

Un équipement d'excellence de mutualisation de ressources et d'outils sur la langue et son traitement informatique

*ORTOLANG bénéficie d'une aide de l'État au titre du programme
« Investissements d'avenir » (ANR-11-EQPX-0032)*



ORTOLANG un équipement d'excellence en réseau réunissant des compétences complémentaires



UMR 7319 - CNRS - ARRU
PAROLE ET
LANGAGE



Laboratoire
Ligérien de
Linguistique



Positionnement institutionnel

- Un **équipement d'excellence** au service de l'ensemble de la communauté
 - en réseau, intégrant entre autres le CNRTL et le SLDR
 - géré par le CNRS,
 - financement assuré jusque fin 2019
 - renouvellement du matériel et des licences assuré jusque 2024
- **Service spécialisé pour la langue**
 - complémentaire de l'offre généraliste proposée par [Huma-Num](#)
- **Une mission complémentaire de développement de ressources et d'outils de base**

Pourquoi un tel équipement ?

- Pour servir de **support aux travaux de recherche** :
 - La **notion de corpus** (annotés ou non) est aujourd'hui **incontournable** spécifiquement **en linguistique et en TAL** pour l'émergence et l'apprentissage de modèles
 - pour les approches stochastiques
 - pour les approches symboliques
- Pour la **valorisation des résultats de recherche** (corpus, lexique, dictionnaires et outils de traitement)
- **La constitution et la normalisation de corpus de qualité est très couteuse**
 - On ne peut pas se permettre de recréer des ressources qui existe déjà
 - Le passage à l'échelle supérieure impose de rassembler des corpus d'origines différentes

Etat actuel du projet

- **Une plateforme matérielle** s'appuyant sur
 - 6 serveurs bi-processeur Dell (3 R630 et 3 R640).
 - 2,3 To de RAM
 - 190 To utile de disques en Raid 6
 - 2 Librairies LTO6 et 50 slots de 300To
 - L'ensemble est hébergé à l'INIST avec un service 24h/24 et 7j/7
- **Une plateforme logicielle** comprenant
 - Un dépôt d'objet numérique fiable et robuste opensource
 - Une sécurisation des données
 - Une gestion des versions
 - Une authentification des utilisateurs permettant l'accès par serveur d'identité (Renater)
 - Une interface de dépôt et de consultation
 - Une création et diffusion de métadonnées

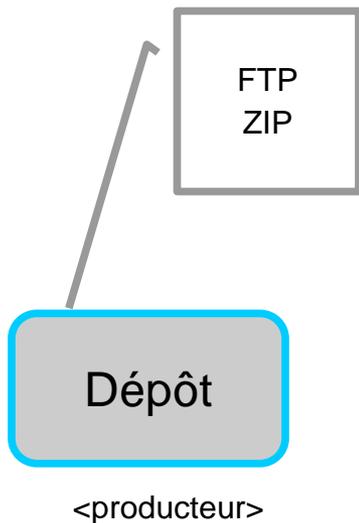
Contexte général

- Les utilisateurs émettent le souhait **de déposer une nouvelle ressource**
- **Ouverture d'un espace dans Ortolang**
- **Dépôt privé** d'**un ensemble de données** (fichiers) et de métadonnées. Ces données peuvent être organisées et formatées comme l'utilisateur le désire
 - Des discussions sur les meilleurs formats et présentations peuvent être ouvertes avec l'équipe Ortolang et la communauté
- Les données déposées sont sécurisées et peuvent être conservées plusieurs années (par exemple pour la durée d'une thèse ou d'un projet)
- Les données prêtes sont **publiées** avec un contrôle (formats, métadonnées, information) des administrateurs de la plateforme
- Les données sont **diffusées** en accord **avec les droits accordés par le déposant**. Des identifiants **pérennes** sont créés pour accéder aux données (qui ne changeront plus). Les données sont ouvertes à la consultation et au téléchargement.
- Les producteurs peuvent **compléter** leur travail et publier une **nouvelle version** de la même ressource.
- Les ressources seront transmises à Huma-Num+CINES pour un **archivage à long terme**.

Accompagnement pour toutes les phases de traitements des données

- Trois Centres de compétences thématiques
 - **Oral** : (Modyco & LPL)
 - **Multi-modal** (Modyco & LPL)
 - **Ecrit** (ATILF)
- Interactions directes et personnelles avec les membres de ces centres thématiques
 - pour l'ATILF : Etienne Petitjean, Cyril Pestel
 - pour le LPL : Gilles Pouchoulin
 - pour MoDyCo : Christophe Parisse
- Des **informations** de suivi **régulier** transmises aux déposants

Le dépôt



- Le producteur dispose d'un moyen **simple** de **dépôt de ses données** même si ces données sont encore impropres à la publication.
- Différentes méthodes sont proposées pour le dépôt ou téléchargement :
 - via **FTP**
 - via une **interface web**
 - par téléchargement de **fichiers compressés**
- Aussitôt déposées, les ressources sont **sécurisées** par l'utilisation de supports fiables (redondance) et par la réalisation de sauvegardes quotidiennes sur bande.

