



## **SYMILA : Una basa de donadas sintaxicas per l'occitan e las autras lengas romanicas de França**

Patric Sauzet, Yaël Champclaux

### **► To cite this version:**

Patric Sauzet, Yaël Champclaux. SYMILA : Una basa de donadas sintaxicas per l'occitan e las autras lengas romanicas de França. Fidelitats e dissidências - XIIeme Congrès de l'Association Internationale d'études occitanes, Jul 2017, Albi, France. pp.271-284. <hal-03174041>

**HAL Id: hal-03174041**

**<https://hal.science/hal-03174041v1>**

Submitted on 22 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

**ACTES** DU **XII<sup>e/n</sup>** **CONGRÈS**  
de l'Associacion Internacionala d'Estudis Occitans  
de l'Association Internationale d'Études Occitanes  
ALBI, 10-15/07 2017

édités par Jean-François Courouau  
en collaboration avec David Fabié  
editats per Joan-Francés Courouau  
en collaboracion amb Dàvid Fabié

**Fid** **elitats**  
**e<sup>t</sup>** **élités**  
**dissi** **déncias**  
**dences**

Section française de l'Association internationale d'études occitanes

**SFAIEO**

Vol. 1



## SYMILA : Una basa de donadas sintaxicas per l'occitan e las autras lengas romanicas de França

L'*Atlas linguistique de la France* (Gillieron / Edmont 1902-1910) gausís pas de la melhora reputacion en cò dels occitanistas que fan sovent sias las criticas qu'Antoine Thomas, en tot que reconegüesse l'ample "gigantal" de l'òbra, formulava en 1904 : manca d'autenticitat lexicala e d'exactitud fonetica mai que mai (Thomas 1904). Sul fons de la sosevaluacion globala de l'ALF pels lingüistas en França, los occitanistas an pas tirat de l'ALF tot çò qu'i aurían pogut trobar, comprés lo lexic. Juli Ronjat l'utilizèt intensivament per redigir la *Grammaire (h)istorique* (Ronjat 1930-1941), mas plan de síntesis o d'introduccions lingüisticas a l'occitan desconisson l'ALF qu'es tanben pasmens un atlàs panoccitan (o gaireben) e lo sol, coma es un atlàs de tot lo domeni primari del francés<sup>1</sup>.

L'ALF permet de sasir d'un còp tot lo domeni occitan quand los atlasses regionals son trocejats e an pas totjorn de questionaris parallèls. Tanben l'ALF, mens ric e mens especific en lexic que sos successors regionals, es probablament mai ric en sintaxi que qualques uns, per sa part publicada almens. Guylaine Brun-Trigaud mostrèt que 25 % del material de l'ALF es format de frases que se pòdon reconstituïr e estudiar (Brun-Trigaud 1990). L'ALF es per tant un document excepcional per l'estudi de la sintaxi occitana (coma, d'autre costat, per la dels parlars d'oïl o francoprovençals e pus mendrament d'unes parlars catalans). Totun la publicacion de l'ALF en mapas lèumens lexicalas rescind aquela dimension e la rend de mal explechar (levat quand la frasa figura tota sus una mapa o un bon tròç). Per de simplas rasons tipograficas, la publicacion papièr podià dificilament restituir un material sintaxic localizat.

Lo projècte SYMILA *Syntactic microvariation in the Romance languages of France* se dona per tòca de bastir una basa documentària amb lo material sintaxic de l'ALF, occitan coma francés, francoprovençal o catalan. Las resultas de la primièira fasa finançada del projècte son consultablas sul sit <http://symila.univ-tlse2.fr/>. La basa d'informacion del projècte SYMILA (per sa partida ALF<sup>2</sup>) es bastida d'una part per aculhir lo material sintaxic de l'ALF, d'autra part per explicitar la natura d'aquel material en i apondent d'informacions gramaticals que ne melhoren l'intelligibilitat e ne permetan una consultacion mai elaborada.

Lo projècte representa una escomessa documentària que demandèt de bastir una basa de donadas de tres dimensions : l'espaci geografic (639 punts d'enquèsta), lo sistèma de cada parlar (180 frases reconstituïdas) e la successiò dels mots de cada frasa (fins a gaireben un vintenat de mots per las frases mai longas). Nos acontentarem aici d'evocar l'estructura de la basa, e en amont lo questionari ALF e la notacion que calguèt gerir, e mai en aval lo ròtle de las notacions

---

<sup>1</sup> L'ALF, tornem-o dire es plan mal nomenat. Cobrís pas tota França, solament las regions romanofònas. Cobrís tanben altra causa que França (la Belgica e la Soïssa romanicas, sens comptar las Illas normandas e qualques punts d'Itàlia). Es pas exactament tanpauc "galloromanic" que, sens parlar del volum còrs, incorpora de punts catalans que segon çò que se fa d'aquela "lenga pont" siá son benlèu pas galloromanics, mas que s'o son, manca un grand tròç del galloromanic a l'atlàs, que lo catalan se redusís pas a Rosselhon. En fach l'ALF es l'atlàs de l'espaci romanic ont lo francés es (o èra al moment de l'enquèsta) la lenga d'educacion, çò que permetiá l'enquèsta per traduccion del francés qu'es al principi de l'ALF.

<sup>2</sup> La valorizaciò de las donadas ALF es una part del projècte Symila (SYMILA-ALF), l'autra part (SYMILA-NES, per "novèlas enquèstas sintaxicas") que se'n tròba pas qu'un escapolon en linha, se fonda sus d'enquèstas novèlas dedicadas a la sintaxi. La plaça manca per las evocar aici.

ortograficas dins l'economia de la basa, que son una illustracion de l'utilitat de la codificacion grafica e lingüistica de la lenga per menar a bon pro son estudi scientific quitament dins una perspectiva purament descriptiva.

### 1. Prealables: questionari e notacion fonetica

Lo questionari de l'ALF es pas estat conservat. La publicacion en mapas alfabeticas e l'utilizacion principalament lexicologica o lexicografica que se faguèt de l'Atlàs (en començar pels trabalhs del quite Gilliéron) an rescoduda la preséncia d'una importanta documentacion sintaxica dins aquela òbra.<sup>3</sup> L'importància ja mençonada de las frasas dins l'elicitacion de l'ALF se pòt establir (Guylaine Brun-Trigaud (1990) o faguèt) per la sintèsi dels títols de las mapas, ont los contèxtes d'elicitacion quand n'i a un son donats. Pro de comprene que cada còp tot lo contèxt es estat notat e cartografiat. Permet de restablir lo questionari e las frasas qu'i cabián. L'utilizacion dels quasèrns d'enquèsta permetèt d'anar mai luènh e tornar trobar completament lo questionari dins l'òrdre que foguèt somés als informators. Los quasèrns d'enquèsta de l'ALF son depausats a la BNF<sup>4</sup>. Compòrtan pas que las responsas, dispausadas totun d'un biais rigorosament constant. En utilizant de quasèrns occitans (que son los mai complets, perque son los de las darrièiras enquèstas ont lo questionari èra estat enriquesit) se pòt inferir la question, l'estimul indicat ça que la pels títols de mapas. Lo questionari reconstituït es disponible sul sit SYMILA ([http://symila.univ-tlse2.fr/alf/questionnaire\\_alf\\_reconstitue](http://symila.univ-tlse2.fr/alf/questionnaire_alf_reconstitue)) e tanben dins la rejoncha de questionaris facha pel projècte TUL-QUEST (<http://tulquest.huma-num.fr/>). Donam en annèxa a la fin d'aquesta presentacion una pagina d'un quasèrn reproducha, e las responsas correspondentas a un punt d'enquèsta (Annèxas 2-4).

