

# LE THESAURUS OCCITAN: UNE BASE DE DONNÉES MULTIMEDIA CONSACRÉE AUX DIALECTES OCCITANS

Guyline Brun-Trigaud

Laboratoire BCL, CNRS UMR 6039  
Université Nice Sophia-Antipolis, MSH de Nice (France)

## Abstract

*The Thesaurus Occitan (abbreviated THESOC) is a multimedia database, which contains, among other things: linguistic and linguistic-related data from field works: maps and survey notebooks from the Atlas linguistiques, monographies, audio records, pictures; linguistic data coming from former analyses: lemmatisation, morphology, etymology, micro-toponymy; bibliographical references; tools for linguistic analyses: maps generator, instruments for diachronic analyses, comparative cartography procedures, morphological analysis instruments; a Morpho-Syntax Module (MMS), detailed by Pierre-Aurélien Georges.*

*Centralised in Nice (France) within the laboratory UMR 6039 «Bases, Corpus, Langage» (attached to the CNRS), this inter-university program associates different teams, upon the direction of Pr. Jean-Philippe Dalbera.*

*One can say the THESOC is a variable geometry database which considers all kinds of exploitations thanks to specific menus, which integrates all kinds of documents. The advantage of such a database lies also in the fact that it can evolve and be updated permanently to satisfy users' needs.*

**Key words:** *Database, mapping, multimedia, dialectology, occitan languages.*



Fig. 1

Logo Thesoc

Le THESAURUS OCCITAN (ou THESOC en abrégé) est une base de données informatique destinée à l'étude des dialectes occitans.

Elle est développée depuis 1992 et centralisée à Nice dans le cadre de l'UMR 6039 du CNRS «Bases, Corpus, Langage», sous la direction de Jean-Philippe Dalbera, il s'agit d'un programme inter-universitaire qui associe différentes équipes.

Le THESOC contient notamment:

- des données linguistiques et péri-linguistiques issues d'enquêtes de terrain: cartes et carnets d'enquêtes des Atlas linguistiques,<sup>1</sup> monographies, enregistrements sonores, documents iconographiques;
- des données linguistiques procédant d'analyses déjà réalisées: lemmatisations, morphologie, étymologie, microtoponymie;
- des données bibliographiques;
- des outils d'analyse: représentations cartographiques, instruments d'analyse diachronique, procédures de cartographie comparative, instruments d'analyse morphologique.

Il s'agit d'un objet à géométrie variable qui envisage différents types d'exploitations grâce à des menus spécifiques, qui intègrent toutes sortes de documents, de telle sorte que le THESOC se présente comme un outil offrant à la fois, mais toujours séparément, des données linguistiques quasi brutes, des données ayant fait l'objet d'analyses et de traitements et des outils d'investigation. L'intérêt d'un tel outil réside également dans le fait qu'il est en permanence amené à évoluer selon les besoins des utilisateurs.

Pour pouvoir figurer dans la base, les données linguistiques brutes doivent satisfaire aux deux critères suivants: d'une part, elles doivent être de nature orale et sont donc saisies dans la base avec leur transcription phonétique en API, pour assurer le partage des données avec les autres chercheurs. De plus, comme le THESOC offre la possibilité de faire entendre les sons enregistrés au cours des enquêtes —lorsque nous les avons—, cela permet à l'utilisateur de contrôler la transcription proposée. D'autre part, les données doivent aussi être précisément localisées, ce qui constitue évidemment une condition essentielle pour l'étude de la variation diatopique.

Actuellement la base comprend plus d'un million de fiches-réponses, plus de mille extraits sonores et plus de 500 documents visuels (photos, vidéos, et dessins) accompagnent et illustrent les données linguistiques, ce qui fait du THESOC un véritable recueil multimédia, qui peut s'inscrire tout autant comme outil de recherche pour les linguistes que comme outil pédagogique pour le grand public.

Une partie des données est consultable en ligne sur Internet à l'adresse <http://thesaurus.unice.fr>

## 1. Les Glossaires

La partie lexicale constitue le «cœur historique» du THESOC, elle peut être consultée à partir d'une interface intitulée «Tableau de bord» (voir fig. 2) qui comporte différents modules:

<sup>1</sup> *Atlas Linguistiques de la France par régions*, Editions du CNRS.

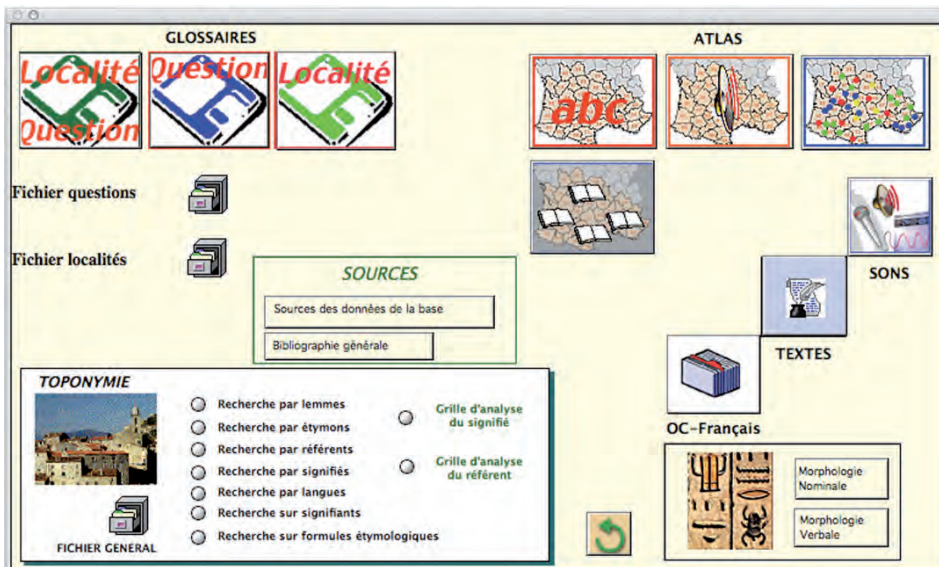


Fig. 2

Tableau de bord

### 1.1. Les fichiers localités et questions

À l'intérieur de la base de données, chaque forme est identifiée par le couple LOCALITÉ-QUESTION, dont voici les deux fichiers principaux: d'une part, le fichier des localités qui contient 831 entrées recouvrant tout le domaine occitan.

