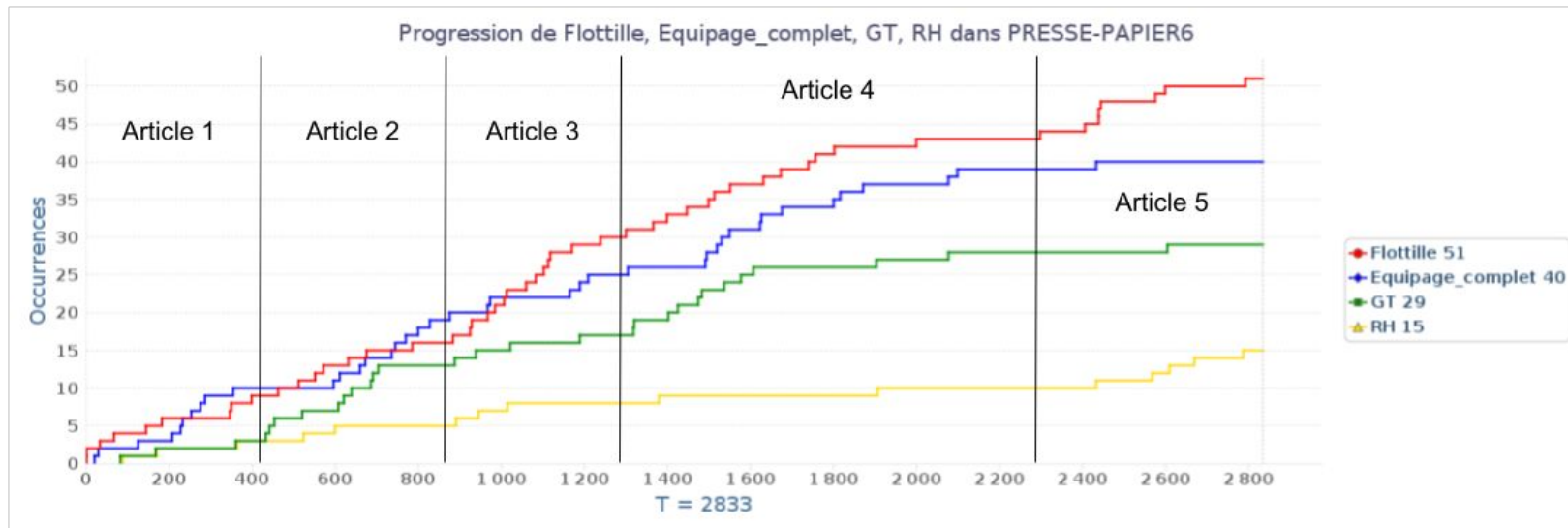
distance médiane : 45

cadence moyenne : $6284 / 135 = 47.97$

cadence médiane : 21

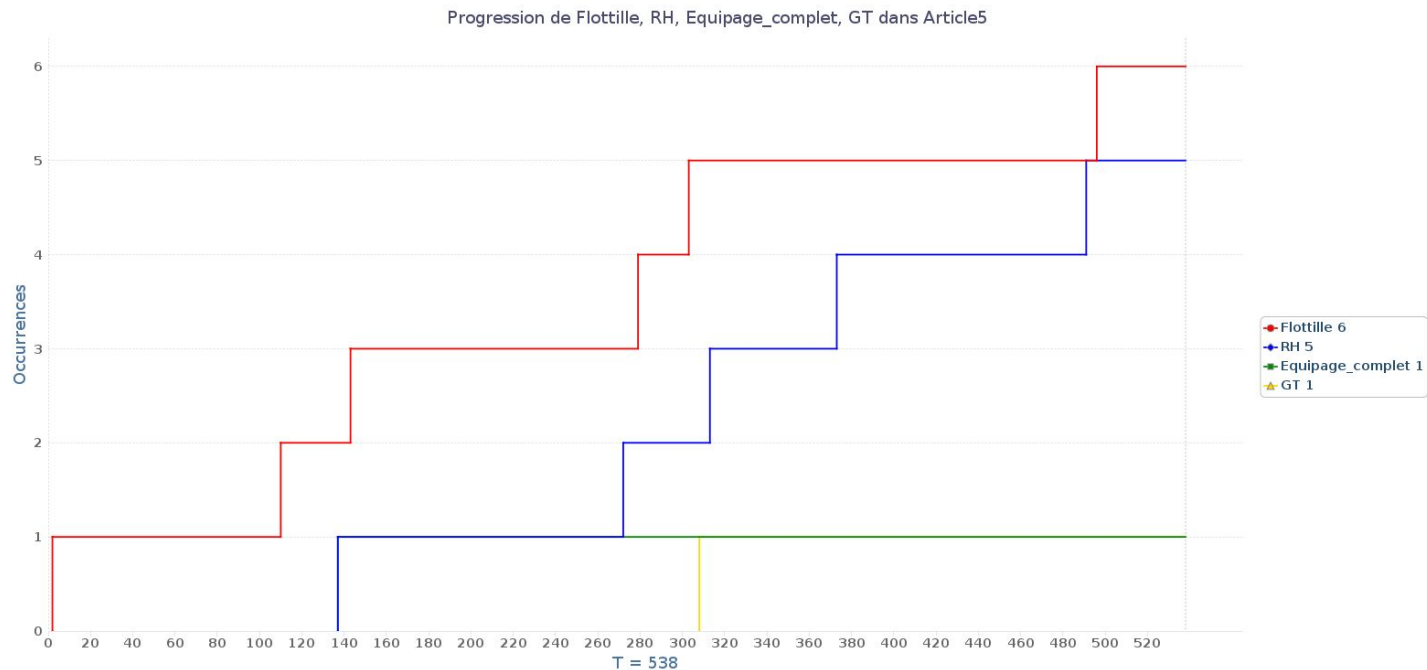
Visualisation

Schémas de progression : à questionner dans ce cas !

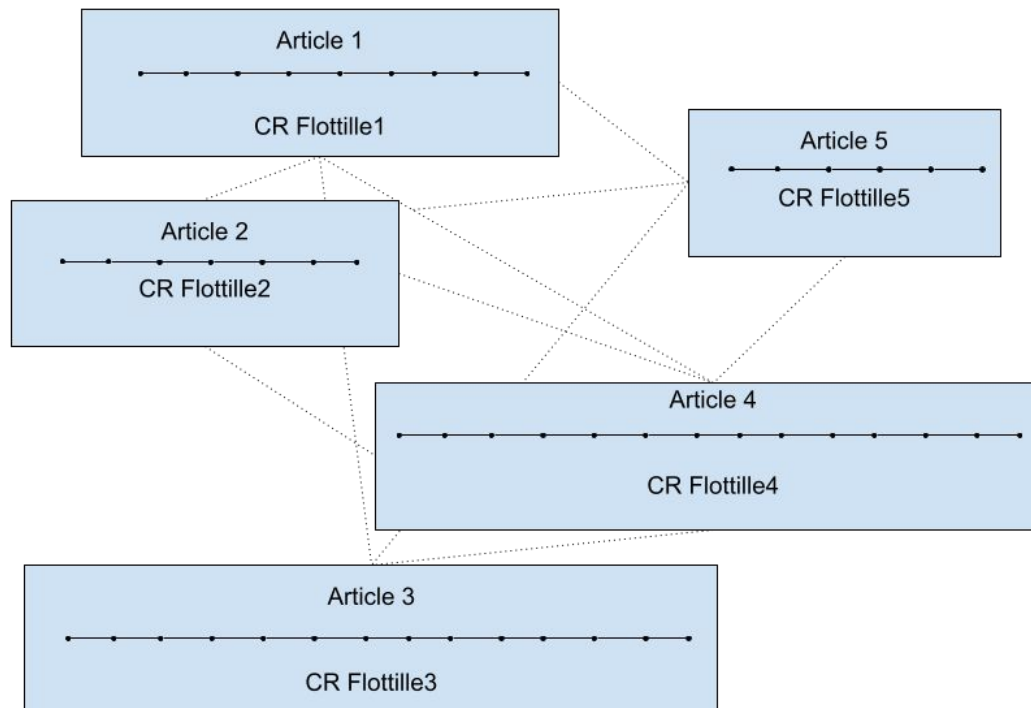


Visualisation

Schémas de progression de l'article 5



Interroger la notion de premier maillon



Flottille pour Gaza (titre) Le Madleen,
le bateau qui tentait de rallier Gaza
avec à son bord notamment l'activiste
suédoise GT et l'eurodéputée RH

Le navire de militants en route pour
Gaza

Le bateau humanitaire de Greta
Thunberg et Rima Hassan

Bateau pour Gaza (titre) / Le bateau
humanitaire Madleen à destination de
Gaza

Bateau pour Gaza (titre) / La "Flottille
de la liberté"

Quelles perspectives ?

