

Typologie des chaînes de référence à la lumière de corpus annotés diversifiés

Sous la direction de Cécile Fabre et Lydia-Mai Ho-Dac

CLLE, CNRS & Université de Toulouse

Silvia Federzoni

Les Sciences des données et leur utilisation en sciences humaines et sociales



18 janvier 2024

Chaînes de référence (CR)

Suite d'expressions d'un texte entre lesquelles l'interprétation établit une identité de référence (Corblin, 1995)

Chaîne de
référence

Mme. Seton est-**elle** aussi connue chez-nous qu'**elle** le devrait être ?
Je ne le crois pas : et, m'aidant de **ses** historiens, je vais donner aux
lecteurs de la Revue Canadienne une biographie de **cette femme**,
l'une des plus accomplies, plus aimantes et des plus aimées qui
aient jamais foulé la terre. **Elle** a beaucoup écrit; et, autant que possible
je **la** laisserai parler **elle-même**.

Les chaînes de référence : définition

Chaînes de référence (CR)

Suite d'expressions d'un texte entre lesquelles l'interprétation établit une identité de référence (Corblin, 1995)

Chaîne de
référence

Mme. Seton est-**elle** aussi connue chez-nous qu'**elle** le devrait être ?
Je ne le crois pas : et, m'aidant de **ses** historiens, je vais donner aux
lecteurs de la Revue Canadienne une biographie de **cette femme**,
l'une des plus accomplies, plus aimantes et des plus aimées qui
aient jamais foulé la terre. **Elle** a beaucoup écrit; et, autant que possible
je **la** laisserai parler **elle-même**.

Référent : entité extra-linguistique dont on parle

Types de référent : être humain, entité abstraite ou concrète, événement

Chaîne de référence

Mme. Seton est-elle aussi connue chez-nous qu'**elle** le devrait être ?
Je ne le crois pas : et, m'aidant de **ses** historiens, je vais donner aux lecteurs de la Revue Canadienne une biographie de **cette femme**, **l'une des plus accomplies, plus aimantes et des plus aimées qui** aient jamais foulé la terre. **Elle** a beaucoup écrit; et, autant que possible je **la** laisserai parler **elle-même**.

Maillons : expressions linguistiques signalant le **référent** et porteuses d'une **valeur instructionnelle**

- leur succession contribue à la création de **liens de cohésion** (Halliday et Hasan, 1976)
- mécanisme fondamental dans l'**organisation** et l'**interprétation** du discours

Maillons : expressions linguistiques signalant le **référent** et porteuses d'une **valeur instructionnelle**

- leur succession contribue à la création de **liens de cohésion** (Halliday et Hasan, 1976)
- mécanisme fondamental dans l'**organisation** et l'**interprétation** du discours

- Plusieurs études :

- Principes qui régissent l'interprétation des expressions référentielles (Ariel, 2001 ; Grosz, Weinstein, & Joshi, 1995)
- Facteurs qui influencent le choix des expressions référentielles (Fossard et al., 2018 ; Vogels, Krahmer, & Maes, 2013)
- Description linguistique des fonctions des expressions référentielles, des variations liées au type de texte, au genre textuel, à la langue, à l'époque, au type de référent (Charolles, 2002 ; Cornish, 2000 ; Kleiber, 1994 ; Schnedecker, 2021)
- Résolution automatique des chaînes (De Marneffe, Recasens, & Potts, 2015 ; Mitkov, 2014 ; Recasens & Hovy, 2010)

- Plusieurs études :
 - Principes qui régissent l'interprétation des expressions référentielles (Ariel, 2001 ; Grosz et al., 1995)
 - Facteurs qui influencent le choix des expressions référentielles (Fossard et al., 2018 ; Vogels et al., 2013)
 - Description linguistique des fonctions des expressions référentielles, des variations liées au type de texte, au genre textuel, à la langue, à l'époque, au type de référent (Charolles, 2002 ; Cornish, 2000 ; Kleiber, 1994 ; Schnedecker, 2021)
 - Résolution automatique des chaînes (De Marneffe et al., 2015 ; Mitkov, 2014 ; Recasens & Hovy, 2010)

- Plusieurs études :
 - Principes qui régissent l'interprétation des expressions référentielles (Ariel, 2001 ; Grosz et al., 1995)
 - Facteurs qui influencent le choix des expressions référentielles (Fossard et al., 2018 ; Vogels et al., 2013)
 - Description linguistique des fonctions des expressions référentielles, des variations liées au type de texte, au genre textuel, à la langue, à l'époque, au type de référent (Charolles, 2002 ; Cornish, 2000 ; Kleiber, 1994 ; Schnedecker, 2021)
 - Résolution automatique des chaînes (De Marneffe et al., 2015 ; Mitkov, 2014 ; Recasens & Hovy, 2010)

- Plusieurs études :
 - Principes qui régissent l'interprétation des expressions référentielles (Ariel, 2001 ; Grosz et al., 1995)
 - Facteurs qui influencent le choix des expressions référentielles (Fossard et al., 2018 ; Vogels et al., 2013)
 - Description linguistique des fonctions des expressions référentielles, des variations liées au type de texte, au genre textuel, à la langue, à l'époque, au type de référent (Charolles, 2002 ; Cornish, 2000 ; Kleiber, 1994 ; Schnedecker, 2021)
 - Résolution automatique des chaînes (De Marneffe et al., 2015 ; Mitkov, 2014 ; Recasens & Hovy, 2010)

- Plusieurs études :
 - Principes qui régissent l'interprétation des expressions référentielles (Ariel, 2001 ; Grosz et al., 1995)
 - Facteurs qui influencent le choix des expressions référentielles (Fossard et al., 2018 ; Vogels et al., 2013)
 - Description linguistique des fonctions des expressions référentielles, des variations liées au type de texte, au genre textuel, à la langue, à l'époque, au type de référent (Charolles, 2002 ; Cornish, 2000 ; Kleiber, 1994 ; Schnedecker, 2021)
 - Résolution automatique des chaînes (De Marneffe et al., 2015 ; Mitkov, 2014 ; Recasens & Hovy, 2010)

① Critères de description

- Nature du maillon : SN_def, SN_dem, pronom, noms propres
- Fonction syntaxique : sujet, objet direct, objet indirect, autre
- Longueur de la chaîne en nb de maillons
- ...

② Facteurs de variation :

- Type de texte
- Genre textuel
- Type de référent : humain - non humain, concret - abstrait, animé - inanimé
- ...

