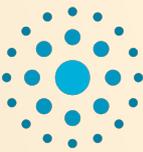


# Enquêtes dialectales et collecte de témoignages oraux



Le projet VERA - Vidéo, Ethnotextes et Ressources Associées,  
de la linguistique de terrain à la valorisation  
du patrimoine culturel immatériel

Pierre-Aurélien GEORGES,  
laboratoire BCL



Huma-Num  
la TGIR des humanités numériques

CORpus  
Langues  
Interactions

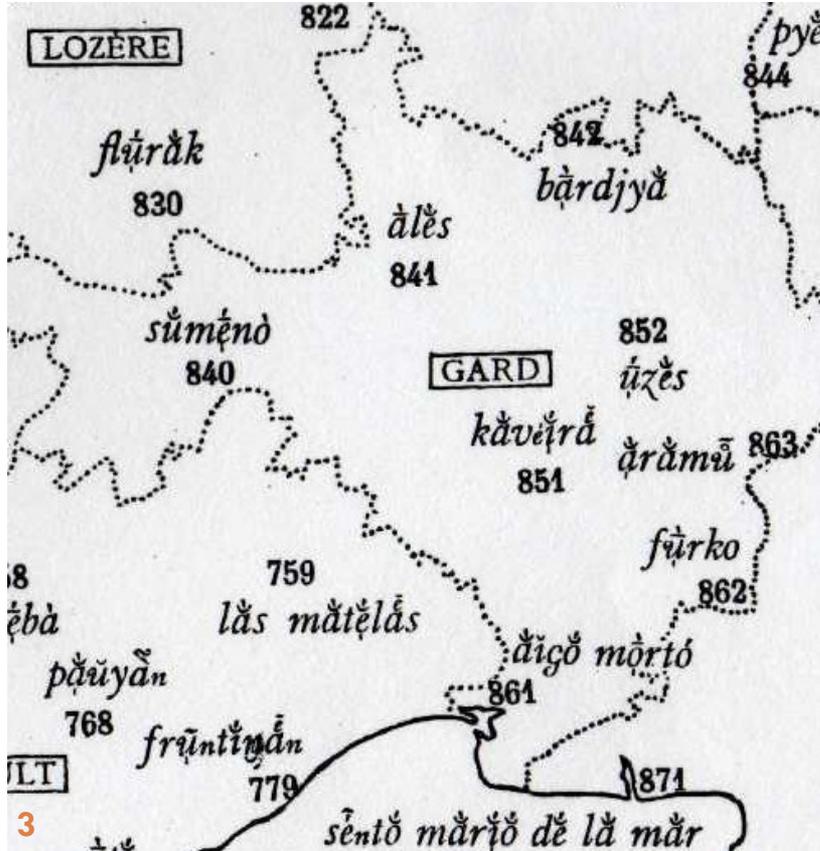




Carte de géographie physique d'Occitanie

© macarel.org

# Thesaurus Occitan



- 900+ points d'enquête
- 8000+ questions
- 1,2 million de réponses

<http://thesaurus.unice.fr>



<http://thesaurus.unice.fr>

vera.project.therokuapp.com/search/lexicon?query=nice&term=Nice&type=cities&uid=56343454-eada-64f4-0001

THESOC  
La base de données dialectales sur l'Occitanie et ses marges

# Nice

Communes

Filtrer résultats

Medias **Lexique (34)**

Thèmes

Thèmes	Notion	Catég.		
Famille Homme	père	[+]		<b>paire</b> (PATER) ALLy 945 ALMC 1607 ALAL 1027 [...] THESOC Q7562
Oiseaux Nature	pinson ( <i>Fringilla coelebs</i> )	[+]	Substantif	<b>*alpin</b> <b>°°balarina</b> <b>pinçon</b> (*PINCIO) [...] ALLy 501 ALMC 316 ALAL 418 [...] THESOC Q7719
Oiseaux Nature	 huppe (oiseau) ( <i>Upupa epops</i> )	[+]	Substantif	<b>puput</b> (ÜPÜPA) <b>huppe</b> (fr) (ÜPÜPA) <b>°°pupa</b> [...] ALMC 319 ALO 421 ALCe 548 [...] THESOC Q5587
Oiseaux Nature	 roitelet ( <i>Regulus regulus</i> )	[+]	Substantif	<b>°°boscarla</b> <b>°°boscarleta</b> [...] ALLy 512 ALMC 316

