

Identité et statuts d'un TEXTE? Rendre visibles des propriétés discursives et leurs paramètres.

- Pierre Fiala
- CEDITEC:
 - Discours: concepts et méthodes
 - Dispositif et inscriptions médiatiques
- <http://textopol.org>

Le texte déstabilisé.

De l'héritage des humanités ...

- Fondements de la philologie: écrit, unicité, auctorialité, intentionalité, historicité.
- Les normes grammatico-rhétoriques, fixées dès l'antiquité classique et déclinées diversement, dans les cultures nationales.
- Des savoir-faire:
 - Etablissement d'index, concordances, citations
 - Recherches de filiation, d'auteurs, de comparaisons
 - Grammaires et stylistiques

... à la déconstruction multiple des SHS

- *Parole et discours*: émergence descriptive des sciences du langage
- La *textualité* des sciences des textes
 - La polyphonie
 - Le sujet mis en cause
 - La déconstruction déridienne
 - La *situation* et le *contexte*
 - Un exemple particulier: la déconstruction lexicométrique par le TLE.
- Le *message et les acteurs*, fondement des SIC.
- La *multimodalité*, coups de grâce de la numérisation audio-visuelle: le WEB2.

Une notion pourtant indispensable, comme le mot, pour...

- L'archivage, l'échantillonnage, la description, la comparaison, les échanges interdisciplinaires, le débat public.
- Identité textuelle... A quel prix?
 - Distinguer le codage numérique
 - Les modèles d'élaboration
 - Les formes de présentation et d'utilisation

Frantext (1962-2010), solution normative minimaliste pour une présentation et une utilisation standard

- **Un auteur** (individu, collectif, ou anonyme)
- **Un titre** (original ou apocryphe)
- **Une date** (originale ou supposée, distincte de l'édition utilisée)
- **Un genre** (prose, poésie, théâtre, technique)

<http://www.frantext.fr/>

La TEI « Text Encoding Initiative » (1996), un standard pour la numérisation des documents textuels.

- Initiateur: Lou Burnard (Oxford, Chicago)
- <http://www-tei.uic.edu/orgs/tei/intros/teiu5.tei>
- <http://www.univ-rennes1.fr/pub/GUTenberg/publicationsPS/24-teilite.ps.gz>

Les métadonnées TEI , solution numérisée à visée exhaustive pour que...

- les descriptions des caractéristiques fondamentales d'un texte puissent être facilement échangées;
- les descriptions spécialisées puissent être facilement ajoutées (ou supprimées) d'un texte;
- la même caractéristique puissent être encodée en parallèle de plusieurs façons;
- la richesse du balisage puisse être déterminée par l'utilisateur de la façon la plus simple possible;
- une documentation relative au texte et à la façon dont il a été codé soit fournie.

La TEI , des contraintes de codage, liées à une description normée et à un format: la DTD

- **il doit comprendre la plupart des éléments** qui appartiennent au *core tag set* de la TEI; ce dernier contient en effet des éléments généraux susceptibles d'être utilisés pour tout type de texte et tout type de traitement;
- il doit permettre de traiter correctement une large gamme de texte, avec un niveau de précision correspondant aux pratiques courantes des chercheurs;
- il doit pouvoir être utilisé aussi bien pour produire de nouveaux documents que pour coder des documents existants;
- **il doit pouvoir fonctionner avec la plupart des logiciels SGML;** et dérivés: XML
- **il doit être dérivé de la DTD TEI** complète en utilisant les mécanismes d'extension décrits dans les *Recommandations* de la TEI;
- **il doit enfin respecter les contraintes énoncées** ci-dessus de la façon la plus simple et la plus concise possible.

