

Unha nova base de datos para a romanística a partir do *Atlas Linguistique Roman*¹

ESTEVE VALLS^[1,2], MANUEL GONZÁLEZ GONZÁLEZ^[1,3]

^[1]*Centro Ramón Piñeiro para a Investigación en Humanidades*

^[2]*Universitat de Barcelona*

^[3]*Universidade de Santiago de Compostela*

RESUMO. O Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) está a elaborar, a partir da explotación do *Atlas Linguistique Roman* (ALiR), unha gran base de datos informatizada con información diatópica, etimolóxica, etnográfica, semántica e mais bibliográfica das distintas falas que integran o contínuum románico. Presentar as características e a utilidade desta base de datos que, malia atoparse aínda nunha fase inicial, xa conta con máis de 57.000 rexistros, é o principal obxectivo deste artigo, que se estrutura en tres apartados: en primeiro lugar, fornécese unha introdución ás principais características do ALiR; en segundo lugar, detállase a estrutura desta base de datos e do formulario que se creou para facilitar as buscas; finalmente, suxírense algunhas posibles explotacións deste novo recurso que se pon ao servizo da romanística. Algunhas destas aplicacións xa se están a desenvolver no CRPIH; outras son unha porta aberta a futuras investigacións.

Palabras chave: romanística, Atlas Linguistique Roman, dialectoloxía, lingüística de corpus, lingüística histórica, semántica.

Data de recepción: 12-09-2014 ▪ Data de aceptación: 16-01-2014.

¹ Este traballo inscríbese nos proxectos «Descrición e interpretación de la variación dialectal: aspectos fonolóxicos y morfolóxicos del catalán» (FFI2010-22181-C03-02), financiado polo MICINN e mais o FEDER, e «Cambio lingüístico no galego actual» (FFI2012-33845), financiado polo Ministerio de Economía e Competitividade; e foi posible grazas á axuda do Centro Ramón Piñeiro para a Investigación en Humanidades.

ABSTRACT. The Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) is working on a huge computerized database with diatopic, etymological, ethnographic, semantic as well as bibliographic information based on the exploitation of the data contained in the *Atlas Linguistique Roman* (ALiR). The aim of this paper is to describe the main features and show the usefulness of this database, which nowadays consists of more than 57.000 items. The paper is divided into three sections: first, we briefly introduce the main characteristics of the ALiR; second, we describe the structure of the database; finally, we suggest several possible investigations which might be carried out with the help of this new tool which is put at the service of the Romance studies.

Keywords: Romance Studies, Atlas Linguistique Roman, Dialectology, Corpus Linguistics, Historical Linguistics, Semantics.

1. INTRODUCCIÓN

O *Atlas Linguistique Roman* (ALiR) constitúe o primeiro intento de levar a cabo unha análise global da *Romania Antiqua* a partir da síntese de todos os atlas lingüísticos –publicados ou aínda inéditos– do ámbito románico. Trátase dun atlas de segunda xeración, interpretativo: en vez de partir de datos recollidos *ad hoc* mediante a realización de novas enquisas dialectais, o ALiR parte dos datos obtidos durante a elaboración da cincuentena de atlas lingüísticos que se confeccionaron ao longo do século XX nos distintos estados, rexións ou ámbitos lingüísticos do espazo románico. Tal como afirman dous dos principais impulsores deste proxecto de alcance europeo, Gaston Tuaille e Michel Contini, o ALiR naceu coa vontade de se converter nun «étude d'ensemble de l'espace roman» (ALiR I 1996: 4) que, dende unha perspectiva esencialmente diacrónica (Contini 1992: 347), pretendía «mettre en évidence les traits linguistiques de différente nature (lexicaux, phonétiques, phonologiques ou morpho-syntaxiques) permettant de définir l'identité des aires dialectales romanes mais aussi les traits qui les rapprochent les unes des autres» (ALiR I 1996: 4).

Para acadar este obxectivo, creáronse dez comités de expertos que, baixo a coordinación científica do Centro de Dialectoloxía da Universidade de Grenoble, recibiron o encargo de realizaren unha *síntese nacional* para cada concepto estudado. Estas sínteses consisten nunha análise detallada da variación lingüística existente nas denominacións dun referente no ámbito de actuación dun determinado comité; todos os comités analizan a variación interna dun estado agás os comités catalán e galego, que se cinguen aos seus respectivos ámbitos lingüísticos. De oeste a leste, os comités encargados da elaboración das sínteses nacionais son: o portugués, o galego, o español, o catalán, o francés, o valón, o suízo, o italiano, o romanés e mais o moldavo.

Unha vez que os comités presentan as sínteses nacionais referidas a un concepto, un ou varios expertos elaboran unha *síntese románica* na que se ofrece unha panorámica da variación lingüística existente no conxunto da Romania verbo do concepto estudado. Como se explicita no volume de introdución ao ALiR, unha síntese románica é

un classement de toutes les formes romanes en fonction de critères de leur choix. Ils [os autores da síntese] présentent en outre un article de commentaire succinct, où les formes sont analysées dans leur diversité, dans leur variabilité, dans leur répartition géographique, en tenant compte toujours de la perspective diachronique dans la mesure où elle facilite la compréhension de la situation actuelle. Ces synthèses sont destinées à la réalisation de la cartographie et contiennent donc des renseignements nécessaires à la réalisation de la carte (ALiR I 1996: 16).

Posto que o ALiR é un atlas lingüístico, o seu obxectivo principal é o de ofrecer unha visualización clara da distribución xeográfica das variantes lingüísticas coas que se denomina un determinado referente ou concepto no espazo lingüístico románico. Porén, tanto as sínteses nacionais como as sínteses románicas ofrecen, tamén, unha inxente cantidade de información complementaria sobre aspectos tan diversos como a orixe etimolóxica ou a motivación semántica de cada unha das denominacións, así como información etnográfica asociada ao concepto estudado en cada ámbito lingüístico ou referencias bibliográficas sobre o tema que puideren resultar útiles ao lector. Malia o interese indiscutible que tería poder consultar todas estas informacións de maneira cruzada, ata o día de hoxe a única forma que había de poñelas en relación era confrontando no papel as distintas sínteses románicas.

Co obxectivo de proporcionar unha nova ferramenta que permita unha mellor explotación dos datos contidos nas sínteses nacionais e románicas, no Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) estase a desenvolver unha base de datos na que se pretende centralizar non só a información diatópica contida no ALiR, senón tamén o corpus analítico complementario –etimolóxico, etnográfico, semántico e bibliográfico– que os expertos dos distintos comités e os autores das sínteses románicas foron achegando, respectivamente, nas sínteses nacionais e nas románicas. Presentar as características e a utilidade desta base de datos que, malia que aínda se está a elaborar, xa conta con máis de 57.000 rexistros dispoñibles, é o principal obxectivo deste artigo, que se estrutura en tres apartados principais: en primeiro lugar, fórnese unha introdución ás principais características do ALiR, poñendo especial atención no corpus que actualmente xa está dispoñible na base de datos que se presenta; en segundo lugar, detállase a estrutura desta base de datos e do formulario que se creou para facilitar as buscas; finalmente, suxírense algunhas posibles explotacións deste

novo recurso que se pon ao servizo da romanística. Algunhas destas aplicacións xa se están a desenvolver no CRPIH; outras son unha porta aberta a futuras investigacións.

2. O ATLAS LINGUISTIQUE ROMAN

Como se mencionou na introdución, o ALiR é un atlas panrománico de segunda xeración, interpretativo; isto quere dicir que a totalidade dos datos do corpus foron tirados de fontes xa existentes –publicadas ou inéditas–, e moi especialmente de atlas lingüísticos dun ámbito xeográfico ou lingüístico específico. Concretamente, as fontes das que proceden a maioría dos datos do ALiR son²:

a) Ámbito galorrománico: o *Atlas linguistique de la France* (ALF), os 21 atlas rexionais –catro deles de áreas non galorrománicas: Córsega, o País Vasco, a Bretaña e mais a Cataluña do Norte– que conforman o *Nouvel Atlas linguistique de la France par régions* (NALF), o *Atlas linguistique de la Wallonie* (ALW), os *Tableaux phonétiques des patois suisses romands* (TP) e mais o *Glossaire des Patois de la Suisse romande* (GPSR).

b) Ámbito italorrománico: o *Atlante Italo-Svizzero* (AIS), o *Atlante Linguistico Italiano* (ALI), o *Atlante Storico-linguistico etnografico friulano* (ASLEF), o *Atlas linguistique de la France. Corse* (ALFCorse), o *Atlante linguistico etnografico italiano della Corsica* (ALEIC), o *Nouvel Atlas Linguistique de la Corse* (NALC), o *Saggio di un Atlante linguistico della Sardegna* (ALSar) e mais o *Vocabolario dei dialetti della Svizzera italiana* (VDSI).

c) Ámbito iberorrománico: o *Atlas Lingüístico de la Península Ibérica* (ALPI), o *Atles lingüístic de Catalunya* (ALC), o *Atles lingüístic del domini català* (ALDC), o *Atlas lingüístico y etnográfico de Andalucía* (ALEA), o *Atlas lingüístico y etnográfico de Aragón, Navarra y Rioja* (ALEANR), o *Atlas lingüístico y etnográfico de las Islas Canarias* (ALEICan), o *Atlas lingüístico y etnográfico de Cantabria* (ALECant), o *Atlas lingüístico y etnográfico de Castilla-La Mancha* (ALECMAN), o *Atlas Lingüístico Galego* (ALGa), o *Atlas lingüístico-etnográfico de Portugal e da Galiza* (ALEPG) e mais o *Atlas Lingüístico de España y Portugal* (ALEP).

d) Ámbito retorrománico: o *Atlante Italo-Svizzero* (AIS) e o *Dicziunari Rumantsch Grischun* (DRG).

