



UNIVERSITÉ
CAEN
NORMANDIE

AXE 2 :

**ÉLABORATION DE RESSOURCES LINGUISTIQUES
ET EXPLOITATION AVANCÉE DE CORPUS ÉCRITS ET ORAUX**



Normandie Université



- **DICTIONNAIRE ÉLECTRONIQUE DES SYNONYMES (DES)**
<http://crisco.unicaen.fr/dictionnaire-electronique-des-synonymes/>
<https://git.unicaen.fr/crisco-des-public>

- **RESSOURCES LINGUISTIQUES POUR LE FINNOIS**
<https://crisco2.unicaen.fr/delfin/index.html>

- **TRAITEMENT AUTOMATIQUE DES TEXTES VERSIFIES**
<https://crisco2.unicaen.fr/verlaine/>
https://crisco2.unicaen.fr/~stage/Verlaine_4/
<https://git.unicaen.fr/malherbe/corpus>

- **PROJET VOVA (VOcal and Verbal Arts)**
<https://vovarts.org/>



Problématiques communes :

- formatage XML-TEI
- étiquetage morpho-syntaxique
- programmation (python...)
- analyse mathématique et statistique
- technologies du web



Dictionnaire Électronique des Synonymes - **Bilan**

Laurette Chardon & Chris Smith

■ **Maintenance :**

plus de 8500 interventions sur la base de données de 2016 à ce jour
([détail site web](#)).

■ **4 publications :**

revue *Syntaxe et Sémantique*, revue *Lexique*, document de travail

■ **Refonte de l'espace sémantique**

■ **Diffusion grand public :**

actualités, projet pédagogique « DidacDES », fête de la science

■ **Nouveaux outils :**

graphe d'adjacence, visualisations 3D

■ **Nouveaux projets amorcés :**

MultiDES, insertion catégorie grammaticale, refonte de l'interface

Dictionnaire Électronique des Synonymes - Perspectives

Laurette Chardon & Chris Smith

- Poursuivre l'**alimentation du DES**
- Traitements automatiques de **nettoyage et/ou de complétude** des liens synonymiques
- **Continuer les projets amorcés** :
 - catégories grammaticales, refonte du DES avec le CERTIC, MultiDES
- Étude statistique et sémantique sur **les mots les plus recherchés**
- **Plan de gestion de données**

Traitement automatique de textes versifiés - **Bilan**

Éliane Delente, Stéphane Ferrari et Richard Renault

■ **Corpus *Malherbæ* :**

- Création d'un dépôt Gitlab en libre accès

- Collaborations :

- Pamela Puntel (doctorante, Université de Lyon II et Università Degli Studi Di Udine)
- *Le Rire des vers* (Anne-Sophie Bories, université de Bâle)
- Benoît de Cornulier (LLing, Université de Nantes)

■ **Outils d'analyse :**

- Évaluation quantitative et qualitative des rimes
- Analyse de la périodicité des strophes
- Intégration de la ponctuation

■ **Étude de la convergence :**

- Réflexion théorique sur le traitement du rythme par le lecteur :
 - une activité mentale
 - se déroule dans le temps
 - s'appuie cruciallement sur les unités prosodiques

Traitement automatique de textes versifiés - Perspectives

Éliane Delente, Stéphane Ferrari et Richard Renault

■ Corpus *Malherbæ* :

- Manuel d'encodage de textes versifiés XML-TEI
- ajouts des textes en version analysée (annotation enrichie)

■ Outils d'analyse :

- Analyse statistique de l'annotation enrichie (poèmes, recueils, auteurs)
- Mise en place d'un dépôt Gitlab des relevés métriques et statistiques
 - Mise en place d'un dépôt Gitlab des programmes et ressources
 - Implémentation de l'approche de *Dell* et *Benini* (2020) de la convergence syntaxe/mètre

■ Étude des relations rythme/sens :

- Extraction automatique d'un corpus exploratoire s'appuyant sur la distribution des signes de ponctuation. Étude envisagée :
 - compatibilité des unités prosodiques avec le mètre
 - les régularités formelles des rejets / contre-rejets
 - les phénomènes de ré-analyse causés par les décalages rythme / sens.



Projet VOVA (VOcal and Verbal Arts) - Bilan

Catharine Mason & Stéphane Ferrari

Développement d'une Archive d'Entextualisations en *Open Source*

■ Recherches interdisciplinaires :

Sociolinguistique - Linguistique ethnographique - Linguistique anthropologique - Ethnopoétique - Philologie et Littérature

■ Enseignements :

Entextualisation (transcription, encodage, annotations, analyse de données) - Protocoles de collecte (chansons, proverbes) - Atelier "Prendre langue !" (locuteurs de langues indigènes)

■ Langues intégrées dans les protocoles VOVA :

Données originales : chercheurs VOVA: 22 ; chercheurs associés : 27 (= 49)
Données tirées de la littérature scientifique: 67

Projet VOVA (VOcal and Verbal Arts) - Perspectives

Catharine Mason & Stéphane Ferrari

■ Développement des ressources en XML-TEI :

Création d'un Code en XML-TEI représentatif des méthodologies et de la nomenclature VOVA, et l'édition d'un Corpus multilingue en XML-TEI en *open source*

■ Développement de bases de données et de logiciels pour les applications dans les domaines suivants :

- recherches comparatistes sur les genres et les styles des traditions orales
- enseignements universitaires dans le recueil et les analyses de données ethnolinguistiques, ethnographiques, ethnopoétiques et interactionnelles
- enseignement des langues par les biais de la stylistique et de la rhétorique
révitalisation des langues indigènes
- valorisation des langues minoritaires (indigènes, dialectes, pidgins, créoles)