L'ALF utiliza una notacion fonetica definida per Pierre Rousselot e utilizada per el dins la *Revue des Patois Galloromans* (RPG) (Rousselot 1887). Per tant, sul sit del projècte SYMILA designam la notacion de l'ALF “notacion RPG”. Aquesta notacion fonetica es nascuda a la meteissa epòca que la notacion definida per Paul Passy<sup>5</sup> e que devenguèt l'*Alfabet fonetic internacional* (AFI, API, IPA) (Passy 1893). Dins las meteissas epòcas se definiguèron d'autres sistèmas de notacion, notadament lo d'Eduard Böhmer (1875), d'ont deriva puèi lo sistèma dich “romanista” mes al punt e difusat per Edouard Bourciez.

Lo sistèma RPG es estat continuat dins los atlases regionals fins a una epòca tota recenta. L'Atlàs de Provença tot escars publicat es lo primièr que rompa amb la tradicion aprèp que la basa THESOC a convertida la notacion dels atlases regionals (Bouvier *et alii* 2016). L'emplec d'una notacion fonetica diferenta contribuís a aprigondir lo valat entre los “dialectològs” e los autres lingüistas. Lo valat es perjudiciable pel domeni francés qu'aluènh l'estudi de l'estandard del dels dialèctes primaris (e fa de l'estandard una lenga fòra-sòl). E per l'occitan, falta d'estandard largament establir, marginaliza tot l'estudi de la lenga vojat dins una “dialectologia” qu'es pas un aspècte de la lingüistica, mas una mena d'autra lingüistica que la lingüistica. Utilizar l'AFI es una condicion per rejónher la comunautat dels lingüistas e rendre las donadas de nòstre domeni accessibles al pus grand nombre de cercaires. Avèm doncas transpausada la notacion RPG en notacion IPA. Pretendèm pas que la correspondéncia siá exacta dins lo sens que se poiriá definir un equivalent acotic absolut comun a las doas notacions. Avèm volgut assegurar d'un costat la mai granda similitud d'intencion de notacion (barradura o apertura vocalica, anterioritat o

<sup>3</sup> L'interès sintaxic de l'ALF es demorat globalament inexpleitat. Senhalem doas excepcions interessantas: l'estudi de Gaston Tuaille sus l'interrogacion a partir d'una mapa-frasa (Tuaille 1975), lo libre de David Heap sul subjècte nul que pren en compte la massa de las donadas mas repausa sus una tipizacion de las donadas (Heap 2000).

<sup>4</sup> Jos la còta : NAF 11971-12030. Lo projècte SYMILA permetèt de finançar la numerizacion d'una part dels quasèrns d'enquèsta que la BNF a puèi mesés en linha. Seria desirable que l'ensem dels quasèrns foguèsson numerizats.

<sup>5</sup> Paul Passy collaborèt ça que la a la *Revue des patois galloromans* (Passy 1891).

posterioritat, vosament o desvosament consonantic...), d'autre costat una transposicion bi-univòca, combinason (signe e diacritics) per combinason, que permeta la reversibilitat. Apondam que la preséncia de las mapas-fonts dins la basa SYMILA permet a cada moment de verificar las transposicions.

Coma per donar rason a Michel de Certeau que plaça dins la vocala l'esséncia de la variacion patesa coma la vesian Grégoire e sos correspondents (Certeau *et alii* 1975, 110 sqq), son las vocalas qu'an demandat lo mai d'elaboracion per transpausar. Avèm comptadas dins las notacions de l'ALF 296 notacions vocalicas (letras amb diacritics) diferentas. Los detalhs de la transposicion, principis e aplicacion, son donats dins lo document en linha sul sit SYMILA ([http://symila.univ-tlse2.fr/alf/notation\\_phonetique](http://symila.univ-tlse2.fr/alf/notation_phonetique)). Evocam pas aici que quelques punts importants.

Dins la transposicion del sistèma vocalic a calgut transpausar las notacions de l'accentuacion, de la longor e de la qualitat vocalica.

Çò de mai simple es estat l'accentuacion<sup>6</sup>: remplaçam lo turet soscrich de l'RPG per un turet auçat abans la vocala 'kasaire'<sup>7</sup> > [kas'aire]. En çò fasent, seguissèm pas la convencion estandard recomandada per l'Associacion fonetica internacionala qu'es de notar l'accentuacion a la debuta de la sillaba, abans l'ataca (dins l'exemple donat : [ka'saire]). Nòstra causida a una rason practica: plaçar la notacion de l'accent tonic dins una posicion atenenta a la vocala accentuada permet la reversibilitat automatica amb la notacion RPG que seriá de plan pus mal definir amb una notacion de l'accent a l'iniciala de sillaba. Un problèma particular se pausariá en cas d'elision davant una vocala tonica notada: coma separam los mots (compres los clitics e las formas embegudas) se notariá l'accent dins un mot diferent de l'original [l'ome] per l'òmè. Se pòt tanben avançar una defensa de principi d'aquesta notacion. L'accent tonic es una mesa en relèu d'una sillaba entre las autras (çò que justifica la convencion de l'AFI) mas en cada sillaba avent pas qu'un cimèl sillabic, es equivalent de notar la prominéncia a l'iniciala o al cimèl de la sillaba. La darrièra causida a l'avantatge de demandar pas de determinar la còpa sillabica per notar l'accent tonic (e la còpa sillabica pòt èsser problematica, aital dins un parlar d'oïl entre [etrãg.r'ɛ] o [etrã.gr'ɛ] per una realizacion d'"étranglerait" o en occitan entre [ka.ler.jo] o [ka.le.rjo] per "caleria" (una forma gascona per *calriá*). Es comòde de poder notar l'accent tonic sens trencar aperabans de questions possiblement espinosas de sillabacion.

