

Annotation manuelle de corpus écrit : *Analec* et analyse outillée.

Frédérique Mélanie-Becquet

Lattice - CNRS ENS Paris3 USPC PSL



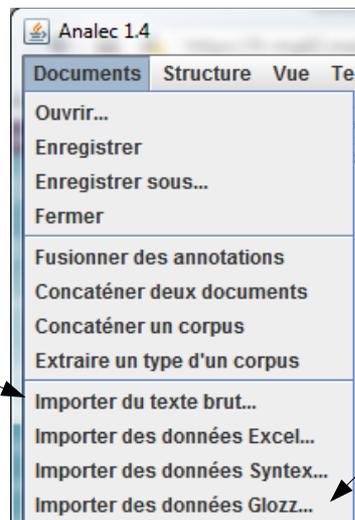
***Collecter et produire des données pour la recherche en SHS
Fréjus, 15-18 novembre 2016***

Les expressions référentielles

MC4 Modélisation Contrastive et Computationnelle des chaînes de Coréférence

texte brut

- 9 textes 20 000 mots
- corpus non homogène
vers ou prose
époques différentes
longueur variable (tout ou partie de l'œuvre)



Genèse Editoriale

Balzac *La bourse*

texte enrichi

- 2 (ou n) éditions
- corpus homogène
- Annotation automatique des phénomène de genèse éditoriale

Genèse Editoriale

Balzac *La bourse*

.ac

La Comédie humaine, Scènes de la vie privée
Paris

La Bourse
1832

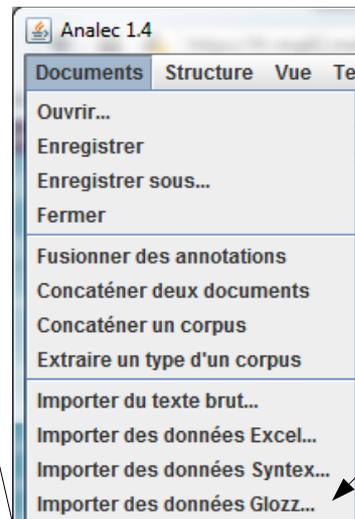
Il est une heure délicieuse aux âmes faciles à s'épanouir, aux âmes fraîches, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore et où le jour n'est plus. La

.aa

```

3513 ▶ <unit id='analec0.2_0'> [17 lines]
3531 ▶ <unit id='analec0.2_0'> [17 lines]
3549 ▶ <unit id='analec0.2_0'> [17 lines]
3567 ▾ <unit id='analec0.2_1'>
3568 ▾ <metadata>
3569 <author>analec0.2</author>
3570 <creation-date>1</creation-date>
3571 </metadata>
3572 ▾ <characterisation>
3573 <type>Seq</type>
3574 ▶ <featureSet> [8 lines]
3583 </characterisation>
3584 ▾ <positioning>
3585 ▾ <start>
3586 <singlePosition index = '56' />
3587 </start>
3588 ▾ <end>
3589 <singlePosition index = '62' />
3590 </end>
3591 </positioning>
3592 </unit>
3593 ▶ <unit id='analec0.2_2'> [25 lines]
3619 ▶ <unit id='analec0.2_3'> [25 lines]
3645 ▶ <unit id='analec0.2_4'> [25 lines]
3671 ▶ <unit id='analec0.2_5'> [25 lines]
3697 ▶ <unit id='analec0.2_6'> [25 lines]
3723 ▶ <unit id='analec0.2_7'> [25 lines]

```



texte enrichi

- 2 (ou n) éditions
- corpus homogène
- Annotation automatique des phénomènes de genèse éditoriale

- Définition d'un schéma d'annotation
- Définition d'une méthode d'annotation
- Exploitation des données annotées

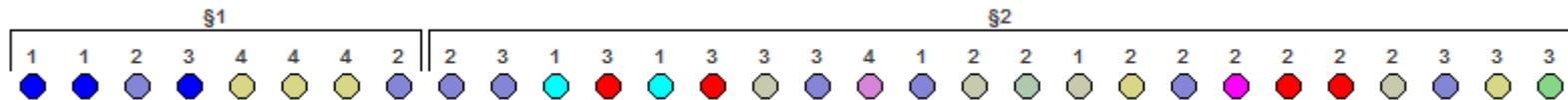
- Définition d'un schéma d'annotation
- Définition d'une méthode d'annotation
- Exploitation des données annotées

Les expressions référentielles

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.

Suite des références :



1 = Sylvie Fabre 2 = Mr. Fabre 3 = Paul Fabre 4 = {Mr. Fabre, Paul Fabre}

Chaîne de coréférence qui concerne Sylvie Fabre : §1 §2

la mère – la mère – sa mère – sa mère – Sylvie Fabre – la

1. Repérer les références

Problème linguistique

2. Délimiter les références

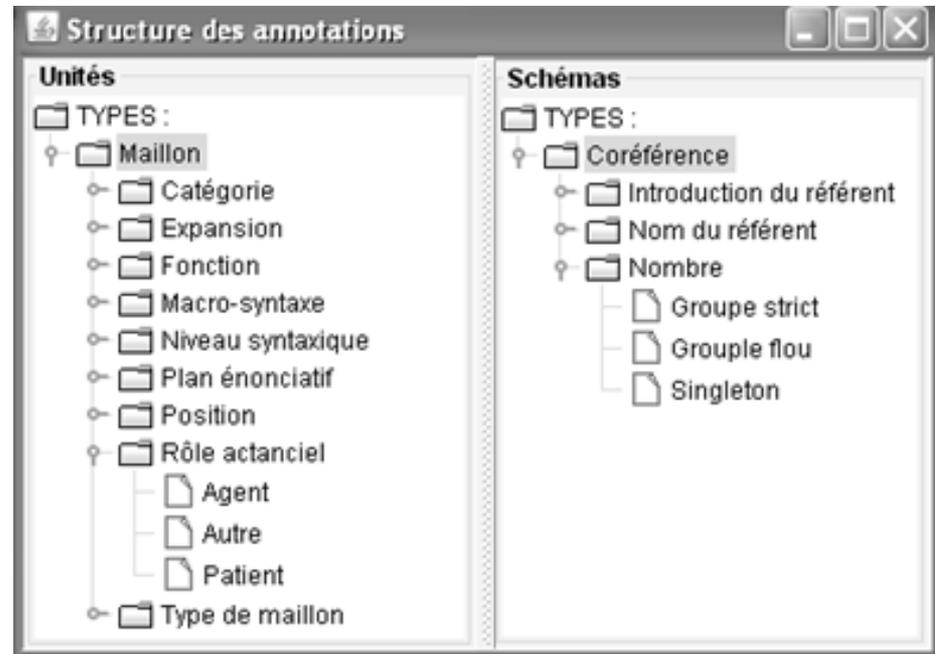
Problèmes d'annotation :