La consultation d'une fiche localité (voir fig. 3) permet d'avoir accès à la liste des enquêteurs et des informateurs associés aux différentes enquêtes qui se sont succédées dans cette localité.

D'autre part, le fichier des questions ou «responsaire» (cf. Oliiviéri 2004) qui est le résultat de la somme des cartes et listes publiées par les différents atlas linguistiques régionaux du domaine occitan, ainsi que les éléments relevés dans des monographies ou des résultats d'enquêtes non publiées.

Il comporte 8.338 questions (voir fig. 4), qui sont regroupées suivant les principaux thèmes traités dans les atlas linguistiques régionaux, comme l'élevage, la nature, l'espace, le temps, l'habitat et la vie quotidienne, etc.


LOCALITE		274
nom	PEZENAS	
indications géographiques	34_HERAULT	
sources	Atlas Rég. ALLOr 34.32	ALF <input type="text"/>
date de l'enquête	1965 / 1979	Autres <input type="text"/>
Informations	<input type="text"/>	
Enquêteurs	BOISGONTIER J. (enquête 1980) (ALLOr) MICHEL L. (enquête 1965) (ALLOr) PETIT J.M. (enquête 1979) (ALLOr)	
Informateurs	ALCOBER Robert [66 ans] (enquête 1979) (ALLOr) LARUE Marcel [60 ans] (enquête 1979) (ALLOr) Mme X (enquête 1965) (ALLOr) SEVERAC Louis [76 ans] (enquêtes 1979 et 1981)	
		

Fig. 3

Exemple de fiche-localité: Pezenas (Hérault)

Liste des questions		8338 fiches			
N°	Intitulé	Scient.	Entrée d'index	Thème	Sous-thème
1	abandonner le nid		abandonner	non spécifique	
3	abasourdi		abasourdi	HOMME	Caractère, sentiments, jugement
4	abat-foin		abat-foin	CULTURES	Prés, foins, plantes textiles
5	abattre un arbre		abattre	NATURE	Forêt
6	abcès		abcès	HOMME	Maladies, affections
7	abeille		abeille	ELEVAGE	Ruches
8	abeille sauvage		abeille	NATURE	Animaux sauvages, insectes
9	abîmer		abîmer	non spécifique	
10	ablette		ablette	NATURE	Animaux sauvages, insectes
11	aboyer		aboyer	ELEVAGE	Chiens, chats

Fig. 4

Liste des fiches-questions

La consultation d'une fiche question (voir fig. 5) permet d'avoir accès à la liste des cartes publiées et des carnets d'enquêtes, concernant cette question. Le cas échéant, d'autres sources éventuelles peuvent également y être consignées.


Question n° 1094 bergeronnette	
dénomination scientifique <i>Motacilla alba</i>	entrée d'index bergeronnette
thème NATURE	sous-thème Oiseaux
SOURCES CARTES PUBLIÉES	
ALF C	1460
ALAL C	414
ALCe C	556
ALG C	28,1208
ALJA C	11L.94
ALLOc C	276
ALLOr C	356
ALLy C	514
ALMC C	314
ALO C	419
ALP C	967
ALEPO C	
SOURCES CARNETS ENQUETES	
ALF	Q
ALAL	Q
ALCe	Q
ALG	Q
ALJA	Q
ALLOc	Q
ALLOr	Q
ALLy	Q
ALMC	Q
ALO	Q
ALP	Q
ALEPO	Q
PAM	Q
SOURCES MONOGRAPHIES PUBLIÉES	
	
SOURCES AUTRES ENQUETES NON PUBLIÉES	

Fig. 5

Exemple de fiche-question: bergeronnette

Les entrées lexicales de la base sont organisées de la manière suivante: chaque entrée lexicale, ou fiche-réponse, est associée à une fiche-question et à une localité donnée.

## 1.2. Les différents modes de consultation

Il existe différentes possibilités d'interrogation de la base pour consulter les entrées lexicales: on peut soit rechercher toutes les fiches-réponses associées à une fiche-question précise, de manière onomasiologique, pour étudier la variation diatopique (voir fig. 6).

Chaque fiche question peut-être associée à une ou plusieurs illustrations (voir fig. 7), ce qui peut s'avérer très utile dans certains cas: par exemple, pour distinguer les différentes variétés, lorsqu'un terme quelconque peut correspondre à plusieurs objets.

