

DOSSIER DE PRESSE



revirada

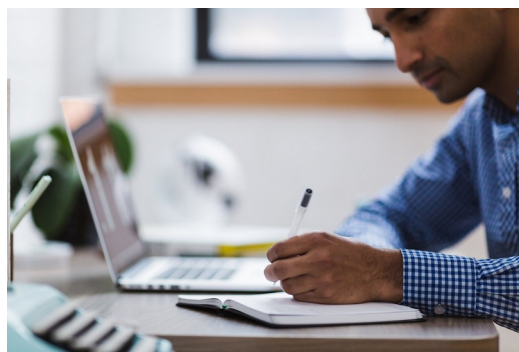
Traducteur automatique occitan

Revirada est le nouveau traducteur automatique occitan-français et français-occitan du Congrès pour les variétés gasconne et languedocienne.

Revirada est le traducteur automatique occitan de référence du Congrès. Il a été construit à partir des meilleures ressources de la plateforme locongres.org, avec des développements innovants pour sélectionner les meilleures paires de traduction.

Un gros travail a été mené tant pour respecter la diversité linguistique infradialectale que pour produire une langue moderne, cohérente et de qualité.

Revirada permet de traduire des textes en copiant-collant, mais aussi des documents en conservant la mise en forme ou des sites web « à la volée » à partir de l'URL. Des applications existent aussi, pour les mobiles et pour les navigateurs. Pour les développeurs, un plug-in et une API permettent d'intégrer des traductions de qualité dans les sites web et les applications.



Accéder au traducteur automatique en ligne *Revirada* : revirada.eu

Dossier de presse numérique (avec des vidéos) à l'adresse premsa.locongres.com/revirada.

Avec le soutien de

Interreg
POCTEFA
LINGUATEC



OC OFICIPUBLIC
de la lenga occitana



PYRENEES
ATLANTIQUES
LE DEPARTAMENT

Partenaires techniques

elhuyar
ezagutuz aldatzea

Apertium

/ *Revirada*, quelques chiffres

170 000 mots occitans dont 250 000 issus du Congrès	1 million formes fléchies occitanes mots conjugués ou déclinés	125 000 mots français dont 26 000 issus du Congrès	540 000 formes fléchies françaises mots conjugués ou déclinés
170 000 paires de traductions français-occitan et/ou occitan- français	400 modèles de conjugaison en occitan gascon et languedocien	35 000 terminaisons verbales représentant la diversité de l'occitan	
1 100 paradigmes de déclinaison noms, adjectifs, pronoms...	1 600 règles de sélection lexicale pour choisir la bonne traduction		

/ Traducteur occitan de référence

Revirada est le traducteur automatique du Congrès permanent de la langue occitane. Il est construit sur le moteur *Open source Apertium* et **intègre les meilleures ressources** de la plateforme occitane de référence locongres.org : 8 dictionnaires français-occitan et occitan-français, la totalité des conjugaisons, ainsi que les bases toponymiques et terminologiques.



Il a fallu plusieurs années de travail à notre équipe pour croiser et harmoniser les données des différents ouvrages. **Des développements innovants** permettent de sélectionner les meilleures paires de traductions pour un résultat optimal.



Le traducteur couvre **l'ensemble de la diversité linguistique infradialectale** dans le sens occitan-français : *Revirada*, en effet, reconnaît plus d'un million de formes fléchies (mots conjugués ou déclinés) représentatives de la diversité des parlers des aires gasconne et languedocienne.

Dans le sens français-occitan, les efforts ont porté sur **la production d'une langue moderne, cohérente et de qualité**. Plusieurs centaines de règles de sélection lexicale et des algorithmes de post-traitement avancés permettent de garantir une langue la plus authentique possible.

Cette première version, limitée pour le moment à l'occitan languedocien et à l'occitan gascon, a comme objectif d'être enrichie et améliorée. L'outil sera **régulièrement mis à jour** au fur et à mesure des retours de ses utilisateurs.



/ Les fonctionnalités de *Revirada*

Formulaire de traduction

Tapez ou copiez-collez du texte dans l'interface web de *Revirada*, choisissez le sens et cliquez sur un bouton.

Traduction de documents

Téléchargez un document texte et *Revirada* vous le passera dans la langue désirée en conservant la mise en forme.