⇒ Pas d'étude à large échelle pour le français

-
- Manque d'une méthode commune d'analyse

⇒ Pas d'étude à large échelle pour le français

⇒ Notion de CR abordée sous des angles différents avec des méthodes d'analyse différentes pour répondre à des objectifs différents

- Diversité en termes de choix d'annotation : hétérogénéité des corpus annotés
- Manque d'une méthode commune d'analyse

⇒ Pas d'étude à large échelle pour le français

⇒ Notion de CR abordée sous des angles différents avec des méthodes d'analyse différentes pour répondre à des objectifs différents

- **Diversité en termes de choix d'annotation : hétérogénéité des corpus annotés**

Peut-on tirer partie de cette hétérogénéité pour appréhender différents aspects des chaînes ?

- Manque d'une méthode commune d'analyse

Objectifs différents : annotations différentes

MUC (MUC consortium 1995c)

AnCora (Taulé et al., 2008)

ACE (Doddington et al., 2004)

Différents modèles d'annotation :

- ① types de référent
- ② critères de délimitation
- ③ annotation des singletons
- ④ informations annotées

AnnoDis (Péry-Woodley et al., 2011)

Ontonotes (Pradhan et al., 2012)

AnCor (Muzurelle et al., 2014)

ARRAU (Poesio et al., 2013)

Différents formats :

- ① SGML (in-line/stand-off)
- ② MMAX2
- ③ Glozz
- ④ XML-TEI-URS

Democrat (Landragin et al., 2019)

WikiCoref (Ghaddar & Langlais, 2016)

E-Calm (Garcia-Debanc et al., 2017)

Objectifs différents : annotations différentes

MUC (MUC consortium 1995c)

AnCora (Taulé et al., 2008)

ACE (Doddingtong et al., 2004)

Différents modèles d'annotation :

- ① types de référent
- ② critères de délimitation
- ③ annotation des singletons
- ④ informations annotées

AnnoDis (Péry-Woodley et al., 2011)

Ontonotes (Pradhan et al., 2012)

AnCor (Muzurelle et al., 2014)

ARRAU (Poesio et al., 2013)

Différents formats :

- ① SGML (in-line/stand-off)
- ② MMAX2
- ③ Glozz
- ④ XML-TEI-URS

Democrat (Landragin et al., 2019)

WikiCoref (Ghaddar & Langlais, 2016)

E-Calm (Garcia-Debanc et al., 2017)

- AnnoDis (2011) : textes longs entièrement annotés, non narratifs, 3 genres textuels (rapports et articles scientifiques, textes encyclopédiques)
 - Sonder l'intuition des locuteurs
 - Annotateurs naïfs

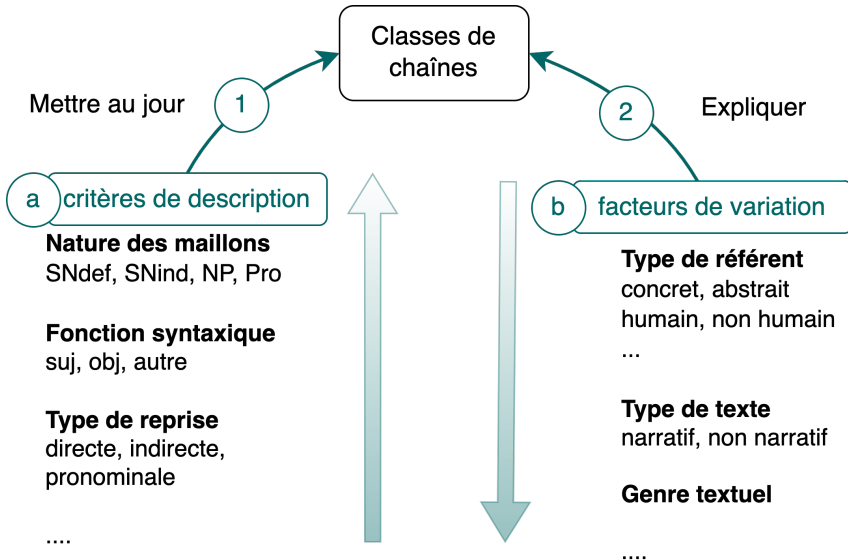
- AnnoDis (2011) : textes longs entièrement annotés, non narratifs, 3 genres textuels (rapports et articles scientifiques, textes encyclopédiques)
 - Sonder l'intuition des locuteurs
 - Annotateurs naïfs
- ANCOR (2014) : trois corpus de parole spontanée transcrite (ESLO, OTG et Accueil_UBS)
 - Annoter les relations de coréférence ou anaphores
 - Annotateurs experts

- AnnoDis (2011) : textes longs entièrement annotés, non narratifs, 3 genres textuels (rapports et articles scientifiques, textes encyclopédiques)
 - Sonder l'intuition des locuteurs
 - Annotateurs naïfs
- ANCOR (2014) : trois corpus de parole spontanée transcrite (ESLO, OTG et Accueil_UBS)
 - Annoter les relations de coréférence ou anaphores
 - Annotateurs experts
- DEMOCRAT (2019) : échantillons des textes (10 000 mots), narratifs (en prose et en vers) et non narratifs, 18 genres textuels, du 12^{ème} au 21^{ème} siècle
 - Annoter toutes les expressions référentielles
 - Annotateurs experts

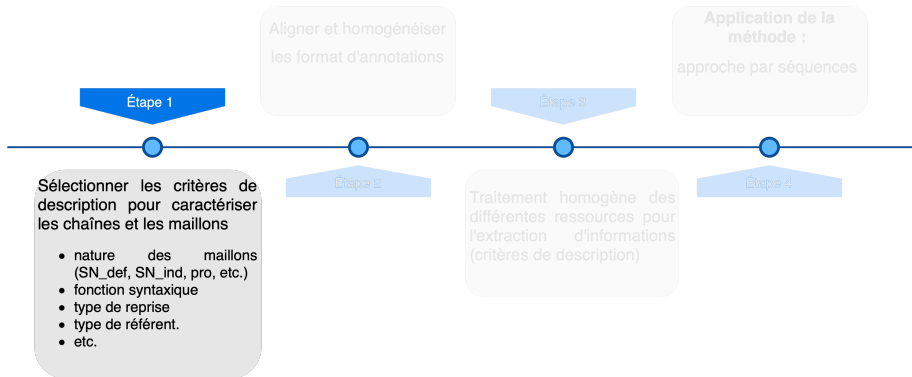
- AnnoDis (2011) : textes longs entièrement annotés, non narratifs, 3 genres textuels (rapports et articles scientifiques, textes encyclopédiques)
 - Sonder l'intuition des locuteurs
 - Annotateurs naïfs
- ANCOR (2014) : trois corpus de parole spontanée transcrite (ESLO, OTG et Accueil_UBS)
 - Annoter les relations de coréférence ou anaphores
 - Annotateurs experts
- DEMOCRAT (2019) : échantillons des textes (10 000 mots), narratifs (en prose et en vers) et non narratifs, 18 genres textuels, du 12^{ème} au 21^{ème} siècle
 - Annoter toutes les expressions référentielles
 - Annotateurs experts
- RésolCo (E-CALM, 2019-2021) : écrits scolaires à différents niveaux de littératie (du primaire à l'Université)
 - Annoter les référents humains principaux introduit dans la consigne
 - Annotateurs experts