Per l'apertura, avèm seguida una logica diferenta segon las classes d'apertura. Per las vocalas nautas, transpausam los accents aguts e grèus de l'RPG en diacritics AFI : 'i se transpausa en [i], 'ɨ se transpausa en [ɨ], e 'ɨ en [ɨ] (e identicament per las 'u simpla o gansada de l'RPG que venon [y] e [u]). Per contra, per las vocalas mejanas, transpausam 'é, 'ó per [e], [o] mas 'è, 'ò per [ɛ], [ɔ] çò que mena a causir una notacion [ɛ] per 'e sens diacritic d'apertura de l'RPG. D'un biais rescontram aici una validacion de nòstra causida. Nos avisam que 'i o 'u sens diacritic que conservam sens diacritic son frequentas dins l'ALF, mentre que 'é, 'è, 'ó, o 'ò, rendudas per de signes simples [e], [ɛ] ... son mai frequentas que 'e o 'o sens diacritic. Ne resulta un aleugeriment de la notacion que se pòt prene coma un signe d'adequacion. Meteissa confirmacion d'adequacion per l'estalvi de diacritics dins la transposicion de 'à en [a] e de 'á en [a], mas de 'a en [a] : 'à e 'á son mai frequentas dins l'ALF que 'a<sup>8</sup>.

La RPG nòta regularament de brèvas (per un breve ˘) e de longas (per un macron ˉ). Transpausam las brèvas de l'ALF per una vocala sens diacritic: ˘ es [i] e las longas pel diacritic

<sup>6</sup> L'accentuacion es la dimension mai criticada e mai criticabla de la notacion fonetica de l'ALF. Ni per aquò la transpausam sens cap de modificacion. Las solas modificacions portadas a la notacion fonetica d'origina concernisson la segmentacion en mots segon las exigéncias de l'establiment de glòsas. A cada còp la modificacion es mençonada dins una anotacion de la frasa.

<sup>7</sup> Plaçam las notacions RPG entre '...' e reservam los cròcs [...] a l'AFI.

<sup>8</sup> Dins lo cas d'a, transpausam (sus la basa dels exemples de l'RPG e de las formas notadas dins l'ALF) 'á definit erronèament coma mai barrat que 'a o 'à coma una vocala posteriora ([ɑ]).

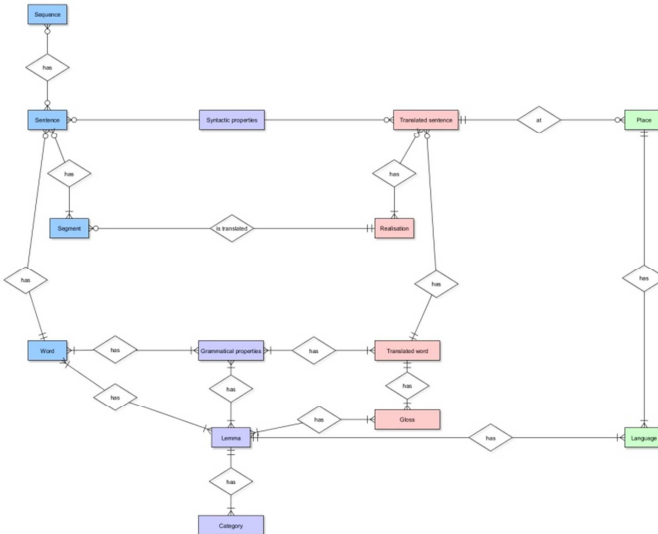
pòstpausat de l'AFI: 'ɪ es [i:]. La vocala sens diacritic de longor de l'RPG es notada pel signe de mièja longor en AFI: [iː] per 'i'. Dins la practica, lo breu, transpausat per l'absència de diacritic es fòrça frequent, lo macron rar e l'absència de diacritic excepcionala. Notam pas de longor sus las vocalas nasalizadas perqu'avèm remarcat que, en tot que siá possible en principi, l'ALF combina pas jamai los dos signes. Es aquí un exemple que demòstra que l'operacion de transposicion se podia pas definir totalament *a priori* mas se deviá establir en foncion de la practica efectiva d'Edmont. Segon los principis de la notacion RPG 'e' sens diacritic es non marcat, mas dins la practica notacionala 'é' o 'è' amb un breu ([e] o [ɛ] de l'AFI) son las valors non marcadadas. E per tal, un còp de mai, la leugieiretat de la transcripcion obtenguda ne valida la pertinéncia.

Las consonantas pausan plan mens de problèmas, primièr per tal que son plan mens nombrosas. Los signes de basa son largament los meteisses. La c crosada ('e') correspond a [ʃ], simpla variacion de simbòl. I a pas de signe simple per las fricativas apicalas non estridentas: 's' e 'z' respectivament valon [θ] e [ð]... Una dificultat qu'avèm resolguda per observacion dels cases reals son las combinasons de simbòls que son per principi dobèrtas dins la notacion RPG ont tot signe se pòt combinar amb tot signe. Las avèm trobadas mai que mai per notar las oclusivas palatalas : 'k' capelat de 't' s'interprèta [c], e per notar de consonantas mièjas vosadas o mièjas devosadas, aital 'd' capelat de 't' (dins la materialitat tipografica: seguit de 't' en expausant) que transpausam per [ɖ] ([d] amb un diacritic de desvosament).

Insistiscam sus un punt per acabar a prepaus de la transcripcion fonetica: assajam de prepausar los equivalents mai probables a las notacions d'Edmont signe per signe. Modificam pas las combinasons. Aital enregistram lo fach que l'RPG desconeis la nocion d'africada e nòta pas que de sequéncias 'ts', 'tc' ... que transpausam per [ts], [tʃ] e non pas per [ts], [tʃ]... Edmont nòta pas tanpauc la sillabicitat e conservam de notacions [au], [aʊ]... per de sequéncias que sabèm monosilabicas: [aʊ], [aʊ]...<sup>9</sup>

## 2. Estructura de la basa

L'estructura informatica de la basa de donadas concebuda per Yaël Champclaux especificament pel projecte SYMILA permet d'integrar las tres dimensions del material (espaciala, paradigmatica e sintagmatica) e navegar entre elas. Se pòt representar dins l'esquema seguent (donat sul sit del projecte).



<sup>9</sup> O dins una grafia fonologizanta [aw] ...

Tot çò en blau dins l'esquema ven del questionari de l'ALF: "sequências de frases" ("sequences") ont figuran las "frases" reconstituïdas de l'ALF ("sentences")<sup>10</sup>. Las frases, objècte central de la basa, son divisadas d'un costat en "segments" (mots o grops de mots cartografiats) e d'autre costat en "mots" ("words") que son a pauc prèp los mots ortografics, levat qualques seguidas de mots categorizadas coma d'entitats lexicalas (coma "parce que" o "peut-être").

