

Semaine du Document
Numérique

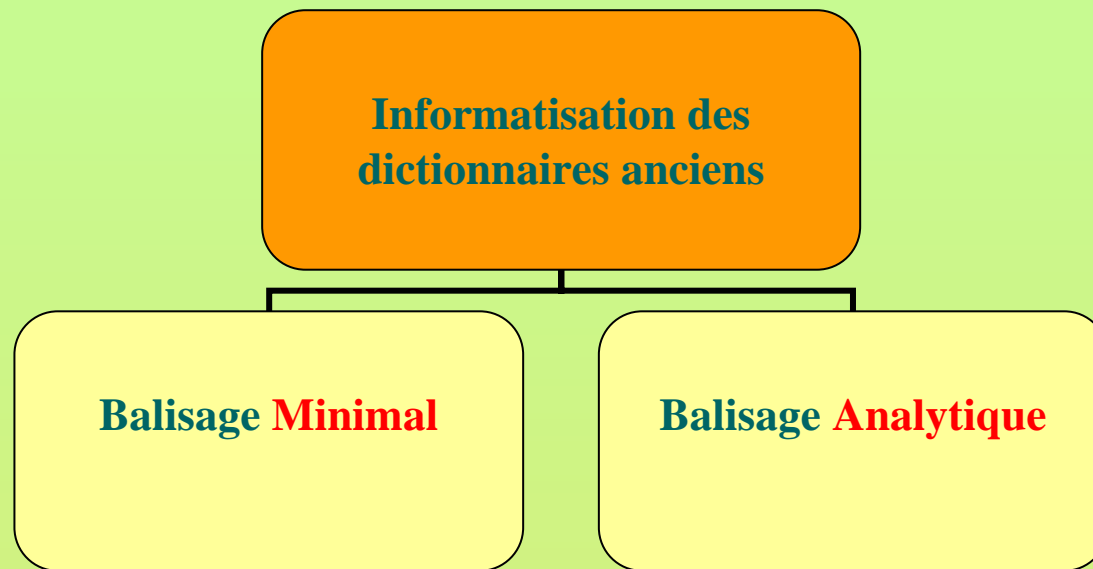
Colloque International sur la
Fouille de Textes

La Rochelle – 22-24 juin 2004

Rey Christophe, Zaoui Corinne
Université de Provence,
Equipe DELIC
Christophe.Rey@up.univ-aix.fr
zaoui@up.univ-aix.fr

Le balisage XML "ciblé": une nouvelle approche dans l'informatisation des corpus

L'informatisation des corpus



Une structure "molle"

Exemple: Information étymologique
« mouvante »:

ACCENT, s.m. Ce mot vient d'*accentum*, supin du verbe *accinere* qui vient de *ad & canere*, [...] (EM)

(N.) **GUTTURAL**, E, adj. Appartenant à la gorge ou au gosier. *Vaisseau guttural. Glande gutturale. Articulations, Consonnes gutturales.*

Ce mot, tiré immédiatement du latin *Gutturalis*, qui a le même sens, vient du nom *Guttur* (Gorge, Gosier). (EM)

LABIAL, E, adj. *Gram.*, qui appartient aux lèvres. Ce mot vient du latin *labia* (les lèvres). (EM)

Balisage Minimal

Information typographique sur la forme

Localisation de la lexie dans la structure
dictionnaire

<page n="589">
<col n="1">

<p><lc> (N.) DENTAL, E.</lc> adj. Appartenant aux dents. Les articulations <i>dentales</i> sont des articulations linguales, dont l'explosion s'opère vers la pointe de la langue appuyée contre les dents. Il y en a des muettes, <i>d</i>, <i>t</i> ; & l'articulation <i>n</i>, outre ce qui la rend nasale, suppose d'ailleurs le même mécanisme que <i>d</i>, & doit être comptée parmi les <i>dentales</i>. Il y en a aussi des sifflantes, <i>z</i>, <i>s</i>. <i>Voyez ARTICULATION</i>. (M. BEAUZÉE.)</p>

Restitution de la mise en page

Aspect italique de certaines
informations

Balisage
typographique

Balisage Analytique

Définition

```

<Entry>
<Form type=LEMMA>
<Orth Rend=CAPS> (N. ) DENTAL, E</Orth>.
<Form>
<GramGrp>
<Pos Type="ADJ">adj.</Pos>
</GramGrp>
<Sense>
<Cdomain></Cdomain>
<Def>Appartenant aux dents.</Def>
<Ency>Les articulations dentales sont des articulations linguales, dont l'explosion
s'opère vers la pointe de la langue appuyée contre les dents. Il y en a des muettes,
d, t ; & l'articulation n, outre ce qui la rend nasale, suppose d'ailleurs le même
mécanisme que d, & doit être comptée parmi les dentales. Il y en a aussi des
sifflantes, z, s. Voyez <Ref Type="VEDETTE">ARTICULATION</Ref>.</Ency>
</Sense>
<Author> (M.BEAUZÉE.) </Author>
</Entry>
  
```

Indication sur la forme

Marque de
domaine

Partie du discours

Délimitation du sens
premier

Référence

Développement
encyclopédique

Auteur

Pourquoi le balisage Ciblé ?