ORTOLANG x Christophe

← → ↻ | Sécurisé | <https://www.ortolang.fr> ☆ ⚙️ 📄 🗨️ ⋮

ORTOLANG Informations Langue Connexion

Accueil Corpus Lexiques Terminologies Outils Projets Intégrés Actualités Informations Producteurs

Outils et Ressources pour un Traitement Optimisé de la LANGUE

ORTOLANG est un équipement d'excellence validé dans le cadre des [investissements d'avenir](#). Son but est de proposer une infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés qui :

- permette, au travers d'une véritable mutualisation, à la recherche sur l'analyse, la modélisation et le traitement automatique de notre langue de se hisser au meilleur niveau international;
- facilite l'usage et le transfert des ressources et outils mis en place au sein des laboratoires publics vers les partenaires industriels, en particulier vers les PME qui souvent ne peuvent pas se permettre de développer de telles ressources et outils de traitement de la langue compte tenu de leurs coûts de réalisation;
- valorise le français et les langues de France à travers un partage des connaissances sur notre langue accumulées par les laboratoires publics.

HN Huma-Num ORTOLANG est un service spécialisé pour la langue, complémentaire de l'offre générale proposée par [Huma-Num](#) (très grande infrastructure de recherche).

La charte d'ORTOLANG définit les modalités d'utilisation et de dépôt de ressources sur la plate-forme. Vous pouvez [consulter la charte](#) ou la télécharger ([fichier au format pdf](#))

 ORTOLANG bénéficie d'une aide de l'Etat au titre du programme « Investissements d'avenir » (ANR-11-EQPX-0032)
ORTOLANG ISSN 2417-7482

Actualités

Toutes les ressources déposées sur ORTOLANG sont désormais moissonnées par le **Virtual Language Observatory** de CLARIN et accessibles via la recherche par facettes.


Common Language Resources and Technology Infrastructure

+ Parcourir le VLO

Toutes les actualités...

Statistiques

1138 utilisateurs
347 ressources
5,7 To de données
325 462 fichiers

Publications récentes

 Présidentielle2017: corpus des tweets de la #présidentielle2017 30 oct. 2017	 CLAPI 12 oct. 2017	 Réseau Lexical du Français (RL-fr) 6 sept. 2017	 Modèle Talismane pour textes littéraires en français moderne 17 août 2017	 Corpus GEACC 29 juil. 2017
--	--	--	---	--

I
d
e
n
t
i
f
i
c
a
t
i
o
n

Connexion à ORTOLANG x

Christophe

Sécurisé | https://auth.ortolang.fr/auth/realms/ortolang/protocol/openid-connect/auth?client_id=market&redirect...



SE CONNECTER À ORTOLANG

Français ▾

Nom utilisateur ou courriel

Mot de passe

Se souvenir de moi
Mot de passe oublié ?

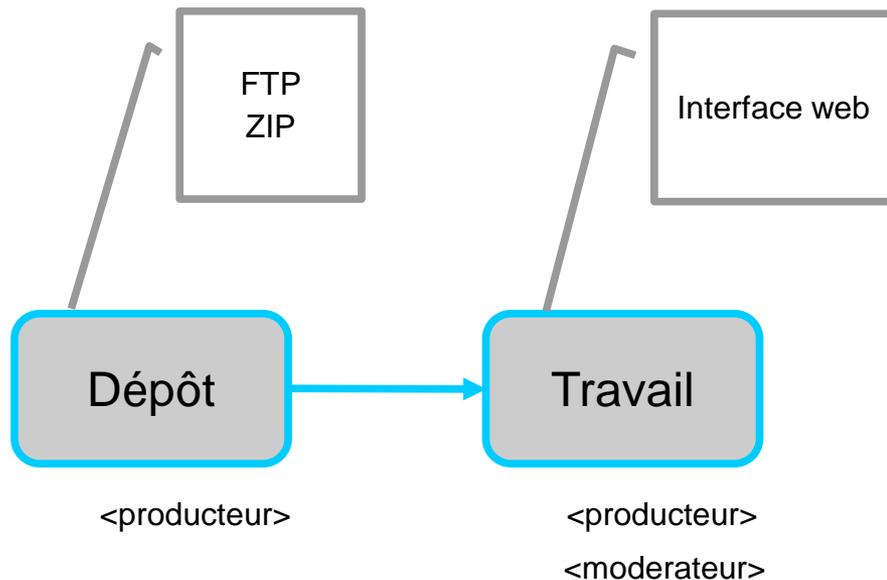
Vous n'avez pas encore de compte ORTOLANG ?
» [Créer un compte maintenant](#)

SE CONNECTER AVEC VOTRE COMPTE ESR

Vous pouvez également vous connecter en utilisant votre compte Éducation / Recherche.