² Nesta listaxe respéctase a clasificación das fontes que se fai na introdución do volume de presentación do ALiR –*vid.* ALiR I (1996). A relación implícita que se establece entre estas fontes e os principais ámbitos supralingüísticos –iberorrománico, galorrománico, etc.– débese interpretar como un intento de sistematización dos diferentes atlas de partida máis que como un posicionamento dos responsables do ALiR no que fai á subagrupación das diferentes variedades románicas.

e) *Ámbito romanés*: os cinco atlas rexionais do *Noul Atlas Lingvistic Român pe regiuni* (Oltenia (NALR Olt), Maramureş (ALRR Mar), Banato (NALR Ban), Moldavia e Bucovina (NALR Mold Bucov) e mais Transilvania (ALRR Trans)), o *Atlasul Lingvistic Moldovenesc* (ALM), o *Atlasul Lingvistic Român pe regiuni Basarabia, Bucovina de nord, Transnistria* (ALRR Bas), o *Aromunischer Sprachatlas. Atlasul lingvistic aromân* (ALiA) e mais o *Meglenoromänischer Sprachatlas* (WILD Megl).

Actualmente a base de datos contén información sobre 50 dos 592 conceptos do cuestionario lexical; son os seguintes: *abeille, alouette, araignée, aujourd'hui, belette, berceau, blatte, bousier, chenille, cigale, cloporte, courtilière, crapaud, demain, essaim, foie, forgeron, fourmi, fourmilière, grenouille, grillon, guêpe, hêtre, hier, libellule, lundi, mante religieuse, mardi, merle, mille-pattes, miroir, moineau, moucheron, moustique, orvet, paille, papillon, perce-oreille, punaise des lits, rouge-gorge, salamandre, serpent, taureau, ténébrion, têtard, toile d'araignée, tortue, ver de terre, ver luisant*.

Os datos contidos neste corpus foron recollidos en 1.037 puntos de enquisa e son xestionados por dez comités de expertos que representan, como se dixo, ben un estado no que se falan varias linguas románicas, ben un ámbito lingüístico determinado, como recolle a táboa 1:

Comité	Puntos de enquisa	Observacións
Catalán (CAT)	41	4 en Francia, 1 en Italia e 1 no Principado de Andorra
Español (ESP)	217	7 nas Canarias
Francés (FRA)	297	8 no ámbito corso e 2 no ámbito céltico da Baixa Bretaña
Galego (GLC)	30	
Italiano (ITA)	171	1 situado na ex-Iugoslavia, 7 situados no ámbito galorrománico, 1 punto serbocroata, 1 punto albanés e 2 puntos gregos
Moldavo (MOL)	18	5 na República de Ucraína
Portugués (PRT)	110	10 nos Azores, 3 en Madeira e 1 en Porto Santo
Romanés (ROU)	129	8 arromaneses, 1 dacorromanés, 1 meglenorromanés e 1 istrio-romanés situados en Grecia, Albania e a ex-Iugoslavia
Suízo (SUI)	15	7 na Suíza galorrománica, 3 na Suíza italiana e 5 no ámbito retorrománico
Valón (WAL)	9	

Táboa 1. Distribución das 1.037 localidades enquisadas por comité, por orde alfabética. Adaptada de ALiR I (1996: 10)

No que fai á densidade da rede de puntos, cómpre ter en conta que nos ámbitos lingüísticos do galego e das variedades galorrománicas e italo-románicas, os puntos de enquisa non representan poboacións, senón pequenas agrupacións de localidades lingüisticamente moi próximas, que denominamos *casas* (en francés, *cases*):

Une case représente en fait une petite aire dialectale dans laquelle peuvent se rencontrer plusieurs autres points ayant fait l'objet d'enquêtes dans le cadre d'atlas nationaux et/ou d'atlas régionaux (voir les cartes ci-jointes), linguistiquement proches, et qui auront été regroupés par les spécialistes, après une étude préalable. Pour les pays qui disposent de plusieurs atlas nationaux (Italie) ou d'atlas nationaux et régionaux (France) cette solution permet de mieux exploiter les données dialectales qu'ils contiennent, par la présence de tous leurs points dans le réseau de cases (ALiR I 1996: 11).

3. ESTRUCTURA DA BASE DE DATOS E DO FORMULARIO

Neste apartado detállase a estrutura da base de datos e do formulario de consulta do ALiR que se está a desenvolver no CRPIH. Malia que a información que proporcionan ambos os dous é a mesma, optamos, por claridade expositiva, por describir a estrutura da base de datos campo por campo. A información que se proporciona no formulario de consulta é idéntica á da base de datos, pero preséntase na pantalla do computador de maneira non lineal. A base de datos consta de 60 campos ou columnas que conteñen as seguintes informacións:

- a) Columna 1: contén un código identificativo da secuencia fónica coa que se denomina un referente nun punto de enquisa determinado. O código 16.299, por exemplo, indica que no punto de fala andaluza ESP111 (Ayamonte), o concepto *abeille* 'abella' recibe o nome de [oβehíta ðe la mjél] 'ovejita de la miel'.
- b) Columna 2: contén o nome –en francés, por ser a lingua de traballo habitual do grupo de lingüistas encargados da explotación do ALiR– co que se identifica o referente na síntese románica –*abeille*, *alouette*, etc. Como se viu no apartado 2, ata hoxe lévanse introducidos na base de datos 50 conceptos, que dan lugar ás 57.461 realizacións que conforman o corpus.
- c) Columna 3: contén unha ligazón a unha imaxe identificativa do referente.
- d) Columna 4: contén o nome científico identificativo do referente; por exemplo, *Mustela nivalis* para a *donicela*.
- e) Columna 5: contén o número da pregunta coa que se obtivo unha determinada resposta no cuestionario; o número 130, por exemplo, remite ao apartado «Fauna

- salvaxe» do cuestionario léxico e corresponde, concretamente, ao concepto *belette* ‘donicela’.
- f) Columna 6: malia que as designacións románicas da maioría de referentes se poden cartografar nun só mapa, os conceptos con maior variación lingüística débense cartografar, por razóns de espazo, en varios mapas parciais. Esta columna contén o número do mapa –de entre os que plasman as designacións dun mesmo referente– no que se atopa cartografada a pronuncia recollida nunha localidade concreta da base de datos –a que recolle a columna 7.
- g) Columna 7: contén un código que remite a un punto de enquisa; o punto CAT201, por exemplo, corresponde á localidade de Taüll, na que se fala o dialecto *pallarès* do catalán.
- h) Columna 8: contén un cadradiño que, de estar marcado, indica que a resposta aparece cartografada no mapa; pola contra, un cadradiño baleiro indica que se trata dunha resposta secundaria, isto é, que non se cartografou no mapa por ser minoritaria ou menos representativa dun determinado lugar ou *casa*. As respostas secundarias, porén, si que se analizan nas sínteses nacionais e románicas.
- i) Columna 9: contén o nome do comité ao que pertence un punto de enquisa; as abreviaturas son as que aparecen na táboa 1.
- j) Columna 10: indica o ámbito lingüístico ao que pertence un determinado punto de enquisa. Empréganse as seguintes abreviaturas:

alb.	albanés	itc.	italiano central
arag.	aragonés	itm.	italiano meridional
ast.-léon.	asturleonés	its.	italiano setentrional
brt.	bretón	lad.	ladino
bsq.	vasco	mir.	mirandés
cast.	castelán	occ.	occitano
cat.	catalán	oïl	oïl
cor.	corso	prt.	portugués
fpr.	francoprovenzal	rhr.	retorrománico
fri.	friulano	rou.	romanés
git.	galoitálico de Sicilia e Sardeña	scr.	serbocroata
glc.	galego	srd.	sardo
gr.	grego		

