

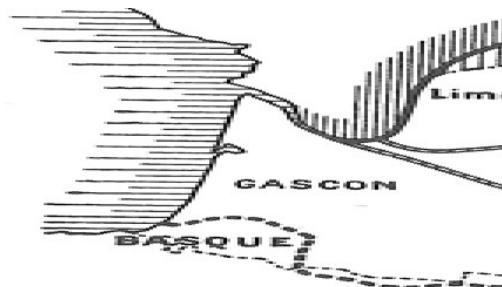


Développement des
ressources et outils
TAL en occitan :
réalisations du
Congrès permanent
de la langue occitane
et applications
possibles pour les
parlers du Croissant

*B. Dazeas, directeur du Congrès,
Aure Séguier, développeuse au
Congrès*



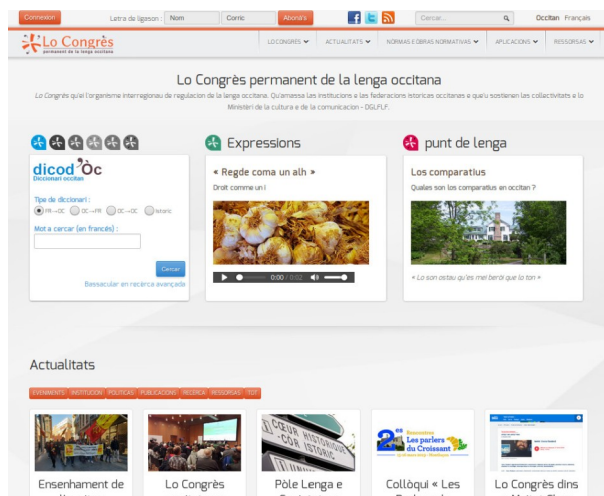
Le Congrès permanent de la langue occitane



- Organisme interrégional de régulation de la langue occitane.
- Deux conseils assesseurs :
 - Conseil linguistique.
 - Conseil des usagers.
- Missions :
 - Production d'outils linguistiques de référence.
 - Régulation linguistique (par la production).
 - Recherche.



Portail lexical occitan



- Multidictionnaire (fr-oc, oc-fr, oc-oc "historiques") : 20 dictionnaires pour environ 500 000 entrées cumulées.
- Conjugueur automatique (languedocien et gascon, provençal en cours de développement).
- Bases terminologique et toponymique
- Normes et œuvres normatives.
- Ressources externes.

Le traitement automatique des langues (TAL)

- C'est « une aire de la recherche et de l'application qui explore comment les ordinateurs peuvent être utilisés pour comprendre et manipuler des textes ou discours en langage naturel pour faire des choses utiles » (Chowdhury, 2003)
- Il permet de traiter de manière informatisée un très grand nombre de données textuelles
- Il permet de rechercher dans des données avec une précision et une rapidité impossibles pour un humain
- Domaines d'application : linguistique de corpus, traduction automatique, classification de documents...