Traduction de sites web

Entrez l'URL du site à traduire, la langue d'origine, la langue de sortie et vous obtiendrez le site traduit.

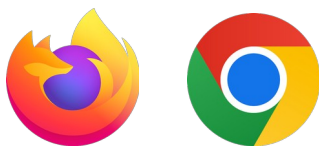
/ Les applications *Revirada*

Pour les mobiles



L'application *Revirada* sur votre mobile : photographiez ou copiez-collez, traduisez et écoutez !

Pour les navigateurs



Téléchargez la barre de navigateur *Linguapir* pour changer la langue des sites que vous visitez.

Webmasters et développeurs



Ajoutez un bouton pour traduire votre site à la volée, traduisez vos articles grâce au Plug-in Wordpress, utilisez notre API dans vos programmes.

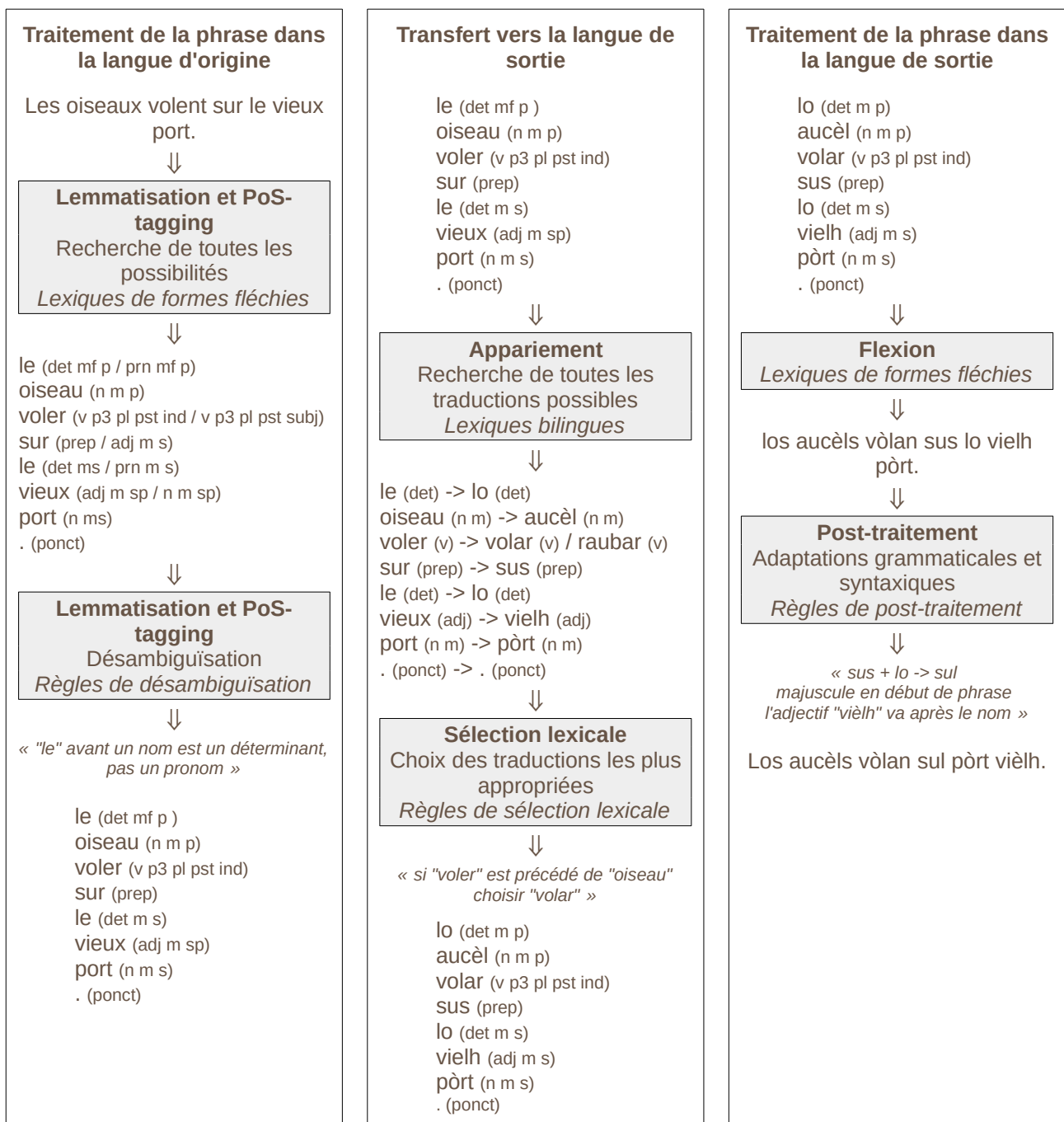
/ *Revirada*, comment ça marche ?