- k) Columna 11: indica a variedade dialectal á que pertence un determinado punto de enquisa. Empréganse as seguintes abreviaturas:

ast.	asturiano	CBPI	portugués de Castelo Branco e Portalegre
léon.	leonés	mad.	portugués da Madeira
and.	andaluz	prtem.	portugués centro-meridional
can.	canario	prtlits.	portugués do litoral setentrional
cant.	cántabro	prtm.	portugués meridional
estr.	estremeño	prts.	portugués setentrional
murc.	murciano	arou.	arromanés
alg.	alguerés	drou.	dacorromanés
gasc.	gascón	irou.	istrio-romanés
aço.	portugués dos Azores	mrou.	meglenorromanés

- l) Columna 12: contén a transcripción fonética da secuencia fónica coa que se denomina un referente nunha determinada localidade. Nesta primeira fase de elaboración da base de datos, optouse por traballar estritamente coas transcripcións que se fornecen nas distintas sínteses románicas; máis adiante, porén, haberá que unificar certas incoherencias que se detectaron nas sínteses á hora de transcribir determinados sons —por exemplo, as consoantes róticas vibrantes [r] e batidas [r̄], que ás veces se transcriben en ambos os dous casos como [r].
- m) Columna 13: contén a forma ortográfica da secuencia fónica coa que se denomina un referente nunha determinada localidade.
- n) Columna 14: contén unha tradución literal ao francés ou unha paráfrase nesta lingua da secuencia fónica coa que se denomina un referente nunha determinada localidade. Por exemplo, no punto de fala castelá ESP094b (Santa Cruz de la Zarza), recolleuse para o concepto *bousier* ‘escaravello’ a forma [tartanéro] ‘tartanero’, para a que se dá a seguinte paráfrase en francés: «celui qui conduit une carriole».
- o) Columna 15: pode conter algunhas notas fonéticas ou ortográficas sobre a secuencia fónica recollida.
- p) Columna 16: contén o xénero –F: feminino, M: masculino– da secuencia fónica recollida.
- q) Columna 17: contén un cadrado que indica, de estar marcado, que o xénero que figura na columna 16 está verificado por un experto e non só polo xestor da base de datos.
- r) Columna 18: contén o número –S: singular, P: plural– da secuencia fónica recollida.
- s) Columna 19: contén un cadrado que indica, de estar marcado, que o número que figura na columna 18 está verificado por un experto e non só polo xestor da base de datos.

- t) Columnas 20-47: conteñen a información etimolóxica da secuencia fónica recollida segmentada morfoloxicamente. Cada morfema preséntase en catro columnas:
- Columna 20: indica a secuencia fónica que constitúe o primeiro morfema da palabra.
 - Columna 21: indica a natureza deste primeiro morfema –polo xeral, un radical.
 - Columna 22: indica a lingua do étimo que dá lugar ao morfema da columna 20.
 - Columna 23: indica o étimo do que provén o morfema da columna 20.
 - As columnas 24-27, 28-31, 32-35, 36-39, 40-43 e 44-47 conteñen, respectivamente, a mesma información que as columnas 20-23 pero para o segundo, o terceiro, o cuarto, o quinto, o sexto e o sétimo morfemas da secuencia fónica recollida.
- u) Columna 48: pode conter algunhas notas explicativas sobre a morfoloxía da secuencia fónica recollida.
- v) Columna 49: pode conter algunha explicación adicional sobre a etimoloxía da secuencia fónica recollida.
- w) Columna 50: contén o número do parágrafo da síntese nacional no que se dá conta da etimoloxía da secuencia fónica recollida.
- x) Columna 51: contén os nomes dos autores que elaboraron a síntese nacional na que se trata o referente estudado nesta liña da base de datos.
- y) Columna 52: contén o título do parágrafo da síntese nacional no que se dá conta da etimoloxía da secuencia fónica recollida nesta liña da base de datos.
- z) Columna 53: contén a transcripción do parágrafo da síntese nacional no que se analiza a secuencia fónica estudada nesta liña da base de datos.
- aa) Columna 54: contén a motivación semántica da secuencia fónica analizada. Nesta base de datos tívose en conta ata o de agora o seguinte abano de motivacións semánticas:

Motivación perdida ou opaca

PERDUE: descoñécese completamente a motivación da palabra. Un referente do que se ignora a motivación en moitas variedades románicas é o dos *cágados* dos anfibios da orde *Anura*; estes pequenos animais adoitan recibir, segundo di Médélice (2009: 222) na síntese románica, nomes «d’origine obscure», como [samarúk] ‘samaruc’ no catalán de CAT219 (Castelló de la Plana), [tutús] no

francoprovenzal de FRA158 (Ambierle *et al.*³), [gwák] no occitano de FRA282 (Gavarnie *et al.*), [bét] no italiano setentrional de ITA006 (Trasquera *et al.*) ou [bolobréʃe] no dacorromanés de ROU062 (Berzasca), entre outros.

OPAQUE: téñense formulado algunhas hipóteses respecto á motivación semántica da palabra, pero non hai consenso sobre ningunha delas. Os *Blattodea* ou *cas-cudas*, por exemplo, reciben nomes moi semellantes nalgunhas variedades do murciano e do valenciano –como [koliána] ‘curiana’ no punto CAT222 (Altea) e [kuriána] ‘curiana’ no punto ESP120b (Puebla de Don Fadrique)–, cuxas orixes son opacas, segundo recolle Franconie (2009a: 17) na síntese románica:

L’aire actuelle de «curiana» provient d’une avancée du terme depuis l’Andalousie, où il s’étend largement, vers le nord et l’est (Murcie, Valence), mais elle a été érodée par «cucaracha». Selon le DCECH, ce terme pourrait provenir d’une allusion au costume sombre des villageois de l’Évêché de Coria (Estremadure). Certes, le terme est attesté dans cette région (cf. RDTP IV, 626), mais il existe aussi en Andalousie une localité du nom de Coria. Or nous venons de voir, avec les désignatifs ethniques, que la péjoration concerne les voisins! Il ne s’agit pas d’une autodérision. Ceci porte à penser que, si la locution «curiana» traduit une moquerie à l’égard des habitants de Coria (d’Estremadure ou d’Andalousie), ce n’est pas dans l’une ou l’autre de ces régions qu’on devrait la trouver, mais chez des voisins. Il faudrait donc envisager d’autres hypothèses.

Le DCECH rejette tout rapport avec les continueurs de CORIUM ‘cuir’, soit direct, soit par l’intermédiaire de «coria», terme scientifique désignant les élytres durs ou la carapace de certains insectes. Pourtant, la blatte pourrait bien être «l’insecte qui ronge les cuirs» (Linné décrit *Blatta laponica* comme rongeur des cuirs des Lapons), et sa couleur fauve peut aussi évoquer celle du cuir.

Enfin, s’agissant d’une région où la présence grecque a été importante, le rapprochement avec gr. κοπίος? ‘punaise’ devrait être étudié, parce qu’il est sémantiquement possible.

Motivación fonosimbólica

PHONOSYMB: o referente vén suxerido pola fonética da palabra, de maneira que os sons non evocan unha categoría acústica, senón unha idea ou unha calidade física ou moral (Ullmann 1972: 95). Un exemplo é o nome que recibe o *Scarabaeus laticollis* ou *escaravello peloteiro* no punto de fala italiana setentrional ITA021 (Vico Canavese *et al.*), [bója]. Para Fidalgo e Álvarez (2009: 49),

³ De agora en diante, cando despois dun punto de enquisa apareza, entre parénteses, un nome de localidade seguido de «*et al.*», debe entenderse que se trata dunha *casa*.

toujours en relation avec des formations à motivation phonosymbolique, on peut ajouter une famille de mots comportant la base «*ba(b)»-, «*bo(b)»-, «*pa(p)»-, «*bau»-, «*bou»-, que l'on retrouve dans les désignations de certains insectes suscitant la peur mais aussi des insectes en général et de certains êtres fantastiques. Nous sommes devant un groupe de formes minoritaires mais très diverses, qui appartiennent (excepté occ. [bóbo]) à l'Italie septentrionale et au sarde; on peut les relever sous une forme simple ou suffixée, mais aussi dans des syntagmes avec modificateur (du type 'noir' ou 'de la merde').

Motivación transparente

ONOMATOP: o nome deriva dunha onomatopea que se relaciona con algún son que fai o referente. Os *Gryllidae* ou *grilos*, por exemplo, reciben nomes que remiten ao son que producen coas súas ás anteriores ao frenalas; así acontece no punto de fala *d'oïl* FRA095 (Ronchamp *et al.*), onde son chamados cun nome de clara motivación onomatopeica: [krikrí] 'cri-cri'. Gargallo e Veny (2001: 202-203) refírense, na síntese románica, ás formas «de base expressive: KRIKK-, GRIKK-, RIKK-», e din que

la base expressive peut être simple ou redoublée. C'est le résultat d'une onomatopée du chant de la bestiole, comme nous avons déjà dit. Le prt., qui emploie [grílu] pour le désigner, se sert de *gri gri* pour imiter son chant: «*Gri gri* sai á porta / que o teu pai vem á horta». L'imitation articulée de son grésillement expliquerait la diversité des variantes occ. [gríp], [rík]; itc. [krikko]; fpr. [trí], [grí]; fri. [grí]; cat. [grí], [gríw], etc. Les formes avec G- pourraient avoir comme base le type lexical héritier de GRILL(I)US. Ajoutons encore la base expressive CHIUR- de l'itm. [tjúru].

