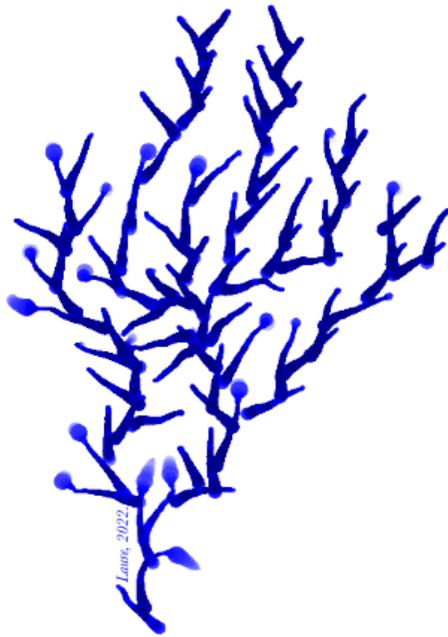


Entre-deux et dialogues disciplinaires, un inconfort stimulant.

Retour d'expériences



Mathilde Dagnat

Université de Lorraine & ATILF-CNRS

<http://mathilde.dagnat.free.fr>

[anga]E

La linguistique à la croisée des disciplines : quelle place pour l'interdisciplinarité ? Lyon, 13 octobre 2023

Introduction

Entre-deux, dialogue ou nouvelle discipline ?

Inconfort productif

Linguistique et ...

1. études littéraires

2. informatique

3. sciences cognitives

Conclusion : enjeux institutionnels

Cotutelles et codirections

Cadres locaux et nationaux

Annexe 4 : Pluridisciplinarité, interdisciplinarité, transdisciplinarité

Pour les définitions de la pluridisciplinarité, de l'interdisciplinarité et de la transdisciplinarité, voir page 19 du Référentiel HCERES disponible à l'adresse suivante :

https://www.hceres.fr/sites/default/files/media/downloads/Referentiel_URinter_%20Vague%20B_RECH%20UR_12nov.pdf

– La **pluridisciplinarité** est une juxtaposition de perspectives disciplinaires qui élargit le champ de la connaissance, en accroissant le nombre des données, des outils et des méthodes disponibles. Les composantes disciplinaires, dans ce cas, gardent leur identité : une discipline, qui se trouve en général en situation de pilotage, utilise la méthodologie et les instruments d'une ou plusieurs autres disciplines pour traiter une question ou faire avancer un projet de recherche qui est propre à son domaine disciplinaire.

– L'**interdisciplinarité** est la coopération de plusieurs disciplines autour de projets communs. Ces projets ouvrent des perspectives de recherche à chacune des disciplines, qui ne sont plus cantonnées pour la plupart à des situations d'application. Les travaux en commun associent des données, des méthodes, des outils, des théories et des concepts issus de disciplines différentes en une synthèse dans laquelle le rôle des composantes disciplinaires va bien au-delà de la simple juxtaposition. Parmi les marques de cette intégration, on retiendra en particulier :

- des combinaisons de modèles ou de représentations qui unifient des approches disparates ;
- un mode de collaboration partenarial et non un simple échange de services, avec un
- investissement coordonné des moyens et une organisation de type coopératif ;
- la création d'un langage commun par hybridation, conduisant à la révision des hypothèses initiales, à la compréhension plus large du problème posé, à l'ouverture de perspectives neuves et à l'élaboration de nouveaux savoirs.

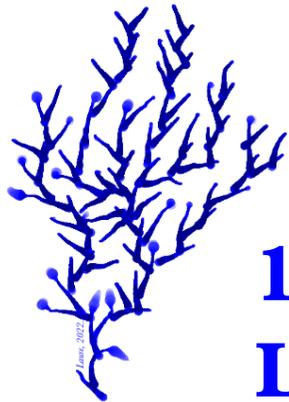
– La **transdisciplinarité** est une approche scientifique qui dépasse les points de vue disciplinaires par l'approche globale d'une question. Elle témoigne d'un degré d'intégration supplémentaire par rapport à l'interdisciplinarité, que les disciplines partenaires atteignent lorsque cette pratique répétée débouche sur la définition de nouveaux paradigmes et sur la formation d'une communauté qui les partage, faisant ainsi émerger peu à peu une nouvelle discipline¹². Ce fut le cas naguère de la biologie des systèmes, de la biologie synthétique, de l'intelligence artificielle et de l'écologie humaine.

Voir également la notion d'interdisciplinarité dans le document édité par l'OCDE disponible à l'adresse suivante :

<https://www.oecd-ilibrary.org/docserver/0ca0ca45-en.pdf?expires=1632382325&id=id&accname=guest&checksum=B9CD0E678106C385AE82A38E0BF35D59>

ANR

Guide AAPG2022_Version 1.2_14/10/2021, p. 42



1 Linguistique et études littéraires

Décrire et mesurer un effet esthétique

Objectif: expliciter le processus de représentation réaliste d'une variété linguistique et l'évolution de son fonctionnement textuel

[Dargnat 2006, 2008, 2012, 2022]

- Lieux de variation et imaginaire linguistique
Oralité **P**opulaire **Q**uébécoise et *joual*
→ Recensement des définitions et des traits linguistiques associés
- Perspective sémiotique et stylistique diachronique
Articuler texte et contexte (social, idéologique et culturel)
- « Sac de données » compartimenté
 - Étape 1 : construction du référent (*Sankoff-Cedergren et Montréal 84*)
 - Étape 2 : 5 pièces de théâtre de M. Tremblay (1968-1998)

- Représenter une « parlure » : choix, conventions et innovations
ex. *bécosses, moé, cibole, enfirouâper*
- « Bricoler » et comparer des données récalcitrantes
Traitement structurel en XML-TEI (*Weblex* → TXM), catégorisation linguistique limitée

Ressources linguistiques : Sankoff-Cedergren et Montréal 84

Transcription : fichier .doc « tout moche »

(No 95'71 1: RS., 2:)

1. Mais: la guerre la guerre mais comment vous avez trouvé ça ? Finalement. Est-ce-que ça été une bonne expérience pour vous ou si:

2. Oui ça été une expérience. Moi: je vais te conter quand je me suis enrôlé: dans notre temps nous-autres ça prenait une huitième année aujourd'hui je pense ' c'est une dixième ou une onzième hein je pense.



Fichier balisé en xml TEI (balises structurales)

```
<HEAD>ORAL POP</HEAD>
<BODY>
<CORPUS TYPE="Sankoff-Cedergren"> [...]
<DIV TYPE="C" DATE="71">
<!-- références originales : No 95'71 1: RS., 2: . Profil
sociolinguistique du LOC 2 : 41/H/marin-chauffeur de
camions/Rosemont/cm=.01-->
<SP NUM="01" WHO="LOC1">mais: la guerre la guerre mais
comment vous avez trouvé ça? finalement ## est-ce-que ça été
une bonne expérience pour vous ou si:</SP>
<SP NUM="02" WHO="LOC2"> oui ça été une expérience ## moi:
je vais te conter quand je me suis enrôlé: dans notre temps nous-
autres ça prenait une huitième année aujourd'hui je pense ' c'est
une dixième ou une onzième hein je pense ##
```

[cf. Martineau et al. 2022 et pour une version numérique récente le *Fonds de Données Linguistiques du Québec*
<https://fdlq.recherche.usherbrooke.ca/corpus>]

Ressources littéraires : 5 pièces de théâtre québécois « joualisantes »

Texte original scanné (1968)

Cuisine. Quatre énormes caisses occupent le centre de la pièce.

premier acte

Entre Linda Lauzon. Elle aperçoit les quatre caisses posées au centre de la cuisine.

in fin out
LINDA LAUZON
Misère, que c'est ça ? Maman !
GERMAINE LAUZON (*dans une autre pièce*)
C'est toé, Linda ?
LINDA LAUZON
Oui. Que c'est ça, les caisses qui traînent dans cuisine ?
GERMAINE LAUZON
C'est mes timbres !
LINDA LAUZON
N Sont déjà arrivés ? Ben, j'ai mon voyage ! Ç'a pas pris de temps !

Entre Germaine Lauzon.

Texte original après océrisation

8

Cuisine. Quatre énormes caisses occupent le centre de la pièce.

premier acte

Entre Linda Lauzon. Elle aperçoit les quatre caisses posées au centre de la cuisine.

LINDA LAUZON Misère, que c'est ça ? Maman !

GERMAINE LAUZON (*dans une autre pièce*) C'est toé, Linda ?

LINDA LAUZON Oui. Que c'est ça, les caisses qui traînent dans cuisine ?

GERMAINE LAUZON C'est mes timbres !