La Feuille de route de développement numérique de l'occitan : objectifs

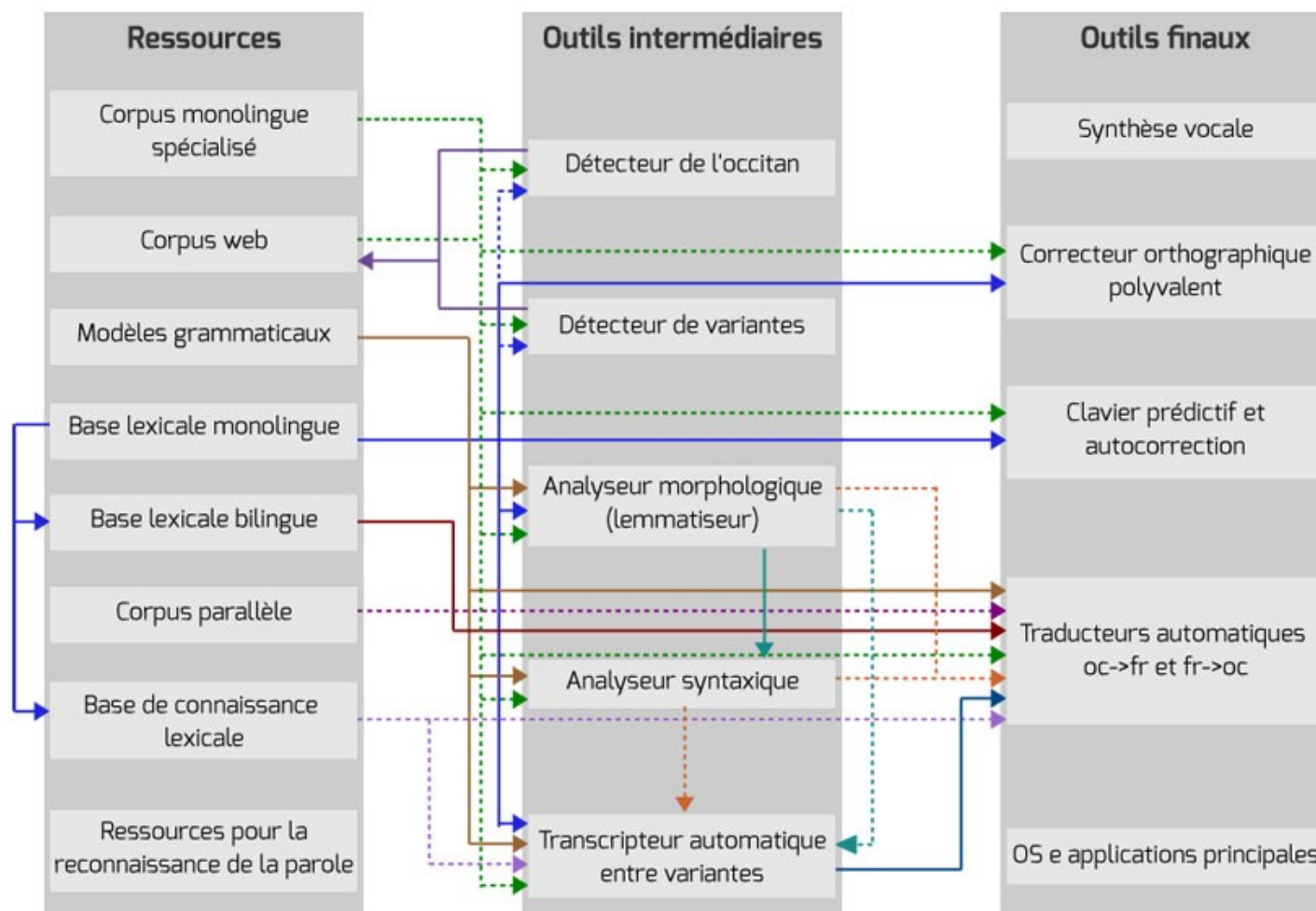
- Définir un ensemble d'objectifs spécifiques et précis pour développer les technologies du langage de l'occitan, à obtenir pour la fin 2019
- Proposer un calendrier pour atteindre ces objectifs pendant la période 2015-2019.

La Feuille de route de développement numérique de l'occitan : objectifs

- Réunion d'une vingtaine de membres de structures actrices du numérique occitan
- Mise en commun d'expertises, de connaissances et d'expériences dans le domaine de l'occitan et des technologies du langage
- 3 étapes :
 - Inventaire et diagnostic des ressources et des outils disponibles pour l'occitan
 - Étude des ressources et des outils qui existent pour d'autres langues (interventions d'experts)
 - Définition des besoins, des objectifs, rédaction du calendrier

Traitement automatique des langues (TAL)

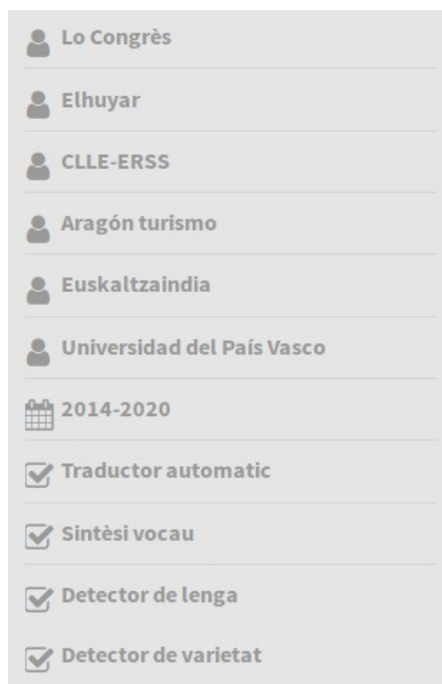
La Feuille de route : ressources et outils à construire



