

Cascades (transcriptions balisées)

Aurélie Belliard
Denis Maurel

PLAN GENERAL



Deux cascades

- Cascade numéro 1
 1. Graphes de balises
 2. Construction de l'identifiant
 3. Balisage interne
 4. Balisage global
 5. Extension du balisage
- Cascade numéro 2
 - Reconstruction des balises XML TEI



Les graphes d'Unitex

- Les graphes d'Unitex permettent
 - de reconnaître certaines parties du document
 - d'y introduire du texte
 - de supprimer du texte
 - d'introduire un « balisage »
 - ouvert et fermé par des accolades
 - avec une catégorie
 - et des traits

Exemple

{partie reconnue,.catégorie+traits+nom du graphe}

```
<placeName type="ville" subtype="fictif"
key="#locf_samagarie">,.baliseXml+nomDplaceName+ville+fictif+key+
grfBaliseXmlPersGeogPlaceName}
```

- La partie reconnue sera ignorée dans la suite, excepté par sa catégorie et ses traits
- La catégorie et les traits peuvent être repris par les graphes suivants
- Le nom du graphe sert au débogage: il permet de vérifier quel graphe a créé le balisage

GRAPHES DE BALISES

Graphes de balise

➤ 3 exemples

1. `<lb/>`

2. `{<lb/>, .baliseXml+nomDFIb+grfbaliseXmlNom}`

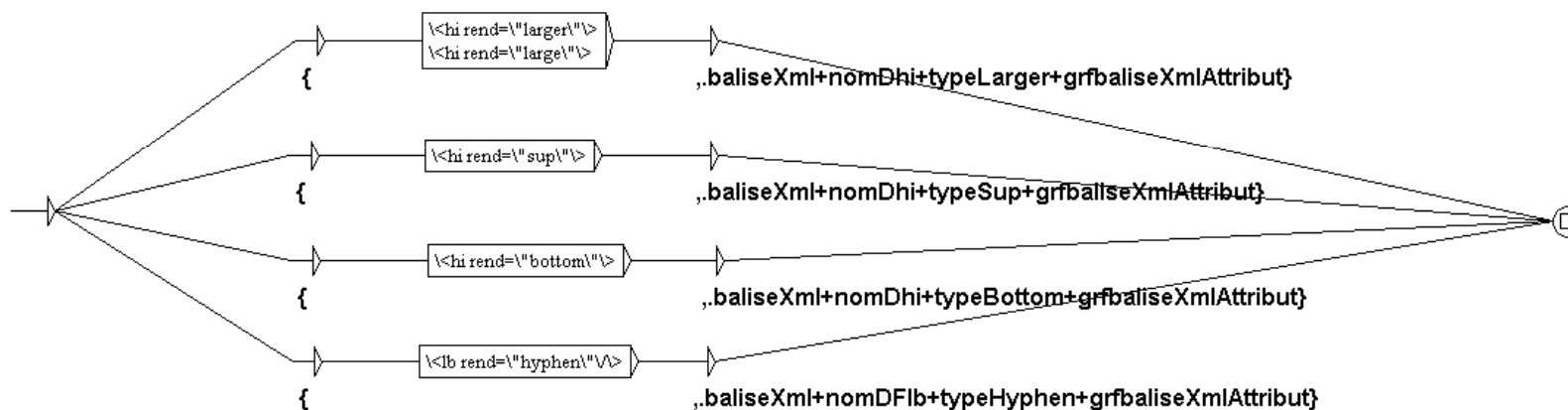
1. `<lb rend="hyphen"/>`

2. `{<lb rend="hyphen"/>, .baliseXml+nomDFIb+typeHyphen+grfbaliseXmlAttribut}`

1. `<placeName key="#locf_samagarie" type="ville" subtype="fictif"> Samagarie </placeName>`