Revirada est construite à partir du traducteur automatique *Apertium* qui traduit en trois étapes :

1. **Traitement de la phrase dans la langue d'origine** : lemmatisation (passer à l'infinitif pour les verbes, au masculin singulier pour les adjectifs...) et analyse morphosyntaxique (indiquer la catégorie grammaticale, le genre, le nombre, la personne...).
2. **Transfert vers la langue de sortie** : pour chaque mot de la phrase originale, proposer la traduction pertinente.
3. **Traitement de la phrase dans la langue de sortie** : flexion (conjuguer les verbes, mettre au pluriel...) et post-traitement (arranger la syntaxe, la grammaire en fonction des spécificités de la langue de sortie).

Pour ce faire, il se base sur des ressources construites par des humains (contrairement aux traducteurs qui utilisent l'intelligence artificielle, pour lesquels le *machine learning* construit automatiquement ce genre de ressources). Dans le cas de *Revirada*, elles ont été créées **à partir des lexiques du Congrès** (*dicod'Òc*, *verb'Òc*, *top'Òc*, lexiques de formes fléchies), des connaissances lexicographiques de l'équipe et de ce qui était déjà présent dans *Apertium*.

- **Lexiques de formes fléchies** : ils donnent des modèles de conjugaison, de passage au féminin ou au pluriel... et des listes de mots associées à un modèle.
- **Règles de désambiguïsation** : elles permettent de choisir la bonne catégorie grammaticale et le bon lemme quand un mot peut en avoir plusieurs (ex. : « avions » forme conjuguée du verbe « avoir » ou pluriel du nom « avion »).
- **Lexiques bilingues** : paires de mots associés à leur traduction.
- **Règles de sélection lexicale** : règles qui indiquent quelle est la bonne traduction, s'il y en a plusieurs possibles, en fonction du contexte.
- **Règles de post-traitement** : quel est l'ordre des mots, les mots qui se contractent ou s'élident, les spécificités idiomatiques...



/ La traduction automatique et son utilité

En l'état actuel des technologies, la traduction automatique n'a pas pour vocation de traduire avec le même niveau qu'un traducteur humain. Elle a cependant une utilité concrète dans plusieurs contextes :

- **Comprendre une langue inconnue** : la traduction automatique peut aider quelqu'un à comprendre un texte dans une langue qu'il ne connaît pas du tout. La traduction aura peut-être quelques erreurs, le style en sera maladroit, mais cela suffira toutefois pour que la personne comprenne le contenu du texte original.
- **Avancer le travail des traducteurs humains** : si un traducteur humain traduit son texte via le traducteur automatique, une partie de son travail sera déjà faite. La personne n'aura plus qu'à corriger le texte généré, plutôt que de devoir tout traduire à partir de zéro.

Avec le **manque de moyens humains et financiers** dont souffrent les langues minorisées, c'est très intéressant. Par exemple, pour une structure qui n'a pas trop de temps, il est à présent possible de ne faire son site web qu'en occitan, et d'utiliser la traduction automatique pour en rendre son contenu accessible aux non-occitanophones. Si, pour une raison ou une autre, il y a besoin que la version en français soit de bonne qualité, elle peut au moins utiliser le traducteur automatique pour pré-traduire ses pages web, et gagner ainsi du temps.

/ La gestion de la variété de la langue

La **question de la variété interne** de l'occitan se pose quasiment à chaque étape de la construction d'un traducteur automatique pour cette langue. Quelle langue doit produire le traducteur ? Comment reconnaître le plus de mots possible ?

L'un des principes d'action du Congrès est le **respect de la diversité** interne de la langue. Le cahier des charges de *Revirada* stipulait donc que le traducteur pour le sens français-occitan devait pouvoir **fonctionner quel que soit le parler** de l'utilisateur (à condition qu'il fasse partie des variétés gasconne ou languedocienne). Un gros travail a été mené pour essayer d'intégrer la diversité du vocabulaire, des conjugaisons, des réalisations grammaticales qui peuvent se trouver sur toute la zone linguistique des deux variétés concernées.