- délimiter les bornes inférieure et supérieure
- problème technique du texte discontinu

3. Ajouter des traits

attribut, valeur

4. Repérer les coréférences



La genèse éditoriale

La Bourse 1832

Honoré de Balzac
La Comédie humaine, Scènes de la vie **privée**
Paris

La Bourse
1832

Il est une heure délicieuse **aux** âmes faciles à s'épanouir, **aux** âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore **et** où le jour n'est plus. La lueur crépusculaire jette ses teintes molles, ses reflets bizarres sur tous les objets, **et, alors**, de douces rêveries **naissent entre ces mille** pièges **confus** que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque pendant cet instant **si fécond en inspirations**, le rend **encore** plus cher aux artistes, aux peintres, aux statuaires. **Alors**, ils se recueillent, se mettent à **dix** pas de leurs œuvres; **et**, ne **pouvant** plus **y** travailler, ils **les** jugent en s'enivrant **de leurs sujets avec délices**.

Celui qui n'est pas demeuré pensif, près d'un ami, dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfiques. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement. **Alors**, l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, **il y a du** sang dans les veines, et les étoffes chatoient. L'imagination aide **merveilleusement** au naturel de chaque détail; **elle** ne voit plus que les beautés de l'œuvre; **et, s'il s'agit d'un tableau, les personnages qu'il représenté** semblent et parler et marcher. A cette heure, l'illusion règne despotiquement; elle se lève avec la nuit; n'est-elle pas pour la pensée une espèce de nuit à

La Bourse 1835

Honoré de Balzac
La Comédie humaine, Scènes de la vie **parisienne**
Paris

La Bourse
1835

Il est une heure délicieuse **pour les** âmes faciles à s'épanouir, âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore, **mais** où le jour n'est plus. La lueur crépusculaire jette **alors** ses teintes molles **et** ses reflets bizarres sur tous les objets. De douces rêveries **ne manquent jamais d'éclorre favorisées par les** pièges que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque **toujours** pendant cet instant fécond en **inspiration** le rend plus **particulièrement** cher aux artistes, aux peintres, aux statuaires. Ils se recueillent, se mettent à **quelques** pas de leurs œuvres **auxquelles ils ne peuvent** plus travailler, **mais qu'ils** jugent en s'enivrant **du sujet dont alors toutes les beautés éclatent aux yeux intérieurs du génie**. Celui qui n'est pas demeuré pensif près d'un ami, dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfiques. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement; **s'il s'agit d'un tableau, les personnages qu'ils représentent** semblent et marcher et parler; l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, **le** sang **coule** dans les veines, et les étoffes chatoient; l'imagination aide au naturel de chaque détail, **et** ne voit plus que les beautés de l'œuvre.

A cette heure, l'illusion règne despotiquement; elle se lève avec

● Remplacements [464]

- 1 privée - parisienne
- 2 1832 - 1835
- 3 aux - pour les
- 4 fraîches - fraîches
- 5 et - , mais
- 6 - et
- 7 naissent entre - ne manquen
- 8 inspirations, - inspiration
- 9 dix - quelques
- 10 ; et, - auxquelles ils
- 11 pouvant - peuvent
- 12 de leurs sujets - du sujet do
- 13 il y a du - le
- 14 ;
- 15 ; elle - , et
- 16 ; et, - ,
- 17 - ,
- 18 et - tout,
- 19 créanciers - misères, les bo
- 20 Donc, à - A
- 21 et - , une
- 22 ; - ,
- 23 ; et, -
- 24 ayant porté - porta
- 25 ; et, - ,
- 26 ; Alors, - ; mais
- 27 le tact des - deux
- 28 ; - mains
- 29 ; - ,
- 30 réalisait, - réalisa
- 31 son - ce
- 32 en a un, - et
- 33 Le - . Ce
- 34 ; de plus, - aussi
- 35 vêtements - vêtements
- 36 fille - inconnue
- 37 celle-ci - la jeune fille
- 38 crimes avoir di - avons cru
- 39 car - et
- 40 ... - ,
- 41 ... - ,
- 42 - -

Précédent Suivant

● Déplacements [26]

1 s'il s'ag	2 semble
4 et	5 marche
7 il	8 silence
10 en le for	11 longue
13 , dont	14 nombre
16 des chene	17 avoir
19 de la	20 jadis

La genèse éditoriale

Medite

<http://www.item.ens.fr/index.php?id=172959>

The Medite interface displays two parallel text windows for comparison. The left window shows the original text with highlighted words in various colors (red, blue, green, yellow). The right window shows the same text with different highlights, indicating a comparison or editing process. The text is from Honoré de Balzac's 'La Comédie humaine, Scènes de la vie privée'.

Original Text (Left):

Honoré de Balzac
La Comédie humaine, Scènes de la vie **privée**
Paris

La Bourse
1832

Il est une heure délicieuse **aux** âmes faciles à s'épanouir, **aux** âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore **et** où le jour n'est plus. La lueur crépusculaire jette ses teintes molles; ses reflets bizarres sur tous les objets; **et, alors**, de douces rêveries **naissent entre ces mille** pièges **confus** que produisent les combats de la lumière et de l'ombre. Le silence qui règne pendant cet instant **si** fécond en **inspirations**, le rend **encore** plus cher aux artistes, aux peintres, aux statuaires. **Alors**, ils se recueillent, se mettent à **dix** pas de leurs œuvres; **et, ne pouvant plus y** travailler, ils **les** jugent en s'enivrant **de leurs** sujets avec délices.