Des solutions pragmatiques intermédiaires. La fiche textopol

- http://textopol.free.fr/Base%20statique/Cadre_consultationBase.htm

Le texte dans tous ses états

- **Les matérialités discursives du texte**
 - **Tout discours fixé par l'écriture (oral ou écrit)**
 - **Des environnements du texte: paratexte, péri-texte, épitexte, hypertexte**
- **Unité sémantique finie présupposant une cohésion et une cohérence.**
- **Organisations grammaticales**
- **Typologies multiples**

Corpus Majorité Plurielle**Emetteur** : Gouvernement Jospin**Domaine** : Gouvernemental**Genre** : Entretien médiatique**Thématique** : Majorité Plurielle

20 ème siècle

Date: 1997-2000

Période : Cinquième République

Langue : Français**Responsabilité de la saisie** : Fiala R**Localisation** : Céditec**Commentaires**

Documents source: retranscriptions de JT et émissions radiotélévisées, archives du monde sur cd rom, site web du premier ministre et des ministères concernés, archives en ligne de libération et du Nouvel Observateur. Constitution du corpus: Textes reproduits sans les questions des journalistes et intervenants autres que les locuteurs considérés. Mise en forme: L'ensemble du corpus est reproduit en minuscules Constitué en juillet 2001 Partition en - 8 locuteurs (allègre, aubry, buffet, chevènement, guigou, DSK, jospin, voynet). - 6 semestres - 5 genres conférences de presse (confpresse), conférences publiques (confpublique), entretiens, libres opinion (lo), vœux. - 10 sources radio (europe 1 , france inter, RTL) tv (tf1, fr2, fr3) presse écrite (Le Monde, libération, Le Nouvel Observateur) et sites ministériels. - 3 supports audiovisuel, presse, discours publics (recueillis sur les sites). -partition en date, format (aaaa-mm-jj) Copier ce fichier et le placer dans le répertoire de Lexico. Lancer ensuite la segmentation. Cette opération terminée, on pourra ouvrir le fichier traité par la commande "ouvrir" en appelant le fichier "gouv6.par" ainsi généré.

Champ d'application : Rapport (S. Bonnafous, P.Fiala, C.Olivier-Yanniv, S.Heiden, S.Vassy, J.Rennes, JM Leblanc)Mots, N°68, P.Fiala, J.Rennes, Majorité plurielle, Trajectoire d'une formule.

Document**source**

10 sources radio (europe 1 , france inter, Ffr3) presse écrite (Le Monde, libération Observateur) et sites

Bibliographie du domaine:**Nature du recueil** : 1**Mots clés**

Majorité plurielle Jospin Voynet Guigo Sautter DSK Aubry Discours gouvernemen

Textopol > Fiche détaillée

- **Code texte 930**
- **Interventions des ministres de la majorité plurielle**
- **Corpus Majorité Plurielle**
 - 20 ème siècle
 - Date: 1997-2000
 - Période : Cinquième République
- **Emetteur** : Gouvernement Jospin
- **Domaine** : Gouvernemental
- **Langue** : Français
- **Genre** : Entretien médiatique
- **Responsabilité de la saisie** : Fiala Rennes Leblanc
- **Thématique** : Majorité Plurielle
- **Localisation** : Céditec
- **Commentaires** :

Documents source: retranscriptions de JT et émissions
19-20 mars 2010 Autour du Texte radiotélévisées, archives du monde sur cd rom, site web du

- **Textopol > Fiche détaillée**
- **Code texte**
930
Interventions des ministres de la majorité plurielle

Corpus Majorité Plurielle

20 ème siècle

Date: 1997-2000

Période : Cinquième République **Emetteur** : Gouvernement Jospin **Domaine** : Gouvernemental **Langue** : Français **Genre** : Entretien médiatique **Responsabilité de la saisie** : Fiala Rennes Leblanc **Thématique** : Majorité Plurielle **Localisation** : Céditec **Commentaires** :