Q saint-jean

- Saint-Pierre-Saint-Jean  
Communes
- bête à saint Jean (fr)  
Lemmes
- corne de saint Jean (fr\*)  
Lemmes
- faire saint Jean (fr)  
Lemmes
- feu de la Saint-Jean  
Questions



# Les spécificités du THESOC

- Oralité (linguistique de terrain)
- Cartographie (toutes les données sont géolocalisées)
- Grande diversité de dialectes  
(versus occitan « standardisé » / « langues nationales »)

# Pourquoi de nouvelles enquêtes ?

6

- Compléter les données des atlas linguistiques
- Données plus récentes (diachronie)
- Multimédia (audio/vidéo et texte synchronisés)
- Questionnaire élargi : syntaxe, discours libre



**Enquête linguistique à Breil-sur-Roya (Alpes-Maritimes), 2008.**  
*Michèle Olivéri & Jean-Claude Ranucci*





L'épreuve de chant  
du Certificat d'Études



La fabrication du  
fromage



L'âge des brebis



Les 5 repas du  
paysan  
limousin



Le métier de  
typographe

1980

2002

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE  
Centre Régional de Publications: Marseille

# TRADITION ORALE ET IDENTITE CULTURELLE

problèmes et méthodes

Jean-Claude BOUVIER, Henry-Paul BREMONDY  
Philippe JOUTARD, Guy MATHIEU, Jean-Noël PELEN



EDITIONS DU CNRS

« *l'ethnotexte est principalement un discours sur le passé qui est recueilli [...] pendant un entretien, ou une enquête, dont la thématique appartient au discours culturel oral d'une communauté.* »

Pierre Chabert

*Ethnotextes : la conscience de l'espace. L'exemple d'un conte, bulletin de l'AFAS n°22, p. 2-8*

<https://journals.openedition.org/afas/1469>

**2006 – 2009 : HOROYA**

**Histoire Orale des vallées de la Roya et Bévéra**

*Resp. Th. Rosso (LAPCOS), 2006-2009.*



**2010 – 2012 : PHC Procope**

**Syntactic Microvariation : Clitic Pronouns in Romance Languages**

*Resp. M. Oliviéri (BCL) & G. Kaiser (Universität Konstanz, Allemagne)*



**2012 – 2015 : DADDIPRO**

**Dialectal, acquisitional, and diachronic data and investigations on subject pronouns in Gallo-Romance**

*Resp. M. Oliviéri (BCL) & G. Kaiser (Universität Konstanz, Allemagne)*



**2013 – 2015 : SyMiLa**

**Syntactic Microvariation in the Romance Languages of France**

*Resp. P. Sauzet (Université Toulouse Jean Jaurès) & D. Sportiche (Paris)*



# 2006 – 2017



**100+ enquêtes de terrain**



**75 localités**

**(France méridionale & Italie)**



**120+ locuteurs interviewés**



**1200+ questions posées**

**(5 questionnaires de phrases)**



**60 h de vidéo inédites à traiter**

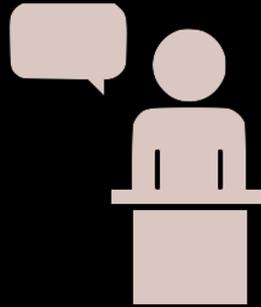


**58 ethnotextes récoltés**

# 3 publics identifiés



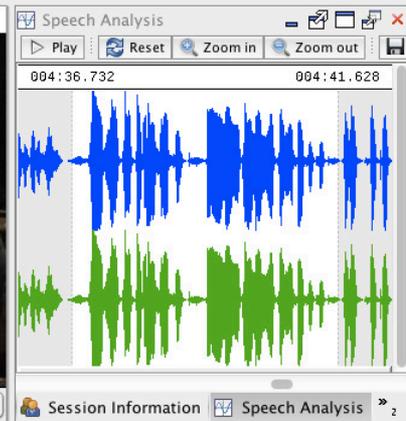
**Grand Public**



**Apprenants  
Enseignants  
Patoisants**



**Linguistes  
Ethnologues  
...**



Record Data

# 65 Speaker: Pascal Martini Exclude from searches

Segment [ 004:36.732-004:41.628 ]

