

CONCEPTION ET DÉVELOPPEMENT D'UNE APPLICATION D'ANNOTATION THÉMATIQUE

dans l'environnement Gate

Charles Follet Roland Bary

Université de Pau et des Pays de l'Adour

- Annotation sémantique : phase en amont de la recherche d'information sémantique

- Annotation sémantique : phase en amont de la recherche d'information sémantique
- Utilisation de la plate-forme GATE Developer

- Annotation sémantique : phase en amont de la recherche d'information sémantique
- Utilisation de la plate-forme GATE Developer
- Problématique : comment annoter de manière automatique un document texte non-structuré ?

Compréhension du domaine

Annotations dans GATE

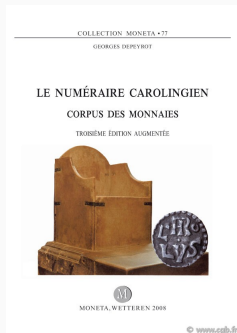
Mise en œuvre

Visualisation

Démonstration

COMPRÉHENSION DU DOMAINE

Recherches sur le Numéraire Carolingien de Georges DEPEYROT.



Pouvoir répondre aux questions suivantes:

- Temporelles : Quelles étaient les pièces en circulation de l'an 859 à l'an 865 ?
- Thématiques : Combien d'exemplaires de la monnaie d'or de Charles le Chauve ont été étudiés ?
Dans quels ateliers, les pièces de type Obole de Charlemagne ont été produites ?

Etape d'OCRisation du numéraire.



Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; **Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50)** Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) **Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);**

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Agen (Lot-et-Garonne)^{ATELIER}

Type de 793/4-812: Charlemagne (768-814)^{TYPE}

Denier de Charlemagne (24 exemplaires étudiés)^{NATURE}

+ CARLVX REX FR croix + AGINNO monogramme^{LÉGENDE}

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50)^{COLLECTIONS} Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex.

(MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);^{TRÉSORS}

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26)^{TROUVAILLES}.

Texte en entrée de notre chaîne de traitement

Agen (Lot-et-Garonne)

Type de 793/4-812: Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Texte en sortie de notre chaîne de traitement

Domaine annoté : Temporel .

Agen (Lot-et-Garonne)

Type de 793/4-812 : Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVS REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45; Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735 (1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60), 794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex. (MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

Texte en sortie de notre chaîne de traitement

Domaine annoté : Thématique .

Agen (Lot-et-Garonne)

Type de 793/4-812 : Charlemagne (768-814)

Denier de Charlemagne (24 exemplaires étudiés)

+ CARLVX REX FR croix + AGINNO monogramme

Gariel XXII 26; Morrison-Grunthal 177-179; Depeyrot (1) (2) 1; Collections: Berlin 1,64, 1,45;

Bruxelles 1,25, 1,25; Charleville-Mézières 1,56; Copenhague 1,31; Garrett 1,73; Grenoble; MEC 735

(1,57), 736 (1,24); Monnaie de Paris 88 (1,55); New York 1,64, 1,57; Prou 792 (1,36), 793 (1,60),

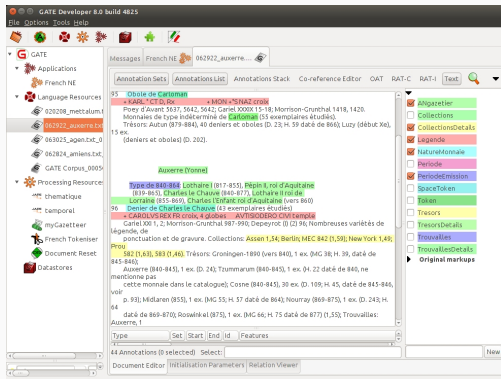
794 (1,50) Trésors: Biebrich (790-814), 2 ex. (MG 11, Völkers, p. 182) (1,60); Ibersheim (814), 1 ex.

(MG 13, Völkers, p. 186) (1,64); Dorestad (822), 4 ex. (MG 18, Völkers, p. 139; H. 7) (1,35);

Trouvailles: Bolsward, 1 ex. (H. 528) (1,26).