LINDA LAUZON Sont déjà arrivés ? Ben, j'ai mon voyage !

Ç'a pas pris de temps !

Entre Germaine Lauzon.

Fichier balisé en xml TEI (balises structurales)

[...]

<PG NUM="8"/>

<IS TYPE="SITU"><I>Cuisine. Quatre énormes caisses occupent le centre de la pièce.</I></IS>

<DIV TYPE="ACTE">

<HEAD>premier acte</HEAD>

<IS TYPE="SITU"><I>Entre Linda Lauzon. Elle aperçoit les quatre caisses posées au centre de la cuisine.</I></IS>

<SP NUM="001" WHO="LINDA LAUZON"> Misère, que c'est ça ? Maman ! </SP>

<SP NUM="002" WHO="GERMAINE LAUZON"> <IS TYPE="ENO"><I>(dans une autre pièce) </I></IS> C'est toé, Linda ? </SP>

<SP NUM="003" WHO="LINDA LAUZON"> Oui. Que c'est ça, les caisses qui traînent dans cuisine ? </SP>

<SP NUM="004" WHO="GERMAINE LAUZON"> C'est mes timbres ! </SP>

<SP NUM="005" WHO="LINDA LAUZON"> Sont déjà arrivés ? Ben, j'ai mon voyage ! Ç'a pas pris de temps ! </SP>

<IS TYPE="SITU"><I>Entre Germaine Lauzon.</I></IS>

[...]

Importation et exploration avec le logiciel *Weblex* (ancêtre de TXM)

Chercher dans la page :

Il y a 5 occurrences de *coudonc* dans le corpus tremblay3 (propriété word)

1	LVM,tot,ALEX II,29,065	mille de vouloir absolument souffrir , '	coudonc	? Aurais - tu aimé mieux que j' vous ba
2	LVM,tot,ALEX I,56,169	s ... même après mon bain ... @S Aïe , '	coudonc	, c' est pas avec toé que j' vas avoir
3	EF,10n2_13a,NANA,27,142	toute ! En as - tu passé des boutes , '	coudonc	? Ben non . Ben t' as compris que quand
4	EF,10n2_13a,NANA,36,250	le roi d' Espagne . Pis en italien ... '	coudonc	, y' a - tu un roi , en Italie ? Y' en
5	IO,act2,LUCILLE,88,370	ez - vous en train de vous réconcilier ,	coudonc	? Jamais de la vie ! Ouf , j' ai eu peu

Corpus [Recherche simplifiée](#)

Source A

Source B

Source C

Général
[Index](#)
[Concordances](#)
[Vocabulaire](#)
[Élagage quantitatif](#)
[Élagage qualitatif](#)
[Lexicogrammes](#)
[Lexicogrammes récursifs](#)
[Segments répétés](#)
[Répartition](#)
[Spécificités](#)

Options générales de l'affichage

- Afficher les résultats dans Listes sous forme de (forcer)
- Afficher les premières lignes Afficher les commentaires :
- Locale Attribut d'occurrence implicite

Affichage de l'index et des éditions

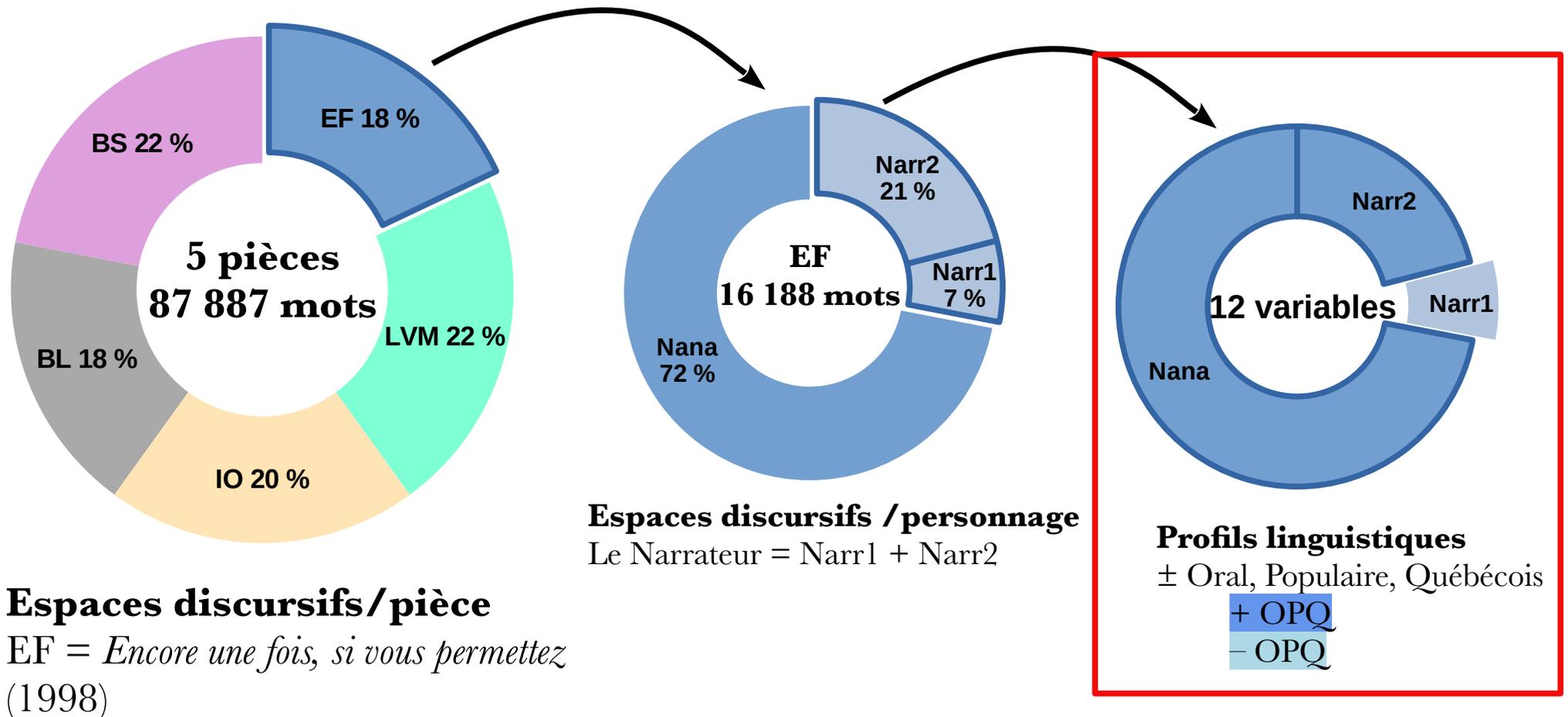
- Composer l'index avec les champs :
word p2 p3 p4 p5 p6 p7 p8 p9 p10
p11 p12 p13 p14 p15 p16 p17 p18 p19 p20
- Format : Tabulé Linéaire



