

Validation `output6.xml`

1. Décomptes

Tableau de contrôle `@DT{dpt}_result.csv` . S'assurer que toutes l'information est conservée.

2. Vérification des vedettes les plus longues

Tableau de contrôle `@DT{dpt}_output6_logs.csv`

Repérer et corriger les **problèmes de segmentation**.

Cas 1. Segmentation

`out:DT01-04543 Faubourg-des-Granges (Grand- et`

remplacer :

```
<vedette>
  <sm>Faubourg-des-Granges (Grand- </sm> et <sm>Petit-),</sm>
</vedette>
```

par :

```
<vedette>
  <sm>Faubourg-des-Granges (Grand- et Petit-),</sm>
</vedette>
```

Cas 2. Segmentation

`out:DT01-07563 Notre-Dame-des-Mares, des Marais ou`

remplacer :

```
<vedette>
  <sm>Notre-Dame-des-Mares, des Marais</sm> ou <sm>de Bresse,</sm>
</vedette>
```

par :

```
<vedette>
  <sm>Notre-Dame-des-Mares, des Marais ou de Bresse,</sm>
</vedette>
```

Cas 3. Les routes

```
out:DT01-09002 Route d'Ambérieu à Lyon par Meximieux
```

remplacer :

```
<vedette>
  <sm>Route d'Ambérieu à Lyon</sm> par Meximieux
</vedette>
<definition>(route départementale n° 5 et route nationale n° 84).</definition>
```

par :

```
<vedette>
  <sm>Route d'Ambérieu à Lyon</sm></vedette>
<definition>par Meximieux (route départementale n° 5 et route nationale n° 84).</de
finition>
```

3. Vérification de l'ordre des mots

Tableau de contrôle `@DT{dpt}_result_commune.csv` : liste de toutes les modifications dans les balises `localisation` .

Permet de voir si l'injection des codes insee n'a pas "cassé" certaines chaînes de caractères.

Cas 1. Erreurs de syntaxe

Exemple

```
DT01-00028 communes réunies du Abergement-le-Grand et du Abergement-le-Petit .
```

remplacer :

```
<definition>nom que l'on donne parfois aux <localisation>communes réunies
  du <commune insee="01176" precision="certain">Abergement-le-Grand</commune> et
  du <commune insee="01292" precision="certain">Abergement-le-Petit</commune>
</localisation>.</definition>
```

par :

```
<definition>nom que l'on donne parfois aux <localisation>communes réunies
    d'<commune insee="01176" precision="certain">Abergement-le-Grand</commune> e
    d'<commune insee="01292" precision="certain">Abergement-le-Petit</commune>
</localisation>.</definition>
```

Cas 2. Ordre des mots (cas des cantons et territoires)

Belley canton de l'arrondissement de

remplacer :

```
<definition>
    <typologie>chef-lieu</typologie> de <localisation>
        <commune insee="01034" precision="approximatif">Belley</commune>
        canton de l'arrondissement de </localisation>.
</definition>
```

par :

```
<definition>
    <typologie>chef-lieu</typologie> de <localisation> canton de l'arrondissement de
        <commune insee="01034" precision="approximatif">Belley</commune></localisation>.
</definition>
```

Cas 3. Ordre des mots (cas des *Saints*-)

t-Jean-de-Gonvillecommune de S

remplacer :

```
<definition>
    <typologie>ancien fief</typologie>,
    <localisation><sup>t</sup>-Jean-de-Gonvillecommune de S</localisation>.
</definition>
```

par :

```
<definition>
    <typologie>ancien fief</typologie>,
    <localisation>commune de S<sup>t</sup>-Jean-de-Gonville</localisation>.
</definition>
```

4. Tests xpath dans output6.xml

4.1. Vérification de la segmentation vedette/définition

```
//vedette//text()[not(ancestor::sm or parent::pg) and normalize-space(.)!='']
```

4.2. Alt Labels

```
//vedette[.//sm[position()>1]]
```

Long, mais très important pour ne pas louper les faux labels du type :

- `<sm>de-Haut (Les),</sm>`
- `<sm>Le Petit-),</sm>`
- `<sm>Petit-),</sm>`
- `<sm>d'en-Haut),</sm>`
- `<sm>Sur la),</sm>`
- `<sm>Les),</sm>`
- etc.

VÉRIFICATION FINALE : `//vedette/sm[position()>1]`

remplacer :

```
<vedette>
  <sm>Châtelard de Broyes, de Brosse</sm> ou
  <sm>de Breul (Le),</sm>
</vedette>
```

par :

```
<vedette>
  <sm>Châtelard de Broyes, de Brosse ou de Breul (Le),</sm>
</vedette>
```

5. Gestion des @precision

Ajout de `@precision="approximatif"` dans le cas de localisations multiples :

Cas 1. @precision manquant pour localisation multiple

- search: `(<commune insee="[^"]+>`
- replace: `$1 precision="approximatif">`
- contexte: `//definition[.//commune[position()>1]]//commune[not(@precision)]`

Cas 3. @precision='certain' pour localisation multiple (erreur ?)

- search: `(<commune insee="[^"]+") precision="certain">`
- replace: `$1 precision="approximatif">`
- context:

```
//definition[.//commune[position()>1]]//commune[@precision='certain']
```

Cas 2. @precision manquant dans les localisation simple

- search: `(<commune insee="[^"]+")>`
- replace: `$1 precision="approximatif">`

ATTENTION: REMPLACER MANUELLEMENT SELON CONTEXTE, PASSER LES CAS OÙ IL FAUT INSCRIRE CERTAIN

Puis faire une passe pour inscrire les `@precision='certain'` qui restent.

Vérification finale : parcourir la liste `//commune[@precision='certain']`

NB. Des erreurs dans la segmentation des communes

```
//commune[@insee = preceding-sibling::commune/@insee]
```

6. Formes anciennes

```
//forme_ancienne[not(i)]
```

 : Cas difficiles des formes anciennes sans label

7. Validation XML

Essayer de valider.

8. Contrôle du formatage

search: `^ +[^<]`

Supprimer les lignes vides et sauts de ligne intempestifs.