

Gilles SOUVAY

REI (24) – *La nouvelle base Frantext-CTLF dans Frantext 2*

[1]

corpus de textes linguistiques fondamentaux

Notices Bibliographie Images Textes Articles

Recherche dans les articles de l'équipe CTLF
| Actualité du CTLF | Actualité de la base d'articles | Fonctionnement du portail |

titre

auteur

rubrique

accès réservé

archive
 non oui

présentation
 abrégée détaillée

par titres par rubriques par dates

1 Pour citer cet article :

SOUVAY Gilles. «REI (24) – La nouvelle base Frantext-CTLF dans Frantext 2». *CTLF – Articles*. Colloque 'Refonte et extension internationale du CTLF: Corpus de textes linguistiques fondamentaux', Université Paris Diderot, UMR 7597, «Histoire des théories linguistiques», Programme Action structurante 2015-2018, 31 mai et 1er juin 2018. Mise en ligne le 01/10/2018 à l'adresse: <http://ctlf.ens-lyon.fr/documents/articles/colloque/REI-24.Frantext.pdf>.

« La nouvelle base Frantext CTLF dans Frantext 2 »

gilles.souvay@atilf.fr

Refonte et extension internationale du CTLF : *Corpus de Textes
Linguistiques Fondamentaux*

31 mai – 1 juin 2018



FRANTEXT : qu'est-ce donc ?

- ▶ Créée dans les années 70 afin de fournir des exemples pour le *Trésor de la Langue Française*
 - TLF
- ▶ Base de données de textes en français
 - différentes déclinaisons
 - accès libre ou sur abonnement
 - textes informatisés au format XML/TEI P5

▶ <http://www.frantext.fr>

Trésor de la Langue Française

- ▶ **TLF : Trésor de la Langue Française**
 - dictionnaire du français littéraire des XIX^e et XX^e siècles
 - 100 000 mots avec leur histoire
 - 270 000 définitions, 430 000 exemples, 350 millions de caractères
- ▶ **Version papier en 16 volumes + 1 supplément**
- ▶ **TLFi : cédérom, version électronique en ligne**
 - www.cnrtl.fr/definition/
 - www.atilf.fr/tlfi/

Un article



PENSÉE¹, subst. fém. ■ PENSÉE², subst. fém. PENSER¹, verbe

A. – BOT. Plante herbacée, de la famille des Violacées, annuelle ou vivace, aux fleurs veloutées tricolores, comptant de nombreuses espèces sauvages ou cultivées, considérée comme le symbole du souvenir. *Pensée des champs; bourriche de pensées.* *Les alouettes sont partout ce soir. Il y a déjà des pensées bleues, blanches, violettes, par bandes dans l'herbe* (POURRAT, Gaspard, 1931, p.109):

- 1. Des **pensées**, des oeillets, des ravenelles, quelques rosiers, agonisaient au fond de ce puits sans air et chauffé comme un four par la réverbération des toits. MAUPASS., *Contes et nouv.*, t.1, Dimanches bourg. Paris, 1880, p.296.

B. – P. méton.

1. La fleur elle-même. *Sur le piédestal s'accumulent les humbles couronnes et les petits bouquets d'immortelles et de pensées* (MÉNARD, *Rév. païen*, 1876, p.218). *Yeux en grand deuil violet comme des pensées!* (LAFORGUE, *Poés.*, 1887, p.207). V. *email* ex. 3.

– [Dans son utilisation médicinale] *Le docteur s'est mis à rire de mes craintes (...) «Tant que vous mènerez votre chaste vie monacale et que vous travaillerez douze heures par jour, prenez tous les matins une infusion de pensée sauvage»* (BALZAC, *Lettres Étr.*, 1834, p.129).

2. Représentation stylisée de cette fleur (relativement au symbole qu'elle évoque). *Un médaillon encadré qui contenait une pensée dessinée en cheveux rouges* (CHAMPFL., *Avent. M^{lle} Mariette*, 1853, p.175). *Dans la paume qu'elle lui tendait il mit une pensée de saphir* (L. DE VILMORIN, *Belles am.*, 1954, p.213). V. *médaille* ex. 2.

3. En empl. apposé inv. [Désigne la nuance violet pourpre d'une variété de pensée] *Ruban pensée.* *Elle portait un tablier de soie violet pensée, avec la bavette* (SAND, *Mare au diable*, 1846, p.200). *Des velours ramagés, couleur pensée, formaient les rideaux et les portières* (BOURGES, *Crépusc. dieux*, 1884, p.102):

- 2. *Toilette de visite (chapeau Figaro).* – Jupe en faille **pensée** avec un grand volant, haut sur la traîne et bas devant. Tunique en velours **pensée** garnie de plumes bleu très-pâle, et de brandebourgs en passementerie **pensée**. Manches ornées de crevés de satin **pensée**. Chapeau Figaro en velours **pensée** et bleu très-clair. MALLARMÉ, *Dern. mode*, 1874, p.779.