Tot çò en ròse correspond al material de las respensas a l'enquèsta d'Edmont. Las traduccions de frases reconstituïdas son analisadas parallèlament en "segments" e en "mots". L'anàlisi en "segments" correspond a la publicacion cartografiada de l'ALF e totes las mapas d'ont es tirada l'informacion sintaxica son disponiblas dins lo sit (cada frasa dona los ligams a las mapas correspondentas çò que permet de verificar la sasida de las formas foneticas<sup>11</sup>). Tot çò ròse ven doncas de las respensas foneticament notadas per Edmond Edmont e cartografiadas per Jules Gilliéron. Levat çò qu'es sonat "glòsa" ("gloss") e qu'i tornarem.

Evidentament las respensas son associadas caduna al punt geografic ont son estadas reculhidas ("place"). Cada punt es estat assignat a una "lenga" ("language"), causida renduda necessària per la lematizacion mesa en plaça que ne parlam un pauc mai luènh. Los punts de l'ALF son estats identificats en longitud e latitud e projectats sus una mapa manejable (OpenStreetMap). De cada punt se pòt navegar cap a las formas qu'i son estadas reculhidas.

L'informacion en malve al centre de l'esquema es tota d'informacion aponduda o almens explicitada, a l'escasença de la bastison de la basa SYMILA. L'informacion es aponduda a partir del questionari: las frases del questionari son estadas sintaxicament caracterizadas (complexitat, modalitats...) e los mots que las compausan lematizats, categorizats e gramaticalament caracterizats (basicament : genre, nombre, animacion, cas quand ne vira, per las formas nominalas, classa morfològica, temps, aspècte, mòde, persona pels vèrbs, qualques indicacions tanben sus la reccion de las preposicions e l'estatut fonologic dels elements: tonicitat, cliticitat). Per default, l'informacion sintaxica globala associada a las frases-estimuls es reportada sus las frases-respensa. *A priori*, una interrogativa se revira per una interrogativa, una negativa per una negativa e una frasa complexa per una frasa complexa. I a qualques cases rars de non isomorfia (coordinacion per subordinacion o l'invers per exemple) e la basa permet la correccion. Al nivèl dels mots, l'informacion gramaticala pòt pas èsser sistematicament inferida de l'estimul. Totun, naturalament, al centre del domeni francés, arriba sovent que la respensa del testimoni es pas qu'una variacion fonetica minima de l'estimul. Per tirar avantatge de la proximitat estrecha de la lenga d'enquèsta amb bon nombre dels parlars documentats, los lèmas dels mots del questionari ALF servisson tanben a lematizar las respensas, e las glòsas de las respensas pòdon tanben èsser identicas a de mots ortografics del questionari.

Lo vector per associar los mots fonetics amb d'informacion gramaticala son los lèmas ('lemmas') e las glòsas ('glosses')

### 3. Lo sistèma de lèmas, de glòsas e de formas ortograficas

L'ALF es un Atlàs mut. I cabon gaireben pas cap d'informacions lingüisticas explicitas. Es lèumens perque lo legeire compren la forma francesa de l'estimul e l'identifica gramaticalament que las formas son analisablas pel legeire.

Semanticament, tot repausa sus la traduccion. Lo francés lenga normada e codificada deliura censat un sens univòc. Aqueste aspècte es estat pro criticat e s'es considerat que mai d'un mot francés èra possiblament ambigües e podiá induire de respensas falsas o pas pro precisas. Tota la reflexion de l'escòla "dels mots e de las causas" s'avodèt a afinar la coneissença del referent per superar las insuficiéncias de la traduccion. Lo novèl atlàs engenhât per Albèrt Dauzat foguèt

<sup>10</sup> Dins lo questionari, las frases venon sovent per sequéncias que se seguisson coma un pichòt tèxt. Avèm conservada aquesta ligason.

<sup>11</sup> E de senhalar d'errors se ne vira a l'autor d'aquestas linhas.

despartit en Atlases regionals per tal d'èsser fondat sus una coneissença exacta de las *realia* localas que permeta l'exactitud lexicala. Per una basa sintaxica es pas un punt essencial.

Gramaticalament l'ALF senhala qualques escarts entre la lenga de l'estimul e las traduccions, pel genre dels noms en particular quand se dedusís pas ni de l'estimul ni de la forma de la responsa. D'autres escarts son pas senhalats, dins las formas verbalas o pronominalas notadament. Lèmas e glòsas permeton d'aportar una informacion gramaticala completa e independenta de l'estimul.

Los lèmas son l'objècte mai conegut per gerir a l'encòp la variacion morfològica (intèrna a la lenga e a cada dialècte) e la variacion dialectala. Es ja çò qu'es mes en plaça dins la basa THESOC que recampa una bona part de las donadas lexicalas dels Atlases regionals occitans (Dalbera *et alii* 2005). La lematizacion (dins una lenga flexionala) es una operacion necessària a tota entrepresa de lexicografia. Es doblament necessària dins una lenga ont la variacion dialectala es integrada a la lexicografia. Lo THESOC a enregistradas (per ara) las donadas lexicalas dels atlases regionals, jos la forma de mots isolats. La lematizacion retenguda met en relacion las formas foneticas (en notacion AFI) e de lèmas que neutralizan a l'encòp la flexion e la variacion dialectala. Los lèmas son preses dins lo diccionari occitan-francés de Loïs Alibèrt. Son doncas en grafia classica (segon la codificacion de Loïs Alibèrt, ara comunament utilizada) e la forma es la forma centrala de l'occitan, una forma lengadociana.

Dins lo THESOC, “galina” lematiza aitanplan una forma [gal'ino] ... que [gar'j'o] o [dʒal'ino]. Las formas del THESOC, de noms mai que mai, donan pauc d'escasença d'illustrar los efèches gramaticals de la lematizacion. Totun qualques formas verbalas son conjugadas (“luisisson” ...) e lo lèma es, coma dins un diccionari, l'infinitiu. Lo THESOC prepausa d'autres camps que la forma fonetica observada e lo lèma que l'interprèta. Un etim es tanben donat e mai una “forma ortografica”. L'etim esperlonga naturalament lo lèma, que tipicament un lèma remanda a un etim unic e presenta una relacion univòca a l'etim (evolucion fonetica o manlèu). La “forma ortografica” del THESOC es prevista (e pas encara mesa en plaça) coma una forma ortografica de tipe mistralian, çò es amb de convencion de notacion mai fonetizantas e de tipe mai francés que la grafia classica, mas pas ortodòxament mistraliana dins la mesura ont deu èsser obtenguda per conversion automatica de la notacion fonetica<sup>12</sup>.