Sélectionnez ou cherchez l'établissement auquel vous appartenez :

Espace de travail sécurisé



- Via un espace de travail en ligne, un partage de fichiers et une API, le producteur peut **travailler** sur ses données.
- Cet **espace de travail est sécurisé** (sauvegarde journalière)
- L'**accès** aux données est **contrôlé** et elles ne sont visibles que par les membres de l'espace de travail et les administrateurs de la plateforme
- Une mise en page HTML des données est possible.

1. PAGE D'ACCUEIL



Se rendre à « MES ESPACES »

ORTOLANG

Langue

Accueil

Corpus

Lexiques

Terminologies

Outils

Projets Intégrés

Actualités

Informations

Producteurs

Mes espaces

Outils et Ressources pour un Traitement Optimisé de la LANGUE

ORTOLANG est un équipement d'excellence validé dans le cadre des [investissements d'avenir](#). Son but est de proposer une infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés qui :

- permette, au travers d'une véritable mutualisation, à la recherche sur l'analyse, la modélisation et le traitement automatique de notre langue de se hisser au meilleur niveau international;
- facilite l'usage et le transfert des ressources et outils mis en place au sein des laboratoires publics vers les partenaires industriels, en particulier vers les PME qui souvent ne peuvent pas se permettre de développer de telles ressources et outils de traitement de la langue compte tenu de leurs coûts de réalisation;
- valorise le français et les langues de France à travers un partage des connaissances sur notre langue accumulées par les laboratoires publics.

 **Huma-Num** ORTOLANG est un service spécialisé pour la langue, complémentaire de l'offre générale proposée par [Huma-Num](#) (très grande infrastructure de recherche).

La charte d'ORTOLANG définit les modalités d'utilisation et de dépôt de ressources sur la plate-forme.

Actualités

Toutes les ressources déposées sur ORTOLANG sont désormais moissonnées par le **Virtual Language Observatory** de CLARIN et accessibles via la recherche par facettes.



Common Language Resources and Technology Infrastructure

+ Parcourir le VLO

Toutes les actualités...

Statistiques

1645 utilisateurs
425 ressources
7,4 To de données
467 063 fichiers

2. ESPACE DE TRAVAIL

➔ Cliquez sur votre corpus « espace de travail »

The screenshot displays the Ortolang web application interface. At the top, the header includes the Ortolang logo and navigation links for 'Informations' and 'Langue'. A sidebar on the left lists various menu items: Accueil, Corpus, Lexiques, Terminologies, Outils, Projets Intégrés, Actualités, Informations, and Producteurs. The 'Mes espaces' item is highlighted in red. The main content area features a search bar labeled 'Rechercher' and a button '+ Créer un espace de travail'. Below this, a workspace titled 'MON CORPUS' is shown, highlighted with a red dashed border. It contains a waveform icon and the text 'Dernière modification il y a ## minutes'.

G
é
n
é
r
a
i

Mes espaces | ORTOLANG x

https://www.ortolang.fr/workspaces/colaje

ORTOLANG Informations Langue Christophe Parisse

colaje Informations Contenu Membres Métadonnées Visibilité Prévisualisation Discussions

Colaje

Créé le 15 janv. 2016 par Administrateur ORTOLANG

Alias colaje Clé 09f220fb-d7dc-4899-9c47-cdeb5dc10e1b

2 membres dans ce projet

2 versions publiées | 5 versions mémorisées

Version 1 Version 1.1 Comparer les versions

Accès

HTTP

Lien vers la fiche de métadonnées (publiée)
<https://www.ortolang.fr/market/item/colaje>

Lien vers le contenu (non publiée)
<https://repository.ortolang.fr/api/content/colaje/head/>

FTP

Utiliser le protocole FTP pour déposer de gros fichiers (supérieur à 1 Go)
<ftp://christopheparisse:409375@repository.ortolang.fr:2221/colaje>

Statistiques de consultation (1031 vues au total)

Month	Views
sept. 2015	20
nov. 2015	25
janv. 2016	30
mars 2016	30
mai 2016	30
juil. 2016	80
sept. 2016	60
nov. 2016	120
janv. 2017	145
mars 2017	70
mai 2017	30
juil. 2017	20
sept. 2017	85
nov. 2017	20