FICHES RÉPONSES : 697		INTITULE : bergeronnette			
localité	commentaire	forme phonique	lemme	graphie phonologisante	étymon
SAINT-PIERRE		berdʒejr'etɔ	bergereta <sup>oo</sup>		VÉRVÉCARIUS
BANON		pastr'esə	pastressa <sup>o</sup>		PASTOR
MEZEL		buvejr'ete	bovaireta <sup>oo</sup>		BŌS
METHAMIS		bardʒejr'etɔ	bergeroneta <sup>oo</sup>		VÉRVÉCARIUS
ANNOT		byjer'etə	bovaireta <sup>oo</sup>		BŌS
FORCALQUIER		rus'etə	rosseta		RŪSSUS
GORDES		pastr'esɔ	pastressa <sup>o</sup>		PASTOR
MENTON		balər'ina	balarina <sup>oo</sup>		BALLĀRE
BOULBON		berʒeron'etɔ	bergeroneta <sup>oo</sup>		VÉRVÉCARIUS
CAVAILLON		berdʒir'etɔ	bergereta <sup>oo</sup>		VÉRVÉCARIUS
MAILLANE		berdʒer'etɔ	bergereta <sup>oo</sup>		VÉRVÉCARIUS
GREOUX		g'ɪŋɔkw'ɔ	guinha-coa <sup>oo</sup>		WINGJAN*
NICE		balər'ina n'egra	balarina <sup>oo</sup>		BALLĀRE
EYGALIERE		pastr'esɔ	pastressa <sup>o</sup>		PASTOR
CADENET		bardʒ'elə	bergièra <sup>o</sup>		VÉRVÉCARIUS
VILLELAURE		g'ɪŋ'yzɔ	guinhusa <sup>oo</sup>		WINGJAN*
EYGUIERES		g'ɪŋɔ kw'a	guinha-coa <sup>oo</sup>		WINGJAN*
SAINT-PAUL-LES-DUR		g'ɪŋɔkw'ɔ	guinha-coa <sup>oo</sup>		WINGJAN*

TRIER PAR :

Fig. 6

Exemple de liste de réponses par question: bergeronnette

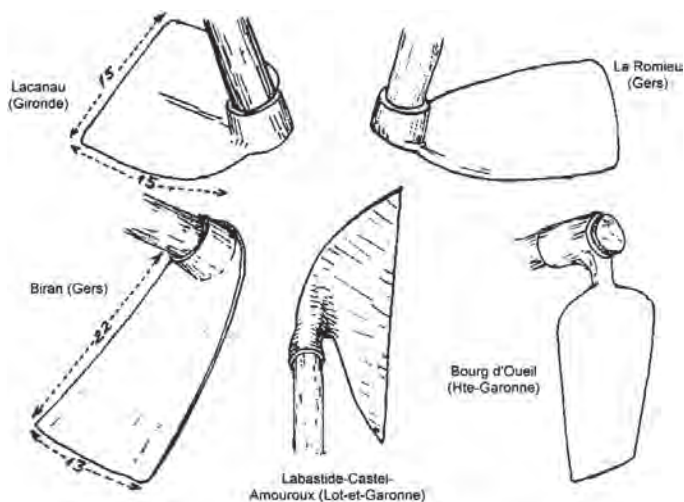


Fig. 7

Exemple d'illustrations: houes

On peut également consulter toutes les fiches-réponses associées à une localité donnée, pour établir une monographie (voir fig. 8).

FICHES RÉPONSES : 2156		PEZENAS		
intitulé	forme phonique	lemme	graphie phonologisante	étymon
chevron	k'uple	coble		CŌPŪLA
latte	tʃaz'eno	jasena		JĀCĪNA*
plancher (s.)	plantʃ'ε	plancher (fr)		PHALANX
Pierre de construction (var)	paβ'at	pavat		PAVĪRE
carreaux du sol de la cuisine	paβ'at	pavat		PAVĪRE
chasse-roue	bœtɔR'odo	butaròda		BUTR
porte découpée dans une p	purtij'u	portilhon <sup>90</sup>		PŌRTA
portail	purt'al	portal		PŌRTA
seuil	suj'et	soihet		SŌLEA
ouvrir	drub'i	dobrir		APERĪRE
ouvrir	durb'i	dobrir		APERĪRE
ouvrir en grand	aland'at	alandar		LANDA*
peinture	palastr'atʃɔ	palastracha		PALA
gond	guf'u	gafon		GŌMPHUS
verrou	bar'ul	verroh		VĒRICŪLUM
déverrouiller	dezbaruj'a	desverrolhar		VĒRICŪLUM
verrouiller	baruj'a	verrolhar		VĒRICŪLUM
fermer	bar'a	barrar		BARRA*
porte	p'ɔrtɔ	pòrta		PŌRTA
fermer à clef	klaβ'a	clavar		CLAVIS
clef	kla <sup>w</sup>	clau		CLAVIS

TRIER PAR :

Fig. 8

Exemple de liste de réponse par localité: Pezenas (Hérault)

On peut enfin rechercher spécifiquement un certain couple localité-question. Par exemple, quel est le terme employé à Pezenas pour désigner “la bergeronnette” (voir fig. 9).