2. `{<placeName type="ville" subtype="fictif" key="#locf_samagarie">, .baliseXml+nomDplaceName+ville+fictif+key+grfBaliseXmlPersGeogPlaceName} Samagarie
{</placeName>, .baliseXml+nomFplaceName+grfBaliseXmlPersGeogPlaceName}`

Graphe des balises <hi> et <lb>



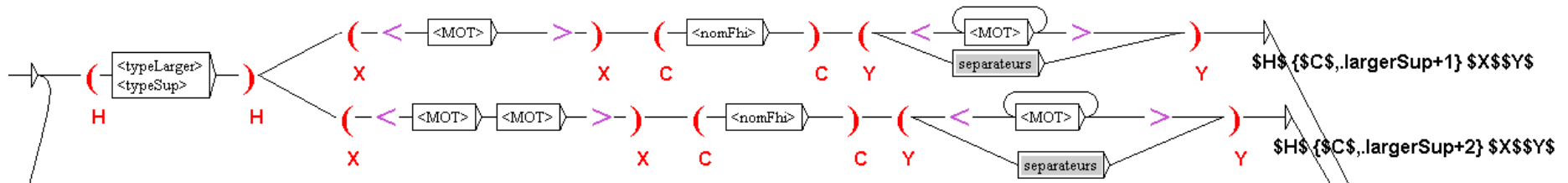
Reconstruction d'un mot

- **Balise <lb rend="hyphen"/>**
 1. <placeName key="#loc_normandie"> Nor <lb rend="hyphen"/> mandie </placeName>
 2. <placeName key="#loc_normandie"> {<lb rend="hyphen"/>, .hyphen+3+grfmotHyphen} Normandie </placeName>

- **Balise <hi rend="larger">**
 1. <hi rend="larger"> C </hi> Es
 2. <hi rend="larger"> {{</hi>\, \.baliseXml\+nomFhi\+grfbaliseXmlNom\}, .largerSup+1} CEs

- **Balise <hi rend="bottom">**
 1. popu- <lb rend="hyphen"/> <hi rend="bottom"> (laire. </hi>
 2. {<lb rend="hyphen"/><hi rend="bottom">, .bottom+4} populaire. </hi>

Graphe de reconstruction: balise <hi rend="larger">

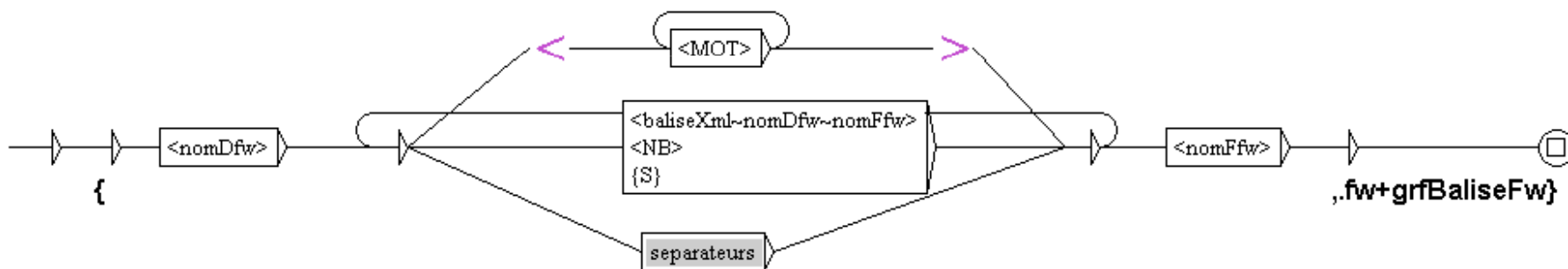


Balises <fw>, correction, régularisation et abbréviation

➤ Encadrement des balises pour en masquer le contenu

1. `<fw place="top-left" type="pageNum"> [97v] </fw>`
 2. `{{<fw place="top-left" type="pageNum">\,\.baliseXml\+nomDfw\+grfbaliseXmlNom\} [97v] {</fw>\,\.baliseXml\+nomFfw\+grfbaliseXmlNom\},.fw+grfBaliseFw}`
-
1. `<choice> <orig> dentre </orig> <reg> d'entre </reg> </choice>`
 2. `{{<choice>} {<orig>} dentre {</orig>} {<reg>},.regularisation} d'entre {</reg>} {</choice>}`

