

REY Christophe

Equipe DELIC

**(DEscription Linguistique
Informatisée sur Corpus)**

Université de Provence

reyc@up.univ-aix.fr

*L'écriture dans tous ses états,
Approches en sciences cognitives*

20-21 mai 2003

**Une expérience d'informatisation
des dictionnaires anciens:
les articles portant sur les sons
dans
L'Encyclopédie Méthodique (1782-1832)**

Informatisation des dictionnaires anciens: 2 questions primordiales

1) Que mettre
en évidence ?

2) Avec quels
outils ?



1) Le type d'encodage



Balisage *minimal*



Balisage *analytique*



2) La nature de l'encodage

❑ Encodage *propriétaire*

❑ Encodage *normatif*



Le type de balisage

Balisage Minimal



- Minimalisation de l'analyse microstructurelle
- Jalons sur les champs informationnels les plus facilement identifiables
- Balisage d'informations typographiques (Edition, pages, colonne, alinéas, etc.)
- Listes de mots-clés métalinguistiques



Thrésor de la langue françoise de Nicot (1606), Dictionnaire de l'Académie Française, etc. (cf. Bibliographie)



Balisage analytique



- Etude approfondie de la microstructure
- Jalons sur l'ensemble des champs informationnels de l'article
- Balisage d'informations typographiques (Petites capitales, italiques, etc.)



Dictionnaire Etymologique ou Origine de la Langue Françoise de Gilles Ménage (1694), Dictionnaire Universel de Furetière revu par Basnage de Bauval (1702). (cf. Bibliographie)



La nature du balisage

Balisages "propriétaires"



Balisages propres à des logiciels de traitement de corpus donnés, développés au sein d'équipes de recherche ou de laboratoires



Non-portabilité



**Wordcruncher
TACT
(Texte Analysis Computing Tools)**

Balisages Normatifs ou standards



Balisages propres à des langages ayant fait l'objet de recommandations par un organisme de standardisation, ou ayant fait l'objet d'études en vue de devenir des standards



Portabilité



**SGML (Standard Generalised
Markup Language)
TEI (Text Encoding Initiative)**

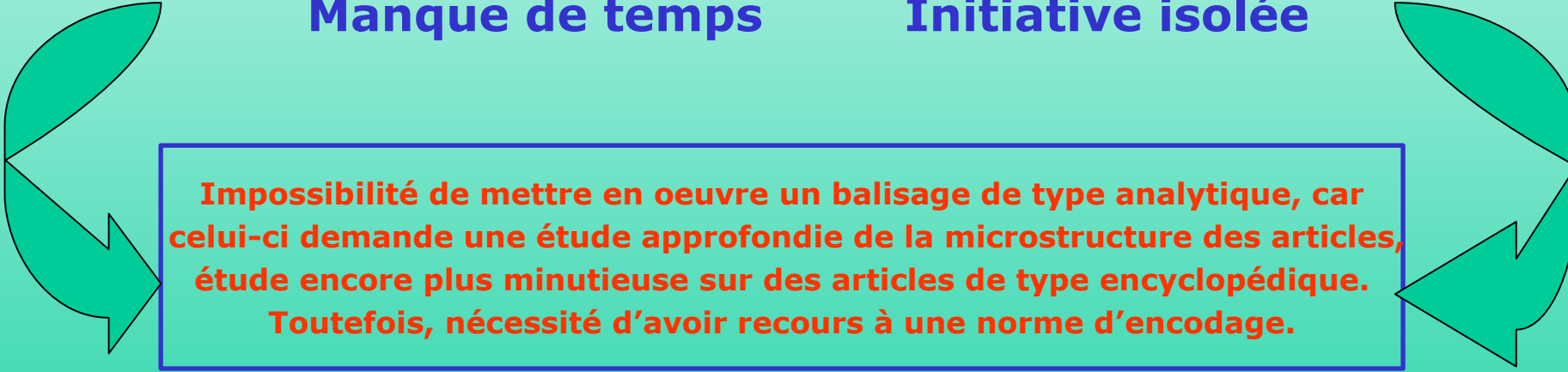
Informatisation des articles sur les sons dans l'*Encyclopédie Méthodique* (1782-1832)



Manque de temps



Initiative isolée



Impossibilité de mettre en oeuvre un balisage de type analytique, car celui-ci demande une étude approfondie de la microstructure des articles, étude encore plus minutieuse sur des articles de type encyclopédique. Toutefois, nécessité d'avoir recours à une norme d'encodage.