Projets structurants pour le TAL

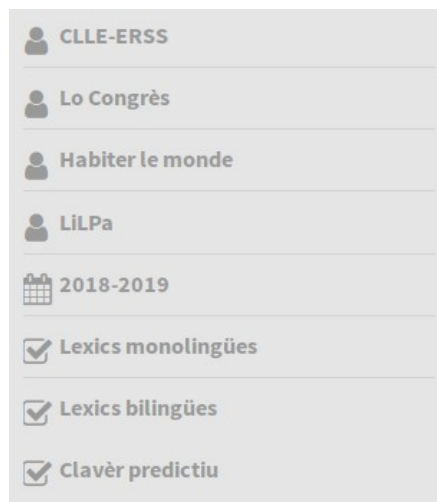


Coopération scientifique et transfert de technologies des langues occitan - basque - aragonais



ROLF

Ressources Outils Langues de France : alsacien, occitan, picard



DGLFLF

N-Aquitaine-Euskadi-Navarre

Développement d'un correcteur orthographique universel en occitan.



Eurorégion Nouvelle-Aquitaine - Euskadi

Formatage en TEI des données du Congrès

- Le format TEI, qu'es aquò ?
 - Un sous-langage du XML (format lisible par les machines)
 - Norme pensée pour le balisage de contenu textuel
 - Propose des balises et des attributs spécifiques pour les dictionnaires et lexiques
 - Le texte original n'est pas modifié. On y ajoute des balises qui qualifient le texte qu'elles encadrent :

`<number norm="plural">plur.</number>`

Formatage en TEI des données du Congrès

aiguille	f.	agulha	Basic (commun gasc-lang)
aiguille	nf	agulha, aguilha ; aijulha (Han). Tròp ponchua es l'agulha que ponha dinque au sang.	Faure (vivaro-alpin)
aiguille	nf.	(a)gulha; bròca (<i>à tricoter</i>).	Atau que's ditz (gascon)
aiguille	nf.	(a)gulha	Limousin
aiguille	nf	• agulha • aiguille à tricoter agulha de debàs, agulha per brocar, bròca.	Laus (languedocien)
aiguille	nf	agulha ADL ; aiguille à tricoter : bròcha LBR ; aiguille d'horloge : <i>det nm Grapina en lo relòtge ont los vielhs dets de fer sabon pus trobar l'ora (L. Peiròche d'ès Baus, 1927)</i> ; aiguille de conifère : <i>espina nf de gama</i> .	Omelhièr (Auvergnat)
aiguille	f.	agulha * ~ à tricoter bròca, agulha de debàs * ~ des heures agulha pichona * ~ des minutes agulha gròssa * chercher une ~ dans une botte de foin cercar una agulha dins un palhièr * de fil en ~ de fil en cordura	Lagarde (languedocien)
aiguille	f.	agulha ; ~ à tricoter bròca ; ~ de conifère garbàia G ; tapis d'~ de pins balòt M	Rei Bèthvéder (gascon)
aiguille	f.	• 1. (gén.) (a)gulha (y compris sur voie ferrée, pointe de rocher...). • 2. (~ à tricoter) bròca. † (~ pour coudre des sacs) agulha saquèra. (~ de bourrelier) agulha botissa. † (grosse ~ à reprendre) agulha de passatge. † (grosse ~ passe-cordon) passacordon m. ; passacòrda. m./f. † (ouvrier qui fait des ~) agulhèr. † (enfiler une ~) engulhar ; en-hialar ; apunterar. • 3. (~ de pin) garbalha.	Per Noste (gascon)

Formatage en TEI des données du Congrès

<u>aiguille</u>	<u>nf</u>	agulha <u>aiguille à tricoter</u> : agulha de debàs
-----------------	-----------	--



```
<entry n="833">
  <form>
    <orth>aiguille</orth>
    <gramGrp>
      <pos norm="noun">n</pos>
      <subc norm="common"></subc>
      <gen norm="feminine">f</gen>
      <gram type="eaglescat" norm="N0210">
        </gram>
      </gramGrp>
    </form>
    <sense n="I-A-1-a">
      <cit type="translation" xml:lang="oc-lengadoc">
        <form>
          <orth>agulha</orth>
        </form>
      </cit>
      <re type="expression">
        <form type="phrase">
          <orth>aiguille à tricoter</orth>
        </form>
        <sense>
          <cit type="translation" xml:lang="oc-lengadoc">
            <form type="phrase">
              <orth>agulha de debàs</orth>
            </form>
          </cit>
        </sense>
      </re>
    </sense>
  </entry>
```