ANNOTATIONS DANS GATE

ANNOTATIONS DANS GATE



- Outil de traitement automatique du langage (en JAVA)
- Université de Sheffield depuis 1995 (version 8)
- Principe: Chaîne de traitement (Pipeline)
- Plusieurs modules dans un pipeline
- ANNIE: Pipeline par défaut pouvant servir de chaîne de départ

ANNOTATIONS DANS GATE

Châînes de traitement

Corpus à annoter

Modules de traitement

Composition d'une chaîne de traitement

Name	Type
Document Reset	Document Reset PR
French Tokeniser	GATE Unicode Tokeniser
myGazetteer	ANNE Gazetteer
temporel	JAPE Transducer
thematique	JAPE Transducer

Corpus: GATE Corpus_0005C

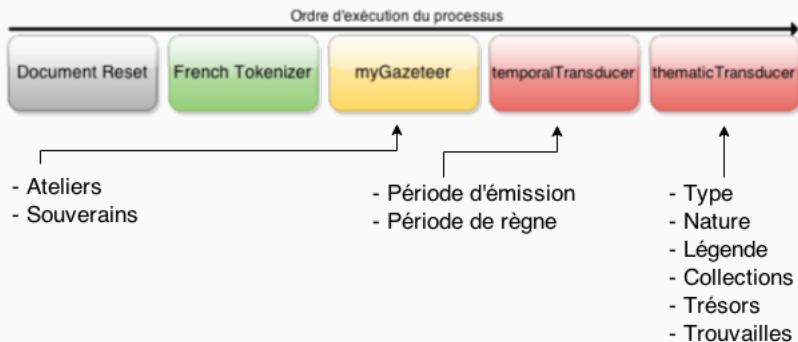
No selected processing resource

Name | Type | Required | Value

Run this Application

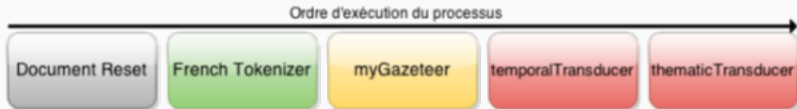
Serial Application Editor Initialisation Parameters

La chaîne de traitement LANG (L'Annotation du Numéraire dans Gate)

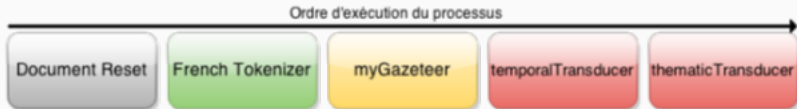




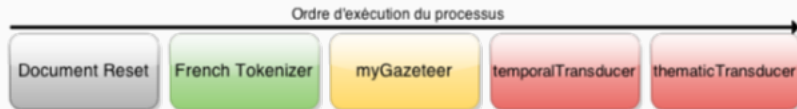
- Document reset : Réinitialisation du document



- Document reset : Réinitialisation du document
- French Tokeniser : Découpage des phrases en mots



- Document reset : Réinitialisation du document
- French Tokeniser : Découpage des phrases en mots
- MyGazetter : Identification de noms d'entités par Gazetier



- Document reset : Réinitialisation du document
- French Tokeniser : Découpage des phrases en mots
- MyGazetter : Identification de noms d'entités par Gazetier
- temporal/thematic Transducer : Repérage d'entités nommées inconnues en fonction de modèles d'extraction écrits en langage JAPE

List name	Major	Minor	Language	Annotation type	Value	Feature 2	Value 2
rois.lst	rois			ANGazetier	Adalbert Lothaire	periode	954-986
					Amoul roi de Germanie	periode	887-899
					Bérenger I	periode	888-924
					Bérenger II	periode	950-961
					Boson roi de Provence	periode	879-887

- Liste de noms d'entités

List name	Major	Minor	Language	Annotation type	Value	Feature 2	Value 2
rois.lst	rois			ANGazetier	Adalbert Lothaire	periode	954-986
					Amoul roi de Germanie	periode	887-899
					Bérenger I	periode	888-924
					Bérenger II	periode	950-961
					Boson roi de Provence	periode	879-887

- Liste de noms d'entités
- Permettant un repérage d'occurrences dans un texte

List name	Major	Minor	Language	Annotation type	Value	Feature 2	Value 2
rois.lst	rois			ANgazetier	Adalbert Lothaire	periode	954-986
					Amoul roi de Germanie	periode	887-899
					Bérenger I	periode	888-924
					Bérenger II	periode	950-961
					Boson roi de Provence	periode	879-887

- Liste de noms d'entités
- Permettant un repérage d'occurrences dans un texte
- Possibilité d'associer des caractéristiques à chaque nom d'entité

List name	Major	Minor	Language	Annotation type	Value	Feature 2	Value 2
rois.lst	rois			ANGazetier	Adalbert Lothaire	periode	954-986
					Amoul roi de Germanie	periode	887-899
					Bérenger I	periode	888-924
					Bérenger II	periode	950-961
					Boson roi de Provence	periode	879-887