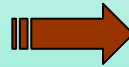


Le balisage "*sélectif XML*"

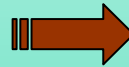
Le balisage *sélectif*



- **Balisage des champs informationnels les plus facilement délimitables**
- **Balisage de champs informationnels importants pour la théorie sur les sons**
- **Balisage de champs informationnels propres à la nature de notre corpus**
- **Non-balisage des informations sur la mise en page des articles**
- **Non-balisage des champs informationnels les moins «stables», susceptibles de se chevaucher.**



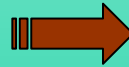
LEMME, INFORMATION GRAMMATICALE,
RENVOIS, LANGUE, etc..



Renvois à des autorités ou à des illustrations



PAIRES MINIMALES, STATUT DE
L'ARTICLE PAR RAPPORT A SON
HOMOLOGUE
DU DICTIONNAIRE RAISONNE)



COLONNE, PAGE, etc. .



DEFINITION, ZONE D'EXEMPLES, etc.

Un balisage normatif: Le balisage *XML*

Le balisage XML



Pourquoi XML?

Volonté d'utiliser des
normes ou standards

Recommandation du W3C
(1998)

Langage de toutes les attentions
scientifiques actuelles

Développement croissant
d'outils (#SGML)

Langage souple et maniable

Langage de balisage adapté à
la restitution de la
structure hiérarchique
d'un dictionnaire

Article "Dental"

(N.) DENTAL, E. adj. Appartenant aux dents. 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 ARTICULATION. (M. BEAUZÉE.)

Balisage de l'article "Dental"

<ARTICLE>

<ENTREE>

<LEMME TYPE="TRES DIFFERENT">(N.) DENTAL, E</LEMME>.

<INFORMATION_GRAMMATICALE TYPE=" ADJECTIF">

<PARTIE_DU_DISCOURS TYPE="ADJECTIF">adj. </PARTIE_DU_DISCOURS>

</INFORMATION_GRAMMATICALE>

</ENTREE>

<CORPS>Appartenant aux dents. 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.

<REFERENCE TYPE="VEDETTE">Voyez ARTICULATION. </REFERENCE>

<SIGNATURE>(M. BEAUZÉE.) </SIGNATURE>

</CORPS>

</ARTICLE>

Élément racine

Information sur le lemme

Information grammaticale

Information
sur la partie du
discours

Corps de l'article, comprenant
des champs divers:
définition, exemples,
développement
encyclopédique, etc.

Renvoi à un autre
article

Auteur de l'article

Eléments de bibliographie

- Leroy-Turcan, I.** 1998. "L'informatisation du Dictionnaire étymologique de Ménage (1694)", *DictA1998 Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20 novembre 1998.*
<http://www.chass.utoronto.ca/~wulfric/DictA/DictA1998/>
- Leroy-Turcan, I.** 1998. "Balisage formel ou balisage fin pour les dictionnaires anciens informatisés : objectifs et implications méthodologiques« , *DictA1998 Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20 novembre 1998.*
- Leroy-Turcan, I.** 1996. "Intérêt d'une base informatisée pour le *Dictionnaire Etymologique ou Origines de la Langue Française, 1694*, de Gilles Ménage: les modalités de mise en oeuvre", *Actes des Journées "Dictionnaires électroniques des XVIe-XVIIe s."*, Clermont-Ferrand, 14-15 juin 1996.
- Wionet, C., Tutin, A.** 2001. Pour informatiser le *Dictionnaire universel* de Basnage (1702) et de Trévoux (1704) Approche théorique et pratique.
- Wionet, C., Tutin, A.** 1998. "Informatisation du *Dictionnaire universel* de Furetière revu par Basnage de Bauval (1702) : premier bilan", *DictA1998 Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20 novembre 1998.*
- Wooldridge, R.** 1998. "L'informatisation du *Dictionnaire de l'Académie française (DAF)*", *DictA1998 Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20 novembre 1998.*
- Wooldridge, R.** 1998. "La rétroconversion des dictionnaires d'Estienne et de Nicot (XVIe s.) pour mise sur Internet", *DictA1998 Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20 novembre 1998.*