Statistiques de téléchargements (depuis novembre 2016)

15 téléchargements complets
451 fichiers ou dossiers téléchargés

Actions

Soumettre à publication

Demands de publication

Activité

Christophe Parisse a ajouté un élément :

- /ELLIE/ELLIE-18-2_02_08/ellie-18-2_02_08-1080p.mp4
il y a une minute

Christophe Parisse a ajouté un élément :

- /ELLIE/ELLIE-17-2_01_18/ellie-17-2_01_18-720p.mp4
il y a 4 minutes

Mes espaces | colaje : / | ORTOLANG x

https://www.ortolang.fr/workspaces/colaje?section=content&root=head&path=%2F

ORTOLANG Informations Langue Christophe Parisse

colaje Informations Contenu Membres Métadonnées Visibilité Prévisualisation Discussions

colaje

Nom ↑	Type	Taille
ADRIEN	collection	-
ANAE	collection	-
ANTOINE	collection	-
aperçu	collection	-
ELLIE	collection	-
JULIE	collection	-
LEONARD	collection	-
licence	collection	-
logo	collection	-
MADELEINE	collection	-
THEOPHILE	collection	-
2012-French-Language-Studies-22-7-12.pdf	application/pdf	462,9 Ko

Auteur root
Type collection
Création 15 janv. 2016
Modification 9 nov. 2017
Clé 623406
Taille 12 éléments (calculer)

Visibilité Pour tous

3. INTÉGRER MES FICHIERS (< 1 Go)

1

Accès

HTTP

Lien vers la fiche de métadonnées (publiée)

<https://www.ortolang.fr/market/item/colaje> 

Lien vers le contenu (non publiée)

<https://repository.ortolang.fr/api/content/colaje/head/> 

FTP

Utiliser le protocole FTP pour déposer de gros fichiers (supérieur à 1 Go)

<ftp://christopheparisse:758741@repository.ortolang.fr:2221/colaje> 

2

FTP

Utiliser le protocole FTP pour déposer de gros fichiers (supérieur à 1 Go)

<ftp://christopheparisse:758741@repository.ortolang.fr:2221/colaje> 

TO KEEP IN MIND !

Cette adresse web se décompose comme suit :

Nom d'utilisateur : christopheparisse

Mot de passe : 758714

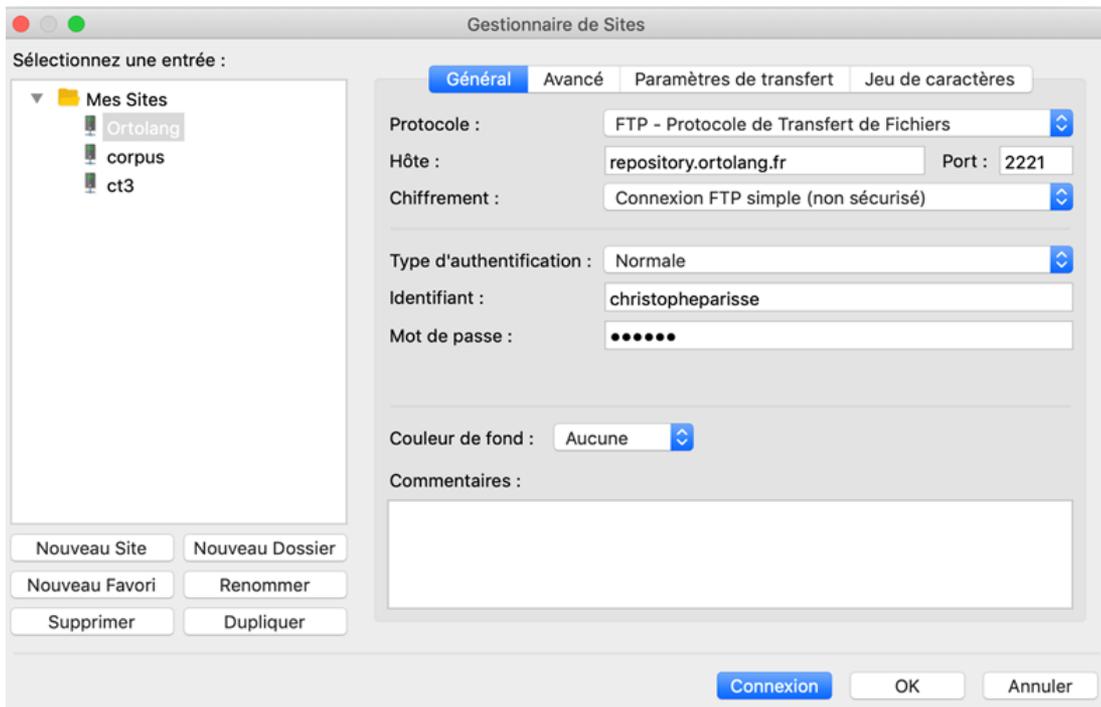
URL : repository.ortolang.fr

Port : 2221

Nom du répertoire distant : colaje

5. INTÉGRATION AVEC FILEZILLA SE

CONNECTER I PROC. 2



PROCEDURE 2
Capture d'écran

5. INTÉGRATION AVEC FILEZILLA

TRANSFÉRER SES FICHIERS

1

Accédez à la **partie gauche** de votre écran pour vous rendre dans **les dossiers et sous-répertoires de votre ordinateur** contenant vos données.