Formatage en TEI des données du Congrès

- Objectifs

- Harmonisation de données en interne
- Échange de données avec les partenaires
- Données lisibles par les machines (intégration dans les outils)
- Création rapide de nouvelles ressources : lexiques de formes fléchies, dictionnaire des rimes, dictionnaire des synonyme
- Rédaction d'un document-guide d'encodage TEI :

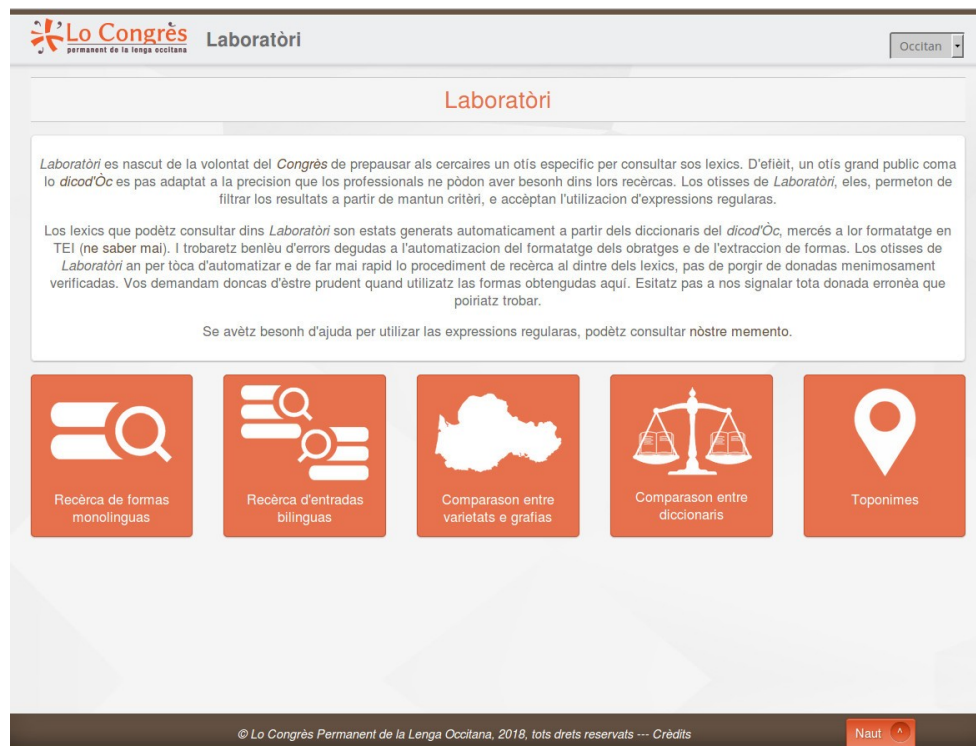
→ http://www.locongres.org/images/docs/choix_encodage.pdf

Dépôt de subtags de variétés et graphies à l'IANA

Varietat	Subtag
Auvernhât	auvern
Gascon	gascon
Lemosin	lemosin
Lengadocian	lengadoc
Provençal	provenc
Vivaroaupenc	vivaraup
Varietats del creissent	creiss
Aranés	aranes
Cisalpenc	cisaup
Niçard	nicard

Grafia	Subtag
Grafia classica (se i a besonh de precisar)	grclass
Grafias mistralenca e derivadas (febusiana, escòla dau Pò)	grmistr
Grafia italianizanta	grital

Laboratòri, outil de consultation avancée de lexiques



→ <http://laboratori.locongres.org>

Laboratòri, outil de consultation avancée de lexiques

Recercar
Los camps d'ortografia accèptan las expressions regularas. ?