Difficultés de découpage des champs informationnels (1)

K, s.m. *Grammaire*. Si l'on confond à l'ordinaire l'*i* voyelle & l'*i* consonne, *K* est la dixième lettre & la septième consonne de notre alphabet ; mais si l'on distingue, comme je l'ai fait, la voyelle *I* & la consonne *J*, il faut dire que *K* est la onzième lettre & la huitième consonne de notre alphabet ; & c'est d'après cette hypothèse très-raisonnable, que désormais je coterai les autres lettres.

K,

s.m.

Grammaire.

Si l'on confond à l'ordinaire l'*i* voyelle & l'*i* consonne, *K* est la dixième lettre & la septième consonne de notre alphabet ; mais si l'on distingue, comme je l'ai fait, la voyelle *I* & la consonne *J*, il faut dire que *K* est la onzième lettre & la huitième consonne de notre alphabet ; & c'est d'après cette hypothèse très-raisonnable, que désormais je coterai les autres lettres.



Forme



Information
grammaticale



Marque de domaine



Définition



Développement
encyclopédique



Définition

Difficultés de découpage des champs informationnels (2)

(N.) OCULAIRE, adj. Relatif à l'œil. On appelle Diphthongue *oculaire*, une voyelle composée de deux voyelles simples réunies pour représenter une voix simple ; comme *ai* dans *j'aimai*, *eu* dans *heureux*, ou dans *coucou*, &c.

(N.) OCULAIRE,

adj.

Relatif à l'œil.

On appelle **Diphthongue oculaire**, une voyelle composée de deux voyelles simples réunies pour représenter une voix simple ;

comme ***ai*** dans ***j'aimai***, ***eu*** dans ***heureux***, ou dans ***coucou***, &c.



Forme

Information
grammaticale

Définition

Exemplification

Développement
encyclopédique

Exemplification

Caractéristiques du balisage Ciblé

o **Balisage XML**

o **Balisage logique**

o **Balisage physique dissocié**

o **Balisage allégé**

Exemple de balisage ciblé

<Article>

<Entree Tpe="EP">

<Forme> (N.) OCULAIRE </Forme> ,

<Information_grammaticale Type="ADJECTIF">

<Partie_du_discours Type="ADJECTIF"> adj.

</Partie_du_discours>

</Information_grammaticale>

</Entree>

<Corps>

Relatif à l'œil.

On appelle Diphthongue oculaire, une voyelle composée de deux voyelles simples réunies pour représenter une voix simple ; comme ai dans j'aimai, eu dans heureux, ou dans coucou, &c. La véritable <Reference type="VEDETTE">diphthongue</Reference> (voyez ce mot) fait entendre à l'oreille deux voix distinctes & consécutives en une seule émission ; & de là lui vient l'épithète d' <Reference Type="VEDETTE">auriculaire</Reference> (voyez ce mot) : les voyelles composées dont il s'agit ici, présentent bien aux yeux les signes de deux voix, mais n'en laissent entendre qu'une dans la prononciation ; & de là leur vient le nom de diphthongues oculaires ; parce qu'elles indiquent aux yeux deux sons, quoiqu'elles n'en expriment qu'un pour les oreilles. on les nomme encore fausses, par rapport aux diphthongues vraies qui font entendre deux sons ; & orthographiques, par opposition avec les vraies, que l'on appelle alors <Reference Type="VEDETTE">syllabiques</Reference>. Voyez ce mot.