Modèles d'annotation

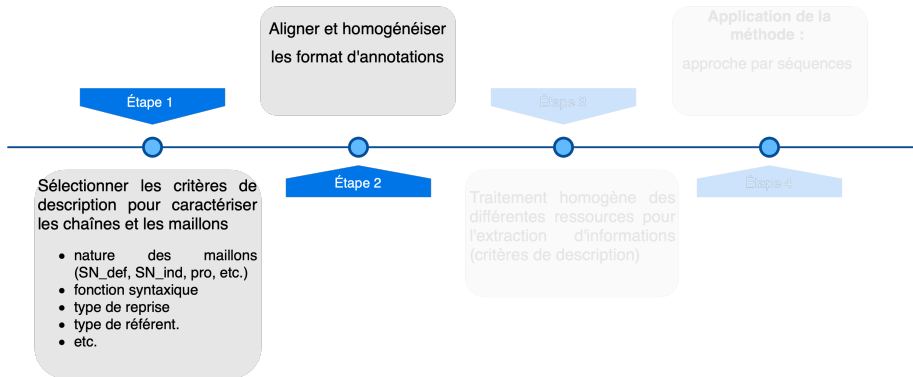
Annotation	AnnoDis	DEMOCRAT	ANCOR	E-Calm
Type de référent	Tout	Pas temporel	Pas spatio-temporel	Humain
Délimitation des maillons	Non précisé	Précisé	Précisé	Précisé
Singletons	X	✓	✓	X
Propriétés annotées				
Nature	✓	X	✓	X
Genre et nombre	X	X	✓	X
Fonction syntaxique	X	X	X	X
Type de reprise	X	X	✓	X
Entité nommée	X	X	✓	X
Type d'emploi	X	X	✓	X



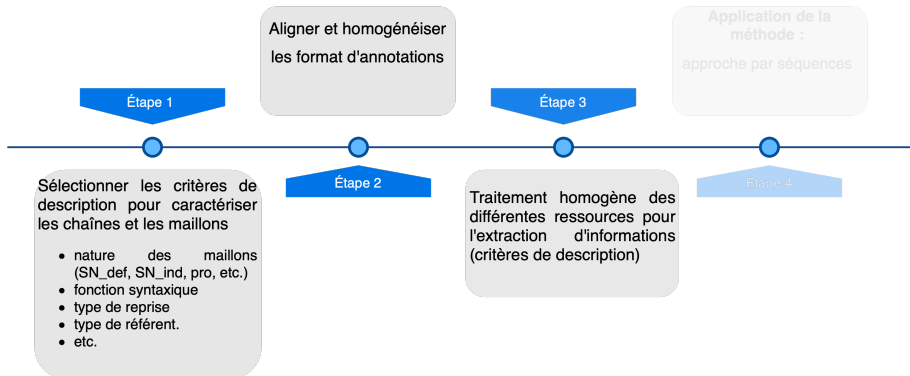
Utiliser des corpus hétérogènes



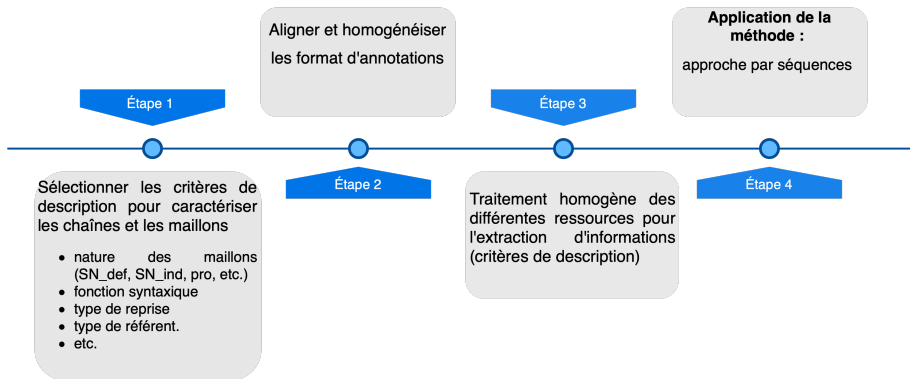
Utiliser des corpus hétérogènes



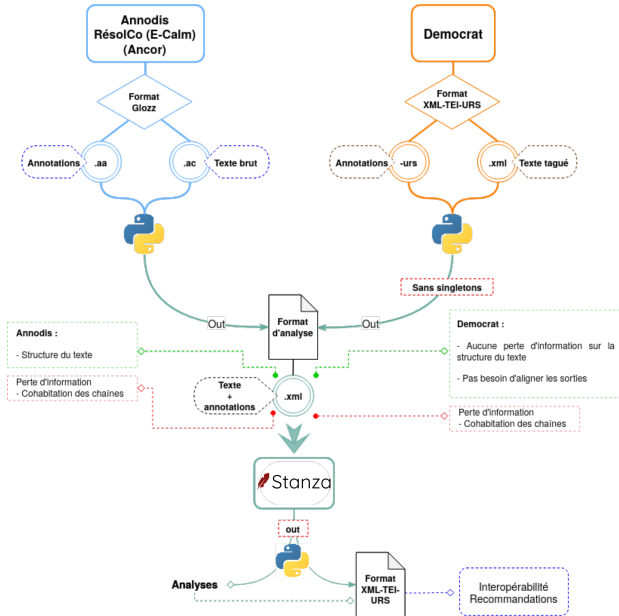
Utiliser des corpus hétérogènes



Utiliser des corpus hétérogènes



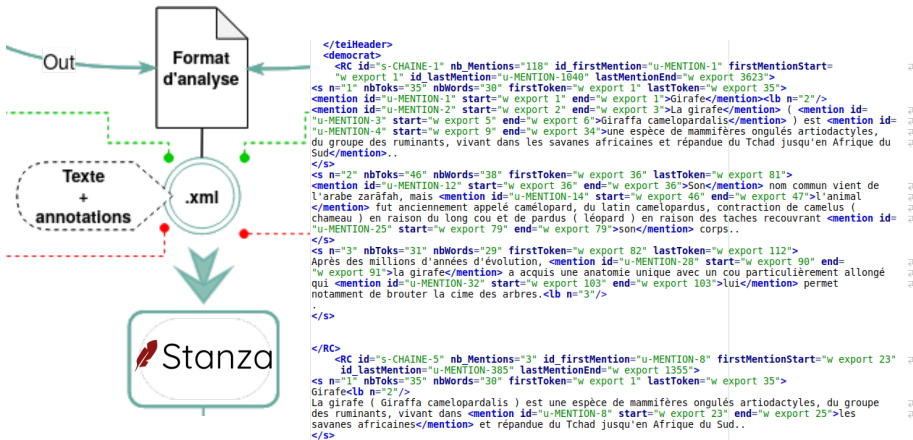
Aligner les formats d'annotation



Aligner les formats d'annotation

Besoin : ne pas perdre d'informations et effectuer des traitements homogènes

Proposition : format intermédiaire, interoperabilité



Informations morpho-syntaxiques

AnnoDis \Rightarrow Homogénéisation des annotations (Federzoni, Ho-Dac, & Rebeyrolle, 2020)