L'informacion gramaticala dins lo THESOC es ligada a cada forma fonetica, que pòt pas èsser estacada al lèma (per definicion o pòt pas èsser tota), ni a la “forma ortografica” qu'es pas concebuda coma una forma autonòma, ni a l'estimul (integrat dins la basa THESOC coma informacion semantica en tant que l'estimul es un mot francés, censat semanticament fixat, e qu'es completat per de precisions discriminantas se ne vira). Non solament las caracteristicas gramaticals de la responsa pòdon diferir de l'estimul a causa de las proprietats de la lenga o del dialècte de traduccion, mas tanben l'elicitacion es facha tot còp dins una forma flechida autra que la que correspond al lèma (aital dins mai d'un punt gascon l'intrada semantica “poule” es renduda per [gar'ios] lematizat “galina”. L'informacion sul nombre es donada dins un camp “morfològia”. Es tanben precisat dins lo meteiss camp que lo mot es un nom femenin, informacion que poiriá èsser portada e aportada pel lèma.

Dins la basa SYMILA, la caracterizacion gramaticala de las formas, qu'es una question marginala per la basa THESOC, es centrala. Coma lo material retengut son de frasas, i cabon regularament de formas verbalas flechidas que demandan una caracterizacion explicita, d'ont mai importanta que pòt coïncidir pas amb las caracteristicas de l'estimul. Dins la basa SYMILA las proprietats gramaticals de las formas son associadas a las formas foneticas per l'intermediari dels lèmas e de çò que sonam “glòsas”. Los lèmas son la meteissa causa que çò que son dins THESOC. Son de formas neutralizadas per la flexion e lo dialècte. Coma dins la basa THESOC los lèmas son restacats a un camp etimologic e un etim es donat (e mai siá pas una tòca centrala de

---

<sup>12</sup> Ai personalament de dobtes sus l'interès d'aquela notacion.



l'entrepresa SYMILA, mas puslèu una pèira d'espera per la mesa en relacion amb d'autres sistèmas s'informacion). Los lèmas especifics a caduna de las lengas que los parlars locals enregistrats per l'ALF s'i restacan, çò es l'occitan (que nomeni en primièr dins aqueste collòqui), lo francés, lo francoprovençal, lo catalan e lo ligurian (per un punt: Fontan ALF 990)<sup>13</sup>. Pel domeni d'oïl se son recuperadas coma basa las formas de l'estimul, de las question de l'ALF<sup>14</sup>. Se son tanben aponduts de lèmas franceses quand la responsa d'un parlar d'oïl usa d'un mot diferent de la question e del francés estandard se ne vira. Aital un lèma "geline" es apondut al lèma "poule" eissit del questionari ALF e lo francés a un lèma "geline" e un lèma "poule" coma l'occitan a un lèma "pola" e un lèma "galina". Los lèmas del francoprovençal son los que pausan mai de dificultat, essent lo pauc d'estandardizacion de la lenga. Seguissèm las solucions graficas del diccionari de Dominique Stich (Stich *et alii* 2003). Lo trabalh de lematizacion per las donadas francoprovençalas es lo mens avançat amb lo ligurian ont es a pena escapolat. Lo catalan pausa pas de problèma particular levada la necessitat de lematizar las formas especificament rosselhoneras.

L'operacion de lematizacion supausa la referéncia a una lenga codificada graficament e mai a una forma pivot, a una forma prototipica, a un estandard (en prenent lo mot dins lo sens de modèl) quina sián las foncions e los usatges de la dicha forma dins la vida culturala e sociala de la lenga. D'aquel punt de vista, las situacions son mai que mai variadas dins l'espaci cobèrt per l'ALF, del francés que sa forma estandard es tament establida que per pauc lo ligam es romput amb l'ensem dialectal que n'es eissit, al francoprovençal ont la quita possibilitat de l'establiment de formas-pivot es discutida e son utilitat (quitament scientifica) denegada per una part dels especialistas. Al mièg, lo catalan es en via avançada de normalizacion, d'installacion sociala e la disponibilitat scientifica de l'estandard es totalament evident. Per l'occitan, coneissèm la situacion. La grafia classica es largament mas pas totalament acceptada. L'estandardizacion, la mesa en circulacion d'un estandard es luènh d'èsser acceptada mas la natura e las formas de l'estandard son relativament claras. L'emplec a la Mistral o "a la Ronjat" del provençal literari (rojanenc elaborat) coma forma estandard panoccitana es pas gaire mai defenduda ni practicada uèi (mas lo provençal literari demòra una forma d'expression productiva e la nòrma classica deu fòrça a la nòrma mistraliana). L'idèa d'un estandard general de l'occitan e, n'i aguèsse un, de l'ample de son emplec, es largament debatuda (e quitament dels estandards regionals) e pro discutida, encara qu'una estandardizacion de fach se faga sovent darrièr la reivindicacion de variacion, es un tèma d'estudi sociolingüistic que s'ameritarà d'aprigondir). Pasmens es admès pro generalament que, se i deu aver un estandard occitan, correspond a una forma de lengadocian, e doncas la forma d'occitan ja codificada per Loïs Alibèrt. Per un prepaus scientific e de lematizacion, per un diccionari general de l'occitan coma per las basas de donadas del tipe THESOC o SYMILA, la causida alibertina elaborada permet d'establir una nomenclatura lingüistica coerenta, a usatge scientific (e mai mai, segon l'utilitat o l'escasença, mas es pas lo debat d'aquesta presentacion).

Per explicitar brèument aquelas causidas, digam que son lengadocianas doncas sud occitanas e conservativas (J.P. Chambon parla a prepaus de lengadocian de son "ultraconservatisme", Chambon 2009, 793). Doncas las evolucions especificament gasconas o nòrd occitanas son pas enregistradas dins los lèmas, coma ja vist amb lo simple exemple del lèma "galina" (vs. "garia" e "jalina"). Tanpauc son pas retengudas las evolucions recentas coma la vocalizcion de l'-l finala simpla en latin dins tota la periferia ocitana (lo lèma es "ostal" e pas "ostau"). Notem que las causidas de la grafia occitana fan una larga part de la besonha e dispensan de causir. La grafia classica "vin" o "pan" es panoccitana e cobrís dirèctament un grand nombre de realizacions (amb o sens betacisme, amb o sens casuda d'-n finala ...). D'unas evolucions especificament lengadocianas

<sup>13</sup> Las donadas còrsas de l'ALF, publicadas a despart son pas integradas dins la basa SYMILA.

<sup>14</sup> Per aqueles lèmas recuperats automaticament, lo camp etimologic es pas estat emplit, lo francés es pro documentat etimologicament per altra part.

(coma “fial”, “abrial” per “fil”, “abril”) son pas tanpauc retengudas. Entre las solucions lengadocianas, ja Pèire Bèc parlava de la causida necessària entre la resolució –ch- e –it- de latin –kt- (*fach* o *fait*). Coma Bèc retenèm la forma en –ch-. Per la vocalizacion de l’-l interna, se reten la solucion tolosana tipica, vocalizacion davant las coronalas (levat s) e pas davant las autras consonantas: *sautar*, *fauç*, *alba*, *qualque*, *fals*.