Le détail d'une fiche-réponse fait apparaître un certain nombre de champs:

- le numéro de question dans la base et son intitulé (1.094/bergeronnette), ainsi que son numéro de saisie (n.° 3.387);
- le numéro de localité dans la base et son nom (274/Pezenas) ainsi que son numéro dans l'Atlas Linguistique du Languedoc Oriental (ALLOr 34.32);
- la transcription phonétique en API (forme phonique);
- la transcription graphique: il s'agit d'une transcription «phonologisante», c'est-à-dire une sorte de forme intermédiaire entre la transcription phonétique et la graphie standardisante d'Alibert, cette forme intermédiaire adopte pour l'essentiel les principes du *Tresor dóu Felibrige* de F. Mistral. Elle est automatiquement générée par un algorithme, appelé *transcripteur*. Celui-ci est basé sur un ensemble de règles de réécriture qui peuvent être configurées par l'utilisateur et qui peuvent varier d'une localité à l'autre pour prendre en compte les systèmes phonologiques des différents dialectes occitans;

question	1094	bergeronnette	n° 3 387
localité	274	PEZENAS	ALLOr 34.32
forme phonique	pastur'elo		source(s) ATLAS
graphie phonologisante	pastourèlo		
lemme	pastorèla		
base morphologique	pastor + ela		
catégorie grammaticale	Substantif Féminin singulier		
	<input type="button" value="Voir Tableau"/>		<input type="button" value="Quitter"/>
étymon	PASTOR		REW 6279
formule étymologique	PASTOR + ELLA		FEW 7, 758b
Commentaire	<input type="text"/>		

Fig. 9

Exemple de fiche localité-réponse: Pezenas/bergeronnette

- le lemme, qui est conçu comme forme de référence (conventionnelle); il sous-tend tout le faisceau de variantes consignées dans la base. Le choix du lemme est effectué en s'appuyant sur le *Dictionnaire occitan-français* d'Alibert; sa notation respecte donc les principes de la graphie alibertine;
- la base morphologique indique si le mot est un composé ou non ((préfixe) + base + (suffixe));
- la catégorie grammaticale;
- l'étymologie, avec des références aux principaux dictionnaires étymologiques (*Französisches Etymologisches Wörterbuch* (FEW) et *Romanisches Etymologisches Wörterbuch* (REW));
- enfin, en cas de besoin, des commentaires peuvent être adjoints.

Dans le cas des fiches contenant un verbe, en cliquant sur le bouton «Voir Tableau», on peut également consulter le paradigme de conjugaison verbale, lorsque celui-ci a été renseigné dans la base (voir fig. 10).

Lorsque le paradigme morphologique d'un adjectif ou d'un article possède des variations phonétiques contextuelles, un tableau supplémentaire permet de visualiser les différentes formes attestées, suivant le contexte phonétique qui suit et/ou qui précède ce terme, le tout étant illustré d'une série d'exemples (voir fig. 11).



Infinitif f'ajre		Classe III c
Participe passé f'atj		NICE
Participe présent f'eʒ		faire
<b>Indicatif présent</b>	<b>Subjonctif présent</b>	<b>Futur</b>
1 f'ow	f'agi	far'aj
2 f'as	f'ages	far'as
3 f'a	f'age	far'a
4 f'eʒ	fag'eʒ	far'eʒ
5 f'es	fag'es	far'es
6 f'agʒ	f'agu	far'agʒ
<b>Indicatif imparfait</b>	<b>Subjonctif Imparfait</b>	<b>Conditionnel</b>
1 fa'iji	fag'esi	far'iji
2 fa'ijes	fag'eses	far'ijes
3 fa'ija	fag'ese	far'ija
4 fajav'agʒ	fagesj'agʒ	farj'agʒ
5 fajav'as	fagesj'as	farj'as
6 fa'iju	fag'esu	far'iju
<b>Impératif</b>	<b>Passé simple</b>	OK
	1 fag'eri	
2 f'aj	2 fag'eres	
	3 fag'e	
4	4 fagerj'agʒ	
5	5 fagerj'as	
	6 fag'eru	

Fig. 10

Exemple de paradigme verbal: Nice (Alpes-Maritimes)/faire

The interface displays a grid of phonetic variants for the article 'le' in Castillon (Alpes-Maritimes). The grid is organized by gender (masc. and fém.) and number (sg. and pl.).

	/-cs	/-voy	/-cs	/-voy
sg.	u	ɹ	a	ɹ
pl.	e	ez	e	ez

Buttons: Voir champ de variation complet, Ok

**Exemples**

01. m.sg/-cs	u kun'i	le lapin
02. m.sg/-voy	lar'ajɹe	l'araire
03. m.pl/-cs	e vez'i's	les voisins
04. m.pl/-voy	ez am'iks	les amis
05. fém.sg/-cs	a f'ea	la brebis
06. fém.sg/-voy	l'arba	l'aube

A red arrow points from the 'Voir champ de variation complet' button to a larger grid:

	/-cs	/-voy	/cs-cs	/cs-voy	/-cs	/-voy	/cs-cs	/cs-voy
sg.	u	ɹ	lu	l	a	ɹ	la	l
pl.	e	ez	le	lez	e	ez	le	lez

Buttons: Ok

**Exemples**

09. m.sg/cs-cs	e <sup>o</sup> lu lj'etɹ	dans le lit
10. m.sg/cs-voy	e <sup>o</sup> lar'ja	en l'air
12. m.pl/cs-voy	e <sup>o</sup> lez t'æs	dans les yeux
13. f.sg/cs-cs	e <sup>o</sup> la k'neva	dans la cave

Fig. 11

Exemple de variations contextuelles: Castillon (Alpes-Maritimes)/le (article)

## 2. Cartographie interactive

La consultation de la base ne permet pas seulement d'obtenir des glossaires, elle permet également de cartographier des faits lexicaux, de différentes manières.

Deux types de cartes sont disponibles dans le THESOC: d'une part, des cartes présentant les faits bruts, et d'autre part, des cartes de synthèse. Aucune carte n'est cependant stockée dans la base de données: elles sont toutes générées dynamiquement à partir des données linguistiques présentes dans la base, en fonction des requêtes demandées par l'utilisateur.

Chaque fois que l'utilisateur modifie sa requête, une nouvelle carte est générée en temps réel. C'est en ce sens que l'on peut dire qu'il s'agit d'une cartographie interactive.