- Liste de noms d'entités
- Permettant un repérage d'occurrences dans un texte
- Possibilité d'associer des caractéristiques à chaque nom d'entité
- Dans Gate: MajorType, MinorType réutilisables dans une règle *JAPE*


```
Macro: CHAINE_DEBUT
(
  ({Token.string == "Type"})({SpaceToken})
  ({Token.string == "de"})({SpaceToken})
)
```

```
// Exemple : Type de 758/9-762
```

```
Rule: PeriodeEmissionRule
```

```
/****** Left Hand Side *****/
```

```
(
  CHAINE_DEBUT ({Periode}):p
):PeriodeEmission
/****** Right Hand Side *****/
-->:PeriodeEmission.PeriodeEmission = {
  Kind = "PeriodeEmission" ,
  D1 = :p.Periode.D1, D2 = :p.Periode.D2
}
/******
```

- Règles JAPE en deux blocs
- Partie gauche (LHS): définition d'un motif d'annotation à repérer
- Partie droite (RHS): Opérations sur un motif (Code JAVA éventuel)
- Utilisation d'un système de Macros

MISE EN ŒUVRE

1. Langage naturel

français

1. Langage naturel
2. Langage régulier

français
expressions régulières

1. Langage naturel
2. Langage régulier
3. Langage de programmation

français
expressions régulières
JAPE

Annotation : La période d'émission

Exemple 1 : *Type de 793/4-812*

Type de 793/4-812

1. LANGAGE NATUREL

Une période est un intervalle entre deux dates. Dans notre travail, les dates sont constituées de trois chiffres. En cas d'ambiguïté, une date peut être suivie d'un / et d'un chiffre traduisant l'indétermination de la date.

Il faut assurer la capture de la période d'émission en ajoutant la contrainte précédée de "Type de" afin de ne pas récupérer toutes les périodes du document.

Type de 793/4-812

1. LANGAGE NATUREL

Une période est un intervalle entre deux dates. Dans notre travail, les dates sont constituées de trois chiffres. En cas d'ambiguïté, une date peut être suivie d'un / et d'un chiffre traduisant l'indétermination de la date.

Il faut assurer la capture de la période d'émission en ajoutant la contrainte précédée de "Type de" afin de ne pas récupérer toutes les périodes du document.

Type de 793/4-812

2. LANGAGE RÉGULIER

Type de : $([0-9]\{3\}(/[0-9])?)-([0-9]\{3\}(/[0-9])?)$

Type de 793/4-812

2. LANGAGE RÉGULIER

Type de : $([0-9]\{3\}(/[0-9])?)-([0-9]\{3\}(/[0-9])?)$

Type de 793/4-812

2. LANGAGE RÉGULIER

Type de : $([0-9]\{3\}(/[0-9])?)-([0-9]\{3\}(/[0-9])?)$

Type de 793/4-812

2. LANGAGE RÉGULIER

Type de : $([0-9]\{3\}(/[0-9])?)-([0-9]\{3\}(/[0-9])?)$

Type de 793/4-812

3. LANGAGE DE PROGRAMMATION

Macro: CHAINE_DEBUT

```
(  
  ({Token.string == "Type"})({SpaceToken})  
  ({Token.string == "de"})({SpaceToken})  
)
```

Macro: TROIS_NOMBRES

```
({Token.kind==number,Token.length == 3})
```

Macro: UN_NOMBRE

```
({Token.kind==number,Token.length == 1})
```

Macro: SLASH

```
({Token.string=="/"})
```

Macro: DATE_PRECISE

```
(TROIS_NOMBRES)
```

Macro: DATE_IMPRECISE

```
(TROIS_NOMBRES SLASH UN_NOMBRE)
```

Macro: DATE

```
(DATE_PRECISE | DATE_IMPRECISE)
```

Rule: PeriodeRule

```
((DATE):d1({Token.string == "-"})(DATE):d2)  
:Periode -->:Periode{"/.../"}
```

Rule: PeriodeEmissionRule

```
(CHAINE_DEBUT ({Periode}):p)
```

:PeriodeEmission

```
-->:PeriodeEmission.PeriodeEmission = {"/.../"}
```

Type de 793/4-812

3. LANGAGE DE PROGRAMMATION

```
Macro: CHAINE_DEBUT
(
  ({Token.string == "Type"})({SpaceToken})
  ({Token.string == "de"})({SpaceToken})
)
```

```
Macro: TROIS_NOMBRES
({Token.kind==number,Token.length == 3})
Macro: UN_NOMBRE
({Token.kind==number,Token.length == 1})
Macro:SLASH
({Token.string=="/"})
Macro:DATE_PRECISE
(TROIS_NOMBRES)
Macro:DATE_IMPRECISE
(TROIS_NOMBRES SLASH UN_NOMBRE)
Macro:DATE
(DATE_PRECISE | DATE_IMPRECISE)
```

```
Rule: PeriodeRule
(({DATE):d1({Token.string == "-"})({DATE):d2})
:Periode -->:Periode{"/"..."/}
```

```
Rule: PeriodeEmissionRule
(CHAINE_DEBUT ({Periode}):p)
:PeriodeEmission
-->:PeriodeEmission.PeriodeEmission = {"/"..."/}
```