Los lèmas pòdon sufir dins una basa de la tòca essencialment lexicala. Dins una basa sintaxica, ont los objèctes manipulats son centralament de frasas, i a clarament una manca que cal comolar entre las formas foneticas e los lèmas. La manca es la d’una forma normalament legibla e de l’informacion gramaticala variabla e per tant absenta del lèma. Tanben, manca als lèmas la tencha locala, que per principi la neutralizan. Las “glòsas” qu’apelam venon comolar lo valat. Las glòsas son de formas ortograficas que tenon compte (segon los principis de la grafia alibertina) de la variacion dialectala de l’occitan e que tanben, en tant que formas flechidas, pòdon portar l’informacion gramaticala de cada forma en context. Las glòsas consèrvan la dialectalitat de basa mas asseguran una melhora lisibilitat comuna que la notacion fonetica. Per melhorar encara la lisibilitat de las donadas las glòsas son dobladas per l’afichatge de formas ortograficas que son provesidas de majusculas e de puntuacion, e mai presentan los embevements tipograficament acostumats. Las formas ortograficas permeton de limitar la lista de glòsas religadas als lèmas e que permeton de seguir cada lèma fins a caduna de sas realizacions flechidas e dialectalizadas dins la basa.

Aplicam al francés los principis de la grafia occitana. D’un costat notam d’especificitat dialectalas ancianas e estructurantas coma lo “cacheur” picard per “chasseur” o “jhournal”, “jhourneau” en santongés per “journal”. Mas d’un autre latz una grafia coma “moi” pòt conservar en sincronica la variacion d’interpretacion qu’a coneguda diacronicment ([moj], [mqœ], [mwe], [mwe], [mwa], [ma]).

En mai de sa foncion de vector d’informacion gramaticala, l’introduccion de las formas ortograficas utilizadas coma de glòsas permet de seguir la practica acostumada de la sintaxi de trabalhar sus de notacions ortograficas dins las lengas qu’an una grafia d’usatge. Dins las lengas qu’an pas d’ortografia fixada, la notacion dels linguistas de frasas o de tèxtes seguits es lèumens l’escasença de metre al punt una notacion practica qu’es potencialment una grafia d’usatge, una ortografia.

Es doncas natural de trabalhar en sintaxi occitana, lenga provesida d’una ortografia (e de qualques autras) sus la basa de notacions ortograficas. Es çò que fa Ronjat dins sa tèsi sus la sintaxi e dins la *Grammaire (h)istorique*. Sul pache, quand ilustra de fenomèns generals, formula los exemples dins l’occitan estandard qu’assumís, çò es en “provençal literari” forma elaborada de rodanenc (Voulard 2004). Lafont procedís parièr dins *La Phrase occitane*. Cita dins sas grafias d’origina (tot còp escalabrosas) los autors que cita, mas en grafia occitana d’usatge sos exemples fargats, e se i a pas de rason de los assignar a un dialècte particular, los formula en occitan estandard.

La consultacion de la basa SYMILA se pòt far segon lo luòc: siá per una pagina dedicada que compòrta una mapa zoomejable e una lista dels luòcs, siá tornar dins la pagina d’accès “per frasa” que permet tanben l’accès per luòc, e mai, i anam tornar un accès crosat. La consultacion se pòt far “per frasas”: totas las frasas, una frasa, qualques frasas causidas. Se pòt far per frasa e per luòc: totas las frasas d’un luòc o de mai d’un luòc, qualques frasas causidas d’unes luòcs causits etc. Se pòdon afichar fins a cent frasas aital mesas en parallèl sus una sola pagina. O se pòt afichar totas las frasas d’un punt çò que permet de sasir sas regularitats (tipicament sus l’emplec dels pronomes subjèctes per exemple).

Se pòt encara (sus tota la basa sasida o per una causida de luòcs) cercar las frasas que presentan tala o tala propietat sintaxica globala (complexitat, ellipsi, interrogacion...) o de propietats particularas sus un mot particular (conjoncion enonciativa, pronom indefinit, negacion febla etc.).

#### 4. Conclusion

La basa SYMILA ja dins son estat de desenvolopament actual permet de manipular una part significativa del material sintaxic contengut dins l'ALF. La notacion fonetica adoptada rend lo material mai legible e partejable dins la comunautat scientifica. L'apondeson de glòsas e de lèmas ortografics a permès l'explicitacion e sovent l'interpretacion grammaticala del material. Aquela explicitacion es possibla solament per la manipulacion d'una forma grafica d'usatge, regularizada e estandardizada de la lenga. Sens aquela ressorsa (o un equivalent *ad hoc* que n'a pas la pertinéncia sociala) la manipulacion de las donadas es esquartairada entre la notacion fonetica e los etimons latins o protoromanics supausats. L'informacion portada pel sistèma lèmas-glòsas es ja disponibla per interrogar la basa. Dins una forma plenament desenvolopada la basa SYMILA completada aculhirà tot lo material ALF, serà provesida d'una interfàcia de consultacion mai plegadissa<sup>15</sup>, permetrà una interrogacion grammaticala mai complèxa e articulada amb lo lexic e la lematizacion. Manca pas que lo finançament.

#### Réferéncias bibliograficas

- BÖHMER, Eduard (1875). « De sonis grammaticis accuratius distinguendis et notandis », *Romanische Studien* 1, 295-302.
- BOUVIER, Jean-Claude / MARTEL, Claude / BRUN-TRIGAUD, Guylaine (2016). *La langue d'oc telle qu'on la parle. Atlas linguistique de la Provence*, Forcalquier, Alpes de Lumières.
- BRUN-TRIGAUD, Guylaine (1990). *Le concept et le mot. Contribution à l'histoire de la dialectologie française au XIX<sup>e</sup> siècle*, Lyon, Université Lyon III, Centre d'études linguistiques Jacques Goudet, Série Dialectologie.
- CERTEAU, Michel de / JULIA, Dominique / REVEL, Jacques (1975). *Une politique de la langue. La Révolution française et les patois. L'enquête de Gergoire*, Paris, Gallimard.
- CHAMBON Jean-Pierre / GREUB, Yan (2009). « L'émergence du protogascon et la place du gascon dans la Romania », in Guy Latry (ed.), *La voix occitane. Actes du VIII<sup>e</sup> congrès de l'Association Internationale d'Études occitanes, Bordeaux, 12-17 octobre 2005*, Bordeaux, PUB, 787-794.
- DALBERA, Philippe / DALBERA-STEFANAGGI, Marie-José / STRAZZABOSCO, Dominique (2005). *THESOC, base de données linguistiques* <https://www.qwant.com/?client=brz-moz&q=thesoc>.
- GILLIÉRON, Jules / EDMONT, Edmond (1902-1920). *Atlas linguistique de la France*, Paris, H. Champion, 35 fasc. de cartes (ALF).
- HEAP, David (2000). *La variation grammaticale en géolinguistique. Les pronoms sujets en roman central*, München, Lincom, 194 p.
- PASSY, Paul (1891). « Patois de Sainte-Jamme (Seine et Oise) », *Revue des patois galloromans* 4, 7-16.
- PASSY, Paul (1893). « L'alphabet phonétique international », *Le maître phonétique*, 2-4.
- RONJAT, Jules (1930-32-37-41). *Grammaire istorique des parlers provençaux modernes*, Montpellier, Société des Langues Romanes, 4 vol.
- ROUSSELOT, Pierre (1887). « Introduction à l'étude des patois », *Revue des patois gallo-romans* 1, 1-22.
- SAUZET, Patrick / DAGNAC, Anne / SPORTICHE, Dominique / CHAMPCLAUX, Yaël (2015). SYMILA *Syntactic variation in the Romance languages of France*, Système d'information en ligne [<http://symila.univ-tlse2.fr>].