IPA Target [ e 'ʃi ke i pɔː faː aː 'brusuː ! me l eʒ ɛ tra'vaju ke nu 'fasu tʃy ]

Graphie [ E sci che i pòo fâa ař brussu ! <i lang="fr">Mais </i> l'ei 'n travayu che nu fassu ciü. ]

Traduction [ Et bien sûr qu'on peut faire le brousse ! Mais c'est un travail que je ne fais plus. ]

Notes [ Le "brousse" désigne un fromage de lactosérum originaire de Provence, à la pâte blanche et à la texture grumeleuse. ]

traduction en Italien [ Ma certo che si puo' fare la brossa! Ma è un lavoro che non faccio più. ]

id: 6f6e49b5-48d8-4d51-8bbf-d9248642ae59 Tier: Group: Character:

Record List :

Segment	Speaker	IPA Actual	Traduction	Graphie	IPA Target	traduction en Italien
002:23.202...	Jean-Cla...	[Je vais vous demander un secret : comment vous faites le fromage ?]				Vi chiederò 'un segret...
002:26.914...	Pascal M...	[ ]	Ah, pour faire le fromage, c'est un secret que je ne le dis à personne.	Ah, peř fâa ař fru...	[aː   peř faː aː fr ...	Ah, per fare il formag...
002:34.066...	Jean-Cla...	[Expliquez-moi un peu...]			[ ]	Mi spieghi un po'...
002:35.856...	Pascal M...	[ ]	Ah, je ne peux, non... Ah non, non, non, non, non...	Ah, nu pòsciü, nu...	[aː   nu 'pœʃu, nu...	Ah, non posso, no...n...
002:39.942...	Pascal M...	[ ]	et après les autres me prennent le brevet, hein !	E dopu i autri me...	[e 'dœpu i 'jœwtri ...	e dopo gli altri mi pr...
002:46.048...	Jean-Cla...	[Ils n'ont plus de vaches ! Il faut faire chauffer le lait ?]			[ ]	Ma no, non hanno pi...
002:51.807...	Pascal M...	[ ]	Eh ! Il faut faire chauffer le lait à environ 35°C,	Eh ! Venta fâa sca...	[eː   'vœta faː ʃka...	Eh, bisogna far scalda...
002:56.760...	Pascal M...	[ ]	et puis il faut y mettre la présure,	e pöi venta meteř...	[e pœʃ 'vœta mete...	e poi bisogna metter...
002:59.426...	Pascal M...	[ ]	puis le laisser reposer, couvert, quarante minutes, à peu près quarante minutes,	pöi lasciařlù paus...	[pœʃ laʃař lu pa...	per lasciarlo riposare...
003:06.261...	Pascal M...	[ ]	puis ensuite il faut le frapper avec le tranche-caillé ;	pöi dopu venta sb...	[pœʃ 'dœpu 'vœta ...	poi in seguito bisogna...
003:11.703...	Pascal M...	[ ]	puis, il doit être rassemblé avec les mains, puis le prendre avec l'étamine,	pöi ař va recampa...	[pœʃ aː va rekæp...	poi, va ricomposto co...
003:18.213...	Pascal M...	[ ]	et puis le mettre à égoutter, le mettre à... le mettre à égoutter.	e pöi u mete a res...	[e pœʃ u 'mete a r...	e poi metterlo a scola...
003:25.292...	Jean-Cla...	[Et après on le met dans la forme ?]			[ ]	E dopo lo si mette nel...
003:29.286...	Pascal M...	[ ]	Eh... moi je ne le mets pas dans la forme, car je le fais carré, avec les mains.	Eh... mi nu u met...	[eː   mi nu u 'met...	Eh... io non lo metto ...
003:36.024...	Jean-Cla...	[Et après il est fait au bout de combien de temps ? Il faut le saler ?]			[ ]	E dopo, in quanto te...
003:39.290...	Pascal M...	[ ]	Ah ! Il faut le saler... le soir il faut le saler d'un côté...	Ah ! Venta sařařlù...	[aː   'vœta sařař l...	Ah, bisogna salarö, la...
003:43.271...	Pascal M...	[ ]	Le lendemain matin, il faut le tourner et le saler de l'autre côté...	Ř'endeman de ma...	[řœdemâ de mař, ...	L'indomani di mattina...
003:49.663...	Pascal M...	[ ]	Et puis il doit être essuyé, il doit être tourné, et de temps en temps il faut s'occuper, ...	e pöi ař va sciügä...	[e pœʃ aː va ʃyğ...	Poi va asciugato, va gi...
003:56.988...	Jean-Cla...	[Et combien de temps il faut le garder avant de le vendre ?]			[ ]	Ma quanto tempo bis...
004:00.347...	Pascal M...	[ ]	Et ça dépend : il y en a qui les vendent même frais, à partir de...	E ça depende : ař ...	[e sa de'pœde   a...	Eh... dipende: ci sono ...

