



Conception d'un outil de visualisation et d'exploration de chaînes de coréférences

Frédéric Landragin

JADT 2016



Plan

- Exemple de départ, contexte de l'étude et objectifs
- Annotation et visualisation de chaînes de coréférences
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- Statistiques sur les chaînes de coréférences
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- Discussion : limites et perspectives pour l'ADT
- Conclusion

Plan

- Exemple de départ, contexte de l'étude et objectifs
- Annotation et visualisation de chaînes de coréférences
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- Statistiques sur les chaînes de coréférences
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- Discussion : limites et perspectives pour l'ADT
- Conclusion

Exemple de départ

- Exemple typique :

« Le village était désert. Il semblait abandonné. La place principale était vide. Elle en paraissait triste. Tout reprendrait vie le lendemain matin : le village s'animerait, la place se remplirait de monde »

- Succession de phrases

- Caractérisée par une certaine cohérence
- Relations anaphoriques et coréférentes qui relient les expressions référentielles

- Pour étudier les liens entre phrases

- Chaînes anaphoriques
- Chaînes de coréférences
- Suite des références

Contexte de l'étude

- Contexte scientifique
 - Étude de la cohérence textuelle et des chaînes de coréférences
 - Tirer parti des avancées de la linguistique de corpus outillée
- Contexte technique
 - Conception d'outils d'annotation, de visualisation et d'analyse des chaînes de coréférences
 - Méthodologie d'analyse assistée par ordinateur des chaînes
- Contexte institutionnel
 - Bilan du projet PEPS **MC4** « Modélisation Contrastive et Computationnelle des Chaînes de Coréférences »
 - Projet ANR **Democrat** « Description et modélisation des chaînes de référence : outils pour l'annotation de corpus et le traitement automatique »

Objectifs

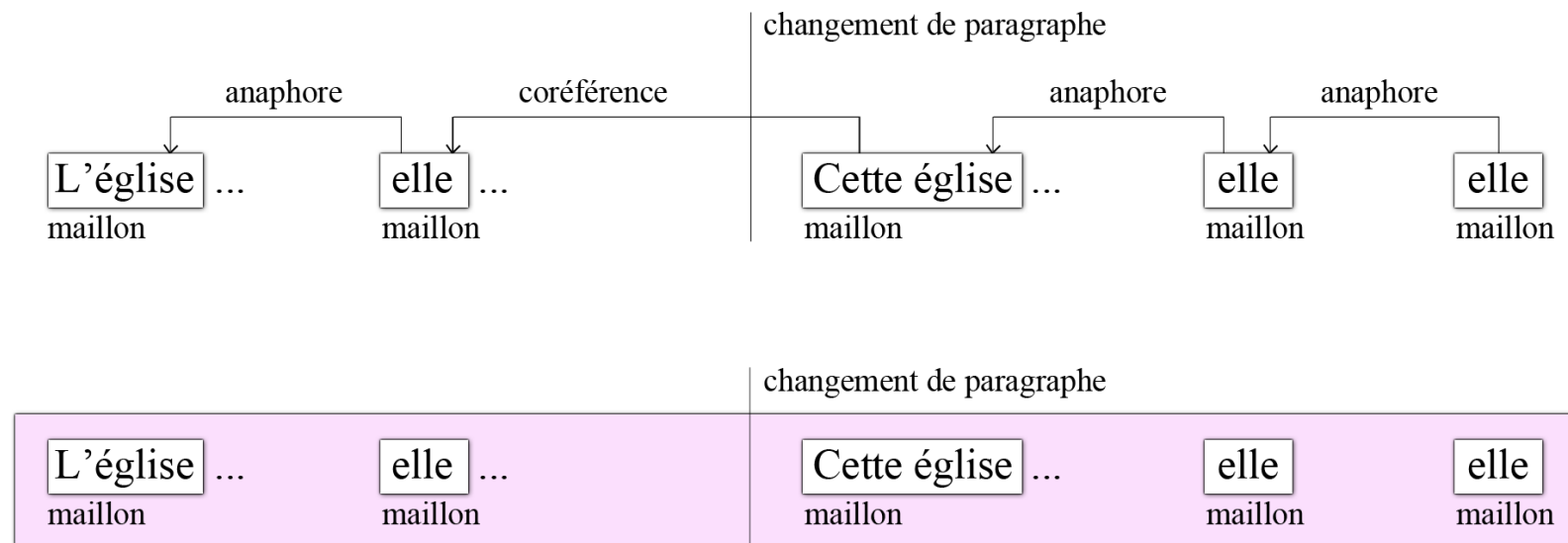
- Conception d'un outil à partir de besoins exprimés par les linguistes du projet MC4
 - Besoin de visualiser – au fur et à mesure de l'annotation – la composition d'une chaîne de coréférences
 - Besoin de tester directement des hypothèses linguistiques :
 - Un paragraphe commence le plus souvent par une référence au dernier référent mentionné dans le paragraphe précédent
 - Le découpage en sous-chaînes (une sous-chaîne commençant par un nom propre ou une description définie singularisante) correspond au découpage en paragraphes
 - Besoin de calculer des statistiques sur les chaînes
- Intégration de cet outil à l'outil d'annotation
 - Car visualiser et interroger les données annotées permet de repérer des erreurs d'annotation et de reprendre celle-ci au bon endroit

Plan

- Exemple de départ, contexte de l'étude et objectifs
- **Annotation et visualisation de chaînes de coréférences**
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- Statistiques sur les chaînes de coréférences
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- Discussion : limites et perspectives pour l'ADT
- Conclusion

Problèmes et méthodologie

- Spécificités des chaînes de coréférences
 - Une chaîne de coréférences couvre potentiellement tout le texte
 - Les chaînes de coréférences s'imbriquent
 - Il existe de nombreuses façons de les aborder et de les annoter



- Relations ou ensembles ? On autorise les deux à la fois...