Les formes redoublées les plus fréquentes ont comme base KRI-, GRI-, GRIU-, RIK- en domaine gallo-roman [krikrí], occ. [griwgríw], cat. [rigrík]; le sarde a développé les bases PIRRI- et SIR- ([piripíri], [sirisíri]).

Unha pequena pescuda na base de datos demostra a relevancia desta motivación transparente para designar os grilos en toda a *Romania Antiqua*. As denominacións de base onomatopeica non son soamente moi diversas, como se desprende de Gargallo e Veny (2001: 202-203), senón que cuantitativamente tamén son as hexemónicas do ámbito lingüístico románico: de feito, case un 94% das denominacións románicas do grilo teñen unha motivación onomatopeica.

Esta motivación é rendible para designar animais moi diversos ao longo da *Romania*: dende o *Passer domesticus* ou *pardal* no punto de fala *d'oïl* FRA130

(Neuvy-Pailloux *et al.*), chamado [pjáf]‘piaf’ por imitación do canto deste paxaro, ata o *Culex pipiens* ou *mosquito* no punto de fala italiana setentrional ITA002 (Piazzola *et al.*), chamado [zanzára] a causa do son que fai este insecto ao voar. Grazas á base de datos creada no CRPIH, agora é posible analizar a distribución areal das motivacións semánticas en toda a Romanía e comprobar, por exemplo, que a distribución regular da motivación onomatopeica en todas as variedades contrasta coa distribución irregular da motivación PRAT_MAGIQUE –aquela pola que se dá nome a un referente pola súa vinculación con algún tipo de práctica máxica ou poder sobrenatural–, que se concentra, como se vai ver máis adiante, en boa medida na Península Itálica. A análise destas diferenzas cuantitativas na distribución das motivacións semánticas é unha das múltiples posibilidades de investigación que se abren grazas á elaboración desta base de datos.

1. Características físicas

COULEUR: o referente é dunha cor característica que lle dá nome. O *Lumbricus terrestris* ou *miñoca*, por exemplo, recibe o nome de [vérmu rúʃʃu], literalmente «verme vermello», no punto de fala italiana meridional ITA092 (Trasacco *et al.*).

ODEUR: o referente cheira dunha maneira característica que lle dá nome. No punto de fala dacorromanesa ROU081, por exemplo, os *Blattodea* ou *casculas* reciben o nome de [gîndák puturós] ‘gândac puturos’, literalmente, «insecto cheirento».

TAILLE: o referente ten un tamaño característico que lle dá nome. No punto de fala aragonesa ESP042 (Hecho), por exemplo, o *Bufo bufo* ou *sapo común* recibe o nome de [tánke] ‘tanque’, designación motivada, segundo di Augusto (2009: 144) na síntese románica, «par l’aspect extérieur, lourdaud et massif du batracien et que l’on a fait remonter au mot anglais ‘tank’».

PARTIES_REF: o referente recibe o nome por sinécdoque dalgunha das súas partes –considerada especialmente representativa da súa morfoloxía. No punto de fala d’oil FRA049 (Sommelonne *et al.*), por exemplo, a *Forficula auricularia* ou *cadela de frade* recibe o nome de [furjét] ‘fourchette’, literalmente «forcadiña», pola similitude entre a súa cola e unha forcada –*cfr.* Carrilho e Lobo (2001: 412-413).

AUTRES_PHYSIQ: o referente ten unha característica física distintiva diferente das anteriores que lle dá nome. O *Anguis fragilis* ou *alacrán*, por exemplo, recibe no punto de fala italiana meridional ITA098 (Gallo *et al.*) o nome de [lutwéjna], que literalmente significa «breco, miope». Dalbera (2001: 382) indica na síntese románica (18.1.3) que

les dénominations analysées brièvement ci-dessus ne doivent pas masquer l'un des ressorts sémantiques essentiels en ce qui concerne l'orvet: le caractère aveugle qu'on lui prête. Toutes les descriptions qui sont faites de ce reptile mentionnent cette propriété: «la disposition de ses yeux, munis d'une paupière clignotante, fait croire qu'il est dépourvu de l'organe de la vue». [...] Le nombre de sources étymologiques indépendantes qui réfèrent à cette propriété est particulièrement élevé: ORBUS 'privé de' et, plus spécifiquement, 'privé de la vue', 'aveugle'; [...] CAECUS 'aveugle' et son dérivé CAECILIA; [...] ABOCULIS 'aveugle'; [...] LUSCUS 'borgne' ou 'myope'; [...] *BORNIUS 'borgne' ou 'aveugle'; [...] THWAIKHS 'oblique, de travers' 'qui louche'; *SLIMB 'oblique, de travers, de biais'.

2. Características comportamentais e fisiolóxicas

MOTRICITÉ: o referente recibe o nome pola maneira que ten de se desprazar ou de voar. No punto de fala andaluza ESP120 (Cúllar Baza), por exemplo, as *serpes* reciben o nome de [arastrás] 'arrastradas' pola súa maneira de arrastrárense polo chan.

CHANT: o referente recibe o nome pola súa maneira de cantar. Os *Cicadidae* ou *carricantas*, por exemplo, reciben no punto de fala galega GLC019 (Fefiñáns *et al.*) nomes como [kantariña] 'cantariña' ou [hajitéira] 'gaiteira'.

HABITAT: o referente recibe un nome que fai referencia ao hábitat no que vive. No punto de fala corsa FRA506 (Coti-Chiavari *et al.*), o *Passer domesticus* ou *pardal* recibe o nome de [atjédqda kazandína], literalmente «pardal da casa», porque adoita vivir en lugares habitados polo ser humano.

NOURRITURE: o referente recibe un nome relacionado cos alimentos dos que se nutre. No punto de fala ladina ITA003 (Pàdola *et al.*), por exemplo, o *Passer domesticus* ou *pardal* recibe o nome de [panegás] 'panegas', literalmente «que come cardos».

COMPORT_HOMME: o referente recibe o nome pola conduta que ten (ou que as persoas cren que ten) con respecto ao ser humano. No punto de fala catalá CAT212 (Barcelona), por exemplo, o *Forficula auricularia* ou *cadela de frade* recibe o nome de [pàpəuréləs] 'papaorelles', literalmente «papaorellas», pola crenza de que este insecto ataca as persoas meténdoselles pola orella. Na síntese románica, Carrilho e Lobo (2001: 406-409) profundan nas orixes desta motivación:

La croyance populaire que le perce-oreille cherche à s'introduire dans l'oreille humaine pour percer le tympan et perforer le cerveau ou manger la cervelle est à l'origine de beaucoup de désignations romanes de l'insecte. [...] Selon Zahler (1898: 24-25) –cité par Poggi (1942: 200)–, à l'origine peut être le fait que les anciens croyaient que beaucoup de maladies, en particulier celles de l'oreille, étaient provoquées par des vers. [...] Les désignations gallo-romanes composées d'un verbe qui véhicule l'idée d'agression, pénétration (à l'origine «percer») auquel s'ajoute le nom «oreille» dateraient, selon le FEW, du XVI^e siècle. Elles auraient été motivées par la ressemblance de l'insecte avec l'instrument de l'orfèvre qui servait à percer le lobe de l'oreille pour y introduire les boucles d'oreille, très à la mode à l'époque. Plus tard, cette désignation aurait été réinterprétée en fonction de la croyance populaire dont il a déjà été parlé, et aurait ainsi permis que surgissent de nouvelles formes verbales, qui ne peuvent plus se rattacher à l'idée de percer le lobe de l'oreille, mais bien à l'idée d'agression à l'oreille interne.