Save as CSV...



Sous-titres au choix :

- français  
- occitan 
- phonétique API 
- etc.

Mon père s'était échappé des allemands ; on était là-bas dans la montagne.



1:13 / 1:57



BASES  
CORPUS  
LANGAGE



ANR-15-CE0007-01





# Contextualisation historique

BASES  
CORPUS  
LANGAGE



Centre de la Méditerranée  
**MMC**  
EA 1193 Nice  
Moderne et Contemporaine



## L'hiver 1956 : le plus froid du XX<sup>e</sup> siècle.

L'hiver 1956.

L'hiver 1956 est marqué, dans l'hémisphère Nord, par une vague de froid sans précédent. Historiquement, il rappelle le Grand Hiver de 1709 qui voit le vin geler dans les coupes de Versailles tout comme l'encrue des encriers de l'école de Pierre Lapomroulle à Sorges dans le Périgord en 1956. Cet épisode n'est pas isolé puisque, de 1999 à 2010, le Grand Hiver intense touche la France. Il en sera de même en 1963, en 1996 et 2009-2010. Cependant, les records de température font le mois le plus froid du XXI<sup>e</sup> siècle.

## Les bombardements à Nice Seconde Guerre mondiale.

Entre 1943 et 1944, plusieurs séries de bombardements sont effectuées sur la ville de Nice. Le 1 et le 3 mai, les premières bombes atteignent Nice. Une nouvelle série de bombardements a lieu le 27 novembre 1943 et détruit l'Arénas. Par la suite, une dizaine de bombardements ont lieu sur les ponts du Var, l'Hippodrome, le Port et le Vieux-Nice.

Le bombardement du 26 mai 1944 compte parmi les plus importants de la Seconde Guerre mondiale. Il provoque de nombreux dégâts dans la ville. Dans le cadre du Transportation Plan, les communications dans le Sud de la France doivent être sécurisées. L'objectif est de détruire les infrastructures ferroviaires notamment la gare de Saint-Roch. Pour cela, l'armée alliée adopte la méthode « américaine » : le choix de largage à haute altitude présente des avantages certains : moins de victimes civiles. Cette opération militaire vise Nice. Le bombardement de la gare de marchandises de Saint-Roch touche également la gare de Riquier, de Pasteur et de République. Le bilan est d'environ 350 victimes.

Plusieurs plaques commémoratives rappellent cet événement au 100 de l'Avenue Séveria, à la Place de la Brigade, à l'École Jean Macé et à Riquier.

Une autre série de bombardement a marqué les esprits des Niciens le 27 août. Un avion, non identifié mais probablement américain, survole et ravage la ville. La nuit du 13 août huit bombes sont ainsi lâchées sur la ville. Les bombardements se poursuivent dans les nuits suivantes et touchent le Mont-Boron, la Rue des Orangers, Massena, le Vieux-Nice et de

## Le Certificat d'études primaires



aires (CEPE) ». Il s'agit d'un « certificat ». Le mode d'évaluation élémentaire en dictée est basé sur des épreuves éliminatoires. De 1882 à 1967, la dictée est la plus active.

Le dernier est en fait un mode de plus de dictée à la dictée éliminatoire. Ce dernier est basé sur la dictée éliminatoire.

Le 16 juin 1880, un arrêté ministériel harmonise les modalités d'évaluation du certificat d'études. Désormais le nombre de cinq fautes à la dictée devient éliminatoire.

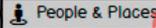
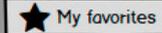




# How to make traditional cheese

St-Martin-du-Var (06670), juin 2014.

Ethnotext



people, place, word, atlas map, ...