Celui qui n'est pas demeuré pensif, près d'un ami, dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfices. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement. **Alors**, l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, **il y a du** sang dans les veines, et les étoffes chatoient. L'imagination aide

Edited Text (Right):

Honoré de Balzac
La Comédie humaine, Scènes de la vie **parisienne**
Paris

La Bourse
1835

Il est une heure délicieuse **pour les** âmes faciles à s'épanouir, âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore, **mais** où le jour n'est plus. La lueur crépusculaire jette **alors** ses teintes molles; **et** ses reflets bizarres sur tous les objets. De douces rêveries **ne manquent jamais d'éclorre favorisées par les** pièges que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque **toujours** pendant cet instant fécond en **inspiration** le rend plus **particulièrement** cher aux artistes, aux peintres, aux statuaires. Ils se recueillent, se mettent à **quelques** pas de leurs œuvres **auxquelles ils ne peuvent** plus travailler, **mais qu'ils** jugent en s'enivrant **du sujet dont alors toutes les beautés éclatent aux yeux intérieurs du génie**. Celui qui n'est pas demeuré pensif près d'un ami, dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfices. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement; **s'il s'agit d'un tableau, les personnages qu'ils représentent** semblent et marcher et parler; l'ombre devient

Juxta

<http://www.juxtasoftware.org/>

The Juxta interface shows two text windows side-by-side, with a comparison tool overlaid. The tool consists of brackets and lines connecting corresponding words or phrases in both windows, highlighting differences. The text is the same as in the Medite screenshot.

Window 1 (Left): B1832.txt

La Comédie humaine, Scènes de la vie **privée**
Paris

La Bourse
1832

Il est une heure délicieuse **aux** âmes faciles à s'épanouir, **aux** âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore **et** où le jour n'est plus. La lueur crépusculaire jette **alors** ses teintes molles, ses reflets bizarres sur tous les objets; **et, alors**, de douces rêveries **naissent entre ces mille** pièges **confus** que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque pendant cet instant **si** fécond en **inspirations**, le rend **encore** plus cher aux artistes, aux peintres, aux statuaires. **Alors**, ils se recueillent, se mettent à **dix** pas de leurs œuvres; **et, ne pouvant plus y** travailler, ils **les** jugent en s'enivrant **de leurs** sujets avec délices.

Celui qui n'est pas demeuré pensif, près

Window 2 (Right): B1835.txt

La Comédie humaine, Scènes de la vie **parisienne**
Paris

La Bourse
1835

Il est une heure délicieuse **pour les** âmes faciles à s'épanouir, âmes **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore, **mais** où le jour n'est plus. La lueur crépusculaire jette **alors** ses teintes molles **et** ses reflets bizarres sur tous les objets. De douces rêveries **ne manquent jamais d'éclorre favorisées par les** pièges que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque **toujours** pendant cet instant fécond en **inspiration** le rend plus **particulièrement** cher aux artistes, aux peintres, aux statuaires. Ils se recueillent, se mettent à **quelques** pas de leurs œuvres **auxquelles ils ne peuvent** plus travailler, **mais qu'ils** jugent en s'enivrant **du sujet dont alors toutes les beautés éclatent**

La genèse éditoriale

1832

VERS la fin du mois d'octobre dernier, quelques temps après l'heure à laquelle s'ouvrent les maisons de jeu, conformément à la loi qui protège, à Paris, une passion essentiellement budgétifiante, un jeune homme vint au Palais-Royal ; et sans trop hésiter, monta l'escalier du tripot établi au numéro 39.

1835

Vers la fin du mois d'octobre dernier, un jeune homme entra dans le Palais-Royal au moment où les maisons de jeu s'ouvriraient, conformément à la loi qui protège, à Paris, une passion essentiellement productive et chère au fisc. Sans trop hésiter, l'inconnu monta l'escalier du tripot désigné sous le nom de Numéro 39.

La genèse éditoriale

La Bourse

18321835

Il est une heure délicieuse **aux** pour les âmes faciles à s'épanouir, **aux** âmes **fraîches** **fraîches**, toujours jeunes et tendres; cette heure, la plus indécise, la plus variable de toutes celles dont se compose une journée, arrive au moment où la nuit n'est pas encore **et**, **mais** où le jour n'est plus. La lueur crépusculaire jette **alors** ses teintes molles **et** ses reflets bizarres sur tous les objets **et**, **alors**, de douces rêveries **naissent** entre ces mille **ne** manquent jamais d'éclorre favorisées par les pièges **confus** que produisent les combats de la lumière et de l'ombre. Le silence qui règne presque **toujours** pendant cet instant **si** fécond en **inspirations**, **inspiration** le rend **encore** plus **particulièrement** cher aux artistes, aux peintres, aux statuaires. **Alors**, ils se recueillent, se mettent à **dix** **quelques** pas de leurs œuvres **et**, **auxquelles** ils ne **pouvant** **peuvent** plus **y** travailler, **mais** qu'ils **les** jugent en s'enivrant **de** leurs sujets avec **délices**.

du sujet dont **alors** toutes les beautés éclatent aux yeux intérieurs du génie. Celui qui n'est pas demeuré pensif **près** d'un ami, dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfices. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement **Alors**, l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, **il** y a **dû** **le** sang **coule** dans les veines, et les étoffes chatoient **;** L'imagination aide **merveilleusement** au naturel de chaque détail **elle**, **et** ne voit plus que les beautés de l'œuvre **et**, **s'il** s'agit d'un tableau, les personnages qu'**il** **représente** semblent **et** parler **et** marcher **.**

La genèse éditoriale

dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfices. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement . Alors, l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, il y a dule sangcoule dans les veines, et les étoffes chatoient . L'imagination aide merveilleusement au naturel de chaque détail ; elle, et ne voit plus que les beautés de l'œuvre ; et, s'il s'agit d'un tableau, les personnages qu'il représente semblent et parler et marcher .

Structure des annotations

Unités

- TYPES :
 - Seq
 - Chunk
 - Medit
 - déplacement(gris)
 - insertion(vert)
 - modification(bleu)
 - suppression(rose)
 - categorie
 - nbrElem
 - refFich
 - 1832
 - 1835
 - refSeq
 - sequence

dans ce moment de songes poétiques, en comprendra difficilement les indicibles bénéfices. A la faveur du clair-obscur, les ruses matérielles, employées par l'art pour faire croire aux réalités de la vie, disparaissent entièrement . Alors, l'ombre devient ombre, le jour est jour, la chair est vivante, les yeux remuent, il y a dule sangcoule dans les veines, et les étoffes chatoient . L'imagination aide merveilleusement au naturel de chaque détail ; elle, et ne voit plus que les beautés de l'œuvre ; et, s'il s'agit d'un tableau, les personnages qu'il représente semblent et parler et marcher .