Dans l'exemple : Le dossier `/colaje/enfants/data/` (cf. Champ « Local Site » à gauche **a** et le site Ortolang pointe sur `/colaje/head` (cf. Champ « Site Distant » à droite). **b**

2

Pour transférer un fichier de l'ordinateur vers l'espace Ortolang, utiliser **glisser-déposer** ou faire clic droit de la souris sur le nom de fichier puis « Envoyer ».

Les transferts réussis sont affichés dans la partie du bas (cliquer sur l'onglet « Transferts Réussis »).

The screenshot shows the FileZilla interface with the following elements:

- Host:** ftp://christopheparisse@repository.ortolang.fr:2221 - FileZilla
- Local site:** /colaje/enfants/data/ (labeled 'a')
- Remote site:** /colaje/head (labeled 'b')
- Local site tree:** Contains folders like 'avantgems', 'chats', 'colaje-old', 'data' (circled in red), 'txm_colaje', 'data-13-07-2018', 'data-chat-25-10-2018', 'data-compound', and 'data-fiction'.
- Remote site tree:** Contains folders for names: ADRIEN, ANAE, ANTOINE, ELLIE, JULIE, LEONARD, MADELEINE, PHILIPPINE.
- File list:** Shows a list of files and directories with columns for Filename, Filesize, Filetype, and Last modified.
- Bottom status bar:** Shows 'Queued files', 'Failed transfers', and 'Successful transfers'.

Annotations in the image include a purple box labeled 'Vos fichiers' pointing to the local site tree, an orange box labeled 'Le serveur ORTOLANG' pointing to the remote site tree, and a red arrow pointing from the local 'data' folder to the remote site.

Métadonnées

Mes espaces | ORTOLANG x

https://www.ortolang.fr/workspaces/colaje?section=metadata

ORTOLANG Informations Langue Christophe Parisse

colaje Informations Contenu Membres **Métadonnées** Visibilité Prévisualisation Discussions



- Informations générales
- Personnes impliquées
- Champs spécifiques
- Informations complémentaires
- Licence
- Sous-parties
- Importer

Appliquer

Documentations

Catégorie *

Corpus Lexique Terminologie

Titre *

Colaje Français

Description *

Français

L'objectif du Projet ANR CoLaJE est de reconstituer l'émergence et le développement de la communication langagière chez le jeune enfant, avec une approche pluridisciplinaire et multimodale. L'analyse simultanée de la phonologie, la prosodie, la morpho-syntaxe, le dialogue et le mimogestuel nous offre une perspective enrichie du développement linguistique de l'enfant.

Notre travail s'appuie sur une base de données commune, comportant pour la première fois des suivis longitudinaux de productions spontanées de 7 enfants, de la naissance jusqu'à l'âge de 7 ans. Les données peuvent être directement visualisées sur le site Thématique 3 d'Ortolang (<http://ct3.ortolang.fr>)

Discussion avec les modérateurs



colaje

Informations

Contenu

Membres

Métadonnées

Visibilité

Prévisualisation

Discussions



Informations générales

Personnes impliquées

Champs spécifiques

Informations complémentaires

Licence

Sous-parties

Importer

Appliquer



MoDyCo - Modèles, Dynamiques, Corpus - UMR 7114 (Nanterre, FR)



STL - Savoirs, textes et langage - UMR 8163 (Lille, FR)



PRISMES - Langues, Textes, Arts et Cultures du Monde Anglophone - EA 4398 (Paris, FR)

Soutien(s) institutionnel(s)



Rechercher dans le référentiel des organismes répertoriés sur ORTOLANG

Vous ne trouvez pas votre organisme ? [Ajouter un nouvel organisme](#)



ANR - Agence Nationale de la Recherche (Paris, FR)

Personne(s) contributrice(s)



Rechercher une personne dans le référentiel ORTOLAN

Sélectionnez son rôle ...