RELAT_HOMME: o referente recibe o nome polo uso que del fai a especie humana (se ten, por exemplo, propiedades medicinais, etc.). No punto de fala francoprovenzal ITA406 (Faeto), por exemplo, os *Oniscidea* ou *cochinillas* reciben o nome de [palóttələ a kkundíj], literalmente, «pelotiña que serve para condimentar» porque se utilizaba para aderezar algúns pratos da cociña tradicional, segundo afirma Franconie (2009b: 111) na síntese románica:

Le parler fpr. de Faeto (ITA 406) présente la locution [palóttələ a kkundíj]. Le terme [palóttələ] 'petite balle' est emprunté aux parlers voisins. C'est le continuateur d'un «*palla» correspondant à l'its. «balla». [...] Mais c'est [kkundíj], le second terme de la locution, qui pose problème. L'hypothèse qui a été privilégiée ici est celle du rapport avec «condire» 'assaisonner', car le cloporte, en tant que remède, était ajouté à certains aliments (pain, pomme, différentes pâtes, etc.). Mais on pourrait aussi chercher une filiation du côté du continuateur de CONDERE 'éloigner de regards, cacher', en le rapprochant de la locution sarde [aβeriyúndʒa] 'ouvre-et-ferme' déjà citée. Mais encore, on peut considérer le cloporte comme une «petite bête des égouts», à partir du continuateur de CONDUCTU 'conduit, égot' (part. passé substantivé de CONDERE d'où it. «condotto»)⁴.

COMPORT_ANIM: o referente recibe o nome pola súa conduta respecto aos outros animais. A *Mustela nivalis* ou *donicela*, por exemplo, recibe o nome de [maɲnabodʒíne], literalmente «papagaliñas», no punto de fala italiana meridional

⁴ Nos casos nos que hai varias hipóteses etimolóxicas ou semánticas sobre un termo, o criterio seguido na base de datos é o de privilexiar sempre a opción preferida polo autor da síntese románica e transcribir, na columna 59, o parágrafo no que este deixa constancia das distintas hipóteses.

ITA083 (Rieti *et al.*) porque, tal como contan Alinei e Dell’Aquila (2009: 339) na síntese románica, «avant de choisir laquelle elle doit manger, la belette essaie de tuer toutes les poules qu’elle trouve dans le poulailler: d’où les désignations qui se retrouvent dans quelques points du Centre et du Sud de l’Italie».

COMPORT_NATURE: o referente recibe o nome pola súa conduta respecto aos outros elementos da natureza. A *Gryllotalpa gryllotalpa* ou *grilo ceboleiro*, por exemplo, recibe no punto de fala dacorromanesa ROU034 (Corbu) o nome de [koropíʃnít̪sə] ‘coropișniță’, literalmente «pequeno becho da remolacha», porque, segundo di Segura da Cruz (2001: 90-91) na síntese románica,

la voracité de la courtilière est bien connue des paysans d’un peu partout, les dégâts qu’elle fait aux cultures constituant pour eux un motif de grand souci (Dembry 1972). Les effets nuisibles sur les cultures dont elle coupe les racines ont donc motivé un grand nombre de désignations qui ne se présentent pas de façon uniforme, ni du point de vue de leur formation, ni du point de vue lexical. Elles ont, quand même, une caractéristique commune: la référence explicite à un produit cultivé. L’*oignon*, la *courge*, le *poireau*, la *pomme de terre*, le *chou*, le *concombre*, la *tomate*, la *betterave*, le *chou-fleur*, le *maïs*, le *blé* et encore (en prenant le contenant pour le contenu ou le tout pour la partie) le *pré* et le *jardin* font partie de l’inventaire des dénominations motivées par des dégâts produits, ce qui donne l’idée de l’étendue des ravages provoqués par cette petite bête, par le biais de la pluralité et de la diversité des types lexicaux trouvés. Malgré cette diversité, on dirait que chaque domaine érige un de ces produits en symbole des dévastations provoquées dans les cultures (Bec 1960, p. 321 et Pellegrini-Benincà 1969, p. 86) et on a *grosso modo* comme noyau d’une série de dénominations: l’*oignon* en Espagne, la *courge* en Italie et la *betterave* en Roumanie; en domaine occitan, l’*oignon* (dans la Provence et le Languedoc oriental), la *courge* et le *maïs* (en Gascogne) et encore le *poireau* (en Provence) se partagent la plupart des dénominations. Il serait intéressant de savoir ce qui détermine ce choix: il est peu probable que ce soit le choix alimentaire de la petite bête à chaque endroit, mais plutôt probablement la culture dominante ou l’intérêt du paysan.

AUTRES_PSYCHO: o referente recibe o nome polo seu comportamento característico, distinto dos que engloban as etiquetas anteriores. No punto de fala istriorromanesa ROU129 (Jeiani), por exemplo, o *Bufo bufo* ou *sapo común* recibe o nome de [ʒ̥ba ɣrumba], literalmente «sapo groseiro». Augusto (2009: 144) explica na síntese románica que esta denominación «est formado por le mot croate ‘zaba’ qui signifie ‘crapaud’ ou ‘grenouille’, modifié par l’adjectif ‘grubo’, toujours croate, qui veut dire ‘grosier’».

3. *Prácticas máxicas e relixiosas*

PRAT_MAGIQUE: o referente recibe o nome porque se cre que ten algún tipo de poder sobrenatural ou de relación con prácticas máxicas. Un dos animais cunha maior variedade de nomes relacionados coa feiticaría é a *Mantis religiosa* ou *barbantesa*, que, por exemplo, recibe o nome de [kavád̥ɖu de makáɾɛ], literalmente «cabalo das meigas», no punto de fala italiana meridional ITA121 (Gallipolli *et al.*). García Mouton (2001: 249-250) refírese a este tipo de denominacións aplicadas a este insecto na súa síntese románica (2.3.1):

Dénominations magico-religieuses: la mante peut accomplir la fonction d'animal qui incarne un aïeul, une âme, un dieu, qui sert d'intermédiaire entre l'homme (l'enfant dans les chants rituels des formulettes) et les forces de l'au-delà (Dieu-diable). Sa principale mission est celle de prier, et c'est pour cela que les croyances des cultures éloignées connaissent toutes l'image de la mante assimilée à une âme en peine (par ex., l'italo-roman «aneme de lu purgatorie») qui fait pénitence en priant ou à une servante des dieux, mais aussi du diable et des sorcières, ce qui rend évident son caractère maléfique. [...] Son caractère maléfique apparaît surtout dans l'Italie centre-méridionale où la désignation «sorcière» est fréquente.

Precisamente, unha das vantaxes da base de datos que se está a presentar é que permite analizar a distribución areal dunha determinada motivación semántica dende unha perspectiva global, de conxunto. Centrándonos na motivación PRAT_MAGIQUE, por exemplo, un pode preguntarse se as denominacións de animais relacionadas con prácticas relixiosas ou poderes sobrenaturais se circunscriben especialmente a algunhas áreas concretas, tal como acontece no caso da parraguesa, segundo di García Mouton.

Unha sinxela pescuda na base de datos –que, insistimos, aínda se está a revisar– amosa que, dos 667 ítems etiquetados coa motivación semántica PRAT_MAGIQUE, 310 ítems –o 46,4%– pertencen a variedades italianas. Tendo en conta que os puntos de enquisa en Italia supoñen soamente o 16,4% do corpus –son 171 puntos ou *casas* dun total de 1.037–, este dato demostra que a base de datos ten un potencial evidente para detectar unha maior densidade areal de certas motivacións semánticas no espazo lingüístico románico. Este tipo de pescudas, que só se poden levar a cabo se o corpus se analiza mediante técnicas globais, podería abrir, por exemplo, novas vías de colaboración entre lingüistas e antropólogos culturais.

RELIG_PRECHR: o referente recibe un nome relacionado con algún aspecto das relixións precristiás que se profesaban en Europa antes de que o cristianismo se

convertese na relixión hexemónica. No punto de fala sarda ITA145, por exemplo, a *bolboreta* recibe o nome de [mamajóla] ‘mamaiola’, literalmente «nai»; Contini (2009: 184) dá conta do que esta denominación significaba no pasado na síntese románica:

Depuis l’antiquité le papillon est considéré comme le symbole de l’âme et donc de l’immortalité. Comme tel il figure dans des bijoux étrusques ou sculpté dans les sarcophages (Riegler 1981). Cette croyance a survécu dans les traditions populaires des pays européens si l’on en croit les nombreux témoignages recueillis par les spécialistes. [...] Cette association du papillon à l’âme des défunts est à rapprocher des zoonymes de parenté qui peuvent être interprétés comme des survivances d’une religion totémique (Alinei 1984). Dans le papillon devenu animal-totem survivent surtout les ancêtres et, parmi ces derniers, le plus fréquemment évoqué est l’ancêtre féminin: la grand-mère ou la mère.

RELIG_MONOTH: o nome do referente ten que ver coa relixión monoteísta hexemónica na Europa moderna, isto é, a cristiá. En moitos puntos de fala andaluza, por exemplo, a *Mantis religiosa* ou *barbantesa* recibe o nome de [san̄taterésa] ou, nas zonas ceceantes, [θan̄tateréθa] ‘Santa Teresa’. Nomes semellantes atópanse por toda a Romania, como detalla García Mouton (2001: 251) na síntese románica:

Dans le processus de christianisation, les saints ont une place de choix, en particulier sainte Thérèse, très présente dans les domaines castillans et les régions environnantes. [...] Des saints, on passe facilement au clergé. Nous avons déjà vu que «celle qui prie» apparaît en cast. et en and., liée à l’idée exprimée par les «louva-a-deus» et «pregadéu». [...] Son rapprochement de la prière et du culte est institutionnalisé dans les cas de «nonne». [...] Les formes cast. [kúra] ‘curé’, l’occ. [kapelãñ] ‘aumônier’ et l’its. [prét] ‘curé’ sont à inclure ici, bien qu’elles puissent être également des restes de la fragmentation d’un syntagme.