Prononc. et Orth.: [pã̃se]. Homon. *panser*. Att. ds Ac. dep. 1694. **Étymol. et Hist.** 1460-66 (MARTIAL D'Auvergne, *Arrêts d'amours*, éd. J. Rychner, p.36, 65). De *pensée*^{1*}, cette fleur étant considérée comme l'emblème du souvenir, cf. 1558 *herbe de la pensée* (L. FUCHS, *Histoire des plantes* ds ROLL. *Flore* t.2, p.173).

FRANTEXT : qu'est-ce donc ?

- ▶ Créée dans les années 70 afin de fournir des exemples pour le *Trésor de la Langue Française*

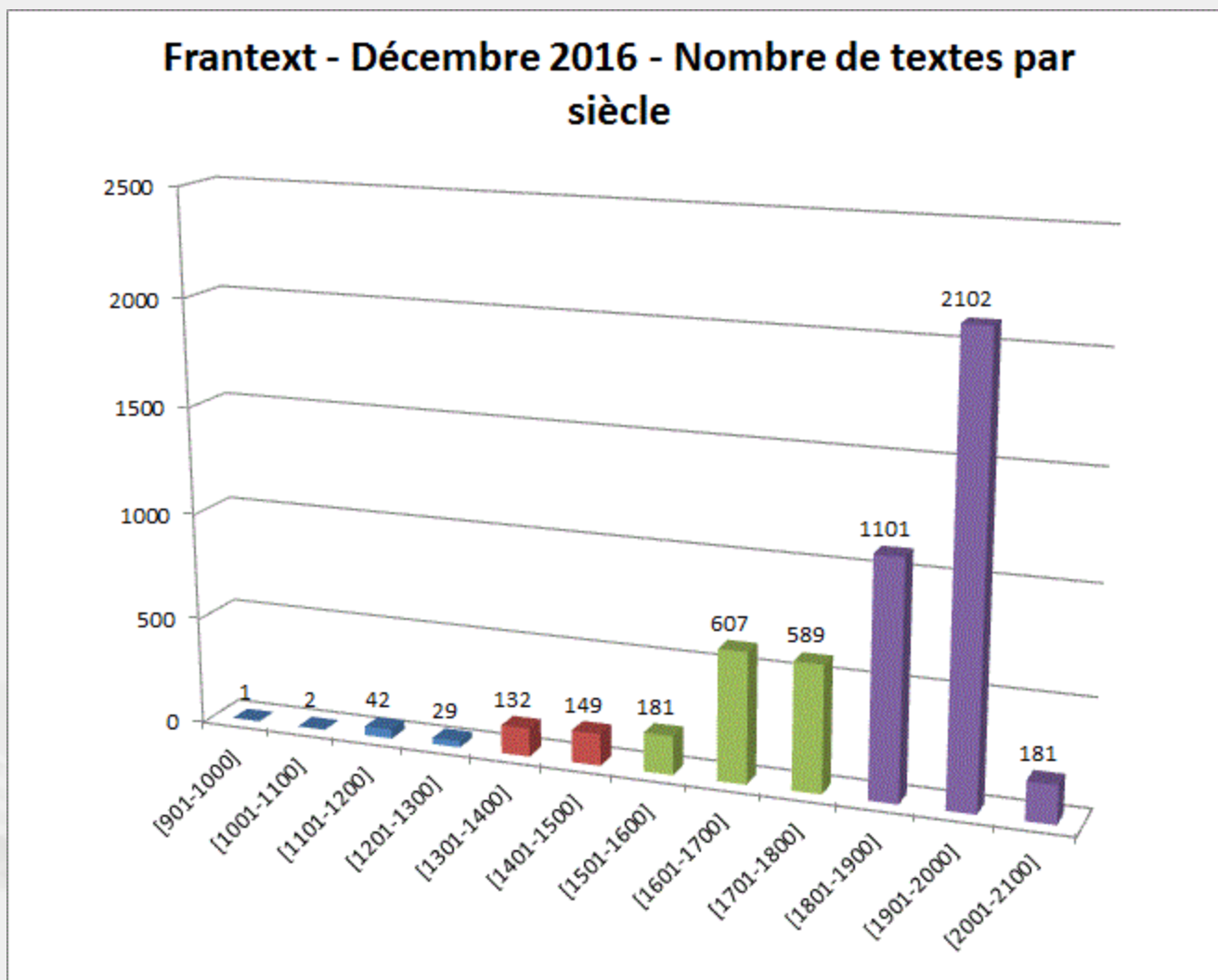
- ▶ Base de données de textes en français
 - pas de traduction
 - différentes déclinaisons
 - accès libre ou sur abonnement
 - textes informatisés au format XML/TEI

- ▶ <http://www.frantext.fr>

FRANTEXT : qu'est-ce donc ?