- Exploration préliminaire et correction des erreurs de délimitation
Ex. : Le [pont des embarcations] / [Le père de **Léonard**]
- Homogénéisation des annotations : nature des maillons

DEMOCRAT :

- Prise en compte des textes du 19^{ème}, 20^{ème} et 21^{ème} siècle
- Sélection des chaînes de référence au moins 3 maillons : singletons et CR 2 maillons pas reproduits
- Correction de la segmentation en phrases
- Stanza : tokeniser, lemmatiser, analyse morpho-syntaxique, dépendances syntaxiques, entités nommées \Rightarrow 4 : PER, LOC, ORG, MISC

Ressource disponible ici : https://hdl.handle.net/11403/democrat_rc_19-21

⇒ Pas d'étude à large échelle pour le français

⇒ Notion de CR abordée sous des angles différents avec des méthodes d'analyse différentes pour répondre à des objectifs différents

- Diversité en termes de choix d'annotation : hétérogénéité des corpus annotés

Peut-on tirer partie de cette hétérogénéité pour appréhender différents aspects des chaînes ?

- **Manque d'une méthode commune d'analyse**

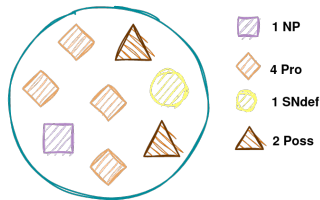
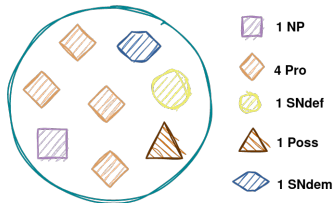
Description des caractéristiques globales des chaînes : longueur en nombre de maillon, distribution de la nature des maillons (noms propres, pronoms, etc.)

⇒ Méthode d'analyse :

- Études existantes : description des chaînes en tant qu'*ensemble* de maillons

Mme Seton est elle aussi connue chez nous qu'elle le devrait être?? Je ne le crois pas : et, m'aidant de ses historiens, je vais donner aux lecteurs de la Revue Canadienne une biographie de cette femme, l'une des plus accomplies, des plus aimantes et des plus aimées qui aient jamais foulé la terre. Elle a beaucoup écrit ; [...]

Les pauvres gens vivaient péniblement des petits appointements du mari. Hector de Gribelin avait été élevé en province, dans le manoir paternel, par un vieil abbé précepteur. Puis, à vingt ans, [...] il était entré, commis à quinze cents francs, au ministère de la Marine. Il avait échoué sur cet écueil comme tous ceux qui ne sont point préparés de bonne heure au rude combat de la vie, [...] Ses trois premières années de bureau furent horribles. Il avait retrouvé quelques amis de sa famille, [...] ; et il s'était fait un cercle de connaissances.

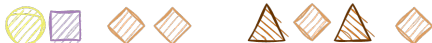


⇒ Méthode d'analyse :

● Mon approche : Description des *enchaînements* de maillons

Mme Seton est elle aussi connue chez nous qu'elle le devrait être?? Je ne le crois pas : et, m'aidant de ses historiens, je vais donner aux lecteurs de la Revue Canadienne une biographie de cette femme, l'une des plus accomplies, des plus aimantes et des plus aimées qui aient jamais foulé la terre. Elle a beaucoup écrit ; [...]

Les pauvres gens vivaient péniblement des petits appointements du mari Hector de Gribelin avait été élevé en province, dans le manoir paternel, par un vieil abbé précepteur. Puis, à vingt ans, [...] il était entré, commis à quinze cents francs, au ministère de la Marine. Il avait échoué sur cet écueil comme tous ceux qui ne sont point préparés de bonne heure au rude combat de la vie, [...] Ses trois premières années de bureau furent horribles. Il avait retrouvé quelques amis de sa famille, [...] ; et il s'était fait un cercle de connaissances.



Analyse des séquences

Identifier les régularités, les ressemblances, construire des typologies de « séquences-types » (Robette, 2011)

- Utilisée en sciences sociales pour étudier les parcours de vie des individus ou pour étudier les séquences de ADN

Analyse des séquences

Identifier les régularités, les ressemblances, construire des typologies de « séquences-types » (Robette, 2011) \Rightarrow Stratégies référentielles

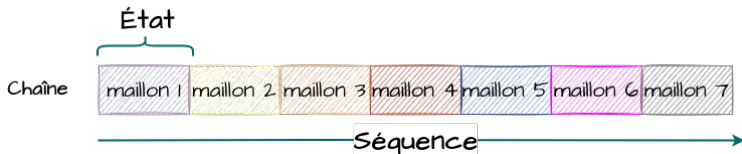
- Utilisée en sciences sociales pour étudier les parcours de vie des individus ou pour étudier les séquences de ADN

Analyse des séquences

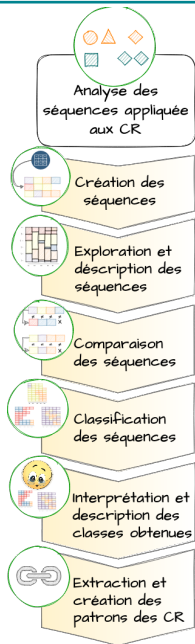
Identifier les régularités, les ressemblances, construire des typologies de « séquences-types » (Robette, 2011) ⇒ Stratégies référentielles

- Utilisée en sciences sociales pour étudier les parcours de vie des individus ou pour étudier les séquences de ADN

Séquence : liste ordonnée d'**états** ou d'événements (Brzinsky-Fay et al., 2006)



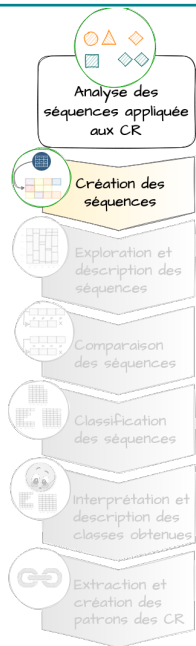
Analyse des séquences appliquée aux chaînes