Type de 793/4-812

3. LANGAGE DE PROGRAMMATION

```
Macro: CHAINE_DEBUT
(
  ({Token.string == "Type"})({SpaceToken})
  ({Token.string == "de"})({SpaceToken})
)
```

```
Macro: TROIS_NOMBRES
({Token.kind==number,Token.length == 3})
Macro: UN_NOMBRE
({Token.kind==number,Token.length == 1})
Macro: SLASH
({Token.string=="/"})
Macro: DATE_PRECISE
(TROIS_NOMBRES)
Macro: DATE_IMPRECISE
(TROIS_NOMBRES SLASH UN_NOMBRE)
Macro: DATE
(DATE_PRECISE | DATE_IMPRECISE)
```

```
Rule: PeriodeRule
((DATE):d1({Token.string == "-"})(DATE):d2)
:Periode -->:Periode{/"..."/}
```

```
Rule: PeriodeEmissionRule
(CHAINE_DEBUT ({Periode}):p)
:PeriodeEmission
-->:PeriodeEmission.PeriodeEmission = {/"..."/}
```

Type de 793/4-812

3. LANGAGE DE PROGRAMMATION

```
Macro: CHAINE_DEBUT
(
  ({Token.string == "Type"})({SpaceToken})
  ({Token.string == "de"})({SpaceToken})
)
```

```
Macro: TROIS_NOMBRES
({Token.kind==number,Token.length == 3})
Macro: UN_NOMBRE
({Token.kind==number,Token.length == 1})
Macro:SLASH
({Token.string=="/"})
Macro:DATE_PRECISE
(TROIS_NOMBRES)
Macro:DATE_IMPRECISE
(TROIS_NOMBRES SLASH UN_NOMBRE)
Macro:DATE
(DATE_PRECISE | DATE_IMPRECISE)
```

```
Rule: PeriodeRule
((DATE):d1({Token.string == "-"})(DATE):d2)
:Periode -->:Periode{ /*...*/ }
```

```
Rule: PeriodeEmissionRule
(CHAINE_DEBUT ({Periode}):p)
:PeriodeEmission
-->:PeriodeEmission.PeriodeEmission = { /*...*/ }
```


Annotation : La nature de la monnaie

Exemple 2 : *Obole de Carloman*

Obole de Carloman

1. **LANGAGE NATUREL**

La nature de la monnaie est toujours un élément de l'ensemble {Denier, Obole, Monnaie D'Or, Faux obole, Monnaies de type indéterminé} suivi du nom d'un souverain. Il suffit donc d'utiliser une expression composée de tous les mots de l'ensemble.

Obole de Carloman

2. LANGAGE RÉGULIER

$[\backslash s]{2,}(\text{Denier}|\text{Obole}|\text{Monnaie d'or}|\text{Faux Obole}|\text{ Monnaies de type indéterminé}) (.*)$

Nom_Souverain

Obole de Carloman

2. LANGAGE RÉGULIER

$[\backslash s]{2,}(\text{Denier}|\text{Obole}|\text{Monnaie d'or}|\text{Faux Obole}|\text{ Monnaies de type indéterminé}) (.*)$

Nom_Souverain

Obole de Carloman

2. LANGAGE RÉGULIER

$[\backslash s]{2,}(\text{Denier}|\text{Obole}|\text{Monnaie d'or}|\text{Faux Obole}|\text{ Monnaies de type indéterminé}) (.*)$

Nom_Souverain

Obole de Carloman

3. LANGAGE DE PROGRAMMATION

```
Macro: MONNAIE
(
  ({Token.string == "Obole"})|
  ({Token.string == "Denier"})|
  ({Token.string == "Monnaie_d'or"})|
  ({Token.string == "Faux_Obole"})|
  ({Token.string == "Monnaies_de_type_indéterminé"})
)
```

```
Macro: LIAISON
({Token.string == "de"})
```

```
Rule: NatureMonnaieRule
(
  ({SpaceToken})?(MONNAIE):monnaie
  ({SpaceToken})(LIAISON)({SpaceToken})
  ({ANGazetier.majorType == "rois"})
):NatureMonnaie
-->
:NatureMonnaie.NatureMonnaie = {/* ... */}
```