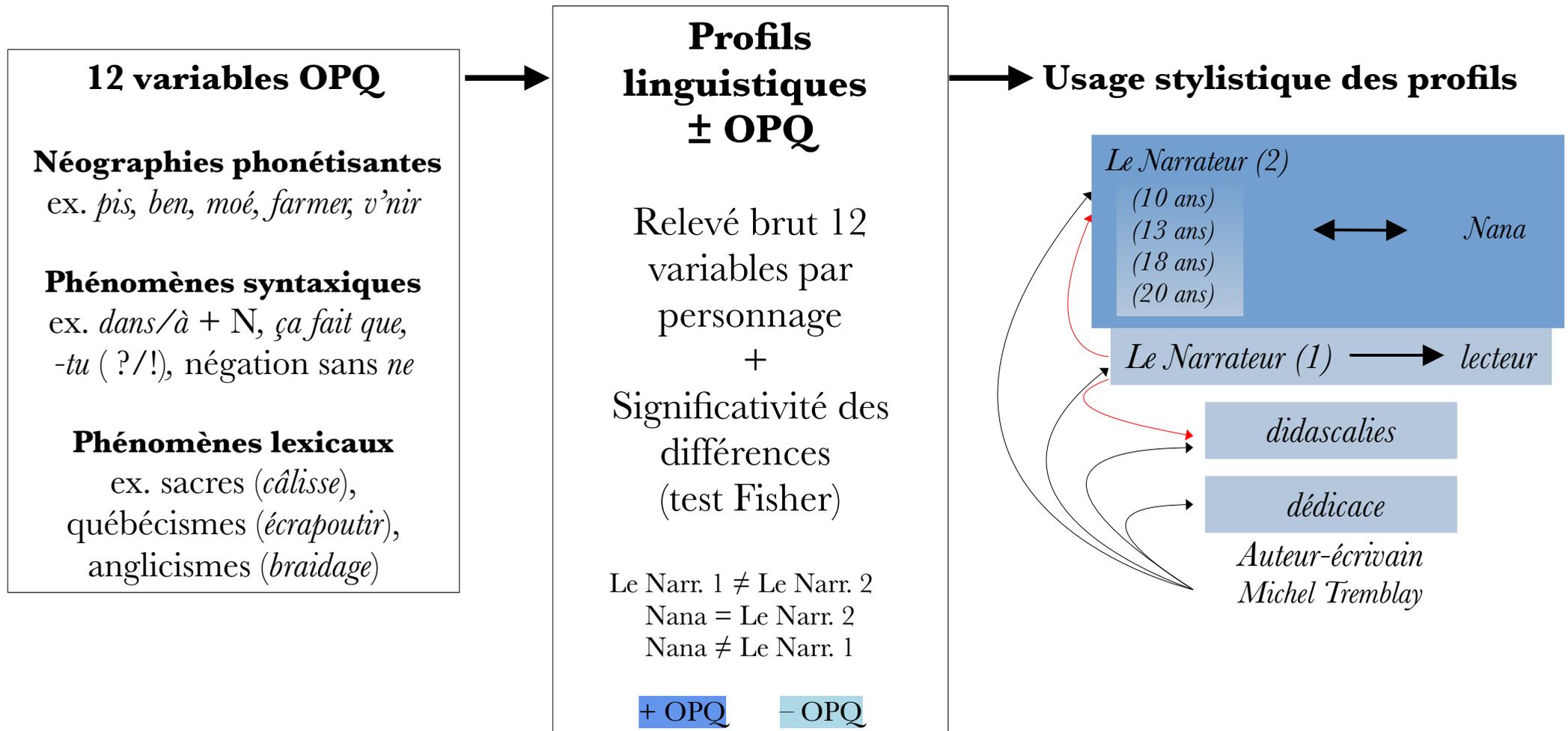
© Serge Heiden

[UMR5191 ICAR CNRS/ENS Lettres](#)

- Quantifier et caractériser la parole et les personnages-locuteurs



- Valeurs stylistiques des différences observées : niveaux énonciatifs et métalepse narrative



Bilan 1. Linguistique et études littéraires : pluri, inter, trans-disciplinarité ?

- Du point de vue théorique : linguistique textuelle, linguistique de l'énonciation, narratologie, pragmatique, théorie de la pertinence, stylistique, sociolinguistique « de la 3^e vague » [Eckert 2012], etc.
- Articulation idéologie, imaginaire et réalités linguistiques, l'exemple de *enfirouâper*

Le cas « enfirouâper »

[cf. Thibault 2008 + TLFQ]

Tromper, duper, rouler, séduire, avaler, envelopper

→ Étymologie proposée

> **enfifrer** (enc-ler) + **rouâper** (râper, gratter, battre, réprimander)

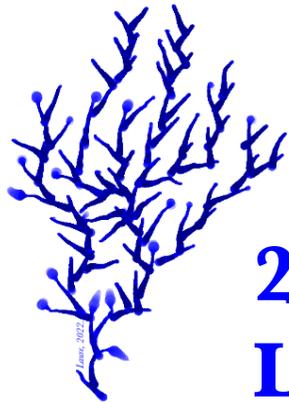
Jeu de mots d'Hector Berthelot en 1878 (?)

→ Réanalyse (erronée) comme anglicisme

Le Devoir, 20-06-1962 + quelques dictionnaires et mentions

> **in fur wrap** (« enrrouler dans la fourrure »)

- Recherche outillée et quantitative (comparaison systématique et statistiques) mais limitée, à cause de la nature des données



2

Linguistique et informatique

Systematiser l'exploration et renouveler la perspective

2.1 Structure du discours et argumentation

[Dargnat 2022, Huber et al. 2019a, 2019b, Tsai 2020]

- **Relations de discours et relations argumentatives**

- Théories (RST¹, SDRT² et argumentation³)

- Typologie des relations de discours*

- [Asher & Lascarides 2003², Renkema 2008¹, Taboada & Mann 2006¹; Freeman 2011³, Walton 2013³]

- Cohérence et intentionnalité* : toutes les relations sont-elles argumentatives ?

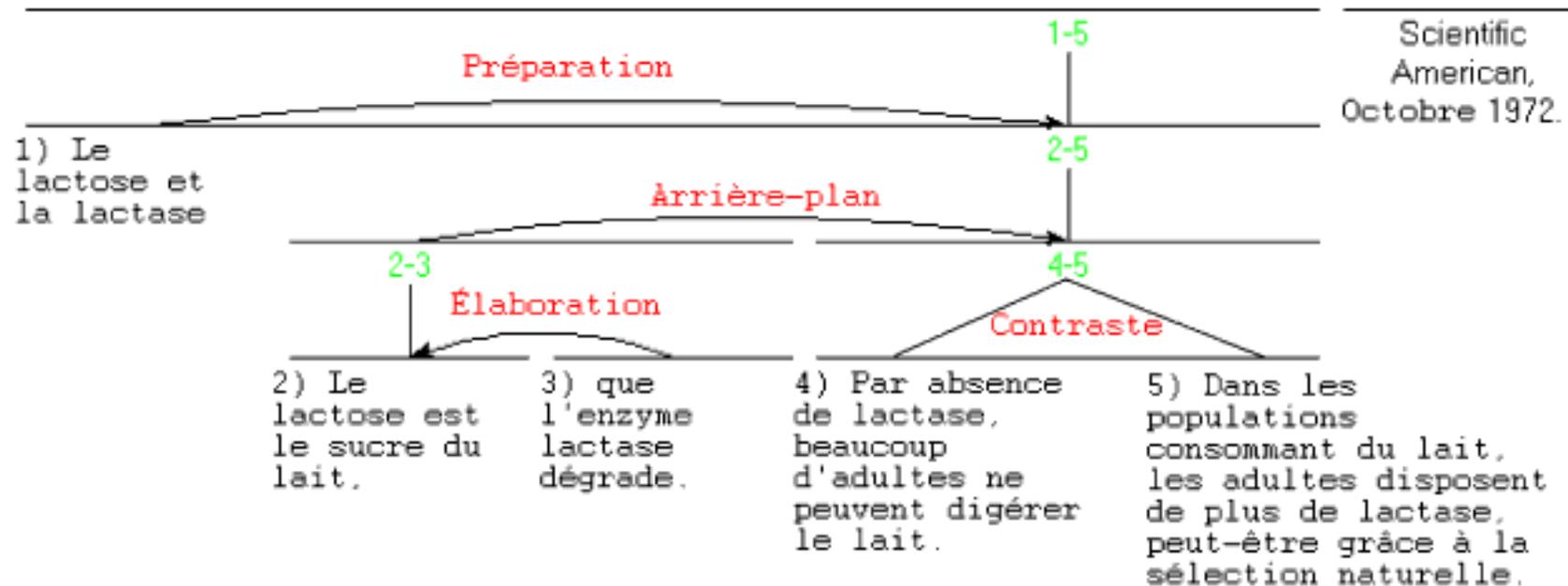
- ex. cause, concession, démonstration, élaboration, contraste, etc.

- Hiérarchie des relations et représentations graphiques, vers l'implémentation

- Outils divers : [RSTWeb](#), [ArgDown](#), [fonctions Latex](#), etc.