Annotation manuelle des expressions référentielles

- Un ensemble de traits pour chaque expression

Analec 1.1 : Echenoz8 - vue : Echenoz8

Documents Structure Vue Texte Unités Relations Schémas Règles Statistiques

Gestion des unités

Unité : Maillon Maillon-17 0 Créer Supprimer Rectifier une borne

Comme tout avait brûlé – la mère, les meubles et les photographies de la mère –, pour Fabre et le fils Paul c'était tout de suite beaucoup d'ouvrage : toute cette cendre et ce deuil, déménager, courir se refaire dans les grandes surfaces. Fabre trouva trop vite quelque chose de moins vaste, deux pièces aux fonctions permutables sous une cheminée de brique dont l'ombre donnait l'heure, et qui avaient ceci de bien d'être assez proches du quai de Valmy.

Le soir après le dîner, Fabre parlait à Paul de sa mère, sa mère à lui Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de Sylvie Fabre, il s'épuisait à vouloir la décrire toujours plus exactement : au milieu de la cuisine naquirent des hologrammes que dégonflait la moindre imprécision. Ça ne se rend pas, soupirait Fabre en posant une main sur sa tête, sur ses yeux, et le découragement l'endormait. Souvent ce fut à Paul de déplier le canapé convertible, transformant les choses en chambre à coucher.

Le dimanche et certains jeudis, ils partaient sur le quai de Valmy vers la rue Marseille, la rue Dieu, ils allaient voir Sylvie Fabre. Elle les regardait de haut, tendait vers eux le flacon de parfum Piver, Forvil, elle souriait dans quinze mètres de robe bleue. Le gril d'un soupirail trouait sa hanche. Il n'y avait pas d'autre image d'elle.

Champs de l'unité : "Maillon-17"

Type de maillon	Catégorie	Expansion	Fonction	Niveau syntaxique
Forme explicite	Pron Indéfini	Aucune	Sujet	Primaire dans enchâssée

Macro-syntaxe	Plan énonciatif	Position	Interprétation	Rôle actanciel	Cas
Préfixe	Plan principal	Initiale	Immédiate	Agent	Non marqué

Nom du référent	Introduction du référent	Nombre
Fabre (le père) et Paul (le fils)	Construction discursive	Groupe strict

Annotation manuelle des chaînes

- Un ensemble de traits pour la chaîne
 - identifiant du référent ; type de référent (individu, groupe)
- Et plus particulièrement pour le premier maillon
 - type d'introduction : par regroupement, extraction, association...
 - c'est là que se font les liens non coréférentiels (anaphore associative)

The screenshot displays the 'Analec 1.1 : Echenoz8 - vue : Echenoz8' window. The top menu bar includes 'Documents', 'Structure', 'Vue', 'Texte', 'Unités', 'Relations', 'Schémas', 'Règles', and 'Statistiques'. The 'Gestion des schémas' section is active, showing a 'Schéma' dropdown set to 'Coréférence', a text field containing 'Sylvie Fabre (la mère)', and a counter set to '0'. Below this, the 'Élément' section shows '< type d'élément >', '< identifiant >', and another counter set to '0'. The main text area contains two paragraphs from a story. The first paragraph mentions 'la mère' and 'la mère'. The second paragraph mentions 'sa mère', 'sa mère à lui', and 'Sylvie Fabre'. The bottom section, 'Champs du schéma : "Sylvie Fabre (la mère)"', contains three fields: 'Introduction du référent' (set to 'Extraction'), 'Nom du référent' (set to 'Sylvie Fabre (la mère)'), and 'Nombre' (set to 'Singleton').

Analec 1.1 : Echenoz8 - vue : Echenoz8

Documents Structure Vue Texte Unités Relations Schémas Règles Statistiques

Gestion des schémas

Schéma : Coréférence Sylvie Fabre (la mère) 0 Créer Supprimer

Élément : < type d'élément > < identifiant > 0 Ajouter Oter Créer

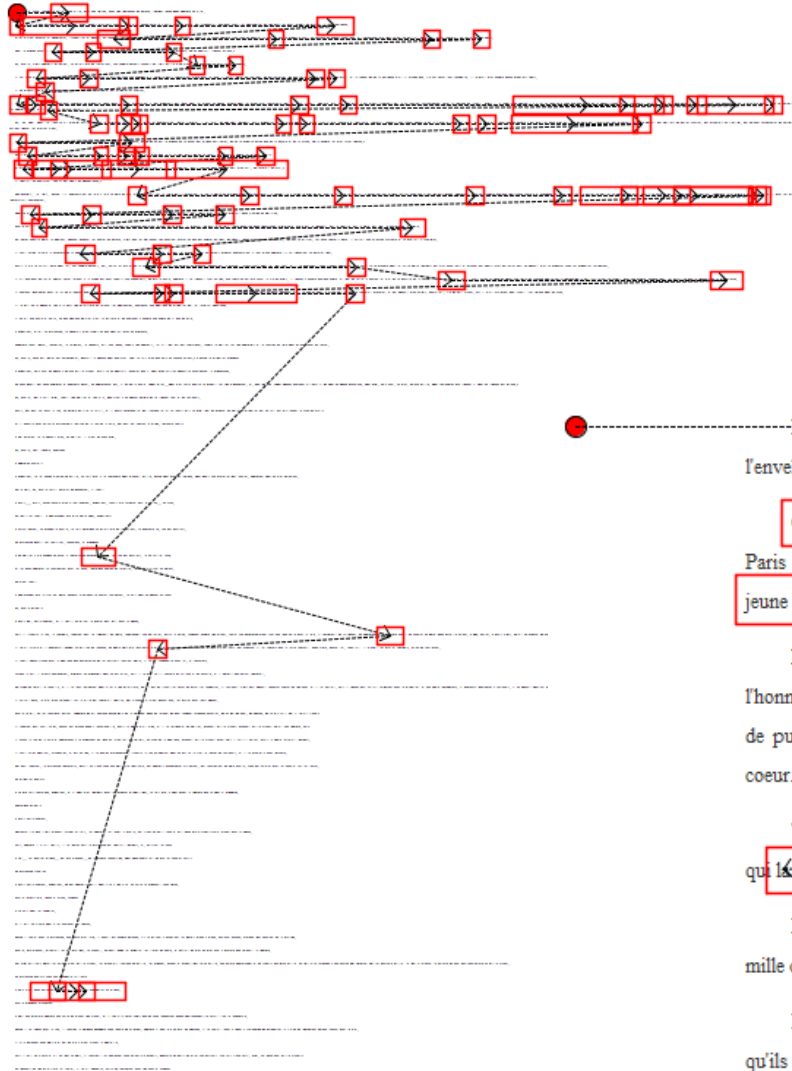
Comme tout avait brûlé – **la mère**, les meubles et les photographies de **la mère** –, pour Fabre et le fils Paul c'était tout de suite beaucoup d'ouvrage : toute cette cendre et ce deuil, déménager, courir se refaire dans les grandes surfaces. Fabre trouva trop vite quelque chose de moins vaste, deux pièces aux fonctions permutable sous une cheminée de brique dont l'ombre donnait l'heure, et qui avaient ceci de bien d'être assez proches du quai de Valmy.