Obole de Carloman

3. LANGAGE DE PROGRAMMATION

Macro: MONNAIE

```
(
  ({Token.string == "Obole"})|
  ({Token.string == "Denier"})|
  ({Token.string == "Monnaie_d'or"})|
  ({Token.string == "Faux_Obole"})|
  ({Token.string == "Monnaies_de_type_indéterminé"})
)
```

Macro: LIAISON

```
({Token.string == "de"})
```

Rule: NatureMonnaieRule

```
(
  ({SpaceToken})?(MONNAIE):monnaie
  ({SpaceToken})(LIAISON)({SpaceToken})
  ({ANGazetier.majorType == "rois"})
):NatureMonnaie
-->
:NatureMonnaie.NatureMonnaie = {/" ... "/}
```

Obole de Carloman

3. LANGAGE DE PROGRAMMATION

```
Macro: MONNAIE
(
  ({Token.string == "Obole"})|
  ({Token.string == "Denier"})|
  ({Token.string == "Monnaie_d'or"})|
  ({Token.string == "Faux_Obole"})|
  ({Token.string == "Monnaies_de_type_indéterminé"})
)
```

```
Macro: LIAISON
({Token.string == "de"})
```

```
Rule: NatureMonnaieRule
(
  ({SpaceToken})?(MONNAIE):monnaie
  ({SpaceToken})(LIAISON)({SpaceToken})
  ({ANGazetier.majorType == "rois"})
):NatureMonnaie
-->
:NatureMonnaie.NatureMonnaie = {/* ... */}
```


Résultat d'annotation

Fichier XML :

<PeriodeEmission>Type de 793/4-812</PeriodeEmission>

<NatureMonnaie>Obole de Carloman</NatureMonnaie>

Résultat d'annotation

Fichier XML :

<PeriodeEmission>Type de 793/4-812</PeriodeEmission>

<NatureMonnaie>Obole de Carloman</NatureMonnaie>

Comment visualiser les annotations en dehors de GATE ?

Résultat d'annotation

Fichier XML :

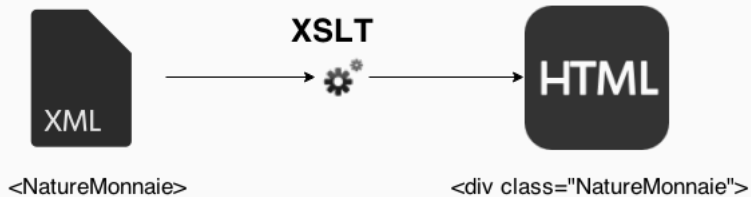
<PeriodeEmission>Type de 793/4-812</PeriodeEmission>

<NatureMonnaie>Obole de Carloman</NatureMonnaie>

Comment visualiser les annotations en dehors de GATE ?

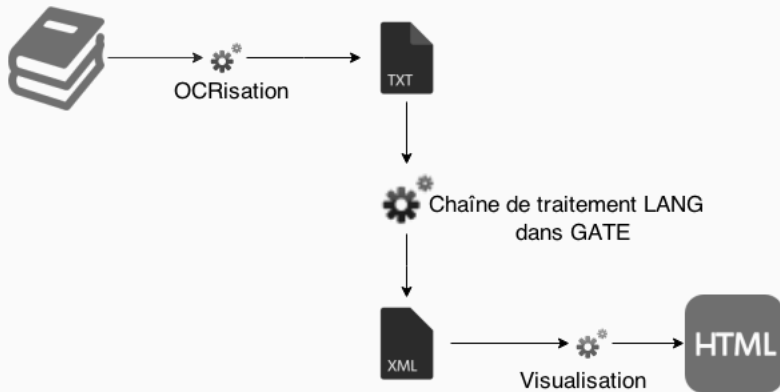
XSLT

VISUALISATION



DÉMONSTRATION

CONCLUSION



- Chaîne de traitement et visualisation fonctionnelles

- Chaîne de traitement et visualisation fonctionnelles
- Besoins du cahier de charges satisfait

- Chaîne de traitement et visualisation fonctionnelles
- Besoins du cahier de charges satisfait
- Améliorations possibles sur :

- Chaîne de traitement et visualisation fonctionnelles
- Besoins du cahier de charges satisfait
- Améliorations possibles sur :
 - L'interface web de visualisation

- Chaîne de traitement et visualisation fonctionnelles
- Besoins du cahier de charges satisfait
- Améliorations possibles sur :
 - L'interface web de visualisation
 - Performance de la chaîne de traitement