

Variación e distancia lingüística na Romania Antiga: unha contribución dialectométrica ao debate sobre o grao de individuación da lingua galega

Esteve Valls Alecha

Universitat de Barcelona / Centro Ramón Piñeiro para a Investigación en Humanidades
e.valls@ub.edu

Manuel González González

Instituto da Lingua Galega – Universidade de Santiago de Compostela / Centro Ramón Piñeiro para a Investigación en Humanidades
manuel.gonzalez.gonzalez@usc.es

Recibido o 02/03/2016. Aceptado o 23/06/2016

Linguistic variation and distance in the Romania Antiga: A dialectometric contribution to the debate about the degree of individuality of the Galician language

Resumo

O obxectivo deste traballo é presentar unha primeira aproximación dialectométrica á organización xeolectal do ámbito lingüístico románico en base a un corpus de nova creación que se está a desenvolver no Centro Ramón Piñeiro para a Investigación en Humanidades a partir dos datos do *Atlas Linguistique Roman (ALiR)*. Alén de permitir analizar a variación e a distancia lingüística entre as variedades da Romania Antiga, o proceso de dialectometrización do *ALiR* debería permitir afrontar, dende unha perspectiva cuantitativa global, determinadas cuestións que teñen xerado controversia ao longo da historia da romanística. Concretamente, este artigo céntrase na cuestión do grao de individuación da lingua galega respecto do portugués e do castelán —unha controversia que entronca directamente co debate sobre a codificación da lingua, hoxe en día aínda non pechado de todo— e chega á conclusión de que reforzar a presenza e a visibilidade dos elementos comúns co portugués durante a planificación do corpus podería repercutir moi positivamente na matriz de trazos valorativos asociada a esta lingua.

Palabras chave

Romanística, dialectoloxía, dialectometría, *Atlas Linguistique Roman*

Sumario

1. Introducción. 2. Corpus. 3. Metodoloxía. 4. A posición do galego no ámbito iberorrománico. 4.1. Un tema menor para a romanística. 4.2. Novas perspectivas baseadas en técnicas de análise cuantitativas. 5. Resultados. 6. Conclusións.

Abstract

This paper presents a first dialectometric approach to the geolinguistic distribution of the Romance varieties using a new corpus based on the *Atlas Linguistique Roman (ALiR)* which is currently being developed at the Centro Ramón Piñeiro para a Investigación en Humanidades. The dialectometrization of the *ALiR* will not only allow us to analyze linguistic distance among the varieties of the Romania Antiga, but may also shed light from a quantitative perspective on several issues that have historically aroused controversy. Specifically, this paper deals with the degree of individuality of the Galician language with respect to Portuguese and Spanish (a controversy closely related to the still ongoing debate about the standardization of Galician), and concludes that reinforcing the presence and visibility of elements shared with Portuguese in corpus planning may have a positive effect on its prestige.

Keywords

Romance studies, dialectology, dialectometry, *Atlas Linguistique Roman*

Contents

1. Introduction. 2. Corpus. 3. Methodology. 4. The position of Galician among the Ibero-Romance languages. 4.1. A minor issue for Romance Studies. 4.2. New perspectives based on quantitative techniques. 5. Results. 6. Conclusions.

Este traballo inscribese no proxecto “Descrición i interpretación de la variación lingüística: aspectos fónicos i morfolóxicos del català i altres llengües romàniques (DIVaL)” (FFI2013-46987-C3-1-P), financiado polo MICINN e o FEDER, e foi posíbel grazas a unha axuda do Centro Ramón Piñeiro para a Investigación en Humanidades da Secretaría Xeral de Política Lingüística da Xunta de Galicia.

1. INTRODUCCIÓN

Malia a análise dialectométrica do *Atlas Linguistique Roman* estar entre os obxectivos iniciais deste proxecto —v. Contini (1992: 345)—, este artigo constitúe —que saibamos— a primeira análise cuantitativa global que se leva a cabo a partir dun corpus de datos que abrangue a totalidade da Romanía Antiqua. Até o de agora, as análises dialectométricas de corpus de datos románicos proviñeron, sobre todo, da Escola Dialectométrica de Salzburgo (EDS), encabezada por Hans Goebel, que empregou o programa *Visual Dialectometry* (VDM) para analizar a variación e a distancia lingüística entre variedades a partir dun número considerábel de atlas lingüísticos do ámbito románico, como o *Atlas Linguistique de la France* (Goebel 2003) ou o *Atlante Italo-Svizzero* (Goebel 2008)¹. A diferenza dos traballos mencionados, porén, neste artigo abórdase a análise da variación e a distancia lingüística entre as variedades románicas a partir dun subcorpus que contén datos de *todos* os atlas lingüísticos —publicados ou inéditos— elaborados até o momento con datos destas falas. Este subcorpus foi deseñado e estase a desenvolver no Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH).