Informant(s) : [Pascal Martini](#)

Source : [DADDIPRO](#)



Export text



Autoplay

IPA transcriptions

Recording



Historical Background

## Les Italiens à Nice

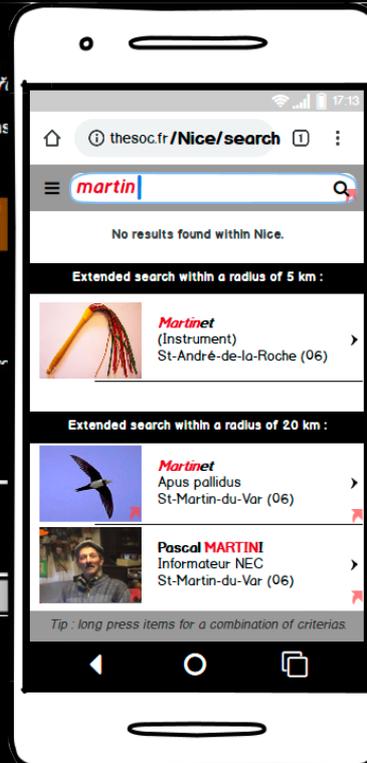
À Partir de l'annexion du Comté de Nice, de nombreux Italiens s'installent à Nice et particulièrement dans la Vieille-Ville. Leur nombre croît jusque dans la première moitié du XXe siècle. En 1975, Paul Caramagna étudie cette communauté présente à Nice en 1921 en s'appuyant sur le recensement de la population. 27% des habitants de Nice sont étrangers et 22,6% sont d'origine italienne dont seulement 48% sont des hommes. Ils habitent majoritairement dans la Vieille-Ville selon un axe Garibaldi-Port. En 1921, ce sont ainsi 12 pêcheurs italiens qui vivent dans le quartier du Malonat. La

User Comments about this ethnotext

F. G. Le su  
grand  
J.-C. R. En ter  
F. G. En ter  
De m  
taure  
J.-C. R. Quel  
F. G. Hé... j  
Le u

1. *tofu* désigne le

de... de mon  
e c'était le *tofu*  
que, en tendas  
cu...  
egran, l'e'a a  
avez-vous ?  
ni...  
a a f me che u



Thesaurus Occitan (THESOC)

https://thesoc.fr/sentences?ipa#map+facets

THESOC - Thesaurus Occitan  
La base de données dialectales sur l'Occitanie et ses marges

Media | Lexicon | Sentences | People & Places

Questions | Productions | Export selection

IPA transcriptions

Ref.	People & Location	Sentence
Question NEC 195	José TRUCHAUD <b>St-Martin-du-Vaï (06)</b> ALF 872 · ALP 34	Quoura eravan jouve, avèn passat de bouòn moumen mé eli. [k' uʁa eʁ' avən dʒ' uve l av' èm pas' at de bwɔ̃m mum' èm me' eli] When (we) were young, (we) have spent some good moments with the Sujet nul   compl. de   aux. avoir   adverbe
Question NEC 195	Jean-Claude PIERRE <b>Eymoutiers (87)</b> ALF 435 · ALAL 34	Quand nos eron juno, nos passerem de brave temps como eli. [kã nu eʁɔ dzyno nu pãʒeʁe] de brav teɔ̃ kumo ji] When we were young, we spent some good time with them. compl. de   aux. avoir   adverbe
Question PAM2 256	José TRUCHAUD <b>St-Martin-du-Vaï (06)</b> ALF 872 · ALP 34	Eh, n'en fan de caua ! Hey, (she_PL) it[PRO] make lot_of things ! Sujet nul   compl. de
Question PAM2 256	Jean-Claude PIERRE <b>Eymoutiers (87)</b> ALF 435 · ALAL 34	[a lœ ne] fɔ̃ dœ la soza ! Hey, she_PL it[PRO] make lot_of the thing. compl. de

Map

Presence of Sujet nul

Eymoutiers (87)  
ALF 435 · ALAL 34  
Jean-Claude Pierre  
de | aux. avoir | adverbe

Facets

- Filter by genre
- Filter by questionnaire
- Filter by tags
  - Sujet : Sujet nul | Expiètf | Impersonnel
  - Verbe : aux. avoir | aux. être