### 2.1. Les cartes de faits bruts

Comme il ne serait pas lisible d'afficher sur une carte de l'Occitanie toute entière l'ensemble des transcriptions phonétiques attestées dans les différentes localités, les cartes présentant les faits bruts sont disponibles à deux échelles, avec un système de

zoom: en premier lieu, au niveau de l'Occitanie toute entière, un simple point rouge signale les localités pour lesquelles la base contient une réponse à la question qui est cartographiée (voir fig. 12).

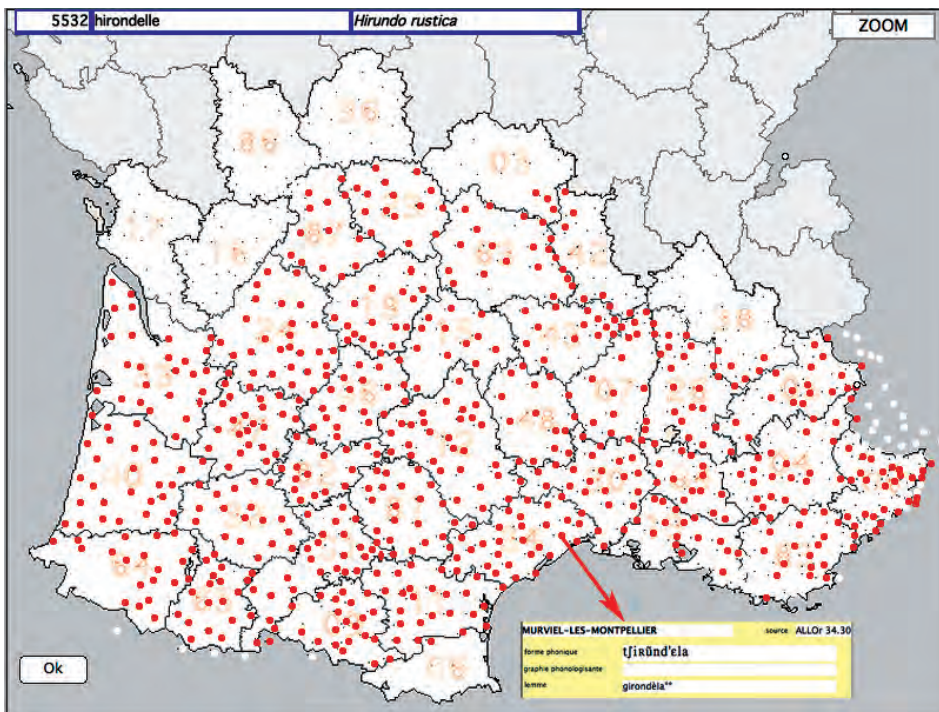


Fig. 12

Carte à symboles: hirondelle

Un clic sur l'un des points permet de visualiser sa ou ses transcriptions phonétiques dans une petite fenêtre.

Puis dans un deuxième temps, il est possible de zoomer à l'échelle d'un département (voir fig. 13).

La carte détaillée affiche la transcription phonétique associée à chaque réponse, à côté du point de la localité concernée, comme dans les atlas linguistiques. Un point rouge indique que l'on peut écouter l'enregistrement sonore associé.

Un module spécifique du Tableau de Bord intitulé «Sons» permet d'écouter les enregistrements regroupés par localités.



Fig. 13

Carte départementale des réponses en phonétique: hirondelle/Gers

## 2.2. Les cartes de synthèse

Voyons à présent à partir d'un exemple, le processus d'élaboration d'une carte de synthèse, concernant la répartition géographique des différents types lexicaux désignant le verbe "scier".

Le logiciel affiche à l'écran la liste des lemmes répertoriés dans les différentes fiches réponses de la base concernant ce terme (voir fig. 14). On peut alors effectuer des groupes contenant un ou plusieurs de ces lemmes, selon nos souhaits, et affecter une couleur à chacun de ces groupes. Un simple appui sur le bouton «symboliser» déclenche la construction et l'affichage quasi-instantanés de la carte ainsi demandée (voir fig. 15).

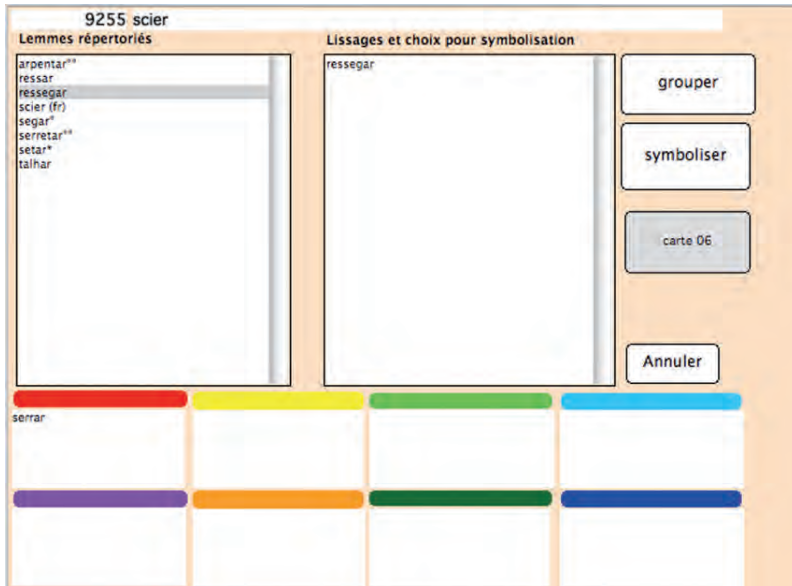


Fig. 14

Tableau de répartition des lemmes pour “scier”