2.1 Argumentation

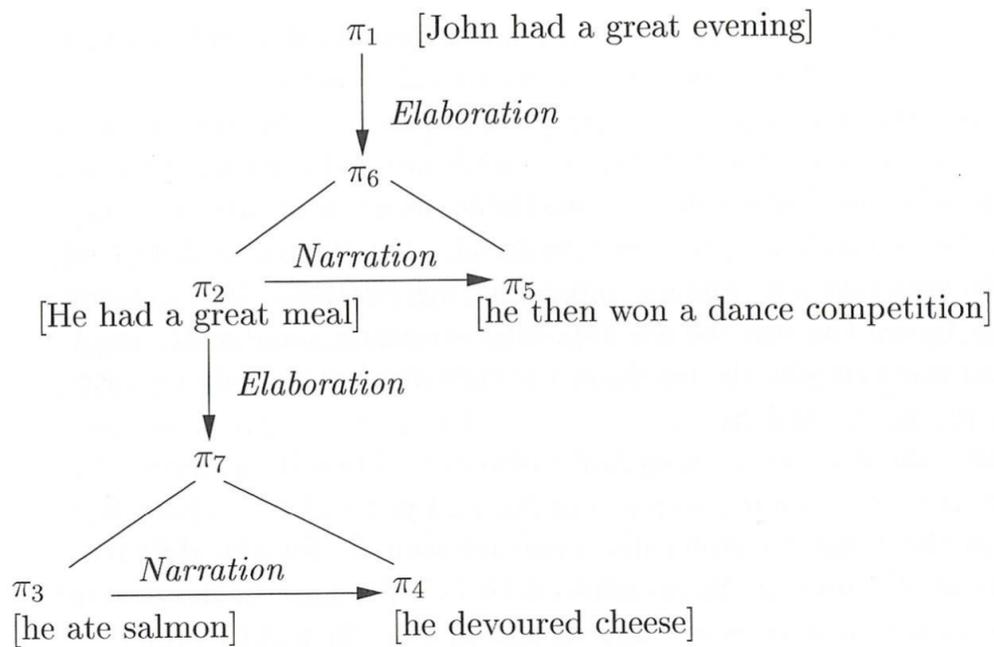
2.2 Marqueurs discursifs

Graphe de la RST (site web) avec RSTWeb

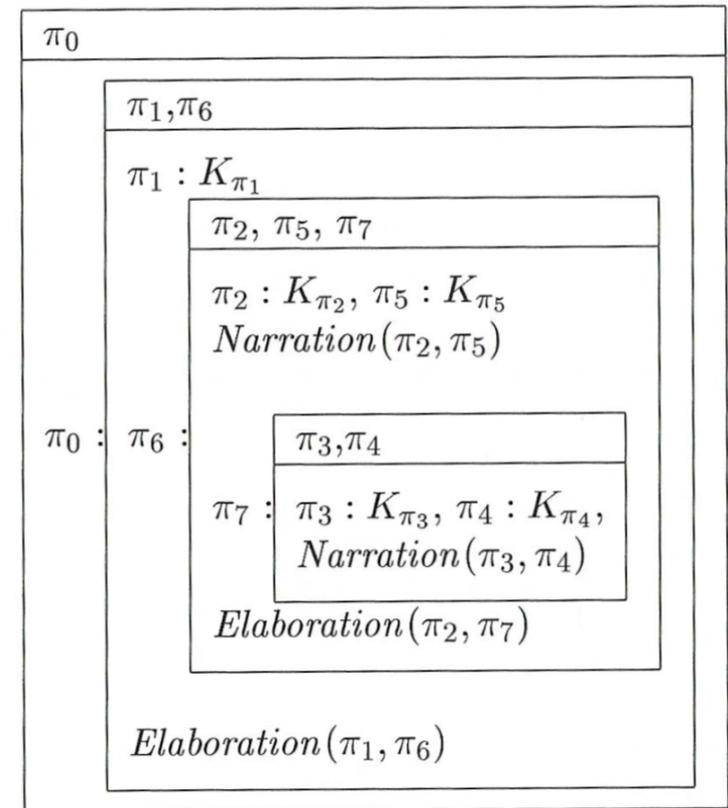
[Mann & Thompson 1988, Taboada & Mann 2006]

2.1 Argumentation

2.2 Marqueurs discursifs

Représentation en SDRT

Graphes



Boîtes

[Asher & Lascarides 2003, p. 104 et 107]

2.1 Argumentation

2.2 Marqueurs discursifs

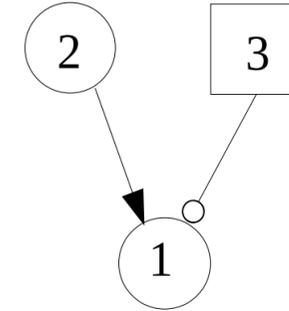
Schémas argumentatifs

[We should tear the building down.]₁

[It is full of asbestos.]₂

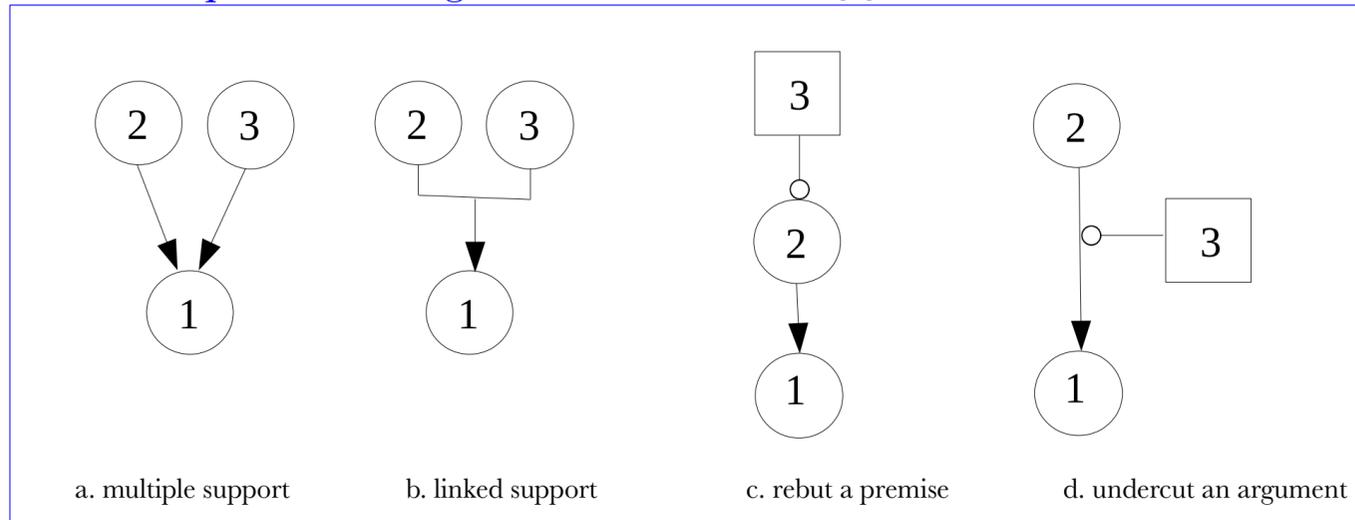
[On the other hand, many people liked the view from the roof.]₃

Il vaudrait mieux démolir le bâtiment. Il est bourré d'amiante. D'un autre côté, beaucoup de gens aimaient la vue depuis le toit.



n (circle) segment n , perspective du promoteur de la thèse —> relation Support

n (square) segment n , perspective de l'adversaire de la thèse —○ relation Attack

Exemples de configurations – **a,b** : Support – **c,d** : Attack

[dans Dargnat 2023, d'après Peldszus & Stede 2013]

- **À partir du projet *arg-microtexts multilayer***

[Peldszus & Stede 2015, Stede et al. 2016]

<https://github.com/peldszus/arg-microtexts-multilayer>

Analyses et représentations d'un ensemble de 112 textes argumentatifs courts :

- allemand et anglais
- trois types d'analyses et leurs représentations : RST, SDRT et ARG (+ sortie xml)
- manuel d'annotation [Stede, Taboada & Das 2017]

→ **Fouille de redescriptions** [Huber et al. 2019a, 2019b]

Objectif: comparer les structures en RST et en ARG au niveau de « motifs » (sous-graphes) et pas seulement des relations (\neq Peldszus & Stede 2015).

Ex. relever toutes les configurations discursives possibles (dans le corpus) des schémas de **Support**.

• Comparaisons de sous-graphes d'après arg-microtexts multilayer

[Huber et al. 2019a, 2019b]

L'Allemagne devrait-elle introduire la peine capitale ?

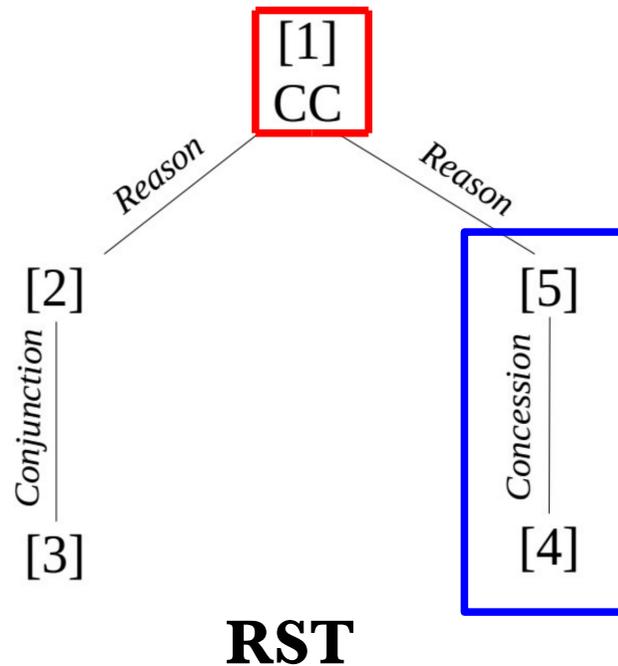
[1] La peine de mort est un moyen légal qui en tant que tel n'est pas applicable en Allemagne.