Champs de l'unité : "Seq-41"

Sequence			
il y a du			
refFich	Medit	nbrElem	refSeq
1832	modification(bleu)	4	00026
catTT		lemmeTT	
PRO:PER_PRO:PER_VER:pres_PRP:det		il_y_avoir_du	

- Définition d'un schéma d'annotation
- **Définition d'une méthode d'annotation**
- Exploitation des données annotées

Une procédure d'annotation ?

1. Import du corpus de travail

- texte brut
- texte enrichi

2. Annotation manuelle de phénomènes linguistiques

- les expressions référentielles
- la genèse éditoriale

3. Annotation manuelle de chaînes

- les chaînes de coréférence

4. Visualisation des annotations

- utilisation de calculs statistiques

Annotation manuelle des expressions référentielles

- Un ensemble de traits pour chaque expression

The screenshot shows the 'Analec 1.4' software interface. The main window displays a text passage with a highlighted unit 'la mère'. Below the text, the 'Gestion des unités' section shows the unit 'Maillon-1' with a count of 0 and buttons for 'Créer', 'Supprimer', and 'Rectifier une borne'. The 'Champs de l'unité : "Maillon-1"' section contains a grid of dropdown menus for various annotation traits.

Type de maillon	Catégorie	Expansion	Fonction	Niveau syntaxique	
Forme explicite	GN Défini	Aucune	Autre	Autre	
Cas	Macro-syntaxe	Plan énonciatif	Position	Interprétation	Rôle actanciel
Non marqué	Préfixe	Parenthèse / incise / incidente	Médiane	Immédiate	Autre
Introduction du référent	Nom du référent	Nombre-1			
Extraction	Sylvie Fabre (la mère)	Singleton			

Annotation manuelle des chaînes

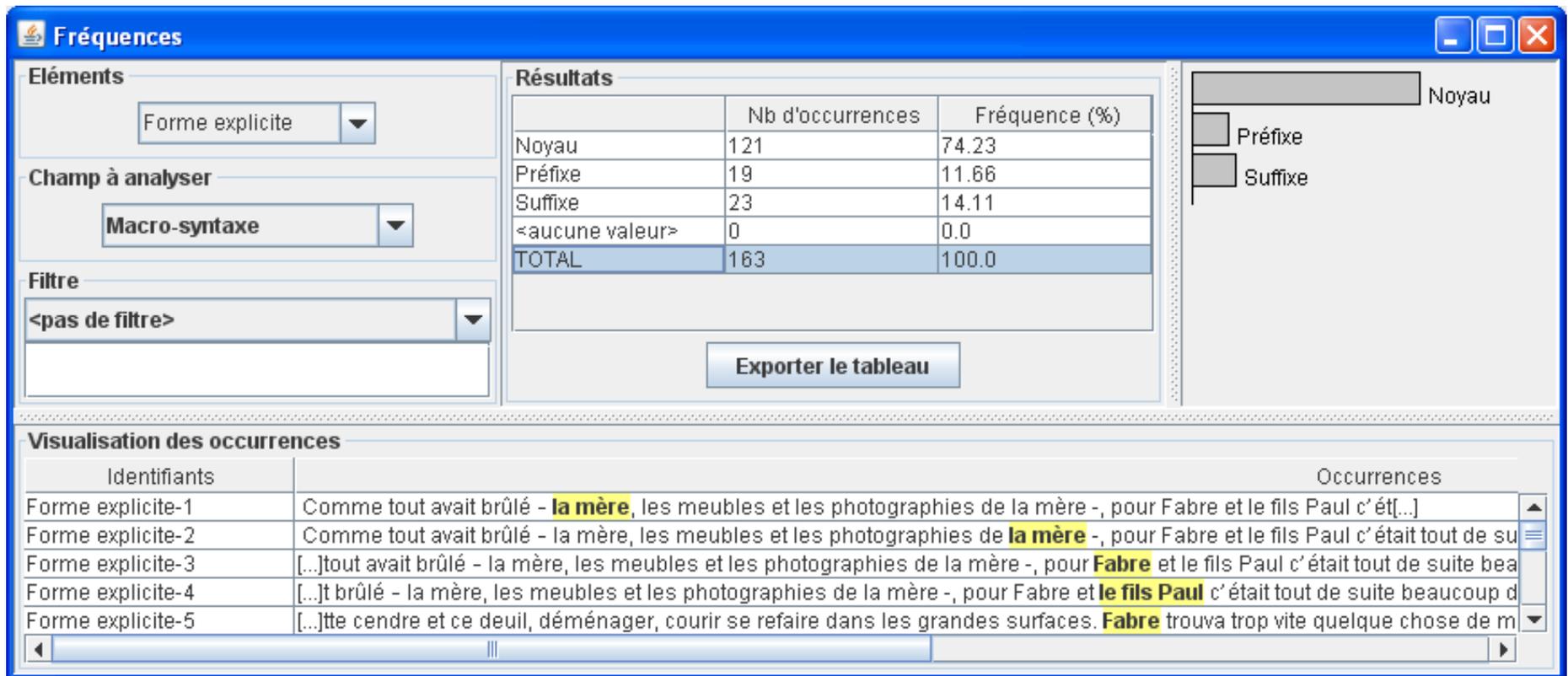
- Un ensemble de traits pour la chaîne
 - identifiant du référent ; type de référent (individu, groupe)
- Et plus particulièrement pour le premier maillon
 - type d'introduction : par regroupement, extraction, association...
 - c'est là que se font les liens anaphoriques non coréférentiels (anaphore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

The screenshot shows the Analec 1.1 software interface. The window title is "Analec 1.1 : Echenoz8 - vue : Echenoz8". The menu bar includes "Documents", "Structure", "Vue", "Texte", "Unités", "Relations", "Schémas", "Règles", and "Statistiques". The "Gestion des schémas" panel is active, showing a schema named "Coréférence" for the text "Sylvie Fabre (la mère)". The "Elément" field is set to "< type d'élément >". The main text area contains two paragraphs from a document by Echenoz8. The first paragraph is: "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." The second paragraph is: "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." The "Champs du schéma : 'Sylvie Fabre (la mère)'" section shows three fields: "Introduction du référent" set to "Extraction", "Nom du référent" set to "Sylvie Fabre (la mère)", and "Nombre" set to "Singleton".