TraMineR (Gabadinho, Ritschard, Studer, & Müller, 2009) :

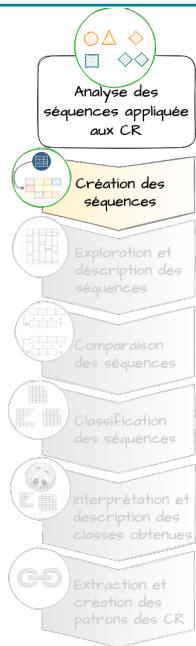
- Visualisation spécifique pour l'analyse des séquences
- Méthodes de classification automatique
- Méthodes statistiques

Création des séquences



- CR une séquence d'états associés à ses critères de description

Création des séquences



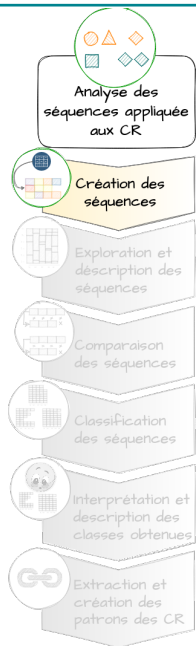
- CR une séquence d'états associés à ses critères de description

Mme Seton est-elle aussi connue chez nous qu'elle le devrait être?

Je ne le crois pas : et, m'aidant de ses historiens, je vais donner aux lecteurs de la Revue Canadienne une biographie de cette femme, l'une des plus accomplies, des plus aimantes et des plus aimées qui aient jamais foulé la terre. Elle a beaucoup écrit ; [...]

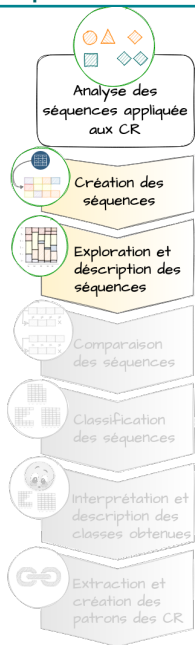


Création des séquences

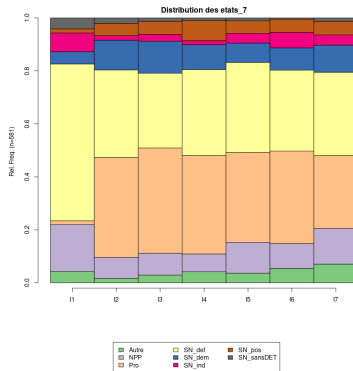


- CR une séquence d'états associés à ses critères de description

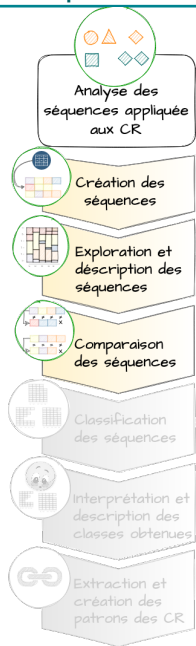
Exploration et description des séquences



- Fréquence
- Distribution des différents états

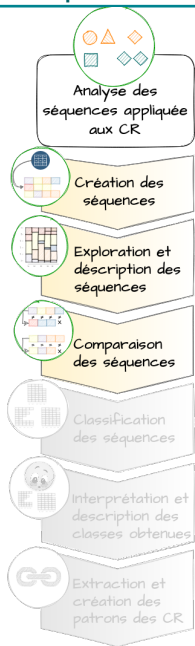


Comparaison des séquences



- Mesure de similarité entre les séquences :
 - Optimal Matching (basée sur la distance de Levenshtein) : coût minimum pour obtenir deux séquences identiques
 - Calculé au moyen d'opération de transformation (insertion, suppression ou substitution)

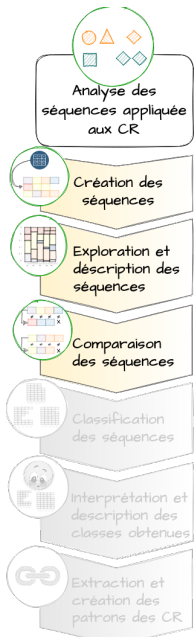
Comparaison des séquences



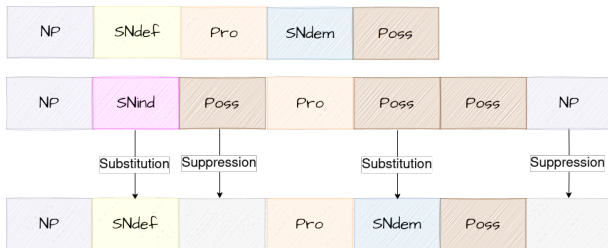
- Mesure de similarité entre les séquences :
 - Optimal Matching (basée sur la distance de Levenshtein) : coût minimum pour obtenir deux séquences identiques
 - Calculé au moyen d'opération de transformation (insertion, suppression ou substitution)



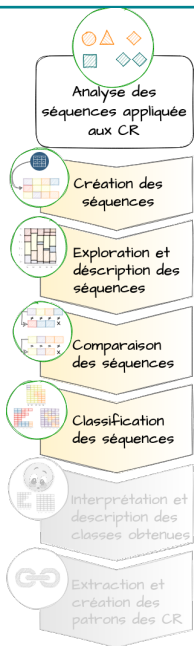
Comparaison des séquences



- Mesure de similarité entre les séquences :
 - Optimal Matching (basée sur la distance de Levenshtein) : coût minimum pour obtenir deux séquences identiques
 - Calculé au moyen d'opération de transformation (insertion, suppression ou substitution)



Classification des séquences



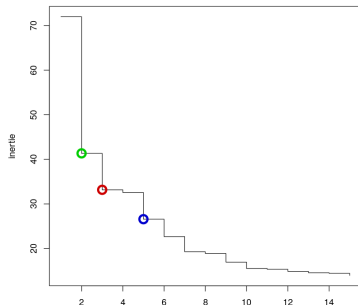
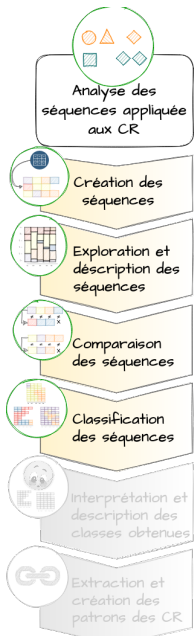
- Clustering :
 - Non supervisé
 - Regrouper les objets similaires en classes