[2] D'une part, l'inviolabilité de la dignité humaine est un pilier de notre constitution

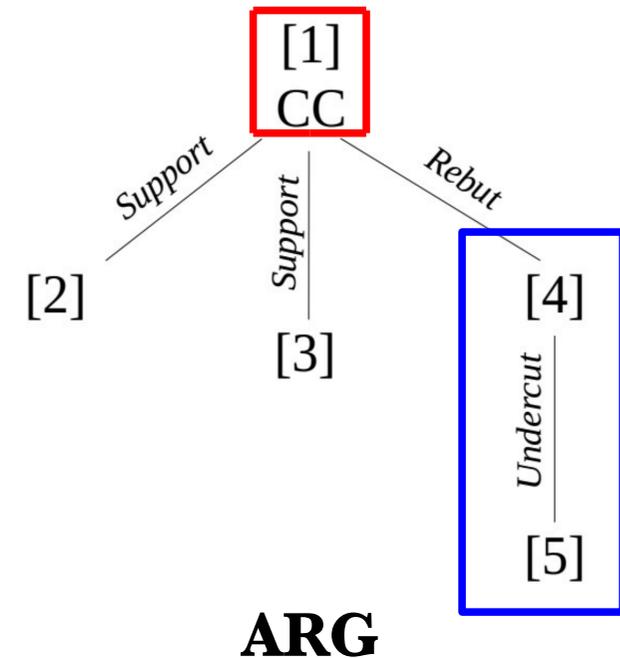
[3] et, d'autre part, personne ne peut avoir le droit de décider de la mort d'un autre humain.

[4] Même si beaucoup de gens estiment qu'un meurtrier a lui déjà décidé de la vie ou de la mort d'une autre personne,

[5] c'est précisément le crime qui ne devrait pas être payé du même prix. [trad. personnelle]



**Graphes « arrangés »
pour la comparaison**

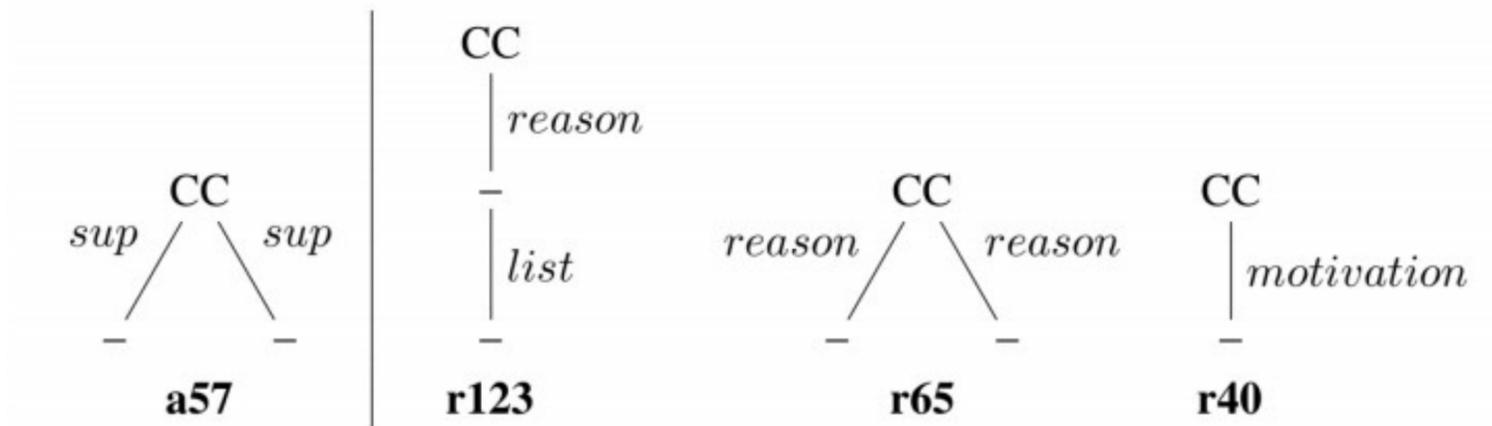


• ***Propriétés de la redescription 1 (qui concerne le + de textes)***

	q1	q2	J(q1,q2)	nbre de textes
Redescription 1	a57	r40 V r65 V r123	0.691	54

a_ = sous-graphe argumentatif (ARG) ; **r**_ = sous graphes discursifs (RST) ; J = indice de Jaccard

Sous-graphes correspondants à la redescription 1



sous-graphe ARG

sous-graphes RST

Bilan de 2.1 Linguistique et informatique, argumentation : inter et trans-disciplinarité ?

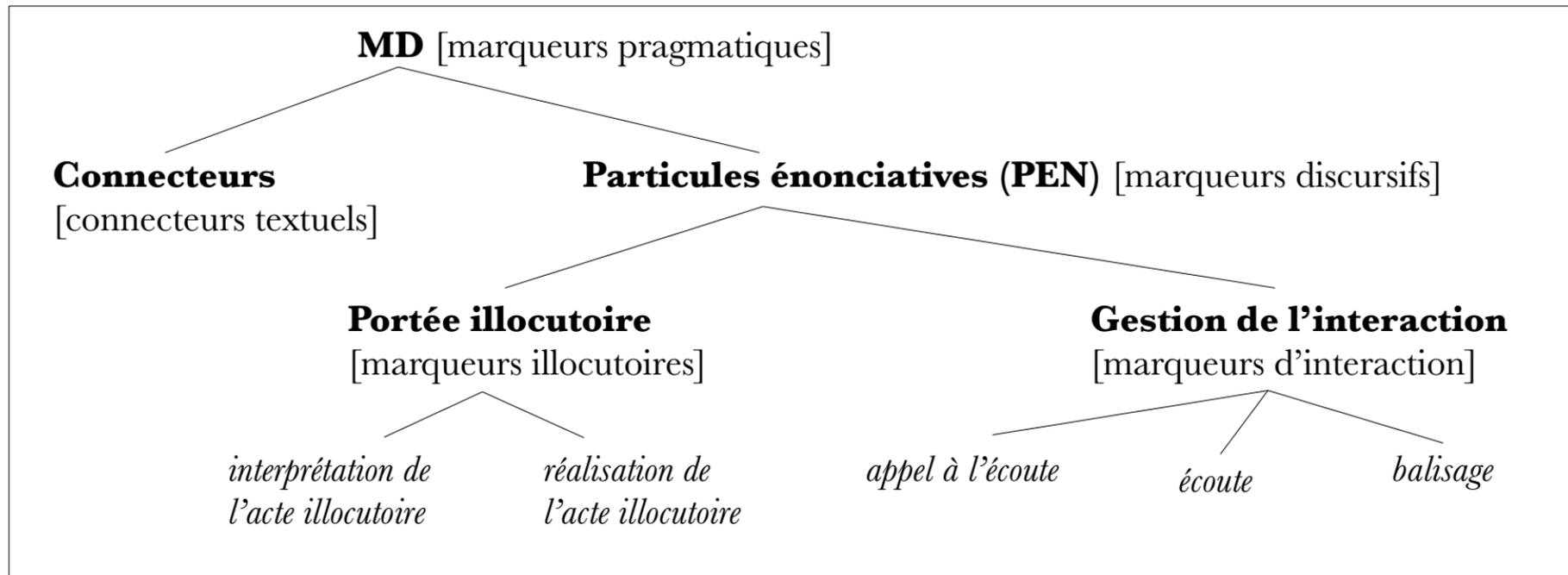
- Méthodologie spécifique qui pose la question du statut du TAL comme discipline
- Apport et biais de la représentation des données langagières (généralisation/limites)
- Données annotées, précision des catégories de relations, comparaison de perspectives
- Inconvénients : pas/peu de retour au texte/ressources linguistiques
- Démarche exploratoire, pas de résultats généralisables

2.2 Marqueurs discursifs (MD)

• Catégorisation linguistique

[Blakemore 2002, Dargnat 2021, 2023, à par., Dostie 2004, Dostie & Pusch 2007, Fraser 1999, Heine et al. 2021, Mosegaard Hansen 1998, Schiffrin 1987, etc., etc., etc.]

Projet [ANR CODIM](#) : Compositionality and Discourse Markers



Dargnat 2023, p. 120, d'après Dostie 2004, p. 46.