- Définition d'un schéma d'annotation
- Définition d'une méthode d'annotation
- **Exploitation des données annotées**

Calcul de fréquences

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



The screenshot shows the 'Fréquences' window with the following settings and results:

- Eléments:** Forme explicite
- Champ à analyser:** Macro-syntaxe
- Filtre:** <pas de filtre>

	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

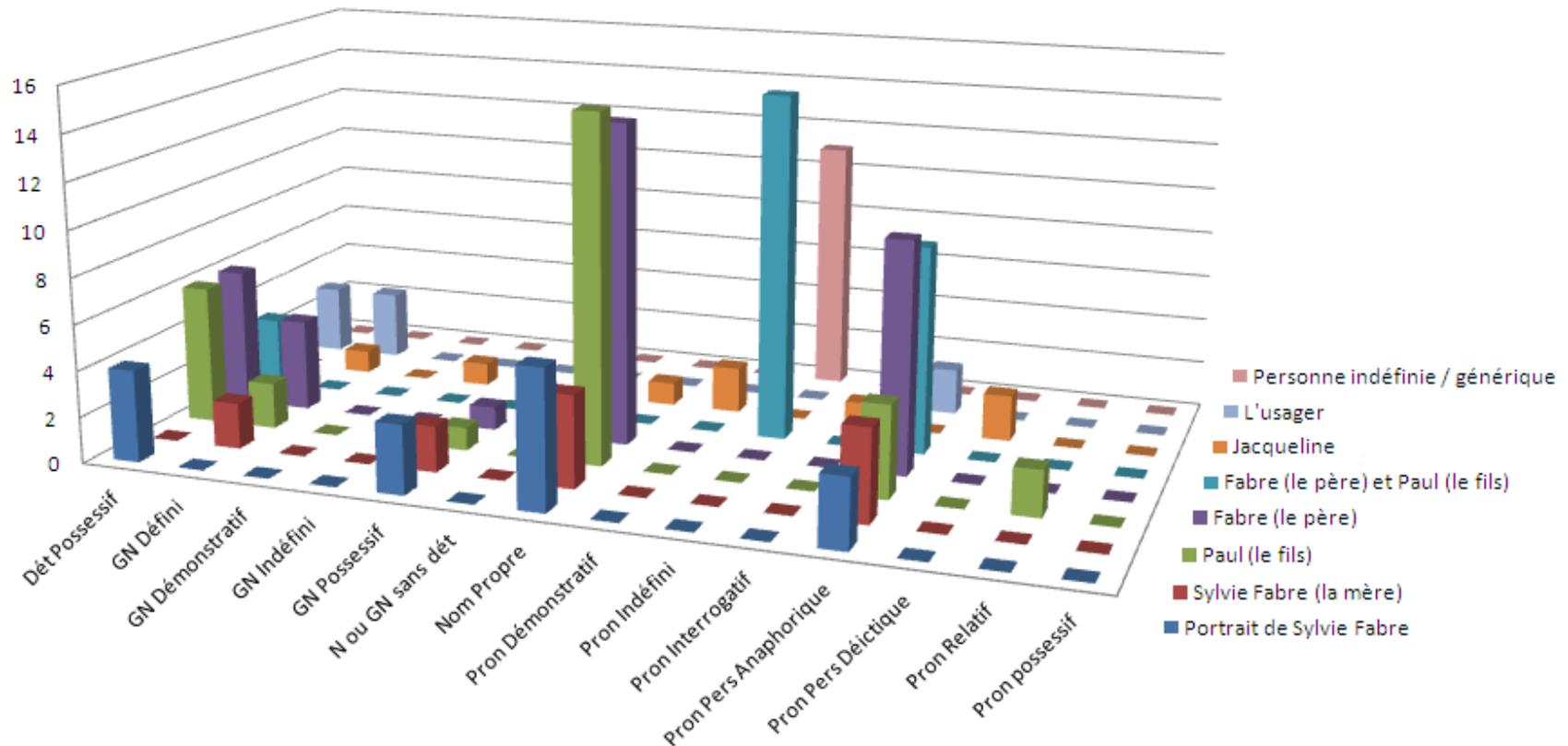
Buttons: Exporter le tableau

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' ét[...]
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

Visualisation de fréquences

- A partir du moment où les fréquences sont calculées, on peut mettre en forme des graphiques



Calcul de corrélations

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

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	0 0 2
Attribut	0 0 0	0 0 0	0 0	0 0 1 0	0 0	0 0 1
Autre	0 0 1	0 0 1	0 2	0 0 0 0	0 0	0 0 4
Circonstant	0 0 1	0 0 1	0 1	0 0 0 4	0 1	0 0 8
Complément Direct	0 0 1	0 0 3	0 6	0 0 0 7	0 0	0 0 17
Complément Indirect	0 0 1	0 1 1	0 1	1 0 0 1	0 0	0 0 6
Complément du nom	0 0 4	0 0 2	0 9	0 0 0 2	0 1	0 0 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 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'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
Forme explicite-150	[...]ize heures - ensuite on s'y met. Un soleil comme celui-ci, développa le père de Paul , donne véritablement envie de foutre le camp. Ils s'exprimèrent égalemen

Visualisation de corrélations

Représentations géométriques

Eléments

Forme explicite ▼

Champs pertinents

- Catégorie
- Expansion
- Fonction
- Niveau syntaxique
- Plan énonciatif
- Macro-syntaxe
- Position
- Interprétation
- Rôle actanciel
- Introduction du référent-1
- Nom du référent-1
- Nombre-1

Mettre à jour le graphe

Graphique

Exporter le graphique en format xml (.svg) Exporter le graphique vers MS-Office (.emf)

Champ à afficher

Catégorie ▼

- 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
- <aucune valeur>

Valeurs typiques

25 éléments sélectionnés

Champ	Valeur	Pourcentage
Catégorie	Dét Possessif	100.0
Expansion	Aucune	100.0
Fonction	Déterminant	100.0
Niveau syntaxique	2e niveau dans pri...	60.0
Plan énonciatif	Plan principal	96.0
Macro-syntaxe	Noyau	68.0
Position	Médiane	84.0
Interprétation	Immédiate	100.0
Rôle actanciel	Autre	96.0
Introduction du r...	Construction discurs...	56.0
Nombre-1	Singleton	84.0