Method name	Parameters	Scalability	Usecase	Geometry (metric used)
K-Means	number of clusters	Very large <code>n_samples</code> , medium <code>n_clusters</code> with <code>MiniBatch</code> code	General-purpose, even cluster size, flat geometry, not too many clusters	Distances between points
Affinity propagation	damping, sample preference	Not scalable with <code>n_samples</code>	Many clusters, uneven cluster size, non-flat geometry	Graph distance (e.g. nearest-neighbor graph)
Mean-shift	bandwidth	Not scalable with <code>n_samples</code>	Many clusters, uneven cluster size, non-flat geometry	Distances between points
Spectral clustering	number of clusters	Medium <code>n_samples</code> , small <code>n_clusters</code>	Few clusters, even cluster size, non-flat geometry	Graph distance (e.g. nearest-neighbor graph)
Ward hierarchical clustering	number of clusters or distance threshold	Large <code>n_samples</code> and <code>n_clusters</code>	Many clusters, possibly connectivity constraints	Distances between points
Agglomerative clustering	number of clusters or distance threshold, linkage type, distance	Large <code>n_samples</code> and <code>n_clusters</code>	Many clusters, possibly connectivity constraints, non Euclidean distances	Any pairwise distance
DBSCAN	neighborhood size	Very large <code>n_samples</code> , medium <code>n_clusters</code>	Non-flat geometry, uneven cluster sizes	Distances between nearest points
OPTICS	minimum cluster membership	Very large <code>n_samples</code> , large <code>n_clusters</code>	Non-flat geometry, uneven cluster sizes, variable cluster density	Distances between points
Gaussian mixtures	many	Not scalable	Flat geometry, good for density estimation	Mahalanobis distances to centers
Birch	branching factor, threshold, optional global clusterer.	Large <code>n_clusters</code> and <code>n_samples</code>	Large dataset, outlier removal, data reduction.	Euclidean distance between points

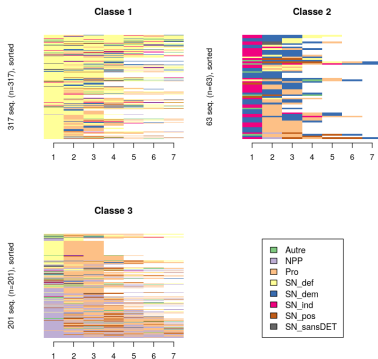
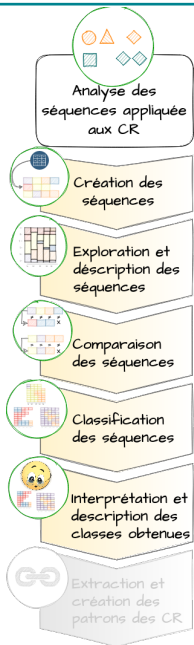
Tazbleau tiré du site <https://scikit-learn.org/stable/modules/clustering.html>

Classification des séquences

- Clustering hiérarchique : paramètres retenus
 - Mesure de similarité entre les séquences : OM
 - Méthode d'agglomération
 - Ward
 - Nombre de clusters : 3 et 5 (dendrogramme et sauts d'inertie)

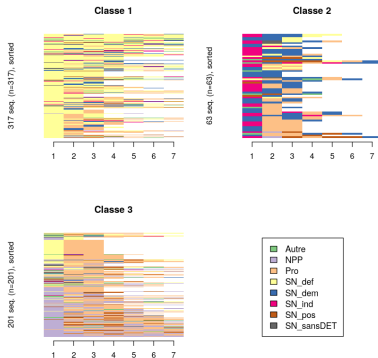
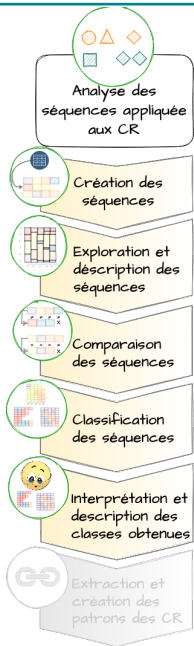


Observation et description des clusters obtenus



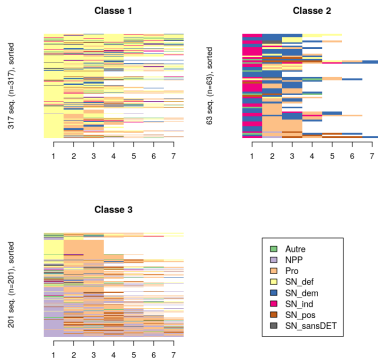
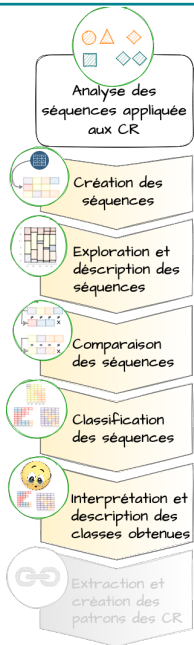
- aperçu global de la composition des chaînes
- chaque couleur correspond à une caractéristique distincte des maillons
- les séquences sont triées par leur similarité

Observation et description des clusters obtenus



- aperçu global de la composition des chaînes
- chaque couleur correspond à une caractéristique distincte des maillons
- les séquences sont triées par leur similarité

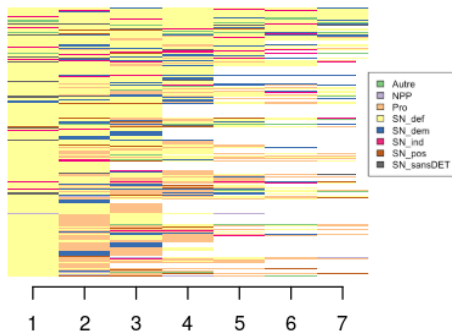
Observation et description des clusters obtenus



- aperçu global de la composition des chaînes
- chaque couleur correspond à une caractéristique distincte des maillons
- les séquences sont triées par leur similarité

Observation des clusters obtenus : résultats

Classe TraMineR 1

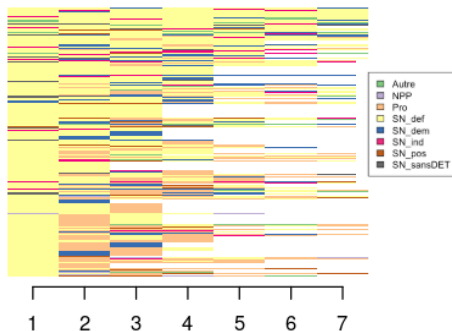


- 371 séquences :
 - 89.6% commence avec un SN_def
 - suivis d'autre SN_def

le vignoble champenois s'étendait sur quelques [...] le vignoble connaît [...] et de la Grande guerre, le vignoble s'est réduit à 12 000 hectares. Aujourd'hui, en 2007, le vignoble champenois s'étend sur 32 341 hectares. (encyclopedic text)