Le soir après le dîner, Fabre parlait à Paul de **sa mère**, **sa mère à lui** Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de **Sylvie Fabre**, il s'épuisait à vouloir **la** décrire toujours plus exactement : au milieu de la cuisine naquirent des hologrammes que dégonflait la moindre imprécision. Ça ne se rend pas, soupirait Fabre en posant une main sur sa tête, sur ses yeux, et le découragement l'endormait. Souvent ce fut à Paul de déplier le canapé convertible, transformant les choses en chambre à coucher.

Champs du schéma : "Sylvie Fabre (la mère)"

Introduction du référent	Nom du référent	Nombre
Extraction	Sylvie Fabre (la mère)	Singleton

Visualiser les chaînes avec Glozz



- Appréciation de la répartition d'une chaîne sur un texte via des possibilités de zoom
- Visualisation « intégrée »

M. Lantin, ayant rencontré cette jeune fille, dans une soirée, chez son sous-chef de bureau, l'amour l'enveloppa comme un filot.

C'était la fille d'un percepteur de province, mort depuis plusieurs années. Elle était venue ensuite à Paris avec sa mère, qui fréquentait quelques familles bourgeoises de son quartier dans l'espoir de marier la jeune personne.

Elles étaient pauvres et honorables, tranquilles et douces. La jeune fille semblait le type absolu de l'honnête femme à laquelle le jeune homme sage rêve de confier sa vie. Sa beauté modeste avait un charme de pudeur angélique, et l'imperceptible sourire qui ne quittait point ses lèvres semblait un reflet de son cœur.

Tout le monde chantait ses éloges, tous ceux qui la connaissaient répétaient sans fin: "Heureux celui qui la prendra. On ne pourrait trouver mieux."

M. Lantin, alors commis principal, au ministère de l'Intérieur, aux appointements annuels de trois mille cinq francs, lui demanda en mariage et l'épousa.

Il fut avec elle invraisemblablement heureux. Elle gouverna sa maison avec une économie si adroite qu'ils semblaient vivre dans le luxe. Il n'était point d'attentions, de délicatesses, de châtiments qu'elle n'eût pour son mari, et la séduction de sa personne était si grande que, six ans après leur rencontre, il l'aimait plus encore qu'aux premiers jours.

Visualiser la suite des références

- Suite linéaire des références annotées, représentée avec un code couleur et un code alphanumérique (alphabet)
- Intérêt = à part illustrer des articles, tenter de détecter des comportements

The screenshot shows a software window titled "Représentation des chaînes". The main area displays a horizontal sequence of colored circles, each with a small alphanumeric code above it. The sequence is divided into sections labeled §2, §3, and §4. Below the sequence, there are three dropdown menus: "Chaînes : Coréférence", "Unités : Forme explicite", and "Champ à afficher : Catégorie". At the bottom, there is a text area containing three paragraphs of text, with some words highlighted in yellow. To the right of the text area is a legend with colored circles and corresponding labels for various grammatical categories.

Représentation des chaînes

§2 §3 §4

2 3 1 3 1 3 3 3 4 1 2 1 2 2 2 2 3 | 4 4 5 5 4 4 5 5 1 | 6 1 6 7 6 6 1 | 6 2 1 1

Chânes : Coréférence Unités : Forme explicite Champ à afficher : Catégorie

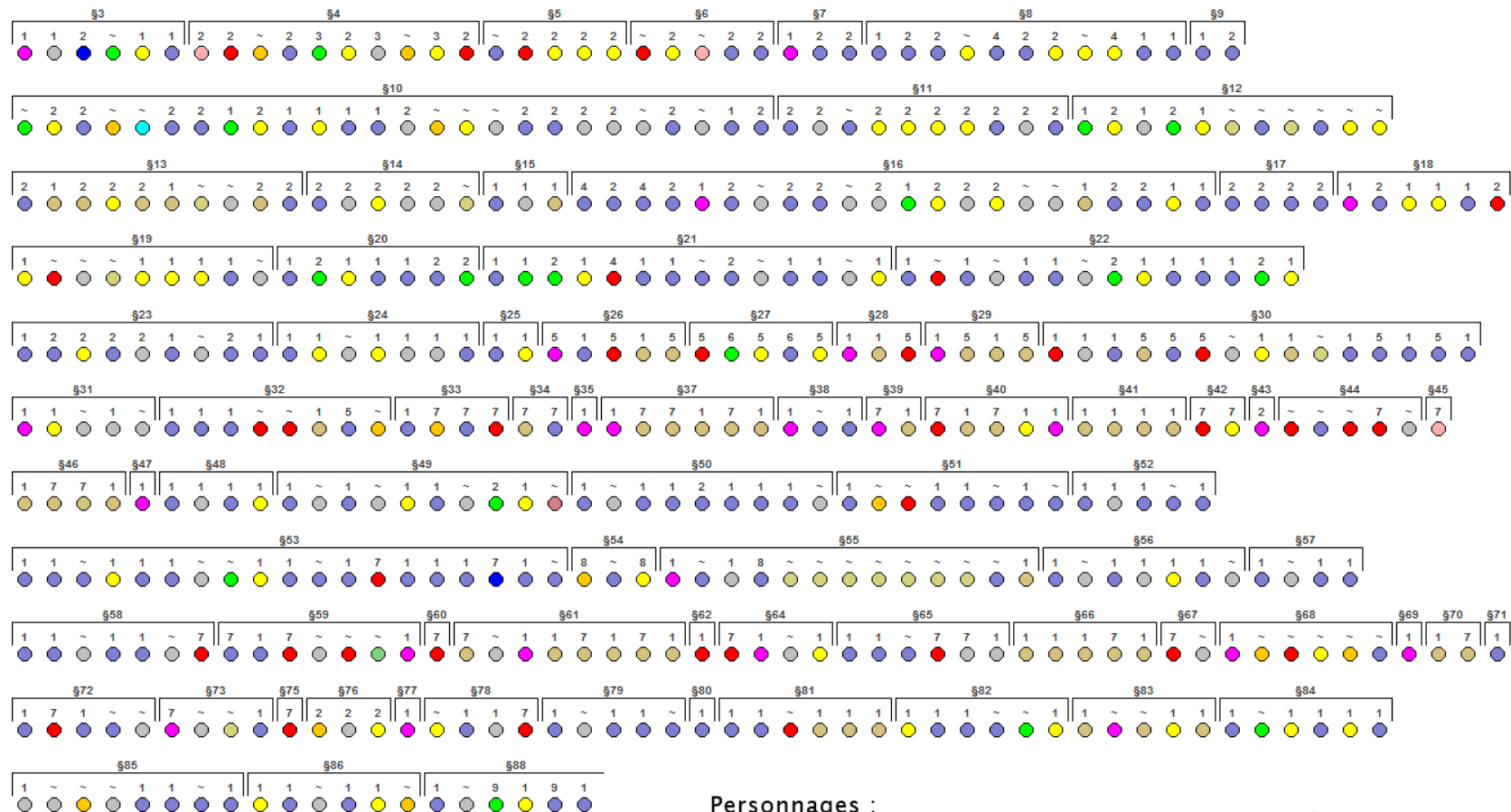
Le soir après le dîner, Fabre parlait à Paul de sa mère, sa mère à lui Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de Sylvie Fabre, il s'épuisait à vouloir la décrire toujours plus exactement : au milieu de la cuisine naquirent des hologrammes que dégonflait la moindre imprécision. Ça ne se rend pas, soupirait Fabre en posant une main sur sa tête, sur ses yeux, et le découragement l'endormait. Souvent ce fut à Paul de déplier le canapé convertible, transformant les choses en chambre à coucher.