Visualisation des occurrences

Identifiants	Occurrences
Forme explicite-9	[...]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-11	[...]ches 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 représenta
Forme explicite-19	[...] 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 d[...]
Forme explicite-20	[...]écision. 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 cana[...]
Forme explicite-30	[...] 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. L'artiste Flers l'avait représen[...]
Forme explicite-36	[...]teur gravé dans un cartouche en haut à droite. Et le mur sur lequel, avec toute son équipe, l'artiste Flers avait peiné pour figurer Sylvie Fabre en pied, surplomb[...]
Forme explicite-46	[...] Paul, pour qui ce mur n'était qu'une tranche de vie antérieure. Regarde un peu ta mère, s'énervait Fabre que ce spectacle mettait en larmes, en rut, selon. Mais [...]
Forme explicite-49	[...]ement hostile à l'endroit de l'effigie contre laquelle, en écho, rebondissaient ses reproches - Paul s'occupant de modérer le père dès qu'un attroupement menaçait [
Forme explicite-58	[...]rd, suffisamment séparé de Fabre pour qu'on ne se parlât même plus, Paul visita sa mère sur un rythme plus souple, deux ou trois fois par mois, compte non tenu de
Forme explicite-63	[...]ne cabine scellée dans le champ visuel de Sylvie Fabre, il avait failli appeler son père lorsqu'on se mit à démolir la vieille chose insalubre qui jouxtait l'immeu[...]
Forme explicite-71	[...]té, l'usager boycotte cet espace rayé du monde chlorophyllien, n'y délègue plus sa descendance, n'y mène plus déféquer l'animal familial. Le trouvant un matin bar[...]
Forme explicite-73	[...]ne palissade, il cautionne cette quarantaine l'œil sec, sans se questionner sur son initiative ; son cœur est froid, sa conscience pour soi. La palissade se dégrad[...]

Visualisation de corrélations

Représentations géométriques

Eléments

Forme explicite

Champs pertinents

- Catégorie
- Expansion
- Fonction
- Niveau syntaxique
- Plan énonciatif
- Macro-syntaxe
- Position
- Interprétation
- Rôle actanciel
- Introduction du référent-1
- Nom du référent-1
- Nombre-1

Mettre à jour le graphe

Champ à afficher

Fonction

- Apostrophe
- Attribut
- Autre
- Circonstant
- Complément Direct
- Complément Indirect
- Complément du nom
- Dislocation
- Déterminant
- Sujet
- adjectif
- agent
- <aucune valeur>

Valeurs typiques

16 éléments sélectionnés

Champ	Valeur	Pourcentage
Expansion	Aucune	100.0
Fonction	Complément du nom	100.0
Niveau syntaxique	2e niveau dans princi...	75.0
Plan énonciatif	Plan principal	100.0
Macro-syntaxe	Noyau	62.5
Position	Médiane	81.25
Interprétation	Immédiate	100.0
Rôle actanciel	Autre	75.0
Introduction du r...	Association	62.5
Nombre-1	Singleton	93.75

Graphique

Exporter le graphique en format xml (.svg) Exporter le graphique vers MS-Office (.emf)

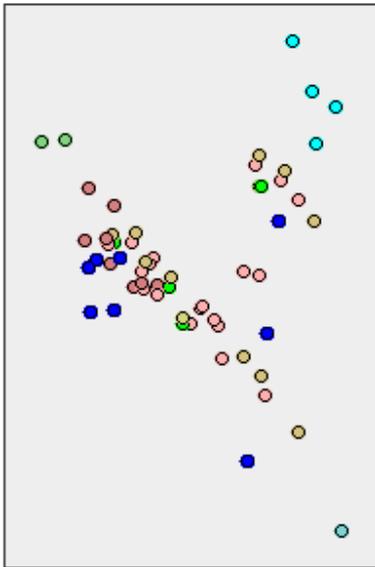
Visualisation des occurrences

Identifiants	Occurrences
Forme explicite-8	[...]assez proches du quai de Valmy. [M] 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	[...]jai de Valmy. [M] 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. [M] L'artiste Flers 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 SyMie. [M] 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 SyMie, on l'approchait d'un autre pas, d'une démarche moins souple. [M] 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

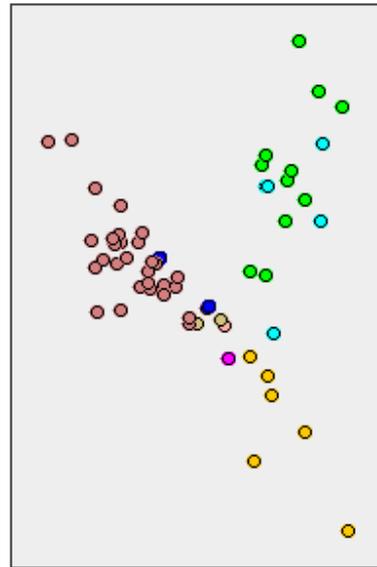
Visualisation de corrélations

- Utilisation 1 : repérer des annotations bizarres, vérifier si tous les champs sont corrects, et corriger si besoin
- Utilisation 2 : repérer des ensembles d'exemples qui se comportent de manière similaire, et identifier les valeurs communes