Observation des clusters obtenus : résultats

Classe TraMineR 1



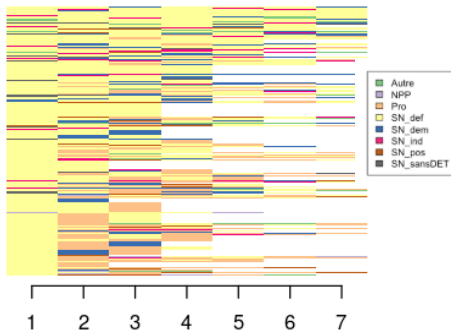
- Stratégies :

- Introduction et maintien des référents non humains
- Introduction et maintien des référents humains

le vignoble champenois s'étendait sur quelques [...] le vignoble connaît [...] et de la Grande guerre, le vignoble s'est réduit à 12 000 hectares. Aujourd'hui, en 2007, le vignoble champenois s'étend sur 32 341 hectares. (encyclopedic text)

Observation des clusters obtenus : résultats

Classe TraMineR 1



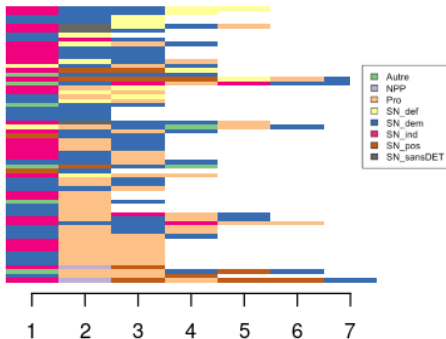
- Stratégies :

- Introduction et maintien des référents non humains
- Introduction et maintien des référents humains

L'Agence de Protection de l'Environnement était en charge [...] Plusieurs mois après, *l'EPA* enregistrait [...] Le 21 août 2003 cependant, *l'Agence* [...] où *l'Agence* se voit reprocher de [...] *L'EPA* envisagea [...]. (encyclopedic text)

Observation des clusters obtenus : résultats

Classe TraMineR 2

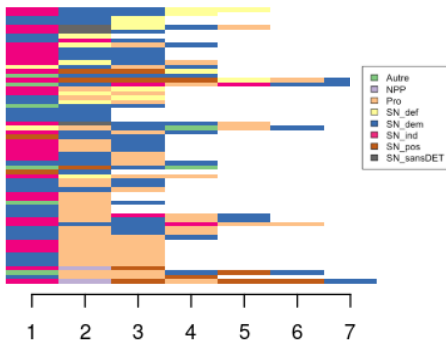


- 63 séquences :
 - 47.6% commence avec un SN_ind
 - 36.5% commence avec un SN_dem
 - suivi par d'autres SN_dem ou pronoms

[...] « le " **communicatif** " présente ainsi fréquemment **une connotation oppositionnelle** (sinon contradictoire) avec le linguistique. **Ceci** est particulièrement crucial lorsqu'on traite de la " **compétence de communication** " [...] **Cette dichotomie** pose problème au psychologue [...]. (scientific article)

Observation des clusters obtenus : résultats

Classe TraMineR 2

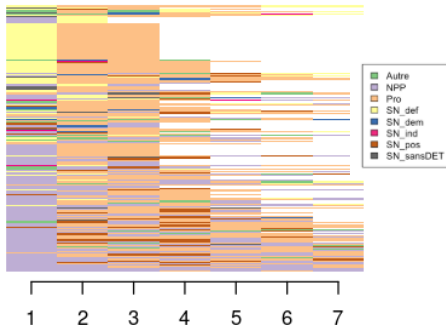


- Stratégies :
 - Référer à des idées ou concepts
 - Encapsuler des portions de texte

[...] « le " communicatif " présente ainsi fréquemment une connotation oppositionnelle (sinon contradictoire) avec le linguistique. Ceci est particulièrement crucial lorsqu'on traite de la " compétence de communication " [...] Cette dichotomie pose problème au psychologue [...]. (scientific article)

Observation des clusters obtenus : résultats

Classe TraMineR 3

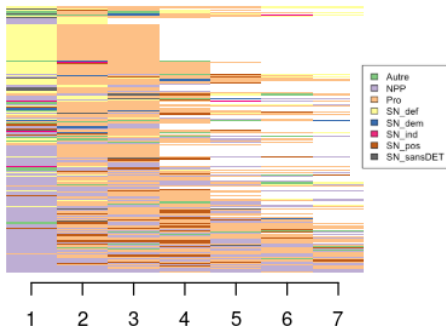


- 201 séquences :
 - 50.7% commence avec un NP ■
 - 28.8% commence avec un SN_def ■
 - suivi d'autres NP ■ ou pronoms ■

Chez F. de Saussure, l'analogie [...], il pose que les facteurs de trouble [...]. Pour lui, cette tendance à l'irrégularité est heureusement contrebalancée par l'analogie [...]. Comme H. Paul, il ramène le concept au calcul de l'équation de la quatrième proportionnelle. (scientific article)

Observation des clusters obtenus : résultats

Classe TraMineR 3

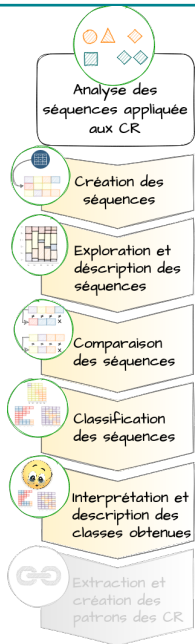


- Stratégies :

- Introduire et maintenir des référents humains

Chez **F. de Saussure** , l'analogie [...], **il** pose que les facteurs de trouble [...]. Pour **lui** , cette tendance à l'irrégularité est heureusement contrebalancée par l'analogie [...]. Comme H. Paul, **il** ramène le concept au calcul de l'équation de la quatrième proportionnelle. (scientific article)

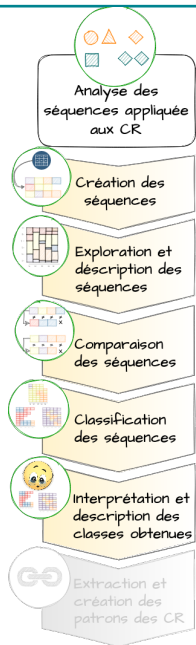
Observation et description des clusters obtenus



Prendre en compte les facteurs de variation

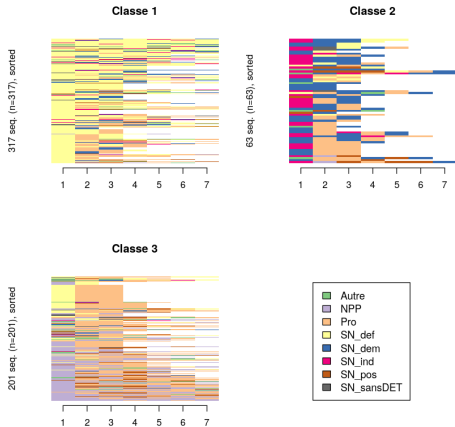
- Test du khi-deux entre les clusters et le facteur de variation (type de texte, genre textuel, type de référent)
- Résidus de Pearson : permettent de décrire la direction de la liaison