Le dimanche et certains jeudis, ils portaient sur le quai de Valmy vers la rue Marseille, la rue Dieu, ils allaient voir Sylvie Fabre. Elle les regardait de haut, tendait vers eux le flacon de parfum Piver, Forvil, elle souriait dans quinze mètres de robe bleue. Le gril d'un soupirail trouvait sa hanche. Il n'y avait pas d'autre image d'elle.

L'artiste Flers l'avait représentée sur le flanc d'un immeuble, juste avant le coin de la rue. L'immeuble était plus maigre et plus solide, mieux tenu que les vieilles constructions qui se collaient en grinçant contre lui, terrifiées par le plan d'occupation des sols. En manque de marquise, son porche saturé de moulures portait le nom (Wagner) de l'architecte-sculpteur gravé dans un cartouche en haut à droite. Et le mur sur lequel, avec toute son équipe, l'artiste Flers avait peigné pour figurer Sylvie Fabre en pied, surplombait un petit espace vert rudimentaire, sorte de

Autre
Dét Possessif
GN Défini
GN Démonstratif
GN Indéfini
GN Possessif
N ou GN sans dét
Nom Propre
Pron Démonstratif
Pron Indéfini
Pron Interrogatif
Pron Pers Anaphorique
Pron Pers Déictique
Pron Relatif
Pron possessif

Visualiser la suite des références



Personnages :

1. Mr. Lantin
2. Mme. Lantin
3. Mère de Mme. Lantin
4. Groupe incluant Mr. et Mme. Lantin
5. Premier bijoutier
6. Commis du premier bijoutier
7. Deuxième bijoutier
8. Flâneurs
9. Deuxième épouse de Mr. Lantin

- Dét Possessif
- GN Défini
- GN Démonstratif
- GN Indéfini
- GN Possessif
- N ou GN sans dét
- Nom Propre
- Pron Démonstratif
- Pron Indéfini
- Pron Interrogatif
- Pron Pers Anaphorique
- Pron Pers Déictique
- Pron Relatif
- Forme atténuée

Plan

- Exemple de départ, contexte de l'étude et objectifs
- Annotation et visualisation de chaînes de coréférences
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- **Statistiques sur les chaînes de coréférences**
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- Discussion : limites et perspectives pour l'ADT
- Conclusion

Rappels sur Analec

- Analec est plus un outil d'annotation qu'un outil d'ADT
- Il propose surtout des fonctionnalités de gestion du schéma d'annotation, de personnalisation de l'interface d'annotation et de gestion des droits et des préférences pour plusieurs annotateurs
- Dans sa version initiale, il propose trois interfaces graphiques simples pour explorer les données annotées (et seulement les données annotées) :
 - Calcul de fréquences
 - Tests de corrélations
 - Visualisations de corrélations

Calcul de fréquences

- Premier des modes d'interrogation de corpus avec Analec = simple calcul de fréquence
- Particularité = vérifier les annotations, corriger...

The screenshot shows the 'Fréquences' application window. It has a blue title bar and standard Windows window controls. The interface is divided into several sections:

- Eléments:** Contains a dropdown menu set to 'Forme explicite'.
- Champ à analyser:** Contains a dropdown menu set to 'Macro-syntaxe'.
- Filtre:** Contains a dropdown menu set to '<pas de filtre>'.
- Résultats:** A table showing the frequency of different elements.
- Visualisation des occurrences:** A section showing the context of the occurrences for the selected element.

Résultats Table:

	Nb d'occurrences	Fréquence (%)
Noyau	121	74.23
Préfixe	19	11.66
Suffixe	23	14.11
<aucune valeur>	0	0.0
TOTAL	163	100.0

Below the table is a button labeled 'Exporter le tableau'.