---

<sup>15</sup> En particular ara s'aficha pas tota l'informacion grammaticala aponduda a las formas. Mas l'afichatge es ja relativament cargat. Seria util de poder gerir e reglar aquel afichatge segon lo besonh de l'utilizaire de la basa.

- STICH, Dominique / GOUVERT, Xavier / FAVRE, Alain (2003). *Dictionnaire des mots de base du francoprovençal. Orthographe ORB supradialectale standardisée*, Thonon-les-Bains, Le Carré.
- THOMAS, Antoine (1904). « L'Atlas linguistique de la France », *Le Journal des savants* 2, 89-96.
- TUAILLON, Gaston (1975). Analyse syntaxique d'une carte linguistique : ALF 25 "Où vas-tu?", *Revue de linguistique romane* 39, 79-96.
- VOULAND, Pierre (2004). *Du provençal rhodanien parlé à l'écrit mistralien*, Aix-en-Provence, Edisud.

Annèxa 1 Exemple consultacion

a) Traducccion de la frasa estimul : « J’ai pris un moineau au piège. » a Vabre (ALF 755)

SYMLA

Syntactic Microvariation in the Romance Languages of France

Accueil

Consultation ALF

Consultation NES

Créer

Consultation

Édition

Le nombre de phrases retournées est 1, vous pouvez accéder à la page d'export : >>> PDF <<<

point n°752 | Montastruc-la-Consellière | Haute-Garonne

point n°753 | Brousse | Tarn

point n°753 | Vabre | Tarn

Choisir le(s) lieu(x) : point n°753 | Lamalou-les-Bains | Hérault

Choisir la (ou les) phrase(s) :  

J'ai cru qu'ils ne viendraient pas.

J'ai pris un moineau au piège.

J'ai une maigreur.

Par stimulus : Je me suis assis sous un arbre, appuyé contre le tronc.

2 Complétée | Complète avec subordonnée

2 Complétée | Complète avec coordonnée

3 Type matrice | Elliptique

4 Prohibitif matrice | Négative

Par phrase contenant un(e) : Nom

ayant la propriété grammaticale : Genre | Féminin

Actualiser

Afficher 100 éléments

Plus

Langue

Propriétés

Contexte

Complète ?

Actions

Phrase 16 J'ai pris un moineau au piège.

e'

Èi

AVER

Verbe

Singulier\_1ère

Indicatif

Passé

Présent

Auxiliaire

prez

pres

PRENE

Verbe

Masculin

Singulier

Participe

Passé

ym

un

UN, UNA

Déterminant

Indéfini

Masculin

Article

Masculin

Singulier

paserak

a

PASSERAK

Nom

Masculin

Singulier

a

A

Adposition

la

la

LO/LA

Déterminant

Défini

Article

Féminin

Singulier

t'endo-

tenda.

TENDA

Nom

Féminin

Singulier

Phrase

Phrase

Phrase

complète,

simple,

1090,

866,

1013

true

Voir

b) Detalh de la traducccion de la frasa estimul : « J’ai pris un moineau au piège. » a Vabre

Phrase 16 J'ai pris un moineau au piège.

e'

Èi

AVER

Verbe

Singulier\_1ère

Indicatif

Passé

Présent

Auxiliaire

prez

pres

PRENE

Verbe

Masculin

Singulier

Participe

Passé

ym

un

UN, UNA

Déterminant

Indéfini

Masculin

Article

Masculin

Singulier

paserak

a

PASSERAK

Nom

Masculin

Singulier

a

A

Adposition

la

la

LO/LA

Déterminant

Défini

Article

Féminin

Singulier

t'endo-

tenda.

TENDA

Nom

Féminin

Singulier

Phrase

Phrase

Phrase

complète,

simple,

1090,

866,

1013

true

Voir

Se legisson sucessivament :

- la frasa estimul en francés
- la frasa responsa reconstituïda en fonetica AFI
- la meteisa frasa responsa en fonetica AFI trocejada en mot
- la transposicion ortografica (version tipograficament afichabla de las glòsas).
- los lèmas (en capitalas)
- las proprietats gramaticalas dels mots donadas pels lèmas e las glòsas.

Dins las colonnas a man drecha, las proprietats de la frasa, e los numèros de las mapas ALF d’ont venon las donadas (tres aici). Los numèros son clicables e permeton de cargar en linha l’imatge de las mapas. La mençon “true” senhala que la frasa es estada exhaustivament glosada e lematizada.

281

paysé	lèi paimyè kaimégu	kou fa dél bèn, lèi
trémblé	dé flauri.	trou (junc) plèò, mè
bèt	lèi sùey ou dèssà	sé kòpò pà.
òrmé	kaimésà yò ywè trèu.	ùnò paimò tëndrò
piulo . f	grèfà (C)	pèrò ; kwén
sàilè	lò bino ; pè dé bino	ginyètso
sèy ; "	" (D) ; "	ròzin ; gràpò
tél	chpimò	nèzé . f ; ròhkolù , "
bùn	sèbò	òuglònò
oussè ; paimèl, - (9)	rèy	kòstònò ; kòstònè ; pèlèu
bui	fiwèyò	èn nial dé prèssè (m)
briso	ruhtò ; ouloun	òkèst ònnàdò , l'ò òtsu
rimé . f.	" "	bèlkwo dé frutsò (f)
gzyè ; (A)	kèpà dél buè, pèr	òuryà déul bèrè
buigou	faire dé fai.	kèsi léiz aubre
tamébrié (B)	mé sùey osyètà trèu	n'èrè kògàr.
paimyè	ùn aubre , opwiyà	sé nòntòm pà
pèryè	kòntro lù pè . ùnò	nòhtròy paimòy ,
nèyè	bròkò pweyridò m'ès	sé mazyèròu lèu.
paimyè	kambàdò sù lò karo	pyèlèlò
sùey	è m'ò fa sònna dél nà	frambwèzò
sùey soubàtse	èi lù kax ké mé ròdò	prèsigè
gzyè ; figò		kumanyè