Graphe <fw>



CONSTRUCTION DE L'IDENTIFIANT

Construction de l'identifiant : avec consultation du dictionnaire

➤ Construction de l'identifiant « **key** » par rapport à l'entrée du dictionnaire

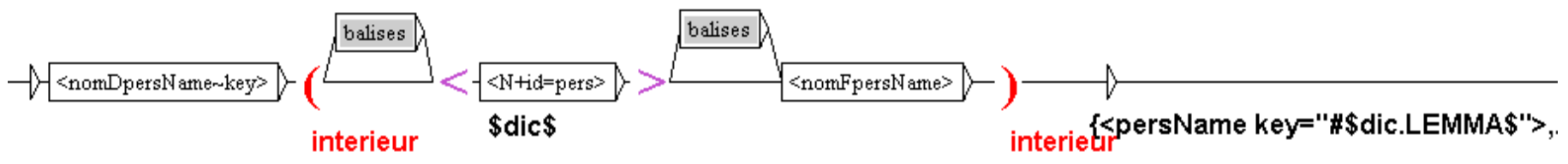
- **<placeName>**:

1. `<placeName type="pays"> Portugal </placeName>`
2. `{<placeName type="pays" key="#loc_portugal">,.baliseXml...}`
`Portugal {</placeName>,.baliseXml...}`

- **<persName>**:

1. `<name> Saint Treignan </name>`
2. `{<persName key="#pers_saintninias">,.baliseXml...}` Saint
Treignan `{</persName>,.baliseXml...}`

Graphe <persName>

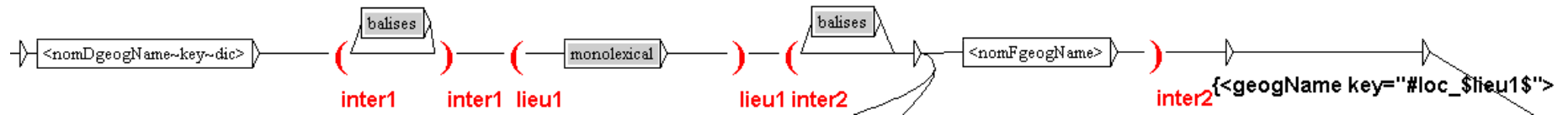




Construction de l'identifiant : sans consultation de dictionnaire

- Construction de l'identifiant « **key** » sans consultation du dictionnaire
 - **<placeName>**:
 1. `<placeName type="pays"> Australie </placeName>`
 2. `{<placeName type="pays" key="#loc_Australie">...} Australie {</placeName>...}`
 - **<geogName>**:
 1. `<geogName> Alpes </geogName>`
 2. `{<geogName key="#loc_Alpes">...} Alpes {</geogName>...}`

Graphe <geogName>

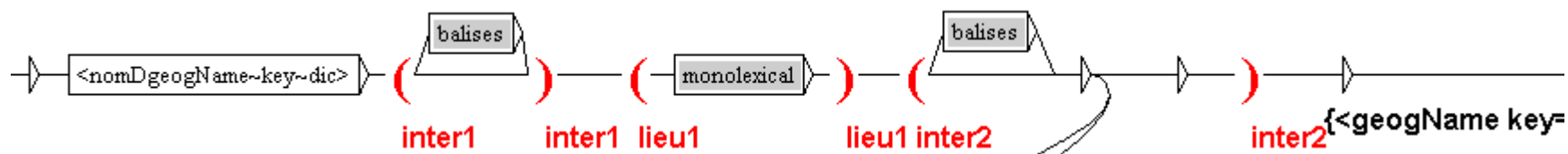


Construction de l'identifiant : imbrication

➤ Construction du premier identifiant lorsqu'il y a une imbrication

- <geogName>:

1. <geogName> isles de <placeName subtype="hypothetique" key="#loch_canaries"> Canarre </placeName></geogName>
2. {<geogName key="#loc_islesde">...} isles de {<placeName subtype="hypothetique" key="#loch_canaries">...} Canarre {</placeName>...},.placeName}{</geogName>...}

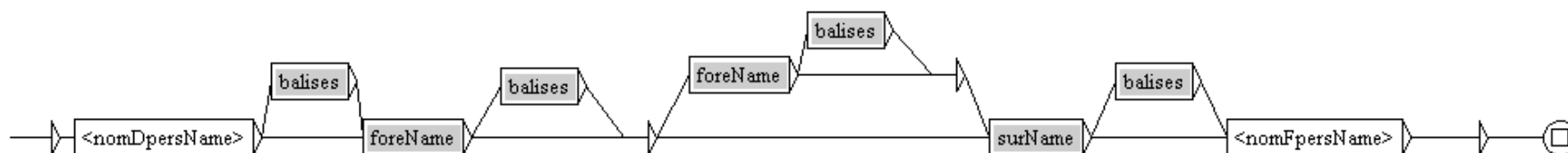


BALISAGE INTERNE

Balisage interne: <persName>

- **Saint + Nom:** <roleName>
 1. <name>Saint Treignan</name>
 2. {<persName key="#pers_saintninius">...} {<roleName type="religion">...} Saint {</roleName>...} Treignan {</persName>...}

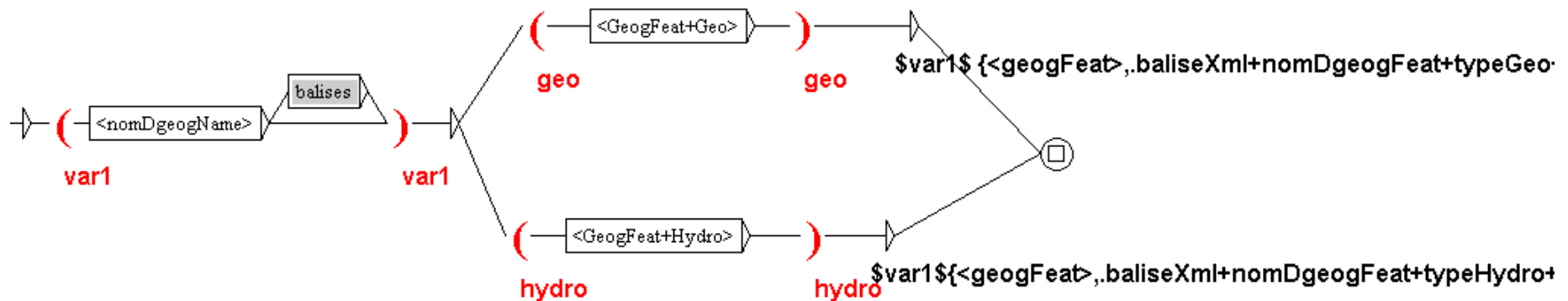
- **Prénom + Nom:** <forename> et <surname>
 1. <name> Jacques Dupont </name>
 2. {<persName key="#pers_Dupont">...} {<forename>...} Jacques {</forename>...} {<surname>...} Dupont {</surname>...} {</persName>...}



Balisage interne: <geogName>

- Balisage interne avec les balises <geogFeat> et ajout de « **typeGeo** » et « **typeHydro** »

1. <geogName key="#loc_montsinai"> mont Sinay </geogName>
2. {<geogName key="#loc_montsinai">...}
{<geogFeat>,.baliseXml+nomDgeogFeat+**typeGeo**+grfGeogFeat}
mont {</geogFeat>,.baliseXml+nomFgeogFeat+grfGeogFeat} Sinay
{</geogName>}