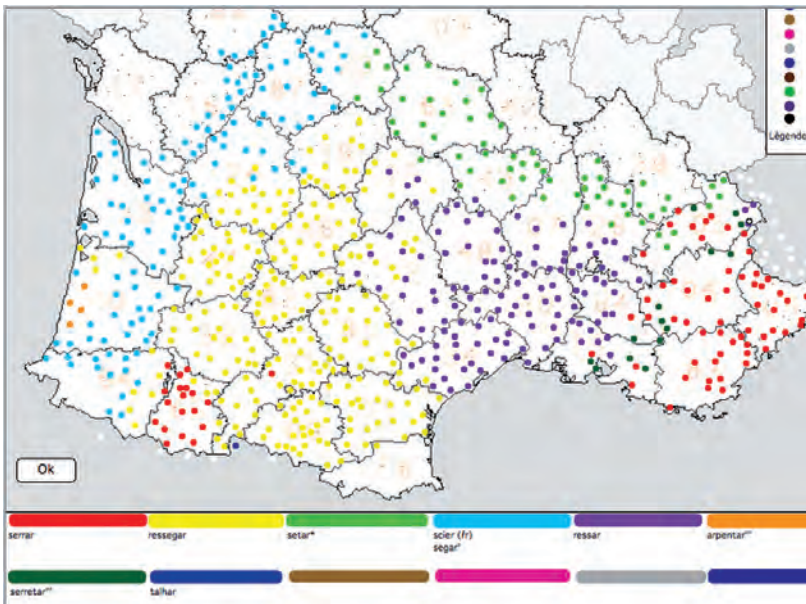


Fig. 15

Carte de synthèse: carte à symboles (“scier”)

Sur cette carte aréale à symboles, on voit très bien se dessiner les différentes zones lexicales très distinctes formées par les termes désignant le verbe «scier» dans l'ensemble des parlers occitans.

L'utilisateur peut ainsi modifier les critères et les regroupements à volonté pour générer autant de cartes qu'il le souhaite. Signalons enfin que ces cartes peuvent être exportées au format vectoriel pour être intégrées dans un document PDF ou pour servir de document de travail.

### 3. Autres ressources

Le THESOC ne se limite pas aux données purement lexicales, mais il permet également d'autres recherches, à divers niveaux.

Nous avons déjà évoqué le module «Sons», mais il faut signaler en outre le module «Morphologie verbale et nominale» qui donne la possibilité de voir les différents paradigmes verbaux et nominaux, comme nous l'avons vu plus haut, le module «Cartes des Atlas» qui permet de visualiser les domaines des différents atlas linguistiques dont les données figurent dans la base, ainsi que le module «Sources» qui donne les références des ouvrages concernés par les travaux du THESOC.

Nous avons également un module spécifiquement dédié à la syntaxe avec de nombreuses fonctions qui seront détaillées par Pierre-Aurélien Georges ici-même.

#### 3.1. Le module Oc-Français ou Dictionnaire inversé

Nous nous attarderons davantage sur le module Oc-Français ou Dictionnaire inversé.

À partir des données lexicales présentes dans la base, le THESOC propose également un dictionnaire inversé occitan / français. Ce module permet une démarche tour à tour onomasiologique et sémasiologique, sur la base des lemmes (voir fig. 16). Il permet ainsi de trouver les différentes notions correspondant à un terme occitan donné. Ce type de requête est particulièrement utile pour des recherches en sémantique lexicale et reconstruction étymologique (cf. Dalbera 2006).

Par exemple, si on choisit comme point de départ, *barbo-*, dans le cadre de droite, s'affichent tous les lemmes commençant par cette séquence. On entre alors dans le tableau principal du module en choisissant dans la liste un lemme occitan donné, par exemple *barbòta* (voir fig. 17).

Dans cette nouvelle fenêtre, s'affichent tous les sens relevés dans la base pour ce mot: lorsqu'ils n'ont pas de signe, c'est que le sens existe dans le dictionnaire d'Alibert qui est notre dictionnaire de référence, et lorsqu'ils sont accompagné d'un signe, c'est que le sens n'avait pas été relevé. On remarque, ici, que les enquêtes des Atlas linguistiques régionaux donnent des indications supplémentaires.

En cliquant sur l'une des lignes, toutes les localités qui utilisent ce terme pour désigner le référent correspondant apparaissent dans la fenêtre inférieure (le mot *barbòta* est utilisé à Peyrat-de-Bellac (Haute-Vienne) pour désigner «la couleuvre»).

Tandis qu'en sélectionnant un signifié (le mot français, par exemple «couleuvre»), on accède à la liste de tous les autres termes dialectaux (les signifiants) renvoyant à la même notion (*bòba*, *cinglant<sup>oo</sup>*, *cingla<sup>o</sup>*, etc.).