Ajouter

Vous ne trouvez pas la personne ? [Ajouter une personne](#)



Christophe Parisse (MoDyCo)
Développeur, Chercheur



Martine Sekali
Chercheur



Aliyah Morgenstern (PRISMES)



Marie Leroy-Collombel



colaje

Informations

Contenu

Membres

Métadonnées

Visibilité

Prévisualisation

Discussions


 Informations générales

 Personnes impliquées

 Champs spécifiques

+ Informations complémentaires

 Licence

 Sous-parties

 Importer

Appliquer

Publications de référence ⓘ

Citation bibliographique de la publication.

Ajouter

Aliyah Morgenstern and Christophe Parisse : The Paris Corpus, Journal of French Language Studies / Volume 22 / Special Issue 01 / March 2012, pp 7 - 12



Numéro spécial de Journal of French Language Studies Guest editor : Martine Sekali First language acquisition of French grammar (from 10 months to 4 years old) Journal of French Language Studies, Vol 22, Issue 1. Cambridge: Cambridge University Press


Aperçu

Sélectionnez un fichier dans votre espace de travail.

Sélectionner

/apercu/MADELEINE-08-1_06_04.subt-816-6.mp4



/apercu/MADELEINE-13-2_01_02.subt-786-61.mp4

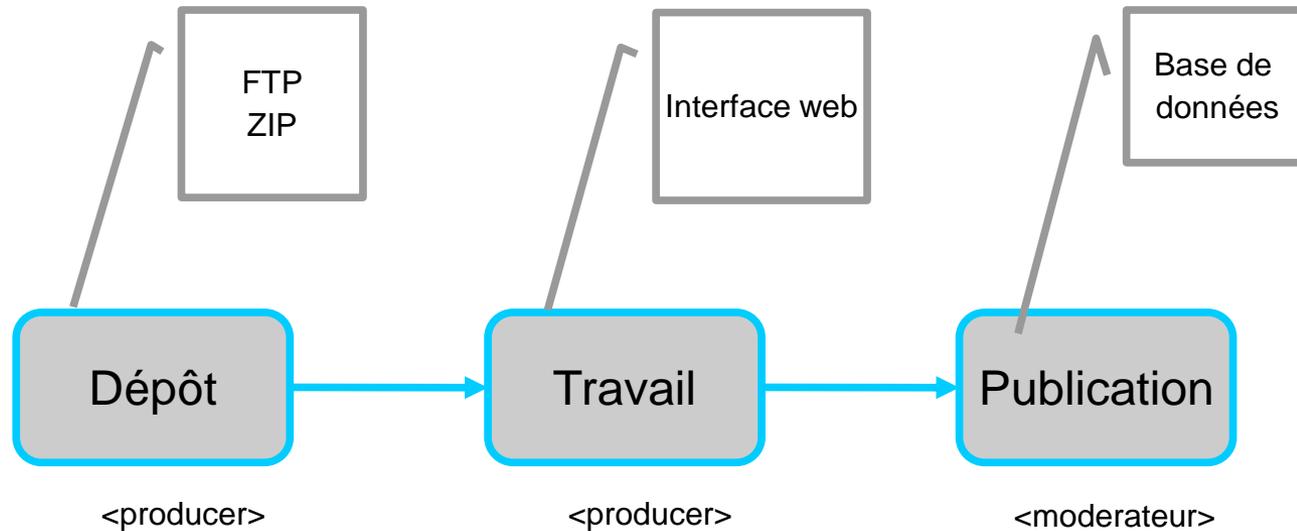

Mots clés

Français ▾

Ajouter un mot-clé.

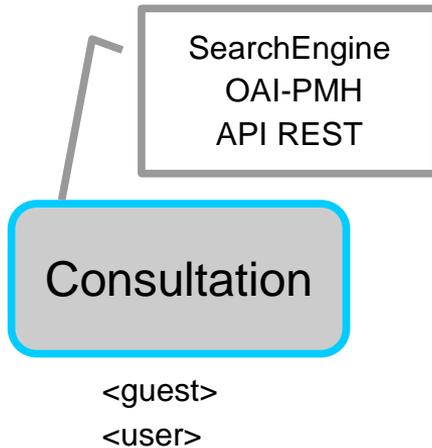
Ajouter

Publication



- Une fois que les données sont prêtes, le producteur peut **soumettre son travail pour publication**.
- Il peut alors **suivre l'état de ses demandes** et, en collaboration avec les administrateurs, aboutir à une version stable de sa ressource.
- **Plusieurs formes de publications** sont possibles (tous, identifié, recherche, privé) – chaque élément de la ressource peut avoir des droits différents