Visualisation des occurrences:

Identifiants	Occurrences
Forme explicite-1	Comme tout avait brûlé - la mère , les meubles et les photographies de la mère -, pour Fabre et le fils Paul c'était [...]
Forme explicite-2	Comme tout avait brûlé - la mère, les meubles et les photographies de la mère -, pour Fabre et le fils Paul c'était tout de su
Forme explicite-3	[...]tout avait brûlé - la mère, les meubles et les photographies de la mère -, pour Fabre et le fils Paul c'était tout de suite bea
Forme explicite-4	[...]t brûlé - la mère, les meubles et les photographies de la mère -, pour Fabre et le fils Paul c'était tout de suite beaucoup d
Forme explicite-5	[...]tte cendre et ce deuil, déménager, courir se refaire dans les grandes surfaces. Fabre trouva trop vite quelque chose de m

Calcul de corrélations

- 2^e type d'interrogation : tester des associations de champs

Corrélations

Type d'unité: **Forme explicite**

Champ 1 (horizontal): **Catégorie**

Champ 2 (vertical): **Fonction**

Résultats

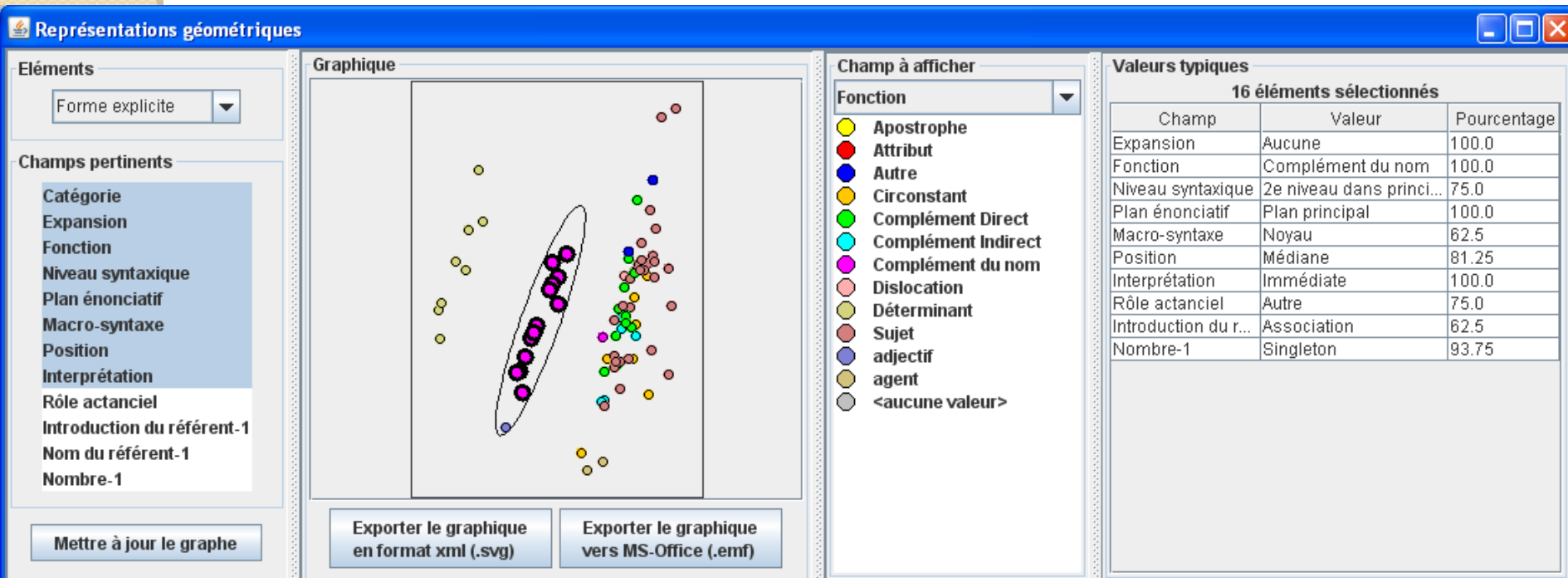
	...	...	GN Défini	...	...	GN Possessif	...	Nom Propre	...	...	Pron Pers Anaphorique	...	Pron Relatif	...	TOTAL
Apostrophe	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2
Attribut	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Autre	0	0	1	0	0	1	0	2	0	0	0	0	0	0	4
Circonstant	0	0	1	0	0	1	0	1	0	0	4	0	1	0	8
Complément Direct	0	0	1	0	0	3	0	6	0	0	7	0	0	0	17
Complément Indirect	0	0	1	0	1	1	0	1	1	0	0	0	0	0	6
Complément du nom	0	0	4	0	0	2	0	9	0	0	0	2	0	1	18

Exporter le tableau

Visualisation des occurrences

Identifiants	Occurrences
Forme explicite-15	[...]lui Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de Sylvie Fabre , il s'épuisait à vouloir la décrire toujours plus exactement : au mil
Forme explicite-43	[...]é de poser. Elle n'avait pas aimé cela. C'était trois ans avant la naissance de Paul , pour qui ce mur n'était qu'une tranche de vie antérieure. Regarde un peu ta
Forme explicite-60	[...]aléas qui font qu'on passe par là. D'une cabine scellée dans le champ visuel de Sylvie Fabre , il avait failli appeler son père lorsqu'on se mit à démolir la vieille
Forme explicite-66	[...]ar la pérennité de l'espace vert, qui formait un gazon subsidiaire aux pieds de Sylvie . Négligence ou manœuvre, on laissait l'espace dépérir. Les choses ver
Forme explicite-98	[...]on ne les confondît pas avec rien. C'était un sépulcre au lieu d'une effigie de Sylvie , on l'approchait d'un autre pas, d'une démarche moins souple. Après l'e
Forme explicite-108	[...]cria Fabre. Paul, dont c'est quand même le nom, se tourna vers celle qui criait Fabre Fabre encore, j'ai du lait. La voix énervante tomba du ciel, La voix énervant
Forme explicite-126	[...]de suite, au quatrième étage côté Wagner, dans un studio situé sous les yeux de Sylvie qui étaient deux lampes sourdes derrière le mur de droite. Selon ses c
Forme explicite-133	[...]lèvres comme dans un hamac ; à son fils il démontra cela sur plans. La voix de Fabre exposait une mission supérieure, relevant d'une cause auprès de quoi l