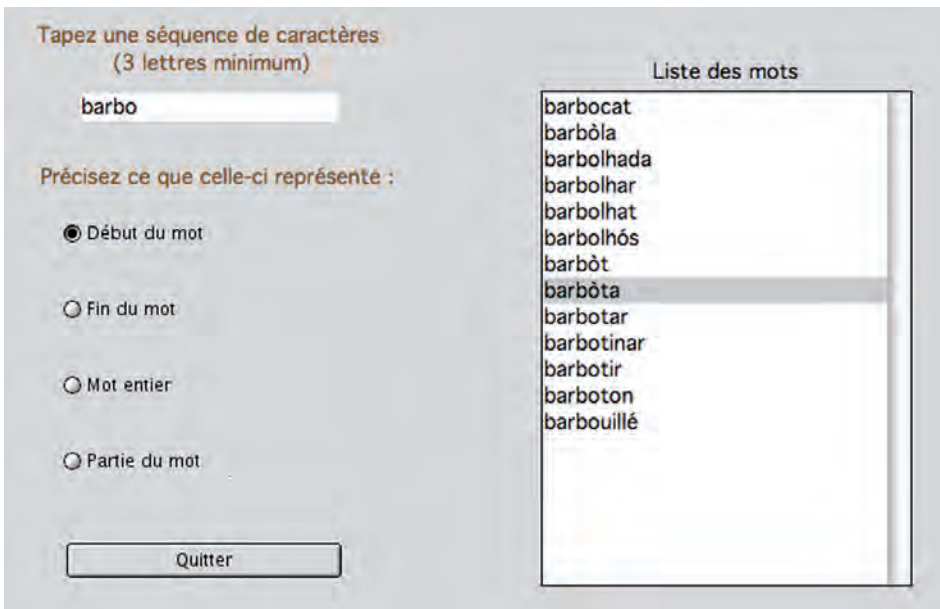


Fig. 16

Tableau de choix du dictionnaire inversé (*barbo-*)

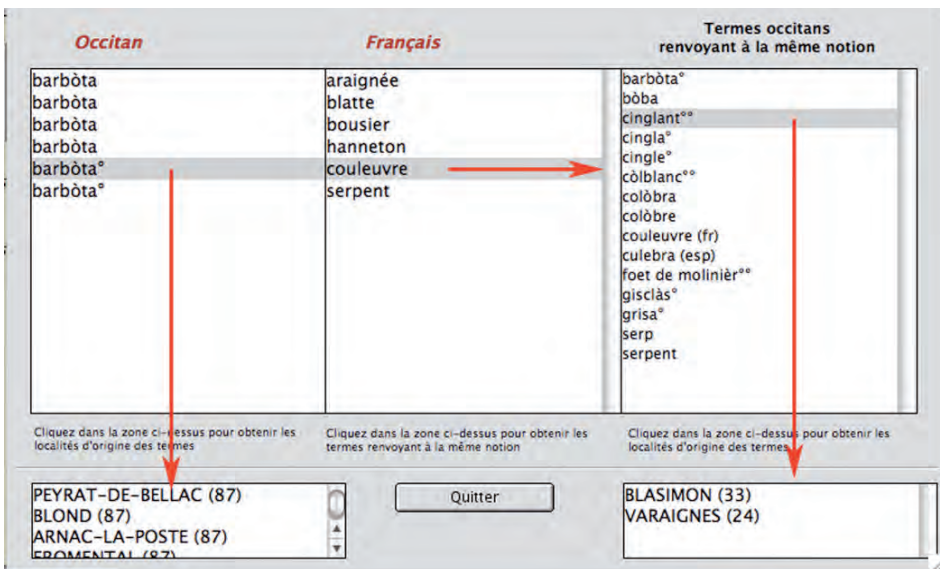


Fig. 17

Tableau de consultation du dictionnaire inversé (*barbòta*)

Enfin, chacune des formes est localisée. Un clic sur l'un des termes dialectaux fait apparaître la liste des localités dans lesquelles ce terme est attesté (Le mot *cinglant*<sup>oo</sup> est utilisé à Blasimon (Gironde) et Varaignes (Dordogne) pour désigner la «couleuvre»).

### 3.2. Recherches étymologiques

Parmi les autres fonctionnalités de recherche de la base que nous n'avons pas encore abordées, il faut noter qu'il est possible d'effectuer une recherche par étymons:

Saisir le début de l'étymon recherché en utilisant le clavier MAJUSCULES :

PASTA  
PASTĪCIUS\*  
PASTINACA  
PASTOR  
PASTŌRIA  
PASTŪRA

Liste des questions 3 fiches

N°	Intitulé	Scient.	Entrée d'index	Thème	Sous-thème
1091	berger		berger	ORGANISATION, PRAT	Groupes humains, relations
1094	bergeronnette	◊	bergeronnette	NATURE	Oiseaux
10462	petit valet		valet	ORGANISATION, PRAT	Groupes humains, relations

Lemmes correspondants

pastor  
pastora  
pastorèla  
pastoreleta<sup>oo</sup>  
pastre  
pastressa<sup>o</sup>  
pastressona<sup>oo</sup>

Réponses correspondantes

forme phonique	question	localité
pastur'ela	1094	SAINT-FIRMIN (05)
pasturel'eta	1094	LARCHE (04)
pasturel'eta	1094	BARCELONNETTE (04)
pastr'esœ	1094	BANON (04)
pastr'esœ	1094	GORDES (84)
pastr'esœ	1094	EYGALIERE (13)
postur'elo	1094	VEYREAU (12)
postur'elo	1094	SAINT-GERMAIN (12)
postur'elo	1094	NANT (12)
postur'elo	1094	CAMPRIEU (30)
postur'elo	1094	AVEZE (30)
pastur'elo	1094	SAINT-AFFRIQUE (12)
pastur'elo	1094	SAINT-GERMAIN (12)

Fig. 18

Tableau de consultation des données étymologiques (PASTOR)



Si l'on saisit par exemple l'étymon latin *PASTOR*, apparaît alors à l'écran la liste des questions en rapport avec cet étymon (voir fig. 18), c'est-à-dire toutes les questions pour lesquelles la base contient au moins une fiche réponse qui possède l'étymon demandé (dans l'exemple les questions «berger» (shepherd), «bergeronnette» (wagtail) et «petit valet» (little farm hand), puis au-dessous la liste des lemmes associés (*pastor*, *pastora*, *pastorela*, etc.), ainsi que dans la fenêtre de droite, la liste des réponses correspondantes pour chaque lemme dans les localités concernées.