- ▶ **Conçue pour permettre des recherches de mots**
 - ni téléchargement, ni lecture des textes en continu
- ▶ **Offrir un répertoire des différents états de la langue française et de la création française**
 - du X^e au XXI^e siècle
 - 5 118 textes (août 2017)
 - 297 600 000 mots
- ▶ **Respect des droits d'auteur et d'édition**
 - Contrat avec le Syndicat National de l'Édition

FRANTEXT : qu'est-ce donc ?



FRANTEXT : qu'est-ce donc ?

- ▶ Frantext propose des textes numérisés en *mode texte*
 - possibilités de recherches plus étendues
 - lemmes, expressions de séquences, expressions régulières
 - tri de vocabulaire, calcul de fréquences

- ▶ L'essentiel du corpus de Gallica et de Google Livres est proposé en mode image ou en océrisation automatique

FRANTEXT : qu'est-ce donc ?

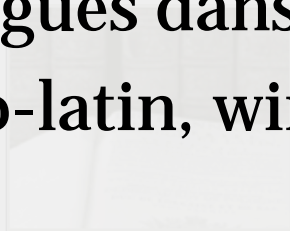
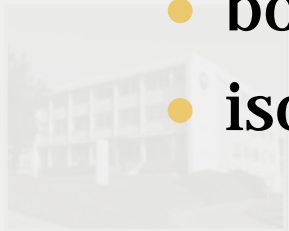
► Différentes déclinaisons

- Frantext intégral
- Frantext catégorisé
- Frantext Moyen Français
 - Dictionnaire du Moyen Français (1330-1500)
- Frantext Agrégation
- Frantext Démonstration
- Frantext CTLF



► Rénovation de Frantext

- procédure entamée depuis de nombreuses années...
- changement du moteur de recherche
 - STELLa > Express
 - procès perdu
 - limites atteintes
- interfaces datent technologiquement
- bogues dans les programmes
- iso-latin, windows, C++



FRANTEXT 2

► Frantext 2

- reprise des fonctionnalités de Frantext 1 et bien plus...
- étiquetage morphologique et lemmatisation
- utf8, linux, C++
- wiki pour une meilleure organisation de l'aide
- téléchargement des textes libres de droits
- ouverture officielle juin 2018 et fermeture de Frantext 1



- ▶ **Constitution du corpus de travail**
 - facettes
 - statistiques
 - conservation d'une session sur l'autre



FRANTEXT 2

▶ 3 modes de recherche

- simple : un ou plusieurs mots
- assistée : interface graphique qui crée les requêtes pour l'utilisateur

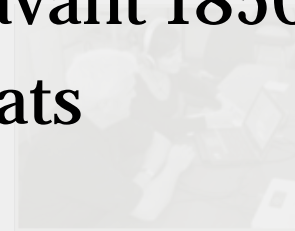


The screenshot shows the search interface for FRANTEXT 2. It features two search boxes, 'Mot 1' and 'Mot 2', each with a search icon. Below each box are three input fields: 'Forme', 'Lemme', and 'Catégorie'. The 'Mot 1' box has 'Forme' empty, 'Lemme' containing 'aimer', and 'Catégorie' empty. The 'Mot 2' box has 'Forme' containing '*.ment', 'Lemme' empty, and 'Catégorie' containing 'adverbe'.

- avancée : requête CQL *Corpus Query Language*
 - [lemma="aimer"][pos="ADV" & word="*.ment"]

FRANTEXT 2

- ▶ Travail en cours
- ▶ Charte graphique provisoire
- ▶ Problèmes
 - segmentation :
 - mauvaise gestion de <lb/>
 - « parce que » 200 réponses ?
 - ergonomie d'utilisation
 - étiquetage des textes avant 1850
 - exportation des résultats
 - mots collés



FRANTEXT 2

- ▶ **Conséquences pour Frantext CTLF**
 - **poursuite avec les interfaces Frantext 1**
 - donc mise à jour 2018 à prévoir
 - **migration vers Frantext 2 à planifier**
 - **étiquetage des textes non français vs lexiques de flexion**



FRANTEXT 2

- ▶ **Démonstrations**
 - **FRANTEXT CTLF**
 - **FRANTEXT 2**