4. Antroponimia

MÉTIIERS: o referente recibe o nome pola semellanza entre o seu comportamento e as accións vinculadas a algún oficio, por paralelismo entre a súa morfoloxía e as ferramentas que se empregan nalgún oficio, etc. No punto de fala occitana FRA277 (Saissac *et al.*), por exemplo, o *Ergates faber* ou *ferreiro* recibe o nome de [fãβre] ‘fabre’, literalmente «ferreiro». Dalbera (1996: 127-128) conta na síntese románica que esta é unha denominación moi frecuente para estes insectos en toda a Romania:

Au premier plan figure le latin FABER FERRARIUS ‘l’artisan qui travaille le fer’. C’est le type statistiquement le mieux représenté puisqu’il est présent dans 60% environ des points (ou cases) romans.

Il s’agit là d’une formation lexicale de modèle courant. En latin, FABER désigne tout ouvrier qui travaille les corps durs (métaux, pierre, bois, ivoire, etc.) et le sens est généralement précisé soit par une épithète, soit simplement par le contexte; [...] mais FABER peut fort bien également s’employer seul pour désigner l’ouvrier en bois ou en fer. Par ailleurs, l’usage de l’adjectif substantivé pour les désignations de métiers est déjà banal en latin: [...] ‘le forgeron’ est FERRARIUS sous la plume de Sénèque. Ce type FABER FERRARIUS donne donc naissance en roman, assez naturellement, à deux séries distinctes selon que c’est le déterminé FABER ou le déterminant FERRARIUS qui se trouve continué.

GÉNÉRIQUES_ANTHROP: o referente recibe un nome xenérico relacionado cun apelativo xeralmente aplicado a persoas. Por exemplo, no punto de fala sarda ITA152 (Mògoro *et al.*) a *Mantis religiosa* ou *barbantesa* recibe o nome de [seɲɲoredd̪a] ‘segnoredda’, literalmente «señorita».

PRÉNOMS: o referente recibe un nome propio de persoa. Por exemplo, no punto de fala d’oïl FRA035 (Vendrest *et al.*) o *Passer domesticus* ou *pardal* é coñecido como [pjero] ‘Pierrot’.

PARENTÉ: o referente recibe un nome de parentesco. Na localidade andaluza ESP128b (Artajate), por exemplo, a *Mustela nivalis* ou *donicela* recibe o nome de [komaðreha] ‘comadreja’, que vén da base latina COMMATER ‘comadre’, que é o apelativo que se lle dá ou ben á madriña dunha persoa con relación ao padriño ou aos pais desta, ou ben á nai dunha persoa con relación ao padriño ou madriña desta.

AUTRES_ANTHROP: o referente recibe un nome que polo xeral se aplica a persoas e que non pertence a ningunha das categorías anteriores. No punto de fala dacorromanesa ROU050 (Cârțișoara), o *Cimex lectularius* ou *chincha* recibe o nome de [fr̪wə puturoásə] ‘fr̪uă puturoasă’, literalmente «muller alemá que cheira mal».

5. Nomes de animais (ou partes do corpo dos animais)

ANIM_DOM: o referente recibe, a causa das semellanzas que comparte con el, o nome dun animal doméstico. No punto de fala dacorromanesa MOL614 (Batir), os *anisópteros* ou *libélulas* reciben o nome de [kálu p̪opi] ‘calul popii’, literalmente «o cabalo do cura».

ANIM_SAUUV: o referente recibe, a causa das semellanzas que comparte con el, o nome dun animal salvaxe. No punto de fala grega ITA502 (Ghorio *et al.*), os *grilídeos* ou *grilos* reciben o nome de [stáλαχο], do grego clásico ἀττάλαβος, co que tamén se designan os saltóns. Gargallo e Veny (2001: 203) indican na síntese románica que isto é frecuente porque «ces insectes «se hallan muy próximos a los verdaderos saltamontes de antenas largas» et sautent aussi, comme les sauterelles, mais avec des bonds plus courts (Zahradnik-Chvála 1990, 60)».

6. *Nomes de plantas (ou partes dunha planta)*

NOMS_PLANTES: o referente recibe o nome por metonimia con algunha planta. No punto de fala grega ITA502 (Ghorio *et al.*), por exemplo, a *abella* é coñecida co nome de [melíssi], literalmente «melisa», como sinala Médélice (2001: 7) na síntese románica:

L'ensemble de la Romania en connaît aujourd'hui une forme moderne pour désigner une plante, la mélisse, que les abeilles apprécient particulièrement. Cet emprunt au grec a éliminé le lat. APIASTER [*apiastrum* chez Pline] qui traduisait la même image.

7. *Nomes de enfermidades*

NOMS_MALA: o referente recibe un nome que en sentido recto se aplica a unha enfermidade. Por exemplo, no punto de fala francoprovenzal FRA159 (Davayé *et al.*) os *Oniscidea* ou *cochinillas* reciben o nome de [grənuzéla] 'grenuzèla', que é o nome que se lle dá tamén á candidiase oral.

8. *Nomes de obxectos*

OBJETS: o referente recibe un nome que en sentido recto se aplica a un obxecto. No punto de fala catalá CAT221 (Ontinyent), por exemplo, os *lampíridos* ou *vagalumes* reciben o nome de [luéerna] 'lluerna', que en sentido recto é o nome que se lles dá ás claraboias das casas.

9. *Referencias temporais*

RÉFÉRENCE_TEMP: o referente recibe un nome que en sentido recto se aplica a unha referencia temporal ou é determinado por algunha referencia temporal.

No punto de fala arromanesa ROU120 (Stejaru), por exemplo, o *Turdus merula* ou *merlo* recibe o nome de [púl di primvǵára], literalmente «paxaro da primavera», xa que, como di González (2009: 392) na síntese románica, «le printemps est l'époque où cet oiseau a la plus grande activité. [...] Il n'est donc pas surprenant que les références au printemps et au mois de mars interviennent dans certaines des dénominations».

10. Outras motivacións

AUTRES_MOTIV: o nome do referente non se debe a ningunha das motivacións anteriores. Por exemplo, os nomes que recibe o *enxame* nalgúns puntos de fala italiana setentrional, [stʃápo de áve], [róz d ávi] ou [bódʒ d ávi], literalmente «nobelo de abellas», «font appel à l'image des insectes agglutinés», segundo Contini e Carpitelli (2009: 165).

A xeito de resumo, a seguinte táboa recolle as distintas motivacións semánticas tidas en conta na base de datos en función da súa rendibilidade no corpus analizado ata hoxe:

ANIM_DOM	1.357	NOMS_PLANTES	718
ANIM_SAU	12.780	NOURRITURE	286
AUTRES_ANIM	69	OBJETS	1.313
AUTRES_MOTIV	854	ODEUR	409
AUTRES_PHYSIQ	1.000	ONOMATOP	1.688
AUTRES_PSYCHO	42	OPAQUE	1.059
CHANT	105	PARENTÉ	288
COMPORT_ANIM	3.237	PARTIES_REF	3.385
COMPORT_HOMME	140	PERDUE	1.498
COMPORT_NATURE	236	PHONSYMBOL	349
COULEUR	5.498	PRAT_MAGIQUE	666
DÉPLACEMENT	129	PRÉNOMS	125
GÉNÉRIQUES_ANTHROP	223	RÉFÉRENCE_TEMP	4.533
HABITAT	2.939	RELAT_HOMME	1
MÉTIER	731	RELIG_MONOTH	546
MOTRICITÉ	1.733	RELIG_PRECHR	162
NOMS_MALAD	1	TAILLE	248

Táboa 2. Rendibilidade das motivacións semánticas –por orde alfabética– no corpus

Como se pode observar, hai unha gran diferenza entre as motivacións máis rendibles –ANIM_SAU, COULEUR, RÉFÉRENCE_TEMP– e as menos rendibles –NOMS_MALAD, RELAT_HOMME. Porén, como ata o momento soamente se analizaron 50 dos 592 conceptos do cuestionario lexical –e hai varios ítems aos que aínda non se lles asignou ningunha motivación semántica–, parécenos que aínda é cedo para tirar conclusións destas diferenzas.