O primeiro obxectivo deste traballo é, xa que logo, presentar unha aproximación dialectométrica á distribución xeolectal do ámbito lingüístico románico. Ademais, enfróntase un dos temas que teñen xerado certa controversia dentro da romanística: o do grao de individuación da lingua galega respecto do portugués e do castelán —unha controversia que entronca directamente co debate, aínda non pechado de todo, sobre a codificación do galego.

Finalmente, e dende un punto de vista estritamente metodolóxico, este artigo quere contribuír á valorización da dialectometría como un conxunto de sistemas de análise necesarios para se aproximar, dende unha óptica cuantitativa global, a cuestións a miúdo tratadas cun enfoque cualitativo. Esta análise será posíbel grazas á aplicación da *distancia de Levenshtein* mediante a interface *Gabmap*, desenvolvida no *Center for Language and Cognition* da Universidade de Groninga.

2. CORPUS

O *ALiR* constitúe o primeiro intento de levar a cabo unha análise global da Romanía Antiqua a partir da síntese de todos os atlas lingüísticos —publicados ou aínda inéditos— do ámbito románico. Trátase dun atlas de segunda xeración, interpretativo: o *ALiR* non utiliza datos novos recollidos ad hoc mediante a realización de novas enquisas dialectais, senón que parte dos datos obtidos dunha cincuenta de atlas lingüísticos que se confeccionaron ao longo do século XX nos distintos estados, rexións ou ámbitos lingüísticos do espazo románico².

Este traballo baséase nun subcorpus do *ALiR* extraído da base de datos que se está a elaborar actualmente no Centro Ramón Piñeiro para a Investigación en Humanidades (CRPIH) para poñer ao dispor do público, dunha forma áxil e en acceso aberto, toda a información —etimolóxica, dialectal, etnográfica etc.— que conteñen as *Sínteses Románicas* que os colaboradores do *ALiR* foron publicando até o momento —v. Tuaille / Contini (1996) e Veny / Contini (2001, 2009). Nestas sínteses, un ou varios autores analizan as designacións románicas dun determinado concepto a partir das informacións proporcionadas polos comités responsábeis de xestionar os datos dos distintos ámbitos estatais ou lingüísticos. Concretamente, o subcorpus a partir do que se realizou a análise dialectométrica consta de 36.808 ítems correspondentes ás designacións románicas de 48 conceptos nun total de 1.010 puntos de enquisa. Os conceptos escollidos son os que xa se incorporaron na base de datos do CRPIH³: *abeille, alouette, araignée, aujourd'hui, belette, berceau, blatte et ténébrion, bousier, chenille, cigale, cloporte, courtilière, crapaud, demain, essaïm, foie, forgeron, fourmi, fourmilière, grillon, guêpe, hêtre, hier, libellule,*

¹ O traballo realizado por Goebel e os seus colaboradores na Escola Dialectométrica de Salzburgo durante as últimas décadas foi inxente e queda fóra das nosas posibilidades repasalo neste traballo. No entanto, pódese atopar unha introdución xeral aos principios da EDS en Goebel (1991, 2006) ou no enderezo www.dialectometry.com [consulta: 27/01/2016].

² V. Tuaille / Contini (1996) para saber cales son os atlas de primeira xeración dos que parte o *ALiR*.

³ Esta base de datos, así como todos os materiais precisos para a súa dialectometrización, pódense descargar no enderezo www.atlaslinguistiqueroman.wordpress.com [consulta: 27/01/2016].

lundi, mante religieuse, mardi, merle, mille-pattes, miroir, moineau, moucheron, moustique, orvet, paille, papillon, perce-oreille, punaise des lits, rouge-gorge, salamandre, serpent, taureau, têtard, toile d'araignée, tortue, ver de terre, ver luisant. En total, o subcorpus deste traballo consta de 273.499 segmentos fonéticos, dos cales 136 son únicos.

Como se desprende da figura 1, o número de respostas dispoñíbeis por localidade varía sensiblemente en función do concepto. Esta figura amosa a distribución dos 36.808 ítems nas 1.010 localidades enquisadas. As localidades cun maior número de ítems dispoñíbeis para o cálculo de interdistancias coloreáronse con tonalidades de azul escuro; pola contra, as localidades cun menor número de ítems dispoñíbeis presentan tonalidades de azul máis claro.