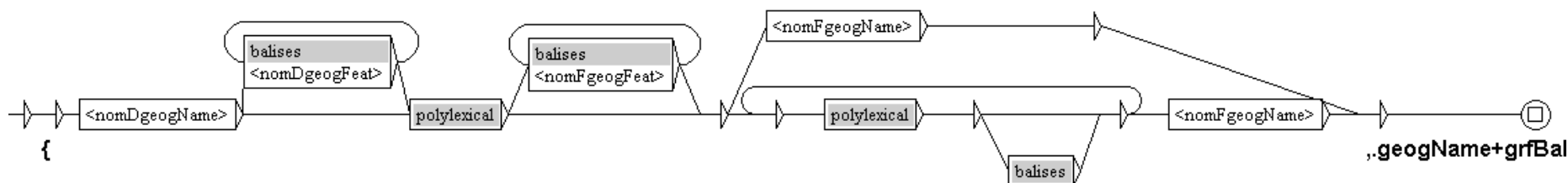
BALISAGE GLOBAL

Fin du balisage interne donc balisage global

➤ Balisage global: <persName>, <geogName>, <placeName> et <orgName>

• <geogName>:

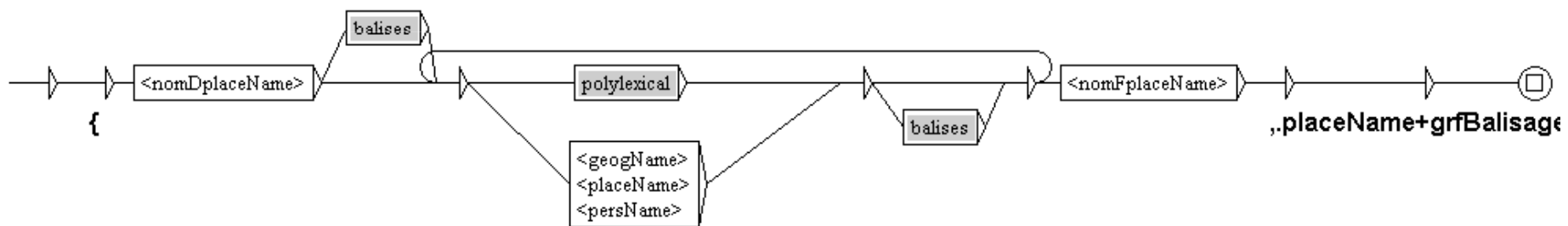
1. {<geogName key="#loc_montsinai">...} {<geogFeat>...} mont {</geogFeat>...} Sinay {</geogName>...}
2. {{<geogName key="#loc_montsinai">...} {<geogFeat>...} mont {</geogFeat>...} Sinay {</geogName>...},.geogName+grfBaliseGeogName}



Balisage global: imbrication

- Balisage global des imbrications
 - Balisage global avec { `.,placeName` } d'un `<placeName>` contenant un `<geogName>` ou un `<persName>` ou un autre `<placeName>`
 1. `<placeName key="#loc_guevede"> gue de`
`<geogName key="#loc_vede"> Vede </geogName>`
`</placeName>`
 2. {
`{<placeName key="#loc_guevede">...} gue de`
`{{<geogName key="#loc_vede">...} Vede`
`</geogName>...},.,geogName}`
`</placeName>...}`
`.,placeName+grfBalisageImbricationPlaceName}`