- ab) Columna 55: pode conter un ou varios etnotextos –unidades fraseolóxicas, proverbios, lendas, etc.– que xiran arredor do referente analizado.
- ac) Columna 56: contén o número do parágrafo da síntese románica na que se trata a secuencia fónica estudada.
- ad) Columna 57: contén o nome do autor ou autores da síntese románica na que se trata a secuencia fónica estudada.
- ae) Columna 58: contén o título da síntese románica na que se trata a secuencia fónica estudada.
- af) Columna 59: contén a explicación que se dá na síntese románica verbo da secuencia fónica estudada.
- ag) Columna 60: contén o símbolo co que se cartografou a secuencia fónica estudada no correspondente mapa do ALiR.

4. UN NOVO RECURSO AO SERVIZO DA ROMANÍSTICA. POSIBILIDADES DE EXPLOTACIÓN

Neste artigo vén de se presentar unha nova ferramenta ao servizo da romanística: unha base de datos que hoxe xa conta con máis de 57.000 rexistros e que pon ao dispor de calquera investigador non só a información diatópica que é a razón de ser do ALiR, senón tamén o corpus analítico complementario íntegro –etimolóxico, etnográfico, semántico e bibliográfico– que os lingüistas que teñen traballado no proxecto foron construíndo durante a elaboración das sínteses nacionais, primeiro, e das sínteses románicas, despois. Esta base de datos, que segue a se elaborar no Centro Ramón Piñeiro para a Investigación en Humanidades, vai ser de acceso libre no enderezo www.atlaslinguistiqueroman.wordpress.com e abre a porta a comezar ou retomar numerosas liñas de investigación no ámbito da romanística:

1. No eido da xeolingüística, unha base de datos como a que se presenta é un material óptimo para se achegar á distribución xeolectal do ámbito lingüístico románico dende unha perspectiva dialectométrica. Unha análise cuantitativa deste tipo

permitiría non só examinar por primeira vez –que teñamos constancia– a variación e a distancia lingüística existentes entre a totalidade das variedades románicas⁵, senón que tamén permitiría, como resultado destas análises, afrontar con novos datos varios temas que teñen xerado controversia ao longo da historia da romanística, a saber: o grao de individuación da lingua galega respecto do portugués e do castelán –unha cuestión que entronca directamente co debate sobre a codificación do galego, hoxe aínda non pechado–; o debate sobre a subagrupación románica da lingua catalá, que Radatz (2012) reabriu recentemente; ou o debate sobre o grao de proximidade que hoxe seguen a manter entre si as variedades friulanas, ladinas e retorrománicas, históricas integrantes dun antigo contínuum lingüístico hoxe moi fragmentado. Malia que se veñen de presentar os primeiros resultados da dialectometrización do ALiR no CRPIH –*vid.* Valls e González (pendente de aceptación)–, a multiplicidade de sistemas de análise e de cuestións que se poden afrontar dende esta perspectiva cuantitativa animámonos a pensar que a análise dialectométrica do ALiR aínda pode dar moitos máis froitos. Por esta razón os materiais necesarios para a súa dialectometrización tamén se poden consultar e descargar libremente no enderezo www.atlaslinguisticuroman.wordpress.com⁶.

2. No eido da lingüística histórica, esta base de datos é unha ferramenta ideal para contrastar a evolución de determinados sons ou morfemas non só entre as variedades que teñen sido con máis frecuencia obxecto de estudo desta disciplina –aquelas máis faladas ou faladas en extensións xeográficas maiores, ademais das variedades estándar–, senón tamén entre estas variedades e outras de ámbito xeográfico ou demográfico máis reducido. Unha operación tan sinxela como a de seleccionar, por exemplo, todas as palabras co sufixo diminutivo feminino -ĪTTA permite contrastar automaticamente as solucións ás que deu lugar este sufixo latino naquelas variedades románicas –das 1.037 dispoñibles no corpus– que presentan algunha palabra

⁵ Malia o inxente traballo de dialectometrización das falas románicas levado a cabo por Goebel dende a década dos anos 70 –*vid.*, por exemplo, Goebel (2010) para unha introdución aos fundamentos teóricos da Escola Dialectométrica de Salzburgo–, a ausencia dun corpus panrománico ata a elaboración do ALiR obrigouno a basear os seus traballos en datos procedentes de atlas de ámbito xeográfico máis reducido, como o *Atlas linguistique de la France* (Goebel 2003) ou o *Atlante Italo-Svizzero* (Goebel 2008). Os traballos deste autor son, porén, unha referencia obrigada para calquera investigador interesado no estudo da variación e a distancia lingüística na Romanía dende unha perspectiva dialectométrica.

⁶ En Valls e González (pendente de aceptación) as interdistancias entre variedades románicas calculáronse mediante a aplicación da *distancia de Levenshtein*, a medida de distancia fonética arredor da que pivota a interface Gabmap do Center for Language and Cognition da Universidade de Groningen –de acceso libre no enderezo www.gabmap.nl.

derivada dun étimo latino con este sufixo. Non fai falla insistir nas posibilidades que ofrece este tipo de busca: pódense analizar os empréstitos do eslavo no dacorrománés ou os do alemán no rético; pódese analizar a adaptación de bases de substratos non románicos –vasco, bretón, xermánico, etc.– ás leis fonéticas e á morfoloxía das distintas variedades románicas etc.

3. No eido da semántica, esta base de datos ofrece un material valioso para aqueles investigadores interesados nas motivacións semánticas que subxacen nas denominacións de moitos referentes. Coa sinxela operación de seleccionar a categoría «ONOMATOP» na columna 54, por exemplo, un pode contrastar que variedades románicas son máis proclives ou máis refractarias a utilizaren designacións de base onomatopeica para denominar determinados elementos da natureza, por exemplo; ou un pode analizar o papel da relixión na zoonimia ao longo da xeografía románica, por citar só outro exemplo.

Estes son soamente tres exemplos das posibilidades que ofrece a base de datos informatizada que, a partir dos materiais do ALiR, se está a desenvolver no CRPIH. Malia que o proxecto aínda se atopa nunha fase inicial, o noso obxectivo é incorporarmos co tempo toda a información contida nas sínteses nacionais e nas sínteses románicas para xerarmos, a partir da análise íntegra do cuestionario do ALiR, unha gran base de datos con información fonética, diatópica, etimolóxica, etnográfica, semántica e mais bibliográfica das distintas falas que compoñen o contínuum románico.

BIBLIOGRAFÍA

- AIS = Jaberg, Karl e Jakob Jud (1928-1940): *Atlante Italo-Svizzero*. Zofingen: Rieger & Co.
- ALC = Griera, Antoni (ed.) (1923-1964): *Atlas lingüístic de Catalunya*. Barcelona: Institut d'Estudis Catalans-La Polígrafa.
- ALDC = Veny, Joan e Lúdia Pons (eds.) (2001-2006): *Atles lingüístic del domini català*. Barcelona: Universitat de Barcelona-Institut d'Estudis Catalans.
- ALEA = Alvar, Manuel (ed.) (1961-1973): *Atlas lingüístico y etnográfico de Andalucía*. Granada: Universidad de Granada-CSIC.
- ALEANR = Alvar, Manuel (ed.) (1979-1983): *Atlas lingüístico y etnográfico de Aragón, Navarra y Rioja*. Madrid: Departamento de Geografía Lingüística-Institución Fernando El Católico de la Excma. Diputación Provincial de Zaragoza-CSIC.

- ALECAN^T = Alvar, Manuel (ed.) (1995): *Atlas lingüístico y etnográfico de Cantabria*. Madrid: Arco Libros.
- ALECMAN = García Mouton, Pilar e Francisco Moreno Fernández (eds.) (2003): *Atlas lingüístico y etnográfico de Castilla-La Mancha*. Alcalá de Henares: Universidad de Alcalá.
- ALEIC = Bottiglioni, Gino (ed.) (1933-1942): *Atlante linguistico etnografico italiano della Corsica*. Pisa.
- ALEICAN = Alvar, Manuel (ed.) (1975-1978): *Atlas lingüístico y etnográfico de las Islas Canarias*. Las Palmas de Gran Canaria: Ediciones del Excmo. Cabildo Insular de Gran Canaria.
- ALEPG = *Atlas lingüístico-etnográfico de Portugal e da Galiza*. Lisboa: Centro de Linguística da Universidade de Lisboa (en preparación).
- ALEP = Alvar, Manuel (ed.): *Atlas Lingüístico de España y Portugal* (en preparación).
- ALF = Gilliéron, Jules e Edmond Edmont (1902-1910): *Atlas linguistique de la France*. Paris: Honoré Champion.
- ALFCORSE = Gilliéron, Jules e Edmond Edmont (1914-1915): *Atlas linguistique de la France. Corse*. Paris: Honoré Champion.
- ALGA = Instituto da Lingua Galega (1990-): *Atlas Lingüístico Galego*. A Coruña: Fundación Pedro Barrié de la Maza.
- ALI = Bartoli, Matteo; Terracini, Benvenuto; Vidossi, Giuseppe; Grassi, Corrado; Genre, Arturo e Lorenzo Massobrio (eds.) (no prelo): *Atlante Linguistico Italiano*. Roma: Istituto Poligrafico e Zecca dello Stato.
- ALIA = Dahmen, Wolfgang e Johannes Kramer (eds.) (1985-1992): *Aromunischer Sprachatlas. Atlasul lingvistic aromân*. Hamburg: H. Buske Verlag.
- ALINEI, MARIO E VITTORIO DELL'Aquila (2009): «Les désignations romanes de la belette», in ALiR IIb, pp. 319-357.
- ALiR I = Tuailon, Gaston; Contini, Michel *et al.* (eds.) (1996): *Atlas Linguistique Roman (ALiR). Volume I Présentation; Atlas Linguistique Roman (ALiR). Volume I Cartes; Atlas Linguistique Roman (ALiR). Volume I Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.
- ALiR IIA = Veny, Joan; Contini, Michel *et al.* (eds.) (2001): *Atlas Linguistique Roman (ALiR). Volume IIA Cartes; Atlas Linguistique Roman (ALiR). Volume IIA Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.
- ALiR IIb = Veny, Joan; Contini, Michel *et al.* (eds.) (2009): *Atlas Linguistique Roman (ALiR). Volume IIb Cartes; Atlas Linguistique Roman (ALiR). Volume IIb Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.