### Annèxa 3 Transpcion API de la pagina 3 del questionari ALF al punt 715

n° p. 3 col 1	n° col 2	n° col 3
183 fr'ajse	215 le <sup>i</sup> pumje / kumɛfɯ / de fl'uri	238 kʰn fa del b'ɛ <sup>n</sup>   / lu tsũ / pl'ɛ:ɔ / mɛ / se k'ɔpɔ / pa
184 tr'ɛ <sup>m</sup> ble		
185 bet		
186 'orme	216 lei sir'ej / ɔu detsa / kumɛsa /	239 ɣno pumɔ t'ɛ <sup>n</sup> drɔ
187 pi'ulɔ	jɔ ju'ɛ tsɯr	
188 s'aulɛ		
189 sɛ:ɟ	217 gr'ɛfa	240 p'ɛrɔ
190 -	218 gr'ɛfu [nòta C]	241 kwɛ <sup>n</sup>
191 te:l	219 lɔ b'ɪnɔ	242 ʃɪrj'etsɔ
192 bɪn	220 pɛ de b'ɪnɔ	243 r'ɔʒi <sup>n</sup>
193 bw'ieu	221 --	244 gr'apɔ
194 pr'ynel	222 p'uda [nòta [D]	245 n'ɥʒɛ ; rɔhk'ɔlu
195 b'ui	223 --	246 ɔugl'ɔnɔ
196 br'usɔ	224 --	247 kɔstɔɟnɔ
197 r'ume [f]	225 --	248 k'ɔstɔɟɛ
198 r'ɥʒɛ	226 ehp'ɪnɔ	249 p'elu
199 egl'ɔtjɛ [nòta A]	227 s'ebɔ	250 ɛn nɯal / de pr'ɛfet [m]
200 gr'ato tju	228 rɛj	
201 bw'ieu	229 fɥ'ejɔ	251 ɔkest ɔnn'adɔ / k'ɔ ɔtsyt / belkwɔ / de fr'ɥ:tsɔ
202 tsɪn'ɛbre	230 r'ɣkɔ	
204 --	231 'oubɣ <sup>n</sup>	252 ɔurjaj / deyt / b'ɛire / kesi / le'z aubre / n ɛrɥ / k'ɔrgat
205 --	232 --	
206 p'umje	233 --	
207 p'erje	234 kupa / del bwe / per fajre / de fa:ɪ	253 se mɔntsɔm pa / nɔhtɔɟ pr'ymɔj / se mouzir'ɔu / l'ɛu
208 n'uyɛ		
209 pr'ymje	235 me sɥɛj / ɔsjeta tsɯz ɣn 'a:ubre / ɔpujja / k'ɔ <sup>n</sup> tro lu pɛ	254 pjel'aɔɔ
210 sɪr'e:ɟ		255 frɯmbw'ɛzɔ
211 sɪr'ej soub'atse	236 ɣno brɔkɔ / pweir'ɪdɔ / m ɛs tɥ <sup>m</sup> bado / sy lɔ k'a:ɔɔ / e m ɔ / fa / s'ɔ <sup>n</sup> na / del na	256 --
212 --		
213 fɪɟjɛ		257 pr'esɪɟɛ
214 fɪɟɔ	237 ej lu ka <sup>k</sup> ke me r'ɔ:ɔ	258 kɯn'asjɛ

## Annèxa 4 La pagina 3 del questionari ALF

(en gras las frasas o segments de frasa, entre parentèsis los n° de mapas)

n°	col 1	n°	col 2	n°	col 3
183	frêne (611)	215	<b>Les pommiers /</b>	238	<b>Quand il fait du vent / le</b>
184	tremble n.m. (arbre) (1329)		<b>commencent / à</b>		<b>roseau / plie / mais / ne</b>
185	bouleau (158)	216	<b>Les cerisiers / ont déjà</b>		<b>/ rompt/ pas.</b> (1370; 1166;
186	ormeau (948)		<b>/ commencé (/ à</b>	239	<b>une pomme tendre</b> (1055)
187	peuplier (1008)		<b>fleurir) / il y a huit</b>		
188	saule (1196)		<b>jours.</b> (218; 383; 313;		
			583; 729)		
189	sureau (1270)	217	greffer (666)	240	poire (1047)
190	hièble (1595)	218	greffon (en note) (1833)	241	coing (1510)
191	tilleul (1303)	219	la vigne (1392)	242	cerise (217)
192	osier (955)	220	cep de vigne (1780)	243	raisin (1129)
193	aubépine (68)	221	sarment	244	grappe de raisin (1832)
194	cenelle (fruit de l'aubépine) (1490)	222	tailler la vigne (1907)	245	noix (920)
195	buis (186)	223	épamprer (1816)	246	noisette (919)
196	bruyère (183)	224	pampre (1654)	247	châtaigne (251)
197	ronce (1163)	225	échalas (rame) (434)	248	châtaignier (1497)
198	rosier (1167)	226	épine (476)	249	bogue de châtaigne (1407)
199	églantier (452)	227	sève (1230)	250	<b>un noyau / de pêche</b> (926;
200	gratte-cul (cynorhodon) (452)	228	racine (1126)		987)
201	buisson (187)	229	feuille (559)	251	<b>Cette année / il y a eu /</b>
		230	écorce (442)		<b>beaucoup / de fruit.</b> (44;
					103; 120)
202	genévrier (636)	231	aubier (1446)	252	<b>Vous auriez / dû / voir /</b>
204	genièvre (baie) (636)	232	provin (1096)		<b>comme / les arbres / en</b>
205	genièvre (boisson) (636)	233	provigner (1682)		<b>étaient / chargés.</b> (99;
					402; 1408; 310; 52; 513;
206	pommier (1058)	234	<b>couper / du bois /</b>	253	<b>Si nous ne mangeons pas</b>
207	poirier (1049)		<b>pour faire / des fagots</b>		<b>/ nos prunes / elles se</b>
			(35; 144; 530; 526)		<b>moisiront / bientôt.</b> (806;
208	noyer (arbre) (927)	235	<b>Je me suis / assis /</b>		1097; 869; 132)
209	prunier (1099)		<b>sous un arbre /</b>	254	pelure de pommes (993)
210	cerisier (218)		<b>appuyé / contre le</b>	255	framboise (609)
			<b>tronc.</b> (500; 63; 51; 48;		
			1334)		
211	cerisier sauvage (218)	236	<b>Une branche / pourrie</b>	256	framboisier (609)
			<b>/ m'est tombée / sur la</b>		
			<b>figure / et m'a / fait /</b>		
			<b>saigner / du nez.</b> (170;		
			1076; 1312; 566; 88; 533;		
			1180; 908)		
212	cerise sauvage (218)			257	pêcher n.m. (1659)
213	figuier (1567)			258	cognassier (1510)
214	figue (1567)	237	<b>La tête me tourne</b>		