- Sortir de l'unité de temps et de lieu
 - “contextes multiples” : énoncés isolés
 - Conserver une approche “syntagmatisante” mais relier des chaînes en réseau ?
- Un pas de côté est sans doute intéressant à faire :
 - multiplication des “premiers maillons”
 - Un ou plusieurs textes ? Un document et des textes ?
- Sans doute des dizaines d'autres questions :
 - multimodalité = gérer les photos, les légendes de photos
 - difficultés liées à la mise à jour permanente (difficulté à retrouver les articles après quelques jours)
 - l'organisation même des sites est sans doute à prendre en compte
 - contraintes techniques :
 - Quels outils ?
 - Quelles mesures utiliser ?

Références

- Chastain, C. (1975). Reference and Context: Language, Mind, and Knowledge (Minnesota Studies in the Philosophy of Science 7).
- Péry-Woodley, M. P., Afantenos, S., Ho-Dac, L. M., & Asher, N. (2011). La ressource ANNODIS, un corpus enrichi d'annotations discursives. *Revue TAL: traitement automatique des langues*, 52(3), 71-101.
- Schnedecker, C. (2021). *Les chaînes de référence en français*. Ophrys.

Les mesures sur les chaînes

Democrat : URS et TXM

	Schema Lengths				Referential Density			
	Min = 2		Min = 3		Min = 2		Min = 3	
	ER/CR		ER/CR		ER/Mots	%	ER/Mots	%
Aden	1588/354	4,485875706	1146/133	8,616541353	1588/13290	11,94883371	1146/13290	8,623024831
Articles Wiki	1761/301	5,850498339	1397/119	11,7394958	1761/10628	16,56943922	1397/10628	13,1445239
Code procédure pénale	2739/656	4,175304878	2087/330	6,324242424	2739/12509	21,89623471	2087/12509	16,68398753
Convention univ	807/182	4,434065934	593/75	7,906666667	807/5384	14,98885587	593/5384	11,0141159
Convention milieu marin	2609/571	4,569176883	2055/294	6,989795918	2609/12860	20,28771384	2055/12860	15,97978227
Convention thon tropic	2036/410	4,965853659	1618/201	8,049751244	2036/11173	18,22250067	1618/11173	14,48133894
Convention transport aérien	1681/360	4,669444444	1241/140	8,864285714	1681/9402	17,87917464	1241/9402	13,19931929
De la ville au moulin	3115/293	10,63139932	2801/136	20,59558824	3115/11945	26,07785684	2801/11945	23,4491419
Diable au corps	2921/375	7,789333333	2479/154	16,0974026	2921/14596	20,01233215	2479/14596	16,98410523
Douce Lumière	2166/265	8,173584906	1856/110	16,87272727	2166/12268	17,6556896	1856/12268	15,12879035
Élisabeth Seton	2574/375	6,864	2168/172	12,60465116	2574/12990	19,81524249	2168/12990	16,68976135
Est Républicain	1336/393	3,399491094	890/170	5,235294118	1336/12254	10,90256243	890/12254	7,262934552
Est Républicain 2	1190/340	3,5	812/151	5,377483444	1190/11605	10,25420078	812/11605	6,996984059
Jean Christophe	2264/310	7,303225806	1952/154	12,67532468	2264/10962	20,65316548	1952/10962	17,80696953
Jean Christophe 2	2054/206	9,970873786	1834/96	19,10416667	2054/10303	19,93594099	1834/10303	17,80064059
Madame de Hautefort	2416/300	8,053333333	2086/135	15,45185185	2416/11758	20,5477122	2086/11758	17,74111243
Nemoville	2528/286	8,839160839	2218/131	16,93129771	2528/12378	20,42333172	2218/12378	17,91888835
Rosalie de Constant	2335/337	6,928783383	1949/144	13,53472222	2335/12688	18,40321564	1949/12688	15,360971