Users comments about this question

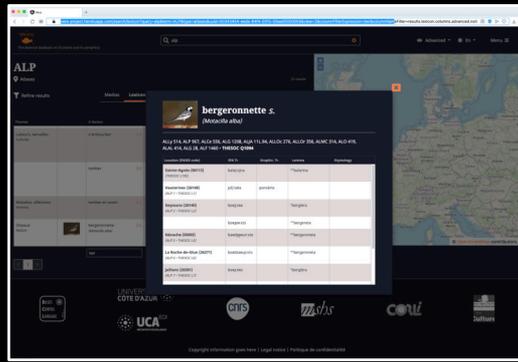


## Annotations linguistiques

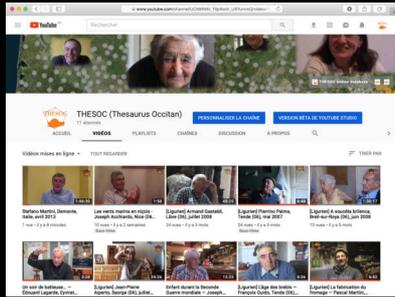
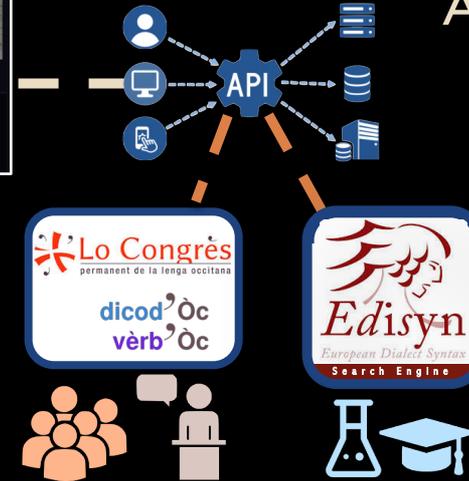
→ cartographie de phénomènes syntaxiques ou lexicaux



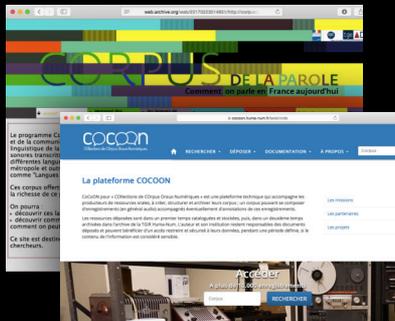
Plateforme web dédiée  
cartes, concordancier, ...