Documents source: retranscriptions de JT et émissions radiotélévisées, archives du monde sur cd rom, site web du premier ministre et des ministères concernés, archives en ligne de libération et du Nouvel Observateur. Constitution du corpus: Textes reproduits sans les questions des journalistes et intervenants autres que les locuteurs considérés. Mise en forme: L'ensemble du corpus est reproduit en minuscules Constitué en juillet 2001 Partition en - 8 locuteurs (allègre, aubry, buffet, chevènement, guigou, DSK, jospin, voynet). - 6 semestres - 5 genres conférences de presse (confpresse), conférences publiques (confpublique), entretiens, libres opinion (lo), voeux. - 10 sources radio (europe 1 , france inter, RTL) tv (tf1, fr2, fr3) presse écrite (Le Monde, libération, Le Nouvel Observateur) et sites ministériels. - 3 supports audiovisuel, presse, discours publics (recueillis sur les sites). -partition en date, format (aaaa-mm-jj) Copier ce fichier et le placer dans le répertoire de Lexico. Lancer ensuite la segmentation. Cette opération terminée, on pourra ouvrir le fichier traité par la commande "ouvrir" en appelant le fichier "gouv6.par" ainsi généré.

Document source :

10 sources radio (europe 1 , france inter, RTL) tv (tf1, fr2, fr3) presse écrite (Le Monde, libération, Le Nouvel Observateur) et sites ministériels.

Champ d'application : Rapport (S. Bonnafous, P.Fiala, C.Olivier-Yanniv, S.Heiden, S.Vassy, J.Rennes, JM Leblanc)Mots, N° 68, P.Fiala, J.Rennes, Majorité plurielle, Trajectoire d'une formule.

Bibliographie du domaine: Nature du recueil : 1 Mots clésMajorité plurielle Jospin Voynet Guigou Allègre Buffet Sautter DSK Aubry Discours gouvernemental **Couverture saisie: 1**

sans questions des journalistes **Disponibilité : Disponible Usage de la casse** : Tout en minuscule **Exceptions casse : aucune**

Nb d'occurrences : 293493 **Nb d'Hapax** : 6396 **Nb formes** : **Freq Max** : 12873 **Logiciel** : Lexico 3 **Extension du fichier** : txt **Nombre des variables** : 8 locuteurs, 6 semestres **Variables / partitions** : Locuteur, semestre, genre, source, support, date **Liste des clés** : **Corpus annoté** : NON **Corpus aligné** : 0 **Notice complémentaire** corpus/Notices/Infogouv6.HTM [Vers le texte](#)

Les propriétés et les paramètres d'un texte politique paradigmatique *J'accuse* de Zola.

- Statut historique de l'événement
- Statut littéraire
- Statut politique
- Statut énonciatif
- Statut rhétorique
- Statut grammatical
- Statut graphique

Les paramètres lexicométriques: un ensemble minimal.

nombre des occurrences	4963
nombre des formes	1372
fréquence maximale	233
nombre des hapax	910

Les paramètres grammaticaux: décomptes et ratios de catégories.

Mots outils :	54%
Articles définis :	11,40%
Articles indéfinis :	2,00%
Adjectifs démonstratifs :	0,90%
Adjectifs possessifs :	1,60%
Adjectifs indéfinis :	0,20%
Prépositions :	11,60%
Conj. coordination :	3,20%
Conj. subordination :	2,50%
Pronoms relatifs :	2,30%
Pronoms possessifs :	0,00%

Les paramètres lexicaux: décomptes et ratios de catégories.

Mots signifiants 46%

Proportion de substantifs :	48,90%
Proportion d'adjectifs :	12,50%
Proportion de verbes :	23,20%
Proportion d'adverbes :	15,40%
Proportion personnels 1-2°p:	2,10%
Proportion personnels 3°p.:	6,30%

Conclusion 1

Les statuts du texte

- Un ensemble de propriétés discursives et de paramètres alpha-numériques opératoires.
- Une présentation tabulaire, à géométrie variable, à visibilité réglable.

Conclusion 2

Ne pas contourner la dimension nécessairement textuelle des sites

- Les sites ne peuvent se ramener à des tas de mots, de sons et d'images
- La textualité généralisée est une composante incontournable des sites
- Cette unité permet de décomposer la complexité des sites et d'organiser véritablement les réseaux sociaux.