<Signature Type="BEAUZEE"> (M.BEAUZÉE.) </Signature>

</Corps>

</Article>

Type d'entrée

Forme

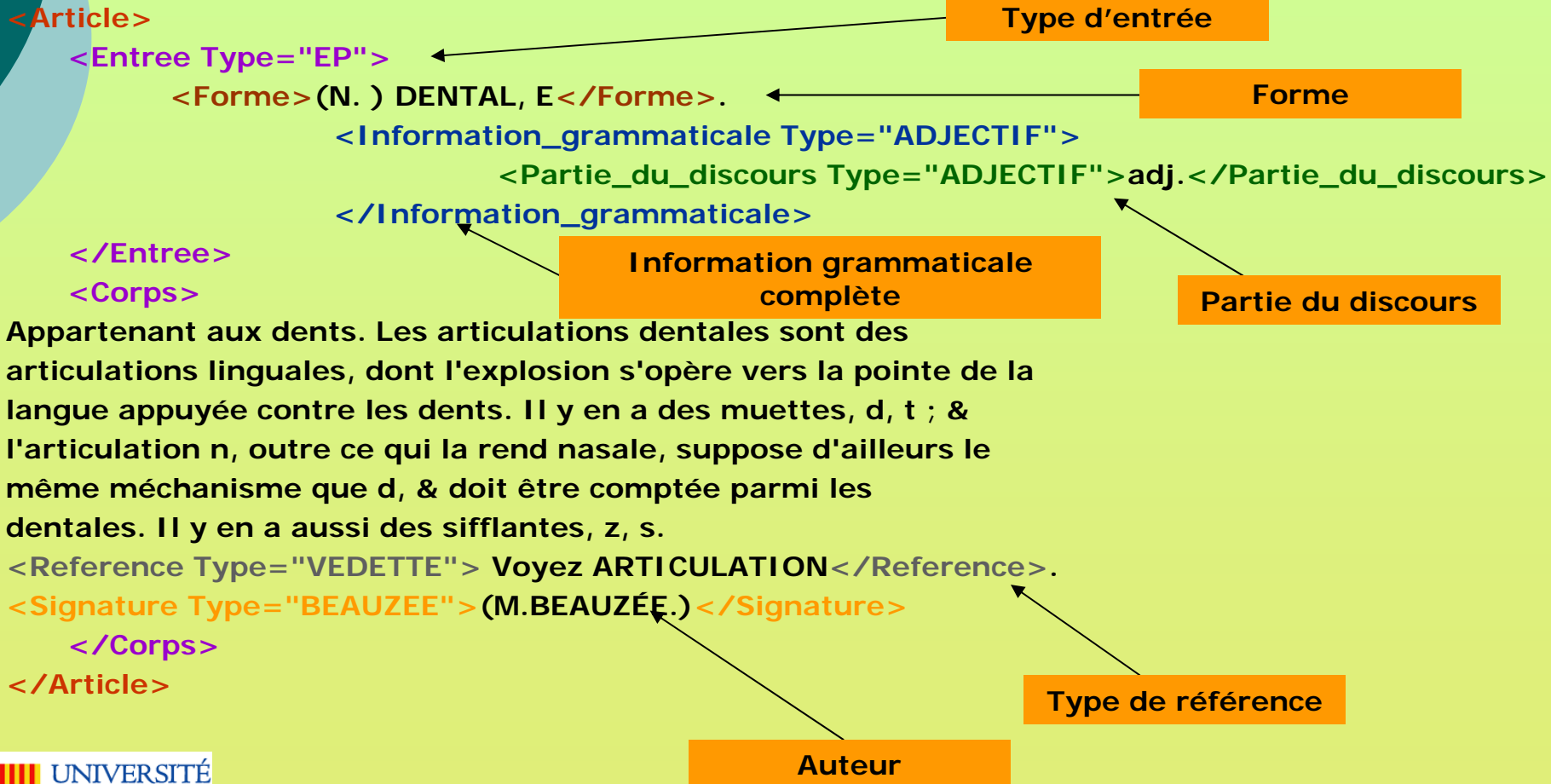
Information grammaticale
complète

Partie du discours

Auteur

Type de référence

Balisage Ciblé de l'article DENTAL



Exploitation du balisage Ciblé

CorpXML

<http://www.up.univ-mrs.fr/delic/perso/rey/methodique/index.htm>

Références bibliographiques

- **LEROY-TURCAN, Isabelle, WOOLDRIDGE, Terence. Russon,** 1998, "L'informatisation du Dictionnaire de l'Académie française", Actes du colloque DictA1998, Table ronde sur l'informatisation des dictionnaires anciens, Limoges, 19-20 novembre 1998.
- **LEROY-TURCAN, Isabelle,** 1996, "Modalités de mise en oeuvre de l'informatisation de la première édition du Dictionnaire de l'Académie française (1694)", Actes des Journées Dictionnaires électroniques des XVIIe- XVIIe s, Clermont-Ferrand, 14-15 juin 1996.
- **WIONET Chantal, TUTIN Agnès,** 2001. Pour informatiser le *Dictionnaire universel* de Basnage (1702) et de Trévoux (1704) Approche théorique et pratique. Honoré Champion.
- **WIONET Chantal, TUTIN Agnès,** 1998, "Informatisation du Dictionnaire Universel de Furetière revu par Basnage de Bauval (1702) : premier bilan", *Actes du colloque-atelier international DictA1998 organisé par le Groupe d'Études sur l'Histoire de la Langue Française (GEHLF) et la Société Internationale d'Études Historiques et Linguistiques des Dictionnaires Anciens (SIEHLDA)*, Université de Limoges, 19-20 novembre 1998.
- **WOOLDRIDGE, T.R.,** 1994, "Projet d'informatisation du Dictionnaire de l'Académie (1694-1935)", *Actes du Colloque international Le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*, Institut de France, novembre 1994; (ed. B. Quemada & J. Pruvost), Paris, Champion: 309-20.