La consultation



- La **consultation** des données peut se faire de plusieurs manières :
 - via une **interface web** qui présente toutes les ressources hébergées organisées par catégories et décrites par une fiche détaillée.
 - une **navigation** dans le contenu des ressources est également possible en ligne
 - un **moteur d'indexation**
 - un connecteur **OAI-PMH**
 - **ISIDORE, VLO CLARIN**
 - une interface **REST**
- Un téléchargement aisé **de la totalité ou de parties des corpus**
 - Filtrage par noms de fichiers
 - Téléchargement de certains répertoires

Outils et Ressources pour un Traitement Optimisé de la LANGue

ORTOLANG est un équipement d'excellence validé dans le cadre des [investissements d'avenir](#). Son but est de proposer une infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés qui :

- permette, au travers d'une véritable mutualisation, à la recherche sur l'analyse, la modélisation et le traitement automatique de notre langue de se hisser au meilleur niveau international;
- facilite l'usage et le transfert des ressources et outils mis en place au sein des laboratoires publics vers les partenaires industriels, en particulier vers les PME qui souvent ne peuvent pas se permettre de développer de telles ressources et outils de traitement de la langue compte tenu de leurs coûts de réalisation;
- valorise le français et les langues de France à travers un partage des connaissances sur notre langue accumulées par les laboratoires publics.

 **Huma-Num** ORTOLANG est un service spécialisé pour la langue, complémentaire de l'offre générale proposée par [Huma-Num](#) (très grande infrastructure de recherche).

La charte d'ORTOLANG définit les modalités d'utilisation et de dépôt de ressources sur la plate-forme.

Vous pouvez [consulter la charte](#) ou la télécharger ( [fichier au format pdf](#))

Actualités

Toutes les ressources déposées sur ORTOLANG sont désormais moissonnées par le **Virtual Language Observatory** de CLARIN et accessibles via la recherche par facettes.



[+ Parcourir le VLO](#)

[Toutes les actualités...](#)

 Statistiques

- Accueil
- Corpus
- Lexiques
- Terminologies
- Outils
- Projets Intégrés
- Actualités
- Informations
- Producteurs

demo

Filtrer

- Type de ressource
- Corpus
 - Lexique
 - Terminologie
 - Outil
 - Projet intégré

- Droits d'utilisation
- Libre
 - Libre sans utilisation commerciale
 - Négociation nécessaire

- Producteur
- Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)
 - Département de linguistique et phonétique générales, Université

Nombre de resultats : 5



ANR Democrat

Democrat

Le corpus Democrat est un corpus textuel annoté en références. Les expressions référentielles sont repérées et annotées avec l'identifiant du référent, ce qui permet de construire les chaînes de ...

Corpus Écrit Libre



Signes en famille

Le langage joue un rôle central dans le développement cognitif de l'enfant : les représentations du monde qu'il se construit sont façonnées par les interactions avec ses parents et la langue d'int...

Corpus Multimodal Libre sans utilisation commerciale



PLPNat

La base de données **Corpus PLPNat**

Référencement / accès

The screenshot shows a web browser window with the URL <https://www.ortolang.fr/market/corpora/colaje>. The page header includes the ORTOLANG logo, navigation links for 'Informations', 'Langue', and a user profile for 'Christophe Parisse'. A left sidebar contains a menu with items: Accueil, Corpus (highlighted), Lexiques, Terminologies, Outils, Projets Intégrés, Actualités, Informations, Producteurs, and Mes espaces. The main content area features a 'Revenir aux résultats' button, a 'Colaje' logo, and a list of 'Laboratoires producteurs' including MoDyCo, STL, and PRISMES. It also lists 'Soutien institutionnel' (ANR) and 'Toutes les versions' (Version 4 and Version 5). A 'Référence à citer' section provides citation information and a BibTeX button. A 'Téléchargement' section details the 'Licence Creative Commons Attribution - Pas du 2019 Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International' and lists usage conditions.

Colaje

Laboratoires producteurs :

- MoDyCo - Modèles, Dynamiques, Corpus - UMR 7114 (Nanterre, FR)
- STL - Savoirs, textes et langage - UMR 8163 (Lille, FR)
- PRISMES - Langues, Textes, Arts et Cultures du Monde Anglophone - EA 4398 (Paris, FR)

Soutien institutionnel :

ANR - Agence Nationale de la Recherche (Paris, FR) [↗](#)

Toutes les versions :

[Version 4](#) [Version 5](#)

Référence à citer [Texte](#) [BibTeX](#)

Aliyah Morgenstern and Christophe Parisse : The Paris Corpus, Journal of French Language Studies / Volume 22 / Special Issue 01 / March 2012, pp 7 - 12

Numéro spécial de Journal of French Language Studies Guest editor : Martine Sekali First language acquisition of French grammar (from 10 months to 4 years old) Journal of French Language Studies, Vol 22, Issue 1. Cambridge: Cambridge University Press

Modèles, Dynamiques, Corpus - UMR 7114 (MoDyCo), Savoirs, textes et langage - UMR 8163 (STL), Langues, Textes, Arts et Cultures du Monde Anglophone - EA 4398 (PRISMES) (2016). *Colaje* [Corpus]. ORTOLANG (Open Resources and TTools for LANGuage) - [www.ortolang.fr](https://hdl.handle.net/11403/colaje/v1.1), <https://hdl.handle.net/11403/colaje/v1.1>.