Visualisation de corrélations



Visualisation des occurrences

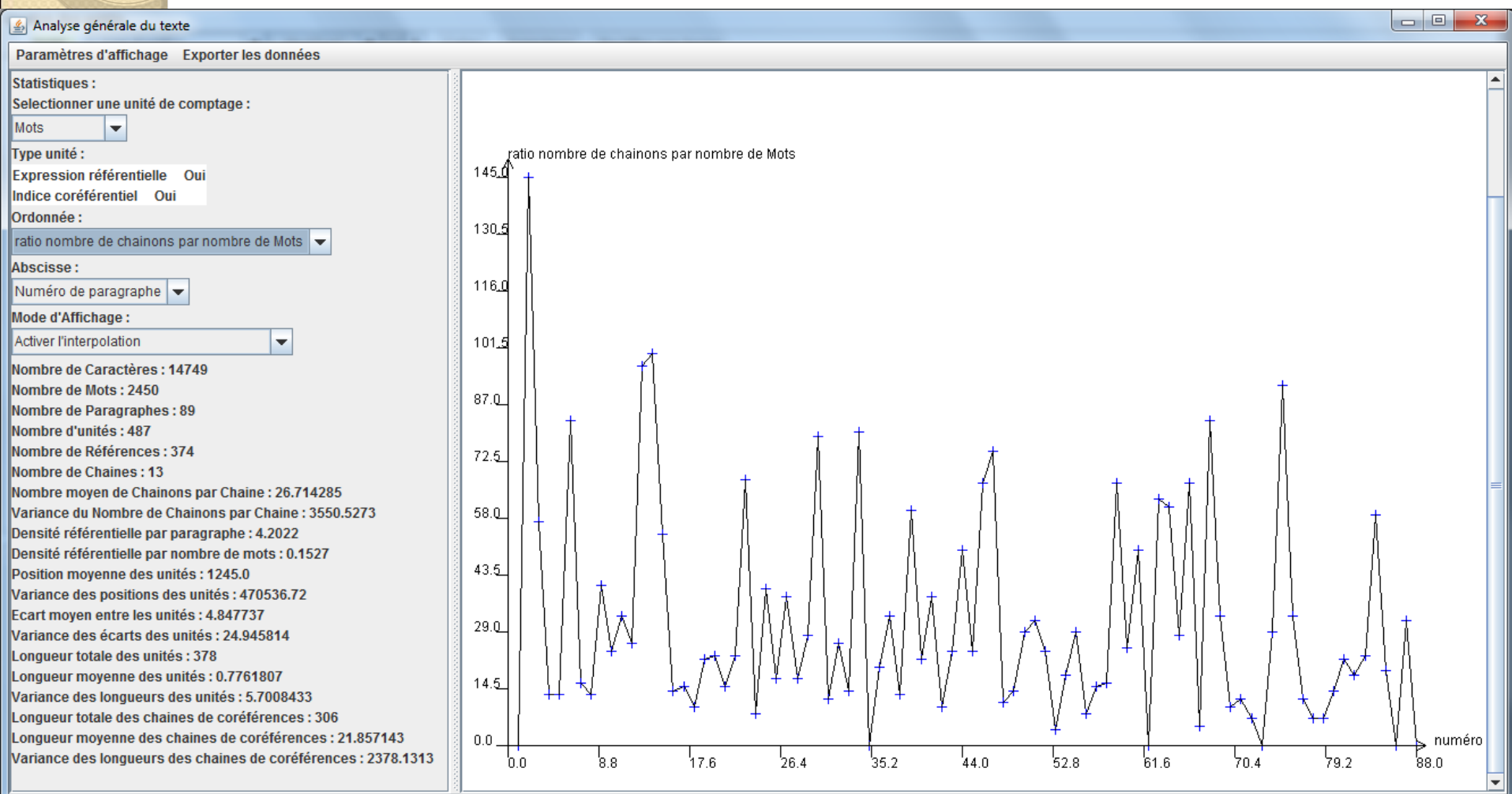
Identifiants	Occurrences
Forme explicite-8	[...]assez proches du quai de Valmy. Le soir après le dîner, Fabre parlait à Paul de sa mère, sa mère à lui Paul, parfois dès le dîner. Comme on ne possédait plus de r
Forme explicite-12	[...]ai de Valmy. Le soir après le dîner, Fabre parlait à Paul de sa mère, sa mère à lui Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de Syl
Forme explicite-15	[...]lui Paul, parfois dès le dîner. Comme on ne possédait plus de représentation de Sylvie Fabre, il s'épuisait à vouloir la décrire toujours plus exactement : au milieu de
Forme explicite-31	[...]Jeue. Le gril d'un soupirail trouait sa hanche. Il n'y avait pas d'autre image d'elle. L'artiste Fiers l'avait représentée sur le flanc d'un immeuble, juste avant le[...]
Forme explicite-34	[...]En manque de marquise, son porche saturé de moulures portait le nom (Wagner) de l'architecte-sculpteur gravé dans un cartouche en haut à droite. Et le mur sur le
Forme explicite-43	[...]é de poser. Elle n'avait pas aimé cela. C'était trois ans avant la naissance de Paul, pour qui ce mur n'était qu'une tranche de vie antérieure. Regarde un peu ta mè[...]
Forme explicite-60	[...]aléas qui font qu'on passe par là. D'une cabine scellée dans le champ visuel de Sylvie Fabre, il avait failli appeler son père lorsqu'on se mit à démolir la vieille chose
Forme explicite-66	[...]ar la pérennité de l'espace vert, qui formait un gazon subsidiaire aux pieds de Sylvie. Négligence ou manœuvre, on laissait l'espace dépérir. Les choses vertes s'y
Forme explicite-90	[...]perçut Fabre une fois sur le chantier, l'immeuble allait atteindre le ventre de sa mère. Une autre fois c'était vers la poitrine, le veuf parlait avec un contremaître [...]
Forme explicite-98	[...]on ne les confondit pas avec rien. C'était un sépulcre au lieu d'une effigie de Sylvie, on l'approchait d'un autre pas, d'une démarche moins souple. Après l'entrée, :
Forme explicite-106	[...]r le trottoir s'arrêta derrière lui, leva les yeux au ciel et cria Fabre. Paul, dont c'est quand même le nom, se tourna vers celle qui criait Fabre Fabre encore, j' [...]
Forme explicite-123	[...]re s'était présenté le premier au bureau de location, avant même l'intervention des peintres, donnant un regard mort à l'appartement témoin. On ne le dissuada pas t