Graphe imbrication: <placeName>



EXTENSION DU BALISAGE

Extension du balisage: <persName>

- **Extension droite: generationnelName:** <genName> + balisage global
 1. <name> Jacques Dupont </name> l'ainé
 2. {{{<persName key="#pers_Dupont">...} {<forename>...} Jacques {</forename>...} {<surname>...} Dupont {</surname>...} {</persName>...},.persName} {<genName>...} l'ainé {</genName>...},.persName}
- **Extension gauche: roleName:** <roleName> + balisage global
 1. roy de <placeName>Paris</placeName>
 2. {{{<persName key="#pers_royde">} {<roleName type="nobiliaire">} roy {</roleName>} de {{{<placeName key="#loc_paris">} Paris {</placeName>},.placeName} {</persName>},.persName}



Extension du balisage: <orgName>, <placeName>, <geogName>

- **Extension gauche:** <orgName>
 1. royaumes de <placeName type="ville"> Tunic </placeName>
 2. {<orgName type="domaine" key="#loc_royaumesde">} royaumes de {{<placeName type="ville" key="#loc_tunisie">} Tunic {</placeName>},.placeName} {<orgName>}

- **Extension gauche:** <placeName>
 1. port de <placeName type="ville">Grace</placeName>
 2. {<placeName type="batiment" key="#loc_portde">} port de {{<placeName key="#loc_Grace">} Grace {</placeName>},.placeName} {</placeName>}

- **Extension gauche:** <geogName>
 1. boys de <placeName>Theleme</placeName>
 2. {<geogName type="geo" key="#loc_boysde">} {<geogFeat>} boys {</geogFeat>} de {{<placeName key="#loc_Theleme">} Theleme {</placeName>},.placeName} {</geogName>}

Graphe <orgName>

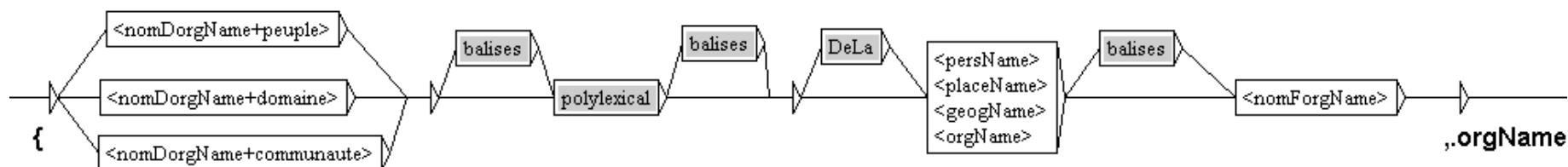


Fin de l'extension du balisage donc balisage global

➤ Balisage global des <orgName>, <placeName> et <geogName>

• <orgName>:

1. {<orgName type="domaine" key="#loc_royaulmesde">} royaulmes de {{<placeName type="ville" key="#loc_tunisie">} Tunic {</placeName>},.placeName} {<orgName>}
2. {{<orgName type="domaine" key="#loc_royaulmesde">} royaulmes de {{<placeName type="ville" key="#loc_tunisie">} Tunic {</placeName>},.placeName} {<orgName>},.orgName}



EXEMPLE RÉCAPITULATIF

Exemple récapitulatif

<persName>

Texte initial

l'empereur <name> Octavian Auguste </name>

1^{er} graphe: namePersName

l'empereur <persName> Octavian Auguste </persName>

Graphes de balises

l'empereur

{<persName>,.baliseXml+nomDpersName+grfBaliseXmlPersGeogPlace
Name} Octavian Auguste

{</persName>,.baliseXml+nomFpersName+grfbaliseXmlNom}



Exemple récapitulatif

Construction de l'identifiant

l'empereur {<persName

key="#pers_octaveauguste">,.baliseXml+nomDpersName+key+dic+g
rfLemmaPersNameAvecBaliseDicNonKey} Octavian Auguste
{</persName>,.baliseXml+nomFpersName+grfbaliseXmlNom}