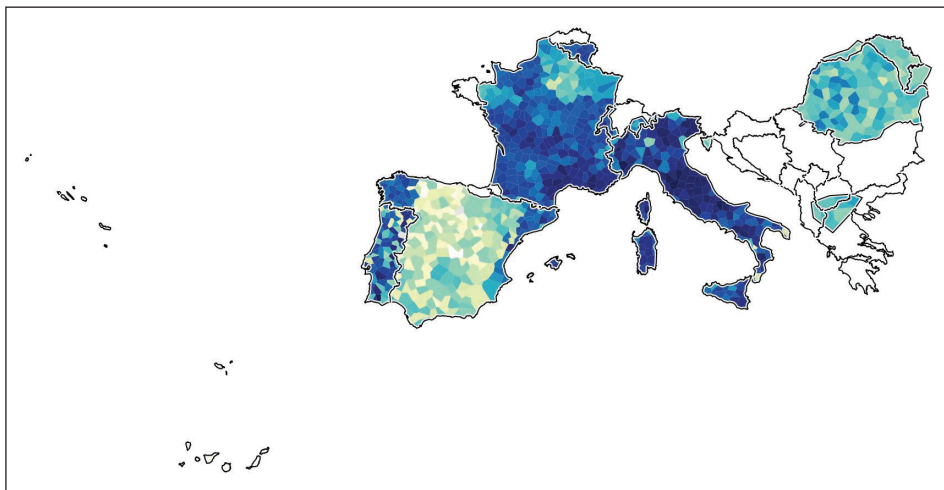


Figura 1. Distribución dos 36.808 ítems do corpus nas 1.010 localidades enquisadas⁴.

Para comprender esta figura cómpre ter en conta tres feitos: por unha banda, que se desestimaron 27 puntos de enquisa dos 1.037 dos que consta o *ALiR* porque practicamente non se dispuña de ningún dato sobre as denominacións dos ítems seleccionados nesas poboacións. Por outra banda, que se mantiveron varias localidades dos ámbitos lingüísticos asturleonés, castelán, aragonés e mais romanés que presentan unha cantidade moi pequena de datos para evitar que a rede de localidades enquisadas nestes ámbitos lingüísticos fose demasiado desequilibrada respecto doutros ámbitos que contan con redes de localidades máis densas. A exiguua información dispoñíbel sobre os mencionados ámbitos lingüísticos vai limitar, por forza, a posibilidade de analizarmos a súa variación interna. E, finalmente, cómpre ter en conta que, nos ámbitos lingüísticos do galego e das variedades galorrománicas e itallorrománicas, os puntos de enquisa non representan localidades, senón pequenas agrupacións de localidades lingüísticamente moi próximas chamadas *casas* —ou, en francés, *cases*.

3. METODOLOXÍA

O cálculo da distancia fonética entre variedades que está na base das análises posteriores levouse a cabo mediante a aplicación da chamada *distancia de Levenshtein* (LD). A distancia de Levenshtein é unha medida de cálculo da distancia fonética entre dúas liñas de datos. Para determinar esta distancia, o algoritmo de Levenshtein busca cal é o menor conxunto de operacións básicas preciso para transformar unha liña de segmentos fonéticos noutra. Estas operacións poden ser

⁴ A cantidade de ítems comparados oscila entre os 47 das localidades coloreadas de azul escuro e os 4 das localidades coloreadas de branco. A pouca información dispoñíbel nalgunhas localidades fai que se teñan que interpretar os resultados cunha certa cautela.

insercións, supresións ou substitucións, e na versión máis simple da LD teñen as tres un custo de 1. A distancia final entre dous dialectos é, polo tanto, a que se obtén de transformar dunha variedade á outra as realizacións fonéticas dun determinado número de palabras. Seguidamente exemplifícase o funcionamento básico da LD a partir das realizacións fonéticas do concepto *abeille* en dúas variedades románicas: unha pertencente ao catalán oriental central barcelonés e outra pertencente ao romanche subselvano. Neste caso, o custo final de transformar a primeira liña na segunda é 5:

Catalán oriental central barcelonés	ə ¹ β ε j ə	substitúe ə / ɐ	1
	ɐ ¹ β ε j ə	substitúe β / v	1
	ɐ ¹ v ε j ə	borra ε	1
	ɐ ¹ v j ə	substitúe ə / ø:	1
	ɐ ¹ v j ø:	insire l	1
Romanche subselvano	ɐ ¹ v j ø: l		
Total			5

Dende outro punto de vista, este procedemento tamén se pode entender como o resultado de aliñar dous conxuntos de segmentos fonéticos. Nestes aliñamentos, a superposición fonética é binaria, de xeito que só contribúen a aumentar a distancia fonética entre liñas os segmentos non coincidentes:

Catalán oriental central barcelonés	ə ¹ β ε j ə				
Romanche subselvano	ɐ ¹ v j ø: l				
Total		1	1	1	1 1 = 5

Co obxectivo de reducir o impacto que a distinta lonxitude das palabras comparadas podería ter nos resultados —dous ítems longos poderían ofrecer máis diferenzas ca dous ítems curtos, malia que en termos relativos os dous ítems longos poderían ser máis semellantes entre eles ca os dous ítems curtos—, procedeuse a calcular a distancia relativa entre cada par de palabras. Isto fíxose dividindo a distancia total entre cada par de pronunciacións polo número de segmentos de cada aliñamento; no exemplo que acabamos de ver, a distancia normalizada sería de 0.8 —as 5 diferenzas divididas por 6, que é o total de segmentos aliñados.

Unha vez obtida a matriz de distancias fonéticas do conxunto de variedades, aplicáronse varios sistemas de clasificación e visualización da paisaxe dialectal: análises xerárquicas aglomerativas, mapas probabilísticos a partir de dendrogramas de consenso e mais análises multidimensionais⁵. Tanto o cálculo da LD como a aplicación destes sistemas taxonómicos e representacionais levouse a cabo mediante a interface *Gabmap*⁶.

4. A POSICIÓN DO GALEGO NO ÁMBITO IBERORROMÁNICO

4.1. Un tema menor para a romanística

O proceso de codificación da lingua galega que culminou no ano 2003 cunha segunda actualización das normas ortográficas e morfolóxicas vixentes dende 1982 —a primeira actualización tivo lugar no ano 1995—, caracterizouse por un agre enfrontamento entre os partidarios do chamado *reintegracionismo* —aqueles que consideran que o galego debería tender á unifica-

⁵ V. Valls (2013: 55-58) para profundar nas características e o funcionamento destes sistemas de clasificación e visualización da distancia lingüística.

⁶ Esta interface —v. Nerbonne *et al.* (2011)— foi desenvolvida no *Center for Language and Cognition* da Universidade de Groninga e é de acceso libre no enderezo <http://www.gabmap.nl/> [consulta: 30/01/2016].

ción coa lingua portuguesa, coa que forma unha mesma realidade lingüística— e os partidarios do chamado *autonomismo* —ou *illacionismo*, dende a óptica lusista—, é dicir, aqueles que consideran que, malia que o galego e o portugués comparten unha orixe común e eran unha soa lingua durante a idade media, hoxe xa deviñeron dúas *linguas por distanciamento* —ou *linguas abstand*, segundo a terminoloxía de Kloss (1967). Como é sabido, as normas do ano 2003 foron froito dun consenso entre os partidarios das teses autonomistas e os partidarios das teses reintegracionistas de mínimos, de xeito que o reintegracionismo se viu obrigado a continuar defendendo os seus posicionamentos á marxe da política oficial.

Un podería pensar, en base á cantidade inxente de páxinas que xerou este debate, que a cuestión da filiación do galego debeu ser central na produción científica que se ocupou da clasificación das variedades románicas ao longo do século XX. A verdade é, porén, que dende o “nacemento da lingüística románica científica” (Fernández Rei 1988: 79) que supón a publicación da *Gramática das linguas románicas* de Friedrich Diez no ano 1836, o interese dos principais tratados de romanística pola lingua galega foi practicamente nulo: oscila entre a indiferenza dos que nin sequera a mencionan —v., por exemplo, Alonso (1943)—, a ollada superficial dos que a consideran un *patois* (Elcock 1960) e a pouca atención que lle dedican os que consideran que é un dialecto do español (Vidos 1956) ou do portugués —o resto, con matices. Feldman (1974) resume no seguinte parágrafo os distintos puntos de vista dos que se teñen ocupado do tema:

La mayor parte de la historia de la filología románica está caracterizada por una casi universal coincidencia de opinión en reconocer cuatro divisiones lingüísticas en la Península Ibérica: la catalana, la española, la vasca y la portuguesa. [...] El gallego, en cambio, no ha gozado de la clasificación de lengua románica independiente en la opinión de la mayoría de los romanistas hasta la fecha, ni por otra parte ha sido unánime aquella opinión respecto del estado dialectal del gallego, ya que algunos investigadores lo han atribuido al ámbito dialectal del portugués, mientras que otros lo colocan dentro del conjunto dialectal del español (Feldman 1974: 38).