Mais cette diversité ne devait **pas porter préjudice** à la cohérence de la langue produite en sortie de la traduction pour le sens français-occitan. On ne pouvait pas avoir, par exemple, un mot de gascon oriental et un mot de gascon landais dans une même phrase. Ainsi, nous avons fait à la fois un travail de recensement de la diversité variétale et un **travail d'homogénéisation linguistique** et de traitement des variantes.

En sa qualité d'organisme de régulation de la langue, Le Congrès se devait de **produire une langue « normalisée »** avec son traducteur. Il a également essayé de suivre son principe d'action qui est, en plus du respect de la diversité interne de la langue cité ci-dessus, **le renforcement de son unité profonde**. Il a donc également essayé de valoriser les mots communs et normatifs au moment de produire les textes en occitan.

Ce travail peut déjà se voir dans cette première version de *Revirada*. Il se poursuivra dans les années à venir pour que la langue générée par *Revirada* soit encore plus homogène et de qualité.

/ Linguatec

Le traducteur automatique occitan *Revirada* a été réalisé dans le cadre de **Linguattec**.

Le **projet européen** EFA 227/16 LINGUATEC « Développement de la coopération transfrontalière et du transfert de connaissances en technologies du langage » vise à développer, démontrer et diffuser de nouvelles ressources, des outils et des applications linguistiques innovantes qui améliorent le niveau de numérisation de l'**aragonais**, du **basque** et de l'**occitan**.



Il est porté par un consortium d'institutions et de centres de recherche **spécialisés dans le TAL** (Traitement Automatique de la Langue) pour les langues aragonaise, basque et occitane, qui pratiquent la **coopération transfrontalière** et le **transfert de compétences**.

Ce projet permet la coopération technologique entre les langues pour développer de nouvelles ressources et des outils linguistiques, ainsi que de nouvelles applications en ligne pour faciliter l'utilisation des trois langues et un accès multilingue au contenu par les utilisateurs finaux.

// Les membres de Linguattec



/ Ils ont construit Revirada

// Lo Congrès

Lo Congrès a construit les ressources nécessaires pour faire fonctionner le traducteur automatique : lexiques, règles de désambiguïsation, de sélection lexicale ou de post-traitement.



Le Congrès permanent de la langue occitane est l'organisme interrégional de régulation de l'occitan. Il œuvre dans les domaines de la linguistique et du TAL (traitement automatique de la langue).

Il produit des outils linguistiques numériques de référence (dictionnaires, conjugueurs, correcteurs orthographiques...), des applications pour le TAL (synthèse vocale, traduction automatique...) et des applications pour les mobiles (claviers prédictifs...).

Il a également des missions de régulation linguistique et de recherche scientifique appliquée.

Il est l'éditeur d'un multidictionnaire occitan (*dicod'Òc*) qui a chaque année plus d'un million de visites.

// Elhuyar

Elhuyar a construit l'interface pour utiliser *Revirada* via un formulaire ou une API, pour la traduction de documents ou de sites web.

L'objectif de la fondation Elhuyar Fundazioa est d'encourager, promouvoir et développer la science et la langue basques. Dans le domaine des technologies de la langue et de l'intelligence artificielle, les secteurs de la recherche sur lesquels travaille la Fondation sont, entre autres, la traduction automatique, la création automatique de ressources linguistiques, l'extraction de textes et les technologies de la parole.



Elhuyar dispose par ailleurs d'une solide expérience dans le domaine du leadership et de la participation à des projets européens de technologies de la langue. La Fondation est membre du Réseau basque de science et technologie.

// Apertium

Apertium est le moteur de traduction utilisé pour construire *Revirada*

Apertium est une plateforme *open source* et gratuite de traduction automatique. Elle propose un moteur de traduction automatique indépendant de la langue, des outils pour aider à construire les données linguistiques nécessaires à son fonctionnement, et des données déjà disponibles pour plusieurs paires de langue (comme occitan-catalan, occitan-espagnol et occitan-français).



// Soutiens



/ Pour plus d'informations

Le Congrès permanent de la langue occitane
Ciutat - Creem
5-7 rue de la Fontaine
64000 Pau

premsa.locongres.com/revirada

info@locongres.org

06 45 57 01 07