Exemple récapitulatif

Balisage interne

l'empereur {<persName

key="#pers_octaveauguste">,.baliseXml+nomDpersName+key+dic+grfL

emmaPersNameAvecBaliseDicNonKey}

{<forename>,.baliseXml+nomDforename+grfPrenomNom} Octavian

{</forename>,.baliseXml+nomFforename+grfPrenomNom}

{<surname>,.baliseXml+nomDsurname+grfPrenomNom} Auguste

{</surname>,.baliseXml+nomFsurname+grfPrenomNom}

{</persName>,.baliseXml+nomFpersName+grfbaliseXmlNom}

Exemple récapitulatif

Balisage global

l'empereur {{<persName
 key="#pers_octaveauguste">,.baliseXml+nomDpersName+key+dic+grf
 LemmaPersNameAvecBaliseDicNonKey}
 {<forename>,.baliseXml+nomDforename+grfPrenomNom} Octavian
 {</forename>,.baliseXml+nomFforename+grfPrenomNom}
 {<surname>,.baliseXml+nomDsurname+grfPrenomNom} Auguste
 {</surname>,.baliseXml+nomFsurname+grfPrenomNom}
 {</persName>,.baliseXml+nomFpersName+grfbaliseXmlNom},.persNa
 me+grfBalisePersName}

Exemple récapitulatif

Extension du balisage

```

l'{{<persName
  key="#pers_empereur">,.baliseXml+nomDpersName+roleName+key
+grfPersNameBalisageExterne} {<roleName
  type="nobiliaire">,.baliseXml+nomDroleName+grfRoleName}
  empereur {</roleName>,.baliseXml+nomFroleName+grfRoleName}
  {{<persName
  key="#pers_octaveauguste">,.baliseXml+nomDpersName+key+dic+
  grfLemmaPersNameAvecBaliseDicNonKey}
  {<forename>,.baliseXml+nomDforename+grfPrenomNom} Octavian
  {</forename>,.baliseXml+nomFforename+grfPrenomNom}
  {<surname>,.baliseXml+nomDsurname+grfPrenomNom} Auguste
  {</surname>,.baliseXml+nomFsurname+grfPrenomNom}
  {</persName>,.baliseXml+nomFpersName+grfbaliseXmlNom},.persN
  ame+grfBalisePersName}
  {</persName>,.baliseXml+nomFpersName+roleName+key+grfPersN
  ameBalisageExterne},.persName+roleName+grfPersNameAvecBalis
  eDicContexte}
  
```

DEUXIÈME CASCADE

Deuxième cascade

- Obtenir un texte balisé XML-TEI
 - Retrouver le texte original tout en gardant les nouvelles balises intégrées
 - Supprimer certaines données introduites par la première cascade
 - Rétablir :
 - les coupures causées par les balises `<hi>`, `<lb rend=« hyphen »>`, etc.
 - les phénomènes de régularisation et de correction
 - Remplacer le balisage "Unitex" par un balisage XML-TEI

Deuxième cascade

- Exemple de sortie : 1^{ère} cascade
- `<name> Gymnaste </name>`

```

<csc>
  <form>
    <csc>
      <form><persName key="#pers_Gymnaste"><\form>
      <code>baliseXml<\code>
      <code>nomDpersName<\code>
      <code>key<\code>
      <code>nonDic<\code>
      <code>grfPersNameAvecBaliseNonDic<\code>
    <\csc>
    Gymnaste
    <csc>
      <form></persName><\form>
      <code>baliseXml<\code>
      <code>nomFpersName<\code>
      <code>grfbaliseXmlNom<\code>
    <\csc>
  <\form>
  <code>persName<\code>
  <code>grfBalisePersName<\code>
<\csc>
<csc>
  <form></persName><\form>
  <code>baliseXml<\code>
  <code>nomFpersName<\code>
  <code>grfbaliseXmlNom<\code>
<\csc>

```

Deuxième cascade

- 2^{ème} cascade
- Suppression des balises <csc> <form> et <code>
- Texte original contenant les nouvelles balises

```
<persName key="#pers_Gymnaste">Gymnaste</persName>
```