Observation et description des clusters obtenus

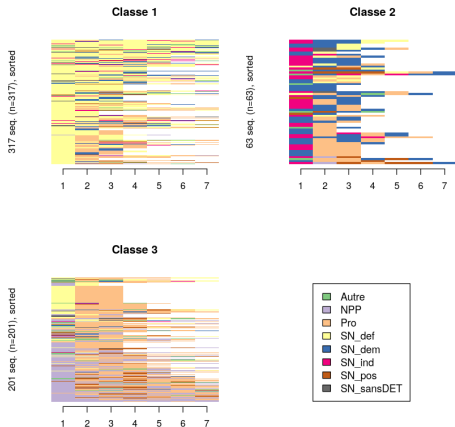


Prendre en compte les facteurs de variation

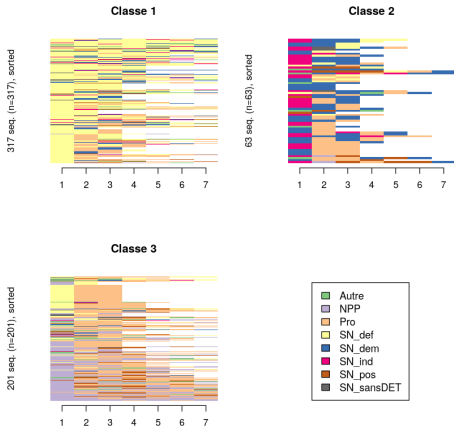
- Test du khi-deux entre les clusters et le facteur de variation (type de texte, genre textuel, type de référent)
- Résidus de Pearson : permettent de décrire la direction de la liaison



- Résultats obtenus directement interprétables



- Résultats obtenus directement interprétables
- Résultats obtenus pertinents d'un point de vue linguistique



- Résultats obtenus directement interprétables
- Résultats obtenus pertinents d'un point de vue linguistique
- Résultats suggèrent l'existence de stratégies différentes pour introduire et maintenir les référents

- Croiser les descriptions des caractéristiques globales et linéaires pour aboutir à une typologie

- Tester l'analyse multidimensionnelle des séquences : prise en compte de plusieurs critères de description

Merci pour votre attention !

Des questions ?

- Antoine, J.-Y., Lefeuve, A., & Schang, E. (2016, juillet). Codage en chaîne ou en première mention de la coréférence : approcher la structure des chaînes de référence par comparaison des deux annotations. In *5ème congrès mondial de linguistique française (cmf'2016)*. Tours, France. Consulté sur <https://hal.archives-ouvertes.fr/hal-01344976>
- Ariel, M. (2001). Accessibility theory : An overview. *Text representation : Linguistic and psycholinguistic aspects*, 8, 29–87.
- Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal*, 6(4), 435–460.
- Charolles, M. (2002). *La référence et les expressions référentielles en français*. Editions Ophrys.
- Corblin, F. (1987). *Indéfini, défini et démonstratif droz*. Genève.
- Corblin, F. (1995). *Les formes de reprise dans le discours. anaphores et chaînes de référence*. Presses Universitaires de Rennes.
- Cornish, F. (1998). Les chaînes topicales : leur rôle dans la gestion et la structuration du discours. *Cahiers de grammaire*, 23(1), 9–40.
- Cornish, F. (2000). L'accessibilité cognitive des référents, le centrage d'attention, et la structuration du discours : une vue d'ensemble.

Verbum, 22(1), 7–30.

- De Marneffe, M.-C., Recasens, M., & Potts, C. (2015). Modeling the lifespan of discourse entities with application to coreference resolution. *Journal of Artificial Intelligence Research*, 52, 445–475.
- Federzoni, S., Ho-Dac, L.-M., & Rebeyrolle, J. (2020, juillet). Les chaînes topicales dans la ressource ANNODIS. In *Cmlf2020 : 7e congrès mondial de linguistique française*. Montpellier, France. Consulté sur <https://hal.archives-ouvertes.fr/hal-02890989>
- Fossard, M., Achim, A. M., Rousier-Vercruyssen, L., Gonzalez, S., Bureau, A., & Champagne-Lavau, M. (2018). Referential choices in a collaborative storytelling task : Discourse stages and referential complexity matter. *Frontiers in psychology*, 9, 176.
- Gabardinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2009). Mining sequence data in R with the TraMineR package : A user's guide. *Geneva : Department of Econometrics and Laboratory of Demography, University of Geneva*.
- Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering : A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2), 203–225.

- Guillot-Barbance, C., & Quignard, M. (2019). Chaînes de référence et structure textuelle dans les essais sur la peinture de diderot. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*(25).
- Kleiber, G. (1994). Anaphores et pronoms. *Louvain-la-Neuve, Duculot*.
- Longo, L. (2013). *Vers des moteurs de recherche "intelligents" : un outil de détection automatique de thèmes*. (PhD Thesis). Université de Strasbourg.
- Longo, L., & Todirascu, A. (2014). Vers une typologie des chaînes de référence dans des textes administratifs et juridiques. *Langages*(195/3), 79-98.
- Mitkov, R. (2014). *Anaphora resolution*. Routledge.
- Obry, V., Glikman, J., Guillot-Barbance, C., & Pincemin, B. (2017). Les chaînes de référence dans les récits brefs en français : étude diachronique (xiii^e-xv^e s.). *Langue française*, 3, 91–110.
- Recasens, M., & Hovy, E. (2010, juillet). Coreference resolution across corpora : Languages, coding schemes, and preprocessing information. In *Proceedings of the 48th annual meeting of the association for*

- computational linguistics* (pp. 1423–1432). Uppsala, Sweden : Association for Computational Linguistics. Consulté sur <https://aclanthology.org/P10-1144>
- Robette, N. (2011). *Explorer et décrire les parcours de vie : les typologies de trajectoires*. CEPED. Consulté sur <https://halshs.archives-ouvertes.fr/halshs-01016125>
- Salles, M. (2015). Chaînes de référence : la deuxième mention. l'exemple des entités inanimées dans les narrations littéraires. *Travaux de linguistique*(2), 111–133.
- Schneedecker, C. (2005). Les chaînes de référence dans les portraits journalistiques : éléments de description. *Travaux de linguistique*(2), 85–133.
- Schneedecker, C. (2021). *Les chaînes de référence en français*. Paris : Éditions Ophrys.
- Vogel, J., Krahmer, E., & Maes, A. (2013). When a stone tries to climb up a slope : the interplay between lexical and perceptual animacy in referential choices. *Frontiers in psychology*, 4, 154.