Description

L'objectif du Projet ANR CoLajE est de reconstituer l'émergence et le développement de la communication langagière chez le jeune enfant. avec une approche pluridisciplinaire et multimodale.

Téléchargement

Licence Creative Commons Attribution - Pas du 2019 Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International [↗](#)

Conditions d'utilisation :

- Les corpus peuvent être utilisés librement pour tout usage à fin scientifique ou éducative.
- L'usage commercial est interdit sans une demande explicite auprès des personnes

Accueil

← Revenir aux résultats

Corpus

Version 7

Version 8

Lexiques

Terminologies

Outils

Projets Intégrés

Actualités

Informations

Producteurs

Référence à citer

Texte BibTeX

Laboratoire de Recherche sur le Langage - EA 999 (LRL) (2017). *CoMeRe (Communication médiée par les réseaux)* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, <https://hdl.handle.net/11403/comere/v3.3>.

Téléchargement

Licence Creative Commons - Attribution 4.0 International

Cette licence permet aux autres de distribuer, remixer, arranger, et adapter votre œuvre, même à des fins commerciales, tant que vous accordez le mérite de la création originale en citant votre nom.

Télécharger

Parcourir

Description

Ce projet vise à rassembler différents corpus représentatifs des formes de communication en français sur les réseaux (Internet, téléphone, etc.), tous structurés et renseignés de la même façon, diffusés en accès libre à des fins de recherche. Le projet CoMeRe a reçu le soutien de **ORTOLANG** et du **consortium national Corpus-écrits**.

Accéder au site internet

Contribution



Thierry Chanier (LRL)
Chercheur

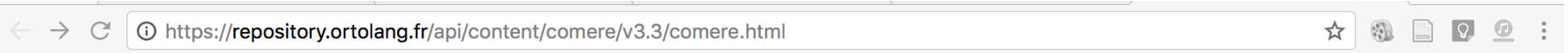
Partager



Contacter



Présentation HTML



This site: <https://hdl.handle.net/11403/comere>

CoMeRe Repository: Corpora of Computer-Mediated Communication in French

This repository includes corpora of mono or multimodal interactions mediated through networks (Internet, Phone, etc.). Three fundamental principles underlie CoMeRe: variety, standards, openness.

- **Variety:** interactions stemming from networks, as well as mono and multimodal, synchronous and asynchronous communications, eg: Email, forums, textchat, Tweets, SMS, Wiki discussions, video-audio conference systems, etc.
- **Standards:** corpora are structured and referred to in a uniform way, i.e. in TEI - Text Encoding Initiative- with a specific extension dedicated to CMC and elaborated within the European TEI-SIGand openness.
- **Openness:** all corpora are open and free access. They can be fully downloaded. CoMeRe follows the [OpenData recommandations for research data](#).

Reference : Chanier,T., Poudat,C., Sagot, B., Antoniadis, G., Wigham,C. R. , Hriba, L.,Longhi, J. & Seddah, D. (2014) « The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres ». Special issue on « Building And Annotating Corpora Of Computer-Mediated Discourse: Issues and Challenges at the Interface of Corpus and Computational Linguistics ». *JLCL (Journal of Language Technology and Computational Linguistics)*. pp1-31. http://www.jlcl.org/2014_Heft2/Heft2-2014.pdf

CMC genres

SMS	Tweets	Email	Text chat	Multimodal
- cmr-smslareunion - cmr-smsalpes - cmr-88milsms	- cmr-polititweets - cmr-intermittent	- cmr-simuligne	- cmr-getalp_org - cmr-favi - cmr-favi (POS tagged)	- cmr-copeas - cmr-tridem06
	Weblog	Discussion forum		Multimodal + 3D

Accès au site www.ortolang.fr

- Ortolang est totalement fonctionnel pour la partie dépôt et accès aux corpus
- Ortolang comprend des corpus et des outils issus des données des précédents centres (CNRTL, SLDR) et des partenaires (Modyco, LLL)
- Ortolang comprend de nombreux corpus ajoutés depuis l'ouverture et également grâce aux appels corpus communs à Ortolang, Corpus Ecrits, IRCOM et CORLI.

Pour en savoir plus
et pour utiliser Ortolang

www.ortolang.fr
contact@ortolang.fr