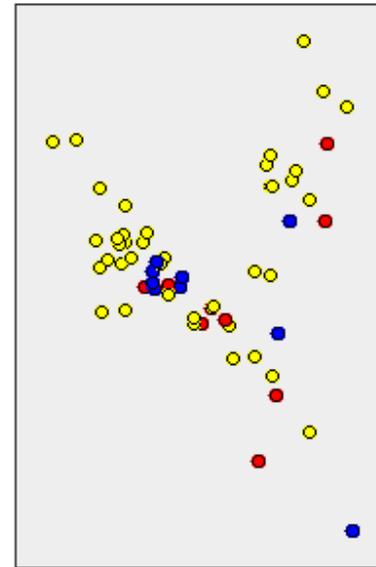
catégories



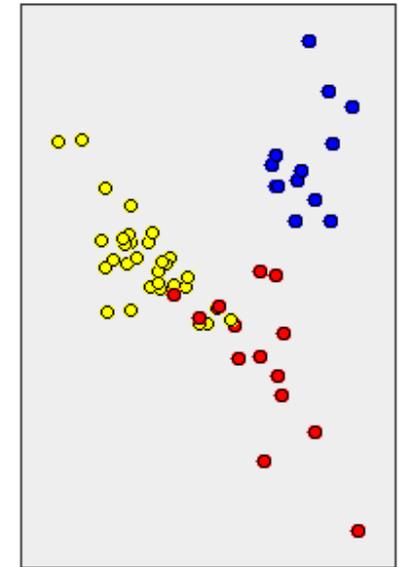
fonctions syntaxiques



macro-syntaxe

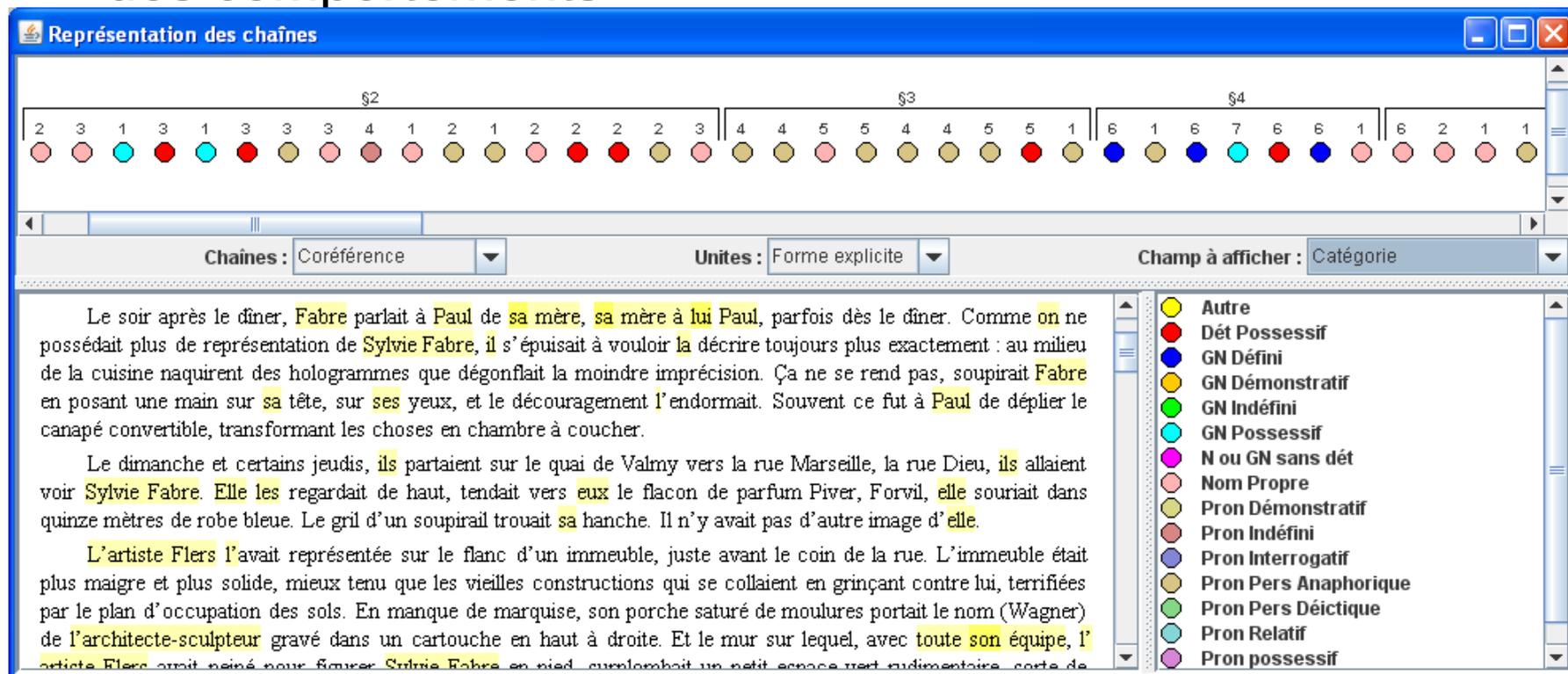


rôles actanciels



Visualiser la suite des références

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



The screenshot shows a software window titled "Représentation des chaînes". At the top, there is a horizontal bar with three sections labeled §2, §3, and §4. Below this bar is a sequence of colored circles, each with a number above it. The colors and numbers correspond to the legend on the right. Below the sequence is a control bar with three dropdown menus: "Chaînes : Coréférence", "Unités : Forme explicite", and "Champ à afficher : Catégorie". The main area of the window contains three paragraphs of text with highlighted words. The legend on the right lists 14 categories with corresponding colored circles.

Code Couleur	Catégorie
Yellow	Autre
Red	Dét Possessif
Blue	GN Défini
Dark Blue	GN Démonstratif
Green	GN Indéfini
Cyan	GN Possessif
Magenta	N ou GN sans dét
Pink	Nom Propre
Light Green	Pron Démonstratif
Brown	Pron Indéfini
Dark Blue	Pron Interrogatif
Light Green	Pron Pers Anaphorique
Light Green	Pron Pers Déictique
Light Blue	Pron Relatif
Purple	Pron possessif

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.

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, surplombeait un petit espace vert rudimentaire, sorte de