### 3.3. Le volet de microtoponymie

Enfin le THESOC comporte aussi un volet de toponymie qui consigne les différents micro-toponymes recueillis lors des enquêtes.

Localité :	121 NICE	n°INSEE	Cadre institutionnel	Université
Forme phonique	Variante	Autre dénomination		
bɔwm'ɛta (li)				
F. graphique	baumeta (li)	Prononciation française	bom'ɛtə (le)	
Lemme	balmetas (las)	Graphie officielle	Baumettes (les)	
Formes provenant de sources écrites	lieu de conservation ou de dépôt (types de sources)	identification précise de la source et date		formes
Référent	Indications complémentaires		quartier de Nice	
Catégorie	quartier			
Signifié	signifié pour l'informateur		"les petites grottes"	
Catégorie	oronyme			
Signifiant	Déterminant	+	Suffixe	-'et-a
			Composé	Syntagme Synthème
Etymologie	formule étymologique	BALM(A)+ÏTTAS (ILLAS)		réf. biblio complémentaire Azaretti (1986)
	étymon BALMA	REW 912	langue ligure	
Discussion	Sur BALMA et ALMA voir, outre l'article d'Azaretti, l'essai de Bessat-Germi.		Commentaires	
	<a href="#">Retour à la liste</a>			

Fig. 19

Exemple de fiche toponymique (*Las Balmetas/Nice* (Alpes-Maritimes))

Une fiche toponyme (voir fig. 19) contient généralement les informations suivantes: la localité dans laquelle le micro-toponyme a été recueilli, les formes graphique et phonétique du toponyme, le signifié associé à ce terme dialectal, les formes graphique et phonétique de sa traduction en français, le type de toponyme (oronyme, hydronyme, etc.), le type de référent (quartier de la commune, chemin,

cours d'eau, forêt, etc.), l'étymon et les références étymologiques du type *REW*, des commentaires, et d'éventuelles références bibliographiques complémentaires.

Lorsque ce toponyme fait par ailleurs l'objet d'attestations écrites, il est également possible de consigner les formes provenant des différentes sources écrites, avec une identification précise de ces sources et de leur date.

À l'instar de la partie lexicale du THESOC, diverses fonctionnalités de recherche et d'analyse proposées par la base permettent ensuite de consulter ces données (voir fig. 20):

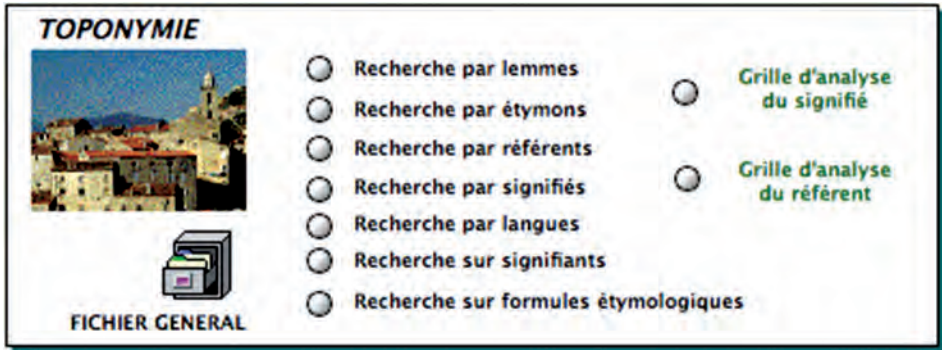


Fig. 20

Toponymie: les différents types de recherches

On retrouve là encore la recherche par localité, pour établir une monographie par exemple, la recherche par lemme, et la recherche par étymon; ainsi que d'autres types de recherche comme la recherche par type de toponyme, par type de référent, par langue d'origine de l'étymon, par type de construction morphologique (avec déterminant ou non, mot composé, syntagme) ou encore par formule étymologique.

En outre, le type de toponyme et le type de référent permettent de proposer deux grilles d'analyse: une grille d'analyse du signifié, et une grille d'analyse du référent.

Nous prévoyons actuellement d'ajouter dans la base la possibilité de joindre des illustrations à une ou plusieurs fiches toponyme.

Voici donc un aperçu de l'ensemble des fonctionnalités de la base de données multimédias du Thesaurus Occitan. D'autres possibilités devraient être prochainement développées, notamment au niveau de la cartographie et de la mise en ligne des données.

## References

- Alibert, L., 1966, *Dictionnaire occitan-français*, Institut d'Etudes Occitanes, Toulouse.  
 Dalbera, J.-P., 2006, *Des dialectes au langage: Une archéologie du sens*, Champion, Paris.  
 Meyer-Lübke, W., 1935, *Romanisches Etymologisches Wörterbuch*, Winter, Heidelberg.  
 Mistral, F., 1878, *Lou Tresor dou Felibrige*, Raphèle-lès-Arles, Paris.  
 Olivieri, M., 2004, «Le responsable du THESOC», *Actes du colloque 8ème Colloque de dialectologie et littérature du domaine d'oïl occidental, Avignon, 12-13 juin 2002*, P. Brasseur (éd.), pp. 23-34.  
 Wartburg (von), W., 1922-..., *Französisches Etymologisches Wörterbuch*, Schroeder, Bonn.