Forma occitana :
ortografia :

categorias gramaticalas :
☒ adjectiu ☒ advèrbi ☒ article ☒ autre ☒ conjoncion ☒ determinant ☒ interjeccion ☒ nom comun ☒ nom pròpri ☒ numeral ☒ preposicion ☒ pronom ☒ vèrb
genres :
☒ feminin ☒ masculin ☒ neutre ☒ sens genre
nombres :
☒ singular ☒ plural ☒ sens nombre
varietats de l'occitan :
☒ auvernhat ☒ comun ☒ gascon ☒ lemosin ☒ lengadocian ☒ provençal ☒ vivaroalpenc
grafias :
☒ grafia classica

Autra forma :
ortografia :

categorias gramaticalas :
☒ adjectiu ☒ advèrbi ☒ article ☒ autre ☒ conjoncion ☒ determinant ☒ interjeccion ☒ nom comun ☒ nom pròpri ☒ numeral ☒ preposicion ☒ pronom ☒ vèrb
genres :
☒ feminin ☒ masculin ☒ neutre ☒ sens genre
nombres :
☒ singular ☒ plural ☒ sens nombre
lengas :
☒ francés

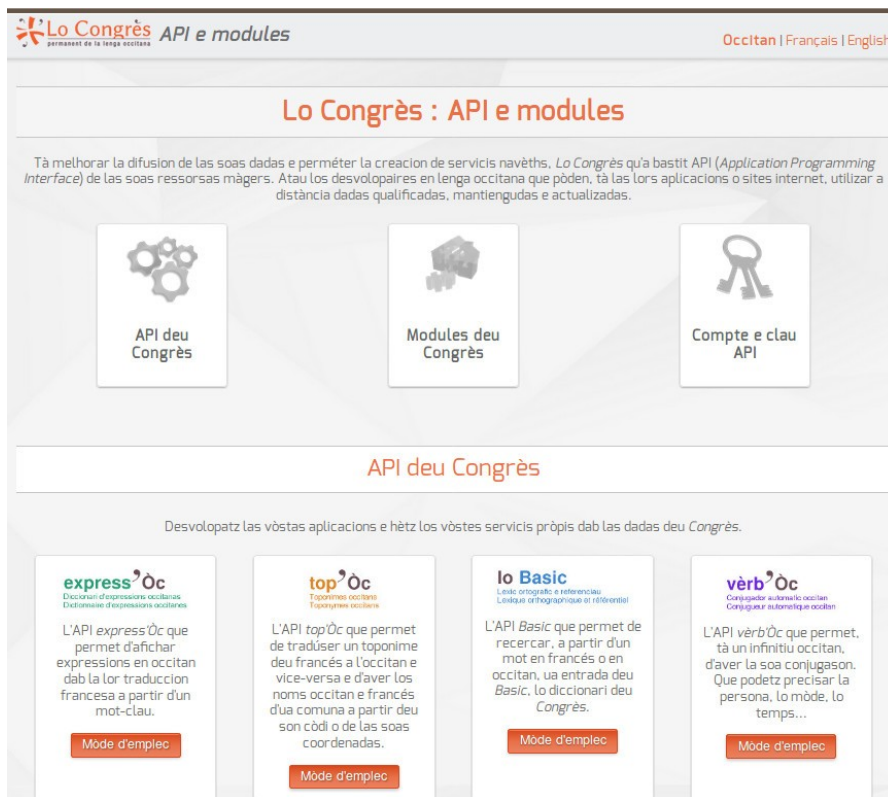
diccionaris sorsas :
☒ ATAU fr-oc ☒ BASIC ☒ BASIC gasc ☒ BASIC leng ☒ CREO fr-oc ☒ FAURE ☒ LAGAR fr-oc ☒ LAUX fr-oc ☒ LEMOSIN ☒ OMLH ☒ PNST fr-oc ☒ RBVD fr-oc

parelhs de mots :

Totes los mots

Filtrar Reinicializar

API pour les développeurs



Lo Congrès API e modules

Occitan | Français | English

Lo Congrès : API e modules

Tà melhorar la difusion de las soas dadas e permèter la creacion de servicis navèths, *Lo Congrès* qu'a bastit API (*Application Programming Interface*) de las soas ressorsas màgers. Atau los desvolopaires en lenga occitana que pòden, tà las lors aplicacions o sites internet, utilizar a distància dadas qualificadas, mantengudas e actualizadas.