Avec notre nouveau module

- Analec propose alors trois interfaces supplémentaires, conçues pour les chaînes de coréférences :
 - Statistiques générales sur le texte et ses annotations
 - Visualisation et statistiques sur les chaînes de coréférences
 - Visualisation et statistiques sur la suite des références
- Ces trois interfaces ont été testées sur le corpus MC4
- Elles ne sont pas encore assez robustes pour être incluses dans la version « officielle » d'Analec
- A priori, elles sont généralisables pour les chaînes d'annotations

Interface I : statistiques générales

- Nouvelle fenêtre d'ANALEC permettant plusieurs calculs : densités référentielles paragraphe par paragraphe, nombre de chaînes et/ou de mentions par paragraphe, ratio nombre de mentions par nombre de mots



Interface 2 : chaînes de coréférences

- Nouvelle fenêtre d'ANALEC pour voir et étudier chaque chaîne

Analyse d'une chaîne de coréférence

Paramètres d'affichage Exporter les données Fusionner des types d'unités

Chaînes :
Coréférence

Unités :
Expression référentielle

☒ Afficher les histogrammes de répartition

Champ à filtrer :
Position

	Initiale	49,711
	Médiane	42,486
	Finale	7,803
	<aucune valeur>	

Champ de la chaîne à analyser :

Nom du référent		
Oui	M. Lantin	58,382
Oui	Mme Lantin	22,254
Oui	Bijoutier n°2	10,116
Oui	Bijoutier n°1	4,335
Oui	Groupe : M. et Mme Lantin 1	1,445
Oui	Mère de Mme Lantin	0,578
Oui	Les flâneurs	0,578
Oui	Commis du bijoutier n°1	0,578
Oui	2e épouse de M. Lantin	0,578
Oui	Sous-chef de M. Lantin	0,578
Oui	Groupe : Mme Lantin et	0,578

Masquer cette valeur de la chaîne

Ne montrer que cette valeur de la chaîne

Paragraphe à filtrer :

Oui	Paragraphe3 " M. Lantin, ayant
Oui	Paragraphe4 " C'était la fille"

M. Lantin :

§3 §7 §8 §9 §10

M. L son l' M. L II il l' II son le sa il la il

exporter en SVG

Statistique :

valeur de la chaîne ^	Finale	Initiale	aucune val...	Médiane
2e épouse de M. Lantin	0	100	0	0
Bijoutier n°1	6,667	53,333	0	40
Bijoutier n°2	17,143	57,143	0	25,714
Commis du bijoutier n°1	50	0	0	50
Groupe : M. et Mme Lantin 1	20	60	0	20
Groupe : Mme Lantin et sa mère	0	100	0	0
Les flâneurs	0	0	0	100
M. Lantin	5,941	50,99	0	43,069
Mère de Mme Lantin	0	0	0	100
Mme Lantin	7,792	45,455	0	46,753
Sous-chef de M. Lantin	0	0	0	100

Chaînes :

Coréférence

Unités :

Expression référentielle

☒ Afficher les histogrammes de répartition

Champ à filtrer :

Catégorie

☐ Pron Pers Anaphorique☐ Dét Possessif☐ Pron Pers Déictique☐ GN Défini☐ Nom Propre☐ GN Possessif☐ GN Indéfini☐ Pron Démonstratif☐ GN Démonstratif☐ Pron Relatif☐ Pron Interrogatif☐ Pron Indéfini☐ N ou GN sans dét☐ <aucune valeur>

Champ de la chaîne à analyser :

Nom du référent

Oui	M. Lantin
Oui	Mme Lantin
Oui	Bijoutier n°2
Oui	Bijoutier n°1
Oui	Groupe : M. et Mme Lantin 1
Oui	Mère de Mme Lantin
Oui	Les flâneurs
Oui	Commis du bijoutier n°1
Oui	2e épouse de M. Lantin
Oui	Sous-chef de M. Lantin
Oui	Groupe : Mme Lantin et sa mère

Mme Lantin :