- Recensement et exploration des données pour le français

- *Uniformisation des données et des métadonnées*

[Sillaire 2023, PGD CODIM]

Oral

CFPP (700 000 mots)

TCOF (1 542 562 mots)

CRFP (440 000 mots)

DÉCLICS (207 442 mots)

FRA80 (200 000 mots)

DinG (11 h.)

ESLO 1 (4 500 000 mots)

ESLO 2 (1 796 818 mots)

Extraits de CLAPI (7h27 min.)

Écrit

Frantext contemporain (44 602 491 mots)

Le Monde (\geq 276 061 791 mots)

Le Monde Diplomatique (1 415 195 mots)

Wikipedia FR (178 900 000 mots au 20/02/23)

Scientext (4 800 000 mots)

- *Recherche de motifs* (*pattern matching*)
- [cf. CODIM, Boczon 2023, Hafed Dahou 2021]

Objectif: isoler les MD dans les données d’après les propriétés des non-MD [pour pouvoir ensuite étudier les combinaisons de MD]

- facilité pour certaines formes (*ah, mais, pourtant, donc, zut, etc.*)
- ambiguïté pour d’autres formes (*tiens, tu parles, bon, quoi, maintenant, après, si, etc.*)
- mal reconnus par les étiqueteurs habituels (*quand même, ça fait que, etc.*)

Suite logicielle : *Unitex-GramLab* [Paumier et al. 2021]

<https://unitexgramlab.org/fr>

Unitex-GramLab

- **Dictionnaires** modifiables
- Attribution de toutes les propriétés possibles (**pas de désambiguïsation**)
- **Automates finis** pour la reconnaissance de motifs dans des textes (sur la base des dictionnaires)
- **Transduction** (étiquetage dans les textes des motifs reconnus)

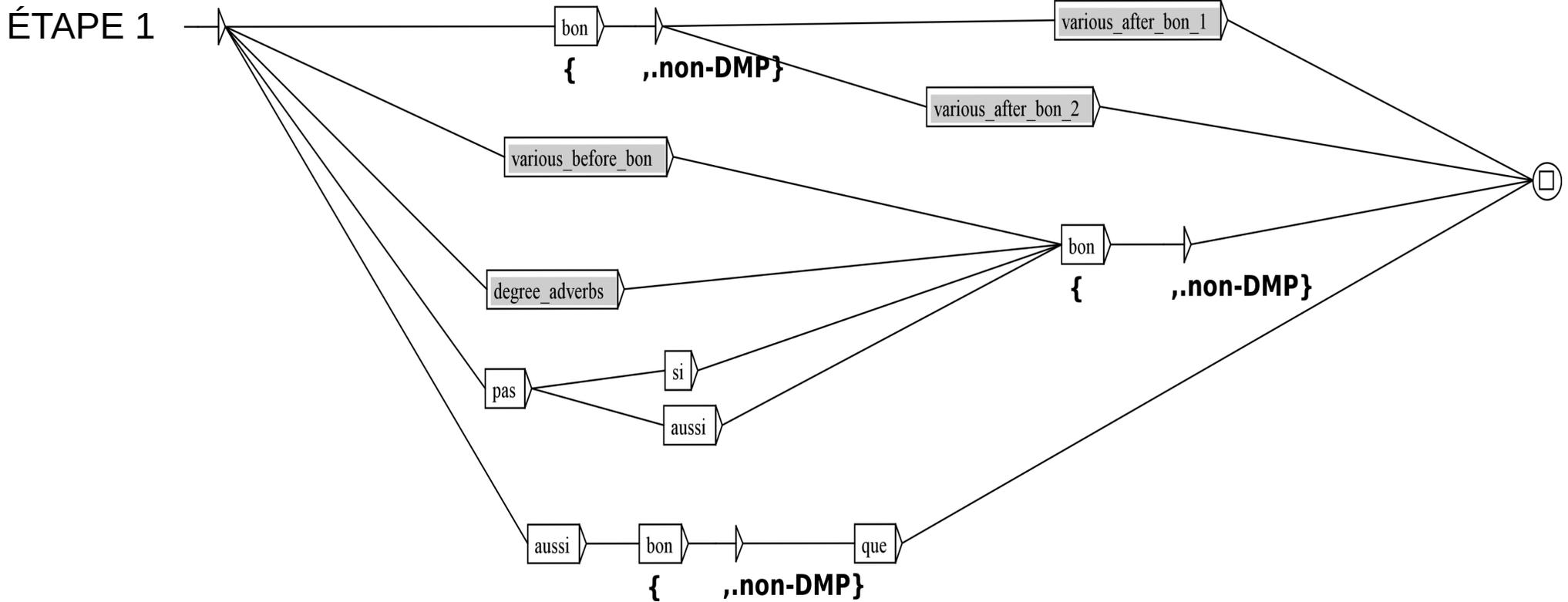
Méthode de type « **cascade** » en deux étapes :

- 1) on isole les non-DM (automate 1)
- 2) ensuite on récupère les items exclus (le « reste ») et on les étiquette MD (automate 2)

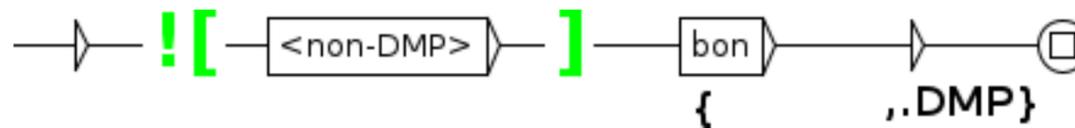
Ex. pour le marqueur *bon*

Unitex-GramLab : cascade pour récupérer et étiqueter bon MD

DMP = discourse marker particle ; DMC = discourse marker connective



ÉTAPE 2



Exemple de reconnaissance des MD dans un texte généré par IA

DMP = Discourse Marker Particle ; DMC = Discourse Marker Connective

A: salut ça va ? {S}
 B: ça va {bien,.DMP} {et,.DMC} toi ? {S}
 A: {oh,.DMP} {tu sais,.DMP} la routine {S}
 B: {bon,.DMP} on fait quoi ce soir ? {S}
 A: {eh,.DMP} {bien,.non-DMP} on pourrait aller au cinéma {tu vois,.DMP} {S}
 B: {ouais,.DMP} c'est une bonne idée {tu parles,.non-DMP} de quel film ? {S}
 A: {regarde,.DMP} il y a ce nouveau film d'action qui vient de sortir {S}
 B: {ah,.DMP} {ouais,.DMP} ? {S}
 A: {attends,.DMP} je pense l'avoir vu en ligne hier {S}
 B: vraiment ? {bon,.DMP} {alors,.DMC_or_DMP} on pourrait peut-être essayer ce nouveau restaurant {à la place,.DMC} {S}
 A: {ouais,.DMP} je l'ai remarqué aussi {tu sais,.DMP} s'ils ont de bonnes critiques ? {S}
 B: apparemment {oui,.DMP} {S}
 A: {tu sais,.DMP} j'ai lu des avis positifs {S}
 B: super allons-y {alors,.DMC_or_DMP} {et,.DMC} {bien,.DMP} sûr on prendra un dessert {S}
 A: {oh,.DMP} {attends,.DMP} j'ai oublié mon portefeuille à la maison {S}
 B: sérieux ? {bon,.DMP} je paierai pour cette fois-ci {mais,.DMC_or_DMP} tu me devras un dessert la prochaine fois {S}

Texte généré par l'IA et étiqueté avec une première version des automates finis déclarés dans Unitex-GramLab. [Boczon 2023, p. 13].

- Combinaisons de MD : compositionnalité ou pas ? (*mais enfin, non mais oh, bon ben alors, quand même etc.*)
[Dargnat 2021, 2022, Waltereit 2007, Cuenca & Crible 2019, etc.]

Exemple de *mais + enfin*

- Mais enfin, bon dieu, qui aurait pu croire un truc pareil ? [D. Pennac, 1995]
- Vous en doutez encore ? M'enfin ! Regardez autour de vous, voyez ? Voyez ? Il est là, partout ! [L. Lang, 2001]
- quand c'est large au départ c'est un peu plus long mais enfin ça va venir à mon avis ça va se refermer aux alentours de deux ans [TCOF, *Pediatrie_lam_08*]

– *Approche qualitative*

Valeurs isolées et contraintes combinatoires (notamment sémantiques, pragmatiques, syntaxiques).

[Bruxelles et al. 1976, Buchi & Städler 2010, Cadiot et al. 1985, Dargnat 2021, Hansen 2005, Némó 2000, Razgoulieva 2002, etc.]

– *Approche quantitative*

Sur la base de données pour lesquelles les éléments étudiés sont des MD (DMC ou DMP dans les automates) → nécessité d'un corpus « propre » et exploitable.