API deu Congrès



Modules deu Congrès



Compte e clau API

API deu Congrès

Desvolopatz las vòstras aplicacions e hètz los vòstres servicis pròpris dab las dadas deu Congrès.

express'Òc
Diccionari d'expressions occitanas
Diccionari d'expressions occitanas

L'API *express'Òc* que permet d'afichar expressions en occitan dab la lor traduccion francesa a partir d'un mot-clau.

Mode d'emplec

top'Òc
Toponimes occitans
Toponimes occitans

L'API *top'Òc* que permet de traduser un toponime deu francés a l'occitan e vice-versa e d'aver los noms occitan e francés d'ua comuna a partir deu son còdi o de las soas coordenadas.

Mode d'emplec

lo Basic
Lèxica orografica e referenciala
Lèxica orografica e referenciala

L'API *Basic* que permet de recercar, a partir d'un mot en francés o en occitan, ua entrada deu *Basic*, lo diccionari deu Congrès.

Mode d'emplec

vèrb'Òc
Conjugador automatic occitan
Conjugador automatic occitan

L'API *vèrb'Òc* que permet, tà un infinitiu occitan, d'aver la soa conjugason. Que podetz precisar la persona, lo mòde, lo temps...

Mode d'emplec

→ <http://api.locongres.org>

Les modules intégrables du Congrès

Lo module express'Òc que'vs permet d'afichar a l'azard, suu vòste site o blòg, ua expression en occitan dab ua illustracion visuau e un son.

Configuracion deu module

Varietat de l'expression :

Longor : px

Hautor : px

Color deu hons : ☐ Transparent

Color deu titol :

Color deu tèxte :

Color de la bordadura : ☐ Pas de bordadura

Còdi a copiar-pegar

```
<iframe title="Module express'Òc" src="https://api.locongres.org/modules/expressoc_frame.php?fond=f5f5f5&titre=ea7148&text=666666" width="300" height="380" frameborder="0" style="border:1px solid #666666;"></iframe>
```

Vista

Jos color de
Sous prétexte de



Mei d'expressions dins l'express'Òc deu Congrès

Lo module top'Òc que'vs permet d'afichar a l'azard, suu vòste site o blòg, un toponime dab los sons noms occitan e francés, a partir o pas d'un numèro de departament.

Configuracion deu module

Departament :

Longor : px

Hautor : px

Color deu hons : ☐ Transparent

Color deu titol :

Color deu tèxte :

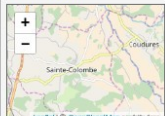
Color de la bordadura : ☐ Pas de bordadura

Còdi a copiar-pegar

```
<iframe title="Module top'Òc" src="https://api.locongres.org/modules/topoc_frame.php?fond=f5f5f5&titre=ea7148&text=666666" width="250" height="300" frameborder="0" style="border:1px solid #666666;"></iframe>
```

Vista

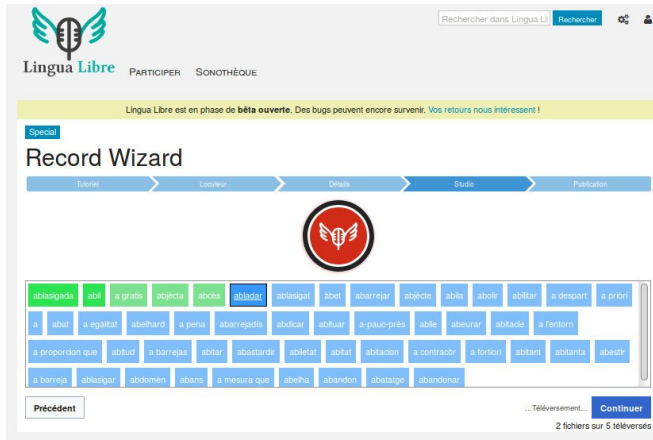
Senta Coloma
Sainte-Colombe (40)



Mei de toponimes dins lo top'Òc deu Congrès

→ <http://api.locongres.org>

Lingua Libre



- Bibliothèque contributive d'enregistrements sonores pour langues de France
- Interface permettant l'enregistrement de mots ou phrases
- Enregistrements documentés et versés à wikimédia commons.
- Pour l'occitan : enregistrement du Basic (environ 20 000 mots en languedocien et gascon)