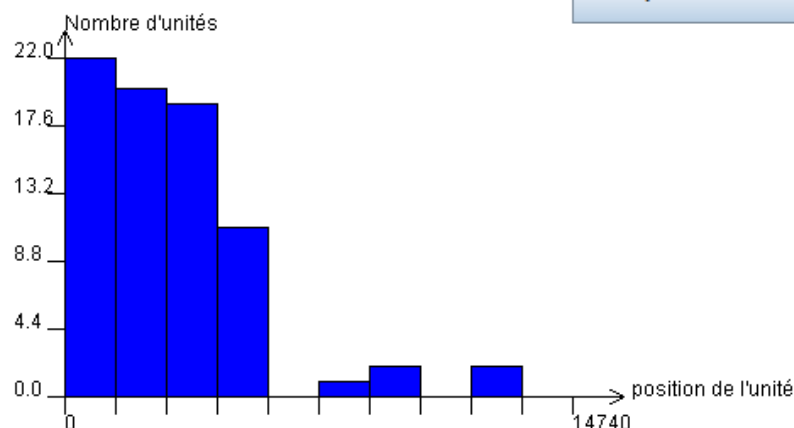
\$3 \$4 \$5 \$6 \$7 \$8

cett C' la f Elle sa la j La j Sa ses son ses la la la l' elle Elle elle

exporter en SVG

exporter en SVG

exporter en CSV

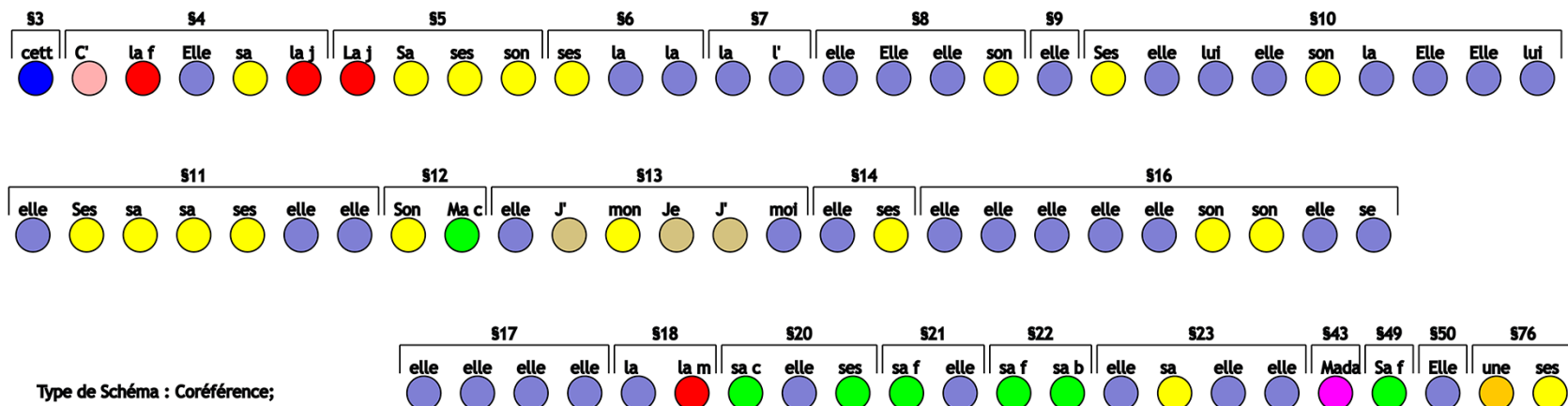


Statistique :

valeur de la chaîne	Pron Pers Anaphorique	Pron Interrogatif	GN Indéfini	aucune valeur	N c
Groupe : Mme Lantin et sa mère	100	0	0	0	0
Groupe : M. et Mme Lantin 1	60	0	0	0	0
M. Lantin	53,465	0	0	0	0
Mme Lantin	51,948	0	1,299	0	0
2e épouse de M. Lantin	50	0	0	0	0
Les flâneurs	50	0	0	0	0
Commis du bijoutier n°1	50	0	0	0	0
Bijoutier n°1	26,667	0	0	0	0
Bijoutier n°2	8,571	0	2,857	0	0

Export des graphiques

- Résultat de l'exportation du graphique en SVG, retouché rapidement avec Adobe Illustrator



Type de Schéma : Coréférence;

Type d'Unités : Expression référentielle;

Champ 1 : Catégorie;

Champ de la Chaîne : Nom du référent;

Champ 1 :

	Pron Pers Anaphorique	46.820812
	Dét Possessif	17.630058
	Pron Pers Déictique	15.028901
	GN Défini	7.2254333
	Nom Propre	6.6473985
	GN Possessif	4.33526
	GN Indéfini	1.1560693
	Pron Démonstratif	0.57803464
	GN Démonstratif	0.57803464

Interface 3 : suite des références

Analyse des suites d'occurrences

Paramètres d'affichage Exporter les données

Chaines :

Coréférence

Unités :

Expression référentielle

Nb de chainons par ligne :

<tous>

☐ Ne pas couper les paragraphes

Champ à afficher :

Catégorie

	Pron Pers Anaphorique	43,885
	Dét Possessif	16,307
	Pron Pers Déictique	12,71
	GN Défini	8,633
	Nom Propre	5,755
	GN Possessif	4,556
	Pron Indéfini	3,357
	GN Indéfini	2,878
	Pron Démonstratif	0,719
	GN Démonstratif	0,48
	Pron Relatif	0,24
	Pron Interrogatif	0,24
	N ou GN sans dét	0,24
	<aucune valeur>	

Champ à afficher :

Fonction

	Sujet	49,161
	Déterminant	16,787
	Complément Direct	11,511
	Complément Indirect	10,791
	Complément du nom	4,077

\$3

04 04 01 04 04

04 04 01 04 04

\$4

04 04 00 04 02 04 00 02 04 01

04 04 00 04 02 04 00 02 04 01

exporter en SVG

exporter en SVG

exporter en CSV

Nombre d'unités

position de l'unité

☒ activer les paragraphes (chiffre 0)
 ☒ activer le mode glissant
 ☐ activer l'affichage des singletons

Choisir une statistique :

Trigramme

Choisir un type de champ :

Unité

Choisir un champ de la chaîne :

Catégorie

1 Dét Possessif
 2 GN Défini
 3 GN Démonstratif

Statistique :

Trigramme	Nombre d'occurrences	Fréquence
bbb	40	9,412
0bb	19	4,471
b0b	17	4
ccc	16	3,765
bb0	15	3,529
1bb	13	3,059
bb1	11	2,588
10b	9	2,118
b1b	9	2,118
cc0	9	2,118
b51	8	1,882
b10	7	1,647

Plan

- Exemple de départ, contexte de l'étude et objectifs
- Annotation et visualisation de chaînes de coréférences
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- Statistiques sur les chaînes de coréférences
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- **Discussion : limites et perspectives pour l'ADT**
- Conclusion

Discussion

- Ce qui nous intéresse dans l'étude des chaînes de coréférences
 - Non seulement l'analyse des expressions référentielles (comme le fait déjà l'analyse de discours outillée par la textométrie)
 - Mais aussi l'analyse des successions d'expressions référentielles
- Perspectives envisagées
 - L'analyse de la suite des références d'un texte pourrait bénéficier d'une recherche de motifs (Mellet & Barthélemy 2007, Longrée et al. 2008) avec les mêmes avantages et les mêmes questions ouvertes :
 - Comment tenir compte des séquences textuelles qui ne comportent pas d'expressions référentielles ?
 - Quelle place accorder à la ponctuation ?...
 - Lexicométrie sur corpus étiquetés (Pincemin 2004), considérant le texte comme une suite de mots mais aussi et en même temps une suite d'annotations référentielles et anaphoriques

Plan

- Exemple de départ, contexte de l'étude et objectifs
- Annotation et visualisation de chaînes de coréférences
 - Problèmes et méthodologie
 - Représentations graphiques de chaînes de coréférences
- Statistiques sur les chaînes de coréférences
 - Statistiques générales
 - Statistiques sur les chaînes de coréférences
 - Statistiques sur la suite des références : bi-grammes et tri-grammes
- Discussion : limites et perspectives pour l'ADT
- Conclusion

Conclusion

- Prototype d'interface graphique pour explorer un corpus annotés en expressions référentielles et en chaînes de coréférences
- Choix d'une GUI simple, sans langage de requête ni langage de script
- Dans le cadre du projet ANR Democrat
 - Finaliser Analec avec les fonctionnalités présentées ici
 - Faire un pont entre Analec et TXM