Objectif : mesure la force d'association (d'attraction) des deux items en jeu.

Mesures statistiques, dites « mesures d'association » [Brezina 2018, Desagulier 2017]

- critères de fréquence et d'exclusivité
- 16 mesures, mais focalisation sur mesures **MI3**, **G** et **DPF**
- Résultats avec *mais* déclaré comme pivot et enfin comme *associé* :

Attraction marquée entre *mais* et *enfin* dans les corpus oraux (CORPAIX, CRFP, DÉCLICS, Fra80), moins net pour ESTER et *Le Monde*.

Attraction directionnelle : *mais* attire *enfin*, mais pas l'inverse.

Commentaires résultats mesures (si besoin)

– **Mesure MI3**

Enfin arrive en tête dans CORPAIX et Fra80, il arrive au 2^{ème} rang dans CRFP, au 9^{ème} dans DÉCLICS et au 32^{ème} dans ESTER et dans *Le Monde*. Dans tous les cas, on voit que les valeurs de *enfin* sont dans les 25% supérieurs (la valeur de la mesure est supérieure au troisième quartile Q3).

– **Mesure G**

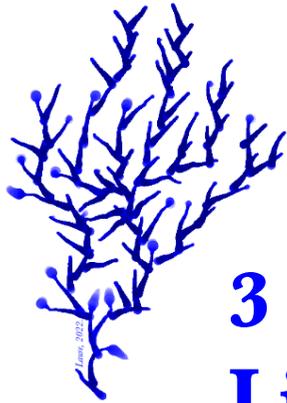
On obtient les rangs suivants pour *enfin* comme associé de *mais* : 1^{er} rang dans CORPAIX et Fra80, 2^{ème} dans CRFP, 14^{ème} dans DÉCLICS, 33^{ème} dans ESTER et 34^{ème} dans *Le Monde*. Les valeurs de *enfin* sont supérieures à Q3 pour DÉCLICS et *Le Monde*, et quasiment égales à Q3 pour ESTER.

– **Mesure DPF**

Parmi les associés de *mais*, *enfin* apparaît aux rangs suivants : 1^{er} dans Fra80, 2^{ème} dans CORPAIX et CRFP, 6^{ème} dans DÉCLICS, 15^{ème} dans ESTER et 18^{ème} dans *Le Monde*. Les valeurs de *enfin* sont toujours supérieures à Q3 sauf pour *Le Monde* où il est quasiment sur la médiane.

Bilan de 2.2 Linguistique et informatique, marqueurs discursifs : pluridisciplinarité ?

- Exploitation de données « massives » et uniformisées : outils (semi)automatiques
- Questionnement systématique sur les propriétés et la représentation de ces propriétés (inventaire des constructions) // catégorisation intuitive
- Articulation analyse qualitative et analyse quantitative
- Mise en relation de la force d'attraction et du degré de figement (grammaticalisation, pragmatification, lexicalisation)



3

Linguistique et sciences cognitives

Objectiver et calibrer le processus interprétatif

- Interfaces disciplinaires et domaines linguistiques
 - *Linguistique* (syntaxe, portée de la négation, items de polarité, incongruités sémantiques, métaphores, résolution d'anaphores, etc.)
 - *Neurosciences et IA* (intelligence artificielle)
 - *Psychologie*
 - *Philosophie (Éthique)*
 - *Informatique et statistiques* (interface et traitement des données)
- Matériel et protocole
 - Différents *protocoles* et *outils* (oculométrie, EEG, IRM fonctionnel, GripForce, etc.)
 - *Plateforme* expérimentale physique (coût) et personnel qualifié/formation aux outils
 - Mise en place d'un *protocole expérimental spécifique* (stimuli, sujets, etc.)
 - Dépôt de *CPP* (Comité de Protection des Personnes)
 - Passation des *expériences*
 - Analyse des *résultats* ... et publication 😊

- Recherches en cours [Dargnat 2023, p. 161-163, Sillaire (thèse UL/UQAM)]

Objectif : étude du traitement cognitif des relations de discours et connecteurs

Méthodologie : oculométrie (et si possible EEG), tâche de lecture à l'écran

Lieux : – ISC-Marc Jannerod à Bron (jusqu'en 2020),
– plateforme ECHO MSH-Lorraine à Nancy (depuis 2022)

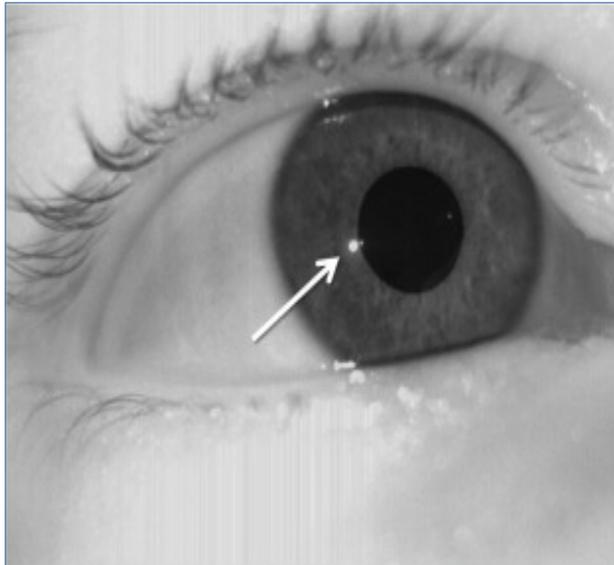
Outils pour l'oculométrie : Tobii ProSpectrum et Tobii ProLab

Expérience 1 : tester le coût cognitif des relations de *concession* et de *causalité*

Expérience 2 : tester la force des connecteurs (*pourtant* vs *quand même*)

– *Oculométrie* (eye-tracking)

→ Hypothèse d'un lien entre cognition et regard



Reflét cornéal de la lumière infrarouge

- **Calibrage** (*apprentissage + calcul de la corrélation entre la position de l'œil et la position de ce que le participant regarde*)
- **Lumière infrarouge**
- **Photo + déduction de la position de la pupille** (*cf. calibrage*) (*de 60/s à 2000/s*)

Difficultés interprétatives

= **fixations** sur la région problématique (un mot, une zone)

= **régressions**

Merci à M. Sillaire pour les éléments de la diapo.

Expérience 1 : tester le coût cognitif des relations de *concession* et de *causalité*

[Contexte de présentation]

Paul a raté/réussi son examen
pourtant/parce qu'il avait beaucoup travaillé

[Continuation en « et + phrase »]

Hypothèse

Le traitement cognitif de la concession est plus coûteux que le traitement cognitif de la causalité

Tâche : lecture

Motivation : fixations et/ou régressions sur la/les région(s) problématique(s)

- *Expérience 2* : Comparer les marques de concession *pourtant* et *quand même*

[Contexte de présentation]

Paul n'avait pas beaucoup de temps [P1],
mais la réunion était très importante [P2]
et/alors il est **quand même**/?**pourtant** venu [C]

[Continuation en « et + phrase »]

Hypothèse

pourtant est plus contraignant (moins naturel, plus fort) que *quand même*

Tâche : lecture

Motivation : fixations et/ou régressions sur la/les régions problématiques

Bilan de 3 : Linguistique et sciences cognitives

- Finesse des stimuli, gestion des contraintes → maîtrise des phénomènes linguistiques et des effets
- Explicitation d'hypothèses (simples) et de moyens techniques pour les tester
- Lourdeur de la passation des expériences (souvent besoin d'une expé. pilote, sujets, CPP, etc.)
- Analyse des résultats et généralisation (« granularité » des observations)

Conclusion : enjeux institutionnels... et contraintes associées

- Cotutelles et codirections

Thèses : Aix-Marseille Université & U. de Montréal ; Université de Lorraine & UQAM (avec G. Winterstein) ; Université de Lorraine & Université Paris-Cité (Paris 7) (avec J. Ginzburg).

M2 : TAL, SDL, orthophonie, MEEF Lettres

Collaborations : ATILF-LORIA, -IDEA , -LLF, -IDMC

- Cadres de réflexion et financements locaux et nationaux

3 contrats doctoraux (UL, LUE, OLKI)

MSH-Lorraine (achat commun Tobii ProLab/Spectrum)

ANR CODIM (UL, Nancy, resp. moi)

ANR PREFAB (Grenoble, resp. A. Tutin)

ANR PROTEXT (Paris 3, resp. G. Cislaru)

ANR sur le gérontolangage... on croise les doigts !

Colloque Metz-Nancy-Augsbourg-Pau, juin 2024. *Discourse Markers : Markers in Discourse, Markers on Discourse*. (I. Gaudy Campbell + moi, A. Fetzer, F. Berthe)

Etc.