→ <https://lingualibre.fr>

BaTelÒc

- Base textuelle en langue occitane
- Plusieurs genres, plusieurs graphies, plusieurs variétés (dont les parlers du croissant)
- Interface de consultation simple ou avancée
- Plus de 3 millions de mots pour une centaine de textes

→ <http://redac.univ-tlse.fr/bateloc>

Développer les ressources pour le nord occitan



- Manque de ressources (lexiques, corpus, etc.), notamment normalisées et utilisables informatiquement.
- Espace nord occitan : développements prioritaires
- Politique linguistique en Auvergne-Rhône-Alpes
- Conjugueur limousin (projet)

Une seule application pour les intégrer toutes...

dicod'Òc
Dictionnaire occitan

Varietats : ☒ auvernhat ☒ gascon ☒ lemosin ☒ lengadocian ☒ provençal ☒ vivaroalpenc

Resultats exactes :

- [oc] **marcha** (nt.)
- [oc] **Marcha** (rp. f.)
- [oc] **marcha** (p3 sg. ind. pres.)
marchar (vt.)
- [fr] **marcha** (p3 sg. ind. pass.)
marcher (nm.)

Autres resultats :

- [oc] **marchà** p3 sg. ind. fut. de **marchar** (vt.)
- [oc] **marchand** (nm.)
- [fr] **marchand** (nm.)
- [oc] **marchar** (vt.)

→ veire mai

Expressions

- Marcher à quatre pattes : *Anà'n a grapas*
- Marcher avec une canne : *Anà'n dab tres camas*

→ Véder las expressions

Punt de lenga
Le -sh- et le -ch- en gascon



Comment écrire le son [j] en gascon ?
→ Léger mei

[fr] marcha
3^{ra} persona del passat simple de **marcher** (vèrb)
→ Cercar dins lo Wiktionnaire

Traduccions

Dens lo **Basíc**, diccionari deu Congrès

marcher (v)	• (général) ◦ comun : caminar ; marchar ◦ auri. : chaminar • (machine) ◦ comun : foncionar ; marchar
--------------------	--

Dens lo **term'Òc**, basa terminologica deu Congrès

marcher (v) [anar per las emparas](#) [Petita enfança \(gasc.\)](#)

Dens los diccionaris francés-occitan

marcher (v)	• marchar ; caminar • (mécanisme, affaire) anar plan ; anar ; marchar ; rotlar (ò) ; foncionar	RBVD (gasc.)
marcher (v)	• (se déplacer) marchar ◦ A quatre chambras, amai pòt pas marchar, da qu'aquò es ? La taula (davinaïha, Chassinhol) ◦ marcher beaucoup ; galopar (LBR) ◦ marcher à vive allure ; marchar a pas ralevats (ADL) ; galopar ◦ marcher à grandes enjambées : cambadar (LBR, ADL) ; cambar long (MAZ) ◦ marcher sur : machicar ◦ vi Z-a machiat un estont de vacha (A. Delieu) ◦ marcher dans l'eau : gasar (vt.) ; gasar l'aiga (ADL) ◦ Per passar l'aiga, nos chauguèt gasar.	Ometièr (auv.)
marcher (v)	• (machine) marchar ; foncionar ; virar • (usuel) marchar ; chaminar ; anar	Lemosin (lem.)
marcher (v)	• (général) marchar ; caminar ◦ marcher devant : davantejar marchar au capdavant ◦ marche ou crève ! : marcha que marcharàs, se l'amèstas que crebaràs ! ◦ l'enfant commence à - : lo mainatge que hè camas ; lo mainatge que hè cametas ◦ il commence à vouloir marcher : que trepeja ; que comença de trepar ; que vòu téisher ◦ il marche en s'appuyant où il peut : que'n va per las emparas • (façons de marcher) ◦ marcher pieds nus : anà'n pèdescauc ; anà'n a sola de pè ◦ marcher sur la pointe des pieds : har puntetas ; anà'n de puntetas ◦ marcher à quatre pattes : anà'n a grapas ◦ marcher vite : marchar de pos ; traçar ; tirar camin ; draihar ; artelhar ; espederà(s) ; escarpinar ; pederiquejar ◦ marcher lourdement : patassejar ◦ marcher en file indienne : cordadejar corderejar	



Mercés plan de vòstra atencion