- ALM = Udler, Rubin; Comarnițchi, Victor; Melnic, Vasile e Vasile Pavel (eds.) (1968-1973): *Atlasul Lingvistic Moldovenesc*. Chișinău.
- ALPI = *Atlas Lingüístico de la Península Ibérica*. Madrid: CSIC, 1962.
- ALRR BAS = Pavel, Vasile (ed.) (1993): *Atlasul Lingvistic Român pe regiuni Basarabia, Bucovina de nord, Transnistria*. Chișinău: Știința.
- ALRR MAR = Neiescu, Petru; Rusu, Grigore e Ionel Stan (eds.) (1969-1973): *Atlasul lingvistic român pe regiuni. Maramureș*. Bucarest: Editura Academiei Române.
- ALRR TRANS = Rusu, Grigore; Loșonți, Dumitru e Bidian, Viorel (ed.) (1992-): *Atlasul lingvistic român pe regiuni. Transilvania I*. Bucarest: Editura Academiei Române.
- ALSAR = Terracini, Benvenuto e Temistocle Franceschi (ed.) (1964): *Saggio di un Atlante linguistico della Sardegna*. Torino: Istituto dell'Atlante Linguistico Italiano.
- ALW = Remacle, Louis; Legros, Elisée; Lechanteur, Jean e Marie-Thérèse Counet (eds.) (1953-1987): *Atlas linguistique de la Wallonie*. Liège: H. Vaillant Carmanne.
- ASLEF = Pellegrini, Giovan Battista (ed.) (1972-1986): *Atlante Storico-linguistico etnografico friulano*. Padova-Udine: Istituto di Glottologia e Fonetica dell'Università di Padova-Istituto di Filologia Romanza della Facoltà di Lingue e Letterature Straniere dell'Università di Udine.
- AUGUSTO, MARIA CELESTE (2009): «Les désignations romanes du crapaud», in ALiR I**b**, pp. 137-160.
- CARRILHO, ERNESTINA E MARIA LOBO (2001): «Les désignations romanes du perce-oreille», in ALiR I**a**, pp. 405-449.
- CONTINI, MICHEL (1992): «L'Atlas Linguistique Roman», *IKER* 7, pp. 339-356.
- CONTINI, MICHEL (2009): «Les désignations romanes du papillon», in ALiR I**b**, pp. 179-213.
- CONTINI, MICHEL E ELISABETTA CARPITELLI (2009): «Les désignations romanes de l'essaim», in ALiR I**b**, pp. 161-177.
- DALBERA, JEAN-PHILIPPE (1996): «Les désignations romanes du forgeron», in ALiR I, pp. 127-139.
- DALBERA, JEAN-PHILIPPE (2001): «Les désignations romanes de orvet», in ALiR I**a**, pp. 377-404.
- DRG = *Dicziunari Rumantsch Grischun*, I-III, Coire: Bischofgerger, 1939-1963; IV, Winterthur: Stamp. Winterthur, 1964 ff.
- FIDALGO, ELVIRA E XOSÉ AFONSO ÁLVAREZ (2009): «Les désignations romanes du bousier», in ALiR I**b**, pp. 46-78.

- FRANCONIE, HÉLÈNE (2009a): «Les désignations romanes de la blatte et du ténébrion», in ALiR Iib, pp. 1-44.
- FRANCONIE, HÉLÈNE (2009b): «Les désignations romanes du cloporte», in ALiR Iib, pp. 93-135.
- GARCÍA MOUTON, PILAR (2001): «Les désignations romanes de la mante religieuse», in ALiR Iia, pp. 239-280.
- GARGALLO, JOSÉ ENRIQUE E JOAN VENY (2001): «Les désignations romanes du grillon», in ALiR Iia, pp. 201-217.
- GOEBL, HANS (2003): «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics XXV*, pp. 59-121.
- GOEBL, HANS (2008): «La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati», *Revue de Linguistique Romane 72* (285-286), pp. 25-73.
- GOEBL, HANS (2010): «Introducción a los problemas y métodos según los principios de la Escuela Dialectométrica de Salzburgo (con ejemplos sacados del 'Atlante Italo-Svizzero', AIS», in Gotzon Aurrekoetxea e José Luis Ormaetxea (eds.): *Tools for Linguistic Variation. Anuario de Filología Vasca «Julio Urquijo» XIII*. Bilbao: Universidad del País Vasco, pp. 3-40.
- GONZÁLEZ, MANUEL (2009): «Les désignations romanes du merle», in ALiR Iib, pp. 389-402.
- GPSR = Gauchat, Louis; Jeanjaquet, Jules e Ernest Tappolet (eds.) (1924): *Glossaire des Patois de la Suisse romande*. Neuchâtel-Paris: Victor Attinger.
- MÉDÉLICE, JEANINE ELISA (2001): «Les désignations romanes de l'abeille», in ALiR Iia, pp. 3-19.
- MÉDÉLICE, JEANINE ELISA (2009): «Les désignations romanes du têtard», in ALiR Iib, pp. 215-253.
- NALC = Dalbéra-Stefanaggi, Marie-José (ed.): *Nouvel Atlas Linguistique de la Corse* (en preparación).
- NALF = Dauzat, Albert (1942): *Le nouvel atlas linguistique de la France par régions. Avec trois cartes linguistiques*. Luçon: Imprimerie S. Pacteau.
- NALR BAN = Neiescu, Petru; Beltechi, Eugen; Faiciuc, Ioan e Nicolae Mocanu (eds.) (1980): *Noul atlas lingvistic român pe regiuni. Banat*. Bucarest: Academiei.
- NALR MOLD BUCOV = Arvinte, Vasile; Dumistrăcel, Stelian; Florea, Ioan; Nuță, Jon e Adrian Turtculeț (eds.) (1987): *Noul atlas lingvistic român pe regiuni. Moldova și Bucovina*. Bucarest: Academiei.
- NALR OLT = Cazacu, Boris; Teaha, Teofil; Ionică, Ion e Valeriu Rusu (eds.) (1967-1990): *Noul atlas lingvistic român pe regiuni. Oltenia*. Bucarest: Academiei.

- RADATZ, HANS-INGO (2012): «Per què els elements gal·loromànics fan ‘heavy’ en català. Arran del clàssic debat sobre la subagrupació del català», *eHumanista / IVITRA* 2, pp. 202-218.
- SEGURA DA CRUZ, LUISA (2001): «Les désignations romanes de la courtilière», in *ALiR* IIa, pp. 89-144.
- TP = Gauchat, Louis; Jeanjaquet, Jules e Ernest Tappolet (eds.) (1925): *Tableaux phonétiques des patois suisses romands*. Neuchâtel: P. Attinger.
- ULLMANN, STEPHEN (1972): *Semántica. Introducción a la ciencia del significado*. Madrid: Aguilar.
- VDSI = *Vocabolario dei dialetti della Svizzera italiana* (en preparació).
- VALLS, ESTEVE E MANUEL GONZÁLEZ (pendente de aceptació): «Variació i distància lingüística a la *Romania Antiqua*. Una aproximació dialectomètrica a l'*Atlas Linguistique Roman* (ALiR)».
- VENY, JOAN E JOÃO SARAMAGO (2001): «Les désignations romanes du moustique», in *ALiR* IIa, pp. 359-376.
- WILD MEGL = Beate, Wild (ed.) (1983): *Meglenorumänischer Sprachatlas*. Hamburg: H. Buske Verlag.