Pérennisation et valorisation des données annotées

```
<p>
Lorsque la guerre fut déclarée, <anchor xml:id="u-Maillon-61-start"
type="AnalecDelimiter" subtype="UnitStart"/>le fils Sauvage, <anchor xml:id="u-
Maillon-62-start" type="AnalecDelimiter" subtype="UnitStart"/>qui<anchor xml:id="u-
Maillon-62-end" type="AnalecDelimiter" subtype="UnitEnd"/> avait alors trente-trois
ans<anchor xml:id="u-Maillon-61-end" type="AnalecDelimiter" subtype="UnitEnd"/>,
<anchor xml:id="u-Maillon-63-start" type="AnalecDelimiter"
subtype="UnitStart"/>s'engagea, <anchor xml:id="u-Maillon-63-end"
type="AnalecDelimiter" subtype="UnitEnd"/>engagea, <anchor xml:id="u-Maillon-64-
start" type="AnalecDelimiter" subtype="UnitStart"/>laissant<anchor xml:id="u-
Maillon-64-end" type="AnalecDelimiter" subtype="UnitEnd"/> <anchor xml:id="u-
Maillon-65-start" type="AnalecDelimiter" subtype="UnitStart"/>la mère<anchor
xml:id="u-Maillon-65-end" type="AnalecDelimiter" subtype="UnitEnd"/> seule au
logis. <anchor xml:id="u-Maillon-66-start" type="AnalecDelimiter"
subtype="UnitStart"/>On<anchor xml:id="u-Maillon-66-end" type="AnalecDelimiter"
subtype="UnitEnd"/> ne <anchor xml:id="u-Maillon-67-start" type="AnalecDelimiter"
subtype="UnitStart"/>la<anchor xml:id="u-Maillon-67-end" type="AnalecDelimiter"
subtype="UnitEnd"/> plaignait pas trop, <anchor xml:id="u-Maillon-68-start"
type="AnalecDelimiter" subtype="UnitStart"/>la vieille<anchor xml:id="u-Maillon-68-
end" type="AnalecDelimiter" subtype="UnitEnd"/>, parce qu'elle<anchor xml:id="u-
Maillon-69-start" type="AnalecDelimiter" subtype="UnitStart"/>elle<anchor
xml:id="u-Maillon-69-end" type="AnalecDelimiter" subtype="UnitEnd"/> avait de
l'argent, <anchor xml:id="u-Maillon-70-start" type="AnalecDelimiter"
subtype="UnitStart"/>on<anchor xml:id="u-Maillon-70-end" type="AnalecDelimiter"
subtype="UnitEnd"/> le savait.
</p>
```

```
<span xml:id="u-Maillon-65"  
from="#u-Maillon-65-start" to="#u-Maillon-65-end"  
ana="#u-Maillon-65-fs"/>
```

Balisateur XML d'un maillon.

```
<fs xml:id="u-Maillon-65-fs">  
<f name="Cas"><string>Non marqué</string></f>  
<f name="Fonction"><string>Complément Direct</string></f>  
<f name="Plan énonciatif"><string>Plan principal</string></f>  
<f name="Macro-syntaxe"><string>Suffixe</string></f>  
<f name="Interprétation"><string>Immédiate</string></f>  
<f name="Rôle actanciel"><string>Patient</string></f>  
<f name="Catégorie"><string>GN Défini</string></f>  
<f name="Position"><string>Médiane</string></f>  
<f name="Expansion"><string>Aucune</string></f>  
<f name="Niveau syntaxique"><string>Primaire dans enchâssée</string></f>  
<f name="Type de maillon"><string>Forme explicite</string></f>  
</fs>
```

Balisateur XML des propriétés et valeurs d'un maillon.

```
<join xml:id="s-Coréférence-14"  
target="#u-Maillon-303 #u-Maillon-424 #u-Maillon-336 #u-Maillon-65 #u-Maillon-267  
#u-Maillon-78 #u-Maillon-67 [...] ana="#s-Coréférence-14-fs"/>
```

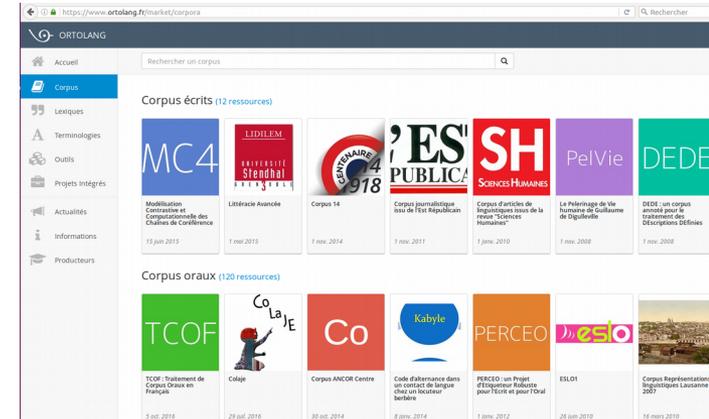
Balisateur XML d'une chaîne de référence.

```
<fs xml:id="s-Coréférence-14-fs">  
<f name="Introduction du référent"><string>Association</string></f>  
<f name="Nombre"><string>Singleton</string></f>  
<f name="Nom du référent"><string>La mère Sauvage</string></f>  
</fs>
```

Balisateur XML des propriétés et valeurs d'une chaîne de référence.

Conclusion :

- Constituer un corpus de référence
- Faire des requêtes (interroger) le corpus annoté
- Déterminer et tester la procédure d'annotation
 - rationnelle, applicable à d'autres états de langue ... et liste de problèmes (de décision, de temps, d'ergonomie, techniques)
 - calculs de fréquences sont très utiles mais restent simples
- ... Vers une annotation automatique



Merci

Bibliographie :

- Fort K., **Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus**, thèse, 2012.
- Fuchs C., Grésillon A., Lebrave J.-L., Peytard J., Rey-Debove J., Culioli A., **La Genèse du texte : les modèles linguistiques** / Préface d'Antoine Culioli, Paris, Éditions du CNRS, « Textes et Manuscrits », 1982, 175p.
- Ganasia J.-G., Fenoglio I. et Lebrave J.-L., **Manuscrits, genèse et documents numérisés. EDITE : une étude informatisée du travail de l'écrivain**, *Document numérique*, 2004/4, Volume 8, pp.91-110.
- Landragin F., **Anaphores et coréférences : analyse assistée par ordinateur**, *Nouvelles perspectives sur l'anaphore. Points de vue linguistique, psycholinguistique et acquisitionnel*, Peter Lang, 2014.
- Landragin F., Poibeau T. & Victorri B., **ANALEC: a New Tool for the Dynamic Annotation of Textual Data**. *Eighth International Conference on Language Resources and Evaluation*, Istanbul, Turquie, 2012, pp. 357-362.
- Mélanie-Becquet F. & Landragin F. , **Linguistique outillée pour l'étude des chaînes de référence : questions méthodologiques et solutions techniques.**, *Langages* 195, Armand Colin, Paris, 2014, pp. 117-137.
- Widlöcher A., Mathet Y. (2009), **La plate-forme Glozz : environnement d'annotation et d'exploration de corpus**, Actes de la conférence *TALN 2009*, Senlis.