Je remercie pour votre attention.

mathilde.dagnat@univ-lorraine.fr

<http://mathilde.dagnat.free.fr>



Références bibliographiques

- Asher N. & Lascarides A. 2003. *Logics of conversation*. Cambridge : Cambridge UP.
- Boczon K. 2023. *Text Mining for Discourse Markers*. M2 TAL Internship report. Université de Lorraine & LORIA. CODIM project. Dir. J. Jayez & Y. Parmentier.
- Brezina V. 2018. *Statistics in Corpus Linguistics : A Practical Guide*. Cambridge : Cambridge UP.
- Bruxelles et al. 1976. « Mais occupe-toi d'Amélie ». *Actes de la recherche en sciences sociales* 2/6, p. 47-62.
- Buchi E. & Städler T. 2010 « La pragmatization de l'adverbe *enfin* du point de vue des romanistes ». In *Actes du CMLF 2008*, p. 159-171.
- Cadiot et al. 1985. « *Enfin*, marqueur métalinguistique », *Journal of Pragmatics* 9, p. 199-239.
- Cuenca M. & Crible L. 2019. « Co-occurrence of discourse markers in English: from juxtaposition to composition », *Journal of Pragmatics* 140, p. 171-184.
- Crible L. & Degand L. 2019. « Domains and functions: A two-dimensional account of discourse markers », *Discours : revue de linguistique, psycholinguistique et informatique* 24, 35 p.
- Dargnat M. 2023. *Lexique et discours*. Habilitation à Diriger des Recherches. Université Paris 8, vol. I : 27 p., vol. II : 200 p., vol. III : 680 p. Garante : C. Beyssade .
- Dargnat M. 2006. *L'oral comme fiction*, thèse de doctorat de Sciences du langage et Ph.D. d'Études françaises (cotutelle), Aix-en-Provence et Montréal, Université de Provence (Aix-Marseille I), Laboratoire Parole & Langage et Université de Montréal, Département d'Études françaises, 2006, 2 volumes (616 p. + 170 p. d'annexes). Dir. M.-C. Hazaël-Massieux & L. Gauvin.
- Dargnat M. 2008. « La catégorisation de la variation linguistique », *Grenzgänge. Beiträge zu einer modernen Romanistik* 15/30, p. 11-41.
- Dargnat M. 2008. « Profils linguistiques et structure textuelle ». In Heiden S., Pincemin B. (éd.), *Actes des 9^e Journées d'Analyse Statistique des Données Textuelles (JADT)* vol. 1. Lyon : Presses Universitaires de Lyon, p. 369-379.
- Dargnat M. 2012. « Textual Stratification and functions of orality in theatre ». In Brumme J., Espunya A. (eds.), *The Translation of Fictive Dialogue*. Amsterdam : Rodopi, p. 81-99.
- Dargnat M. 2021. « *Mais enfin* : construction et association », *Langages* 225, p. 49-63.
- Dargnat M. & Jayez J. 2021. « *Mais* et associés », *Lexique* 29, p. 211-228.

- Dargnat M. 2022. « Explication et argumentation ». In Blasco M. (éd.), *Parler à l'hôpital : écouter ce qui est dit, décrypter ce qui se dit*. Münster : NODUS Publikationen, p. 88-204.
- Dargnat M. à par. « Les particules discursives ». Notice pour l'*Encyclopédie Grammaticale du Français*, en ligne, 42 p.
- Desagulier G. 2017. *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. Cham : Springer.
- Dostie G. 2004. *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. Liège : De Boeck/Duculot.
- Dostie G. & Pusch C. D. 2007. « Présentation. Les marqueurs discursifs. Sens et variation », *Langue française* 154/2, p. 3-12.
- Fraser B. 1999. « What are discourse markers ? ». *Journal of Pragmatics* 31, p. 931-952.
- Fraser B. 2009. « An account of discourse markers ». *International Review of Pragmatics* 1, p. 293–320.
- Hafed Dahou M. 2021. *Discourse Marker Identification in French Spoken Corpora : Using Rule-Based Method and Machine Learning*. Mémoire de stage de M2 TAL, IDMC Nancy. Dir. M. Dargnat & M. Constant.
- Freeman J. B. 2011. *Argument Structure: Representation and Theory*. Cham : Springer.
- Hansen M.-B. M. 1998a. *The Function of discourse particles*. Amsterdam-Philadelphia : John Benjamins.
- Hansen M.-B. M. 2008. *Particles at the Semantics/pragmatics Interface: Synchronic and Diachronic Issues, a Study with Special Reference to the French Phrasal Adverbs*, Elsevier.
- Heiden S. 2002. *Weblex, Manuel utilisateur, version 4.1*, Lyon, CNRS/ENS-LSH.
- Heine B., Kaltenböck G., Kuteva T. & Long H. 2021b. *The Rise of Discourse Markers*. Oxford : Oxford UP.
- Huber L., Toussaint Y., Roze C., Dargnat M. & Braud C. 2019a. « Alignement de structures discursives et argumentatives par fouille de graphes et de redescriptions ». In Couceiro M., Napoli A. (éd.), *Actes des XXVI^e rencontres de la Société Francophone de Classification (SFC)*, Nancy, p. 83-88.
- Huber L., Toussaint Y., Roze C., Dargnat M. & Braud C. 2019b. « Aligning Discourse and Argumentation Structures using Subtrees and Redescription Mining ». In Stein B., Wachsmuth H. (eds), *Proceedings of the 6th Workshop on Argument Mining*, ACL 2019, Florence, Association for Computational Linguistics, p. 35-40.
- Eckert P. 2012. « Three waves of variation study : The emergence of meaning in the study of sociolinguistic variation », *Annual Review of Anthropology* 41, 87-100.

- Mann W. & Thompson S. 1988. « Rhetorical Structure Theory : Toward a functional theory of text organization », *Text* 8(3), p. 243-281.
- Martineau et al. 2022. *Le Français au Québec et en Amérique du Nord*. Paris : Ophrys.
- Némo F. 2000. « *Enfin, encore, toujours* entre indexicalité et emplois ». *Actes du XXII^e Congrès international de linguistique et de philologie romanes* (Bruxelles, juillet 1998). Tübingen : Max Niemeyer Verlag, p. 499-511.
- Paumier S. et al. 2021. *UNITEX 3.3 Manuel d'utilisation*. <https://hal.science/hal-03589598/document>
- Peldszus A., Stede M. 2015. « An annotated corpus of argumentative microtexts ». *First European Conference on Argumentation: Argumentation and Reasoned Action*, Portugal, Lisbon, June 2015.
- Peldszus A. & Stede M. 2016. « Rhetorical structure and argumentation structure in monologue text ». *Proceedings of the 3rd Workshop on Argumentation Mining*, ACL, Berlin, August 2016, p. 103-112.
- Razgoulieva A. 2002. « Combinaison des connecteurs *mais enfin* », *Cahiers de Linguistique Française* 24, p. 143-168.
- Renkema J. 2009. *The texture of discourse*. Amsterdam-Philadelphia : John Benjamins.
- Schiffrin D. 1987. *Discourse Markers*. Cambridge : Cambridge University Press.
- Stede M., Afantenos S., Peldszus A., Asher N., Perret J. 2016. « Parallel Discourse Annotations on a Corpus of Short Texts ». *LREC 2016*, Slovenia, Portorož 2016.
- Stede M., Taboada M. & Das D. 2017. *Annotation Guidelines for Rhetorical Structure*. Version 1.0. <https://www.sfu.ca>
- Sillaire M. 2023. *Beyond Corpus Blobs*. M2 TAL Intership report. Université de Lorraine & LORIA. CODIM project. Dir. M. Dargnat & J. Jayez.
- Sillaire M. (en cours). *La concession du quotidien*. Thèse de doctorat en Sciences du langage, Université de Lorraine & ATILF. Sous la dir. de M. Dargnat & G. Winterstein.
- Taboada M. & Mann W. 2006. « Rhetorical Structure Theory: Looking Back and Moving Ahead ». *Discourse Studies* 8/3, p. 423-459.
- Tai Y. T. 2020. *A Corpus-Based Comparative Analysis of Argumentation and Rhetorical Structure Theory*. M2 TAL Intership report. Université de Lorraine & LORIA. Dir. M. Dargnat & Y. Toussaint.
- Waltereit R. 2007. « A propos de la genèse diachronique des combinaisons de marqueurs. L'exemple de *bon ben* et *enfin bref* ». *Langue française* 154, p. 94-109.
- Walton D. N. 2013. *Methods of Argumentation*. Cambridge : Cambridge University Press.