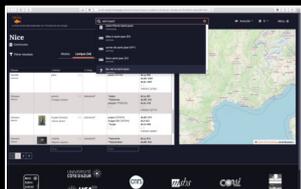
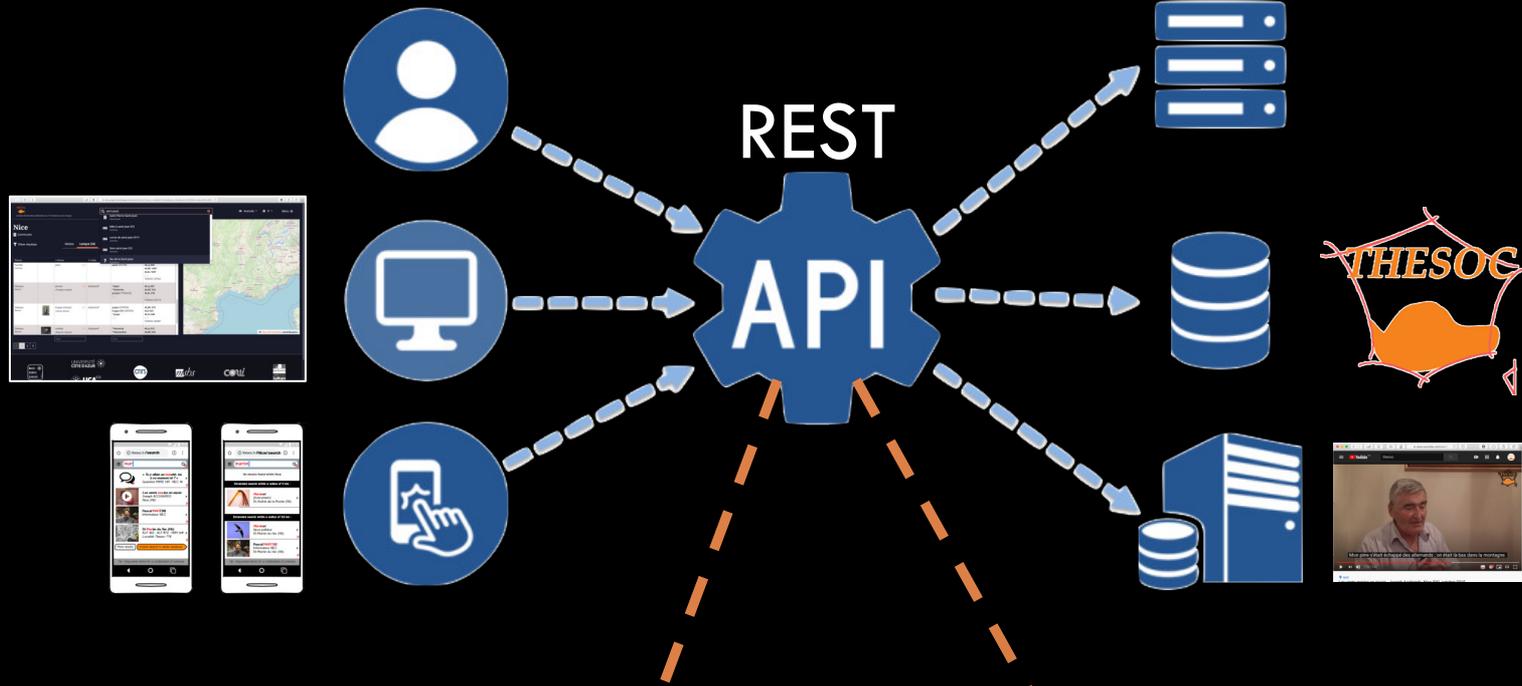
Interopérabilité  
API REST



Chaîne YouTube  
ethnotextes sous-titrés



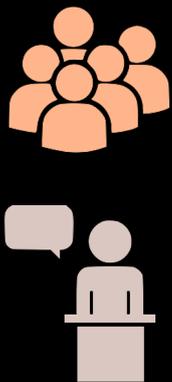
Archivage pérenne  
+ Corpus de la parole



Lo Congrès  
permanent de la lenga occitana

dicod'Òc  
vèrb'Òc

Edisyn  
European Dialect Syntax  
Search Engine





## Formats de fichier standards

Enregistrements en MPEG H264 + .WAV

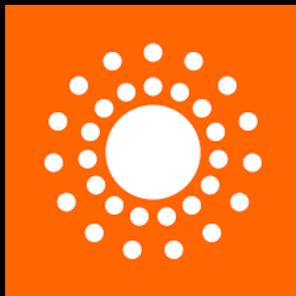
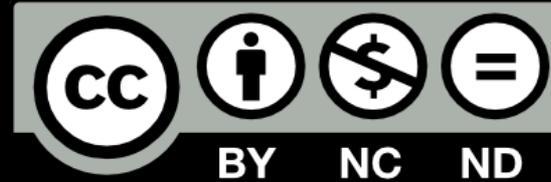
Ressources en fichiers XML / TEI P5



## Licences ouvertes

Vidéos en CC-BY-NC-ND

Tout le reste en **LGPL-LR**



## Métadonnées

Dublin Core & OLAC

# multiples applications possibles

- **Traitement Automatique du Langage**  
(traduction automatique, reconnaissance vocale, etc.)
- **Linguistique** (dialectologues, phonologues, syntacticiens ;  
mais aussi communication non verbale, etc.)
- **Autres disciplines scientifiques** (historiens, sociologues, ...)
- **Apprentissage des langues régionales**
- *Big-data / Deep Learning*

# Quelques conseils

- Ne pas filtrer les données à priori  
(difficile de prévoir toutes les exploitations possibles)
- Témoins s'exprimant dans leur dialecte vernaculaire  
(+ naturel et spontané ; patrimoine linguistique intéressant)
- Qualité des enregistrements et des données récoltées  
Manipulation du matériel, déroulement de l'enquête
- Penser à récolter les données annexes / métadonnées  
Photos, croquis, relevés GPS, etc.
- Autorisations des témoins  
(droit à l'image, vie privée, anonymat, etc.)