Tal como se recolle na táboa 1 —un repaso polos principais manuais de referencia da lingüística románica—, até a segunda metade dos anos 80 do século XX non aparecen os primeiros traballos que describen o galego de forma sistemática e que o consideran unha lingua autónoma independente das variedades próximas:

F. Diez (1836)	Considera que hai 6 linguas románicas: o italiano, o valaco, o español, o portugués, o provenzal e mais o francés. O galego considérase unha variedade do portugués.
G. Gröber (ed.) (1888)	Considera o galego como unha variedade dialectal da lingua portuguesa. Así e todo, e como se verá a continuación, esta é a única obra que contén unha descrición sistemática do galego até a publicación do <i>Lexikon der Romanistischen Linguistik</i> (1988-2005).
W. Meyer-Lübke (1901)	Considera que hai 9 linguas románicas: o romanés, o dálmata, o rético, o italiano, o sardo, o provenzal, o francés, o español e mais o portugués. O galego, como o mirandés, é considerado unha variedade do portugués.
A. Alonso (1943)	Acuña o concepto de <i>Romania continua</i> que engloba, de oeste a leste, o portugués, o leonés, o castelán, o aragonés, o gascón, o catalán, o provenzal e o italiano. O francés e o romanés quedan fóra pola súa maior distancia lingüística. Como se observa, Alonso non ten en conta o galego neste continuum románico, aínda que si toma en consideración outras variedades neolatinas —leonés, aragonés— que non eleva á categoría de linguas autónomas.
B. E. Vidos (1956)	Considera que, a diferenza do portugués, que se converteu nunha lingua autónoma a partir da independencia de Portugal, a variedade galega da antiga unidade galaicoportuguesa se converteu, por falta desta independencia política, nun dialecto do español.

C. Tagliavini (1949)	Estabelece catro grandes bloques de linguas románicas unidos por <i>linguas-ponte</i> : o balcanorrománico, o italoorrománico, o galorrománico e mais o iberorrománico. Neste último grupo, o autor inclúe o portugués, o castelán e o catalán. O galego considérase unha variedade neolatina tan afín ao portugués que non chega á categoría de lingua autónoma. De feito, o autor inclúe o galego entre as variedades setentrionais do portugués, xa que segue a clasificación dos dialectos portugueses de Leite de Vasconcelos do ano 1901.
W. D. Elcock (1960)	Estabelece un bloque hispanorrománico semellante ao de Tagliavini (1949), que integran o español, o portugués e mais o catalán. Segundo o autor, e a diferenza do portugués, "Galician, in the meantime, degenerated to a patois status" (Elcock 1960: 428), de xeito que non se considera unha lingua autónoma.
H. Lausberg (1963)	Só fai referencia ao galego cando se refire ás orixes do portugués.
R. A. Hall (1974)	Considera que o bloque iberorrománico occidental está constituído por dous estándares lingüísticos: o galego e o portugués. As variedades dialectais do galego, porén, forman parte, para este autor, do grupo de falas portuguesas setentrionais.
L. Renzi (1985)	Aínda que no texto considera que hoxe galego e portugués son dúas linguas independentes que forman un grupo lingüístico, nos mapas do apéndice o galego inclúese entre os dialectos do portugués. Estes mapas amosan a extensión das seguintes 11 linguas románicas en Europa: o portugués, o español, o catalán, o francés, o francoprovenzal, o provenzal, o sardo, o italiano, o romanche —que inclúe o ladino e mais o friulano—, o dalmata e o romanés.
M. Harris / N. Vincent (eds.) (1988)	Aínda que consideran que o galego e o portugués son dúas linguas independentes, os autores só describen 9 linguas románicas —ademais do latín e os crioulos románicos— con detalle: o español, o portugués, o catalán, o francés, o occitano, o italiano, o sardo, o retorrománico e o romanés.
G. Holtus / M. Metzeltin / Ch. Schmitt (eds.) (1988-2005)	Trátase da primeira obra importante da historia da romanística na que o galego recibe un trato similar ao do resto de linguas románicas sen estado propio.
G. Tuaillon / M. Contini (eds.) (1996)	No <i>Atlas Linguistique Roman</i> o galego recibe a mesma consideración ca o resto de linguas iberorrománicas.

Táboa 1. Consideración do galego nas principais obras de referencia da historia da romanística (creada a partir de Fernández Rei 1988).

Actualmente, entre os que pensan que o galego é unha variedade do portugués e os que cren que galego e portugués xa son dúas linguas *abstand* diferentes, semella que o posicionamento maioritario entre os romanistas é que, malia que estas variedades forman parte dunha mesma realidade lingüística, sociolingüísticamente xa son dúas linguas distintas. Muljačić (1986), por exemplo, considera que o galego e o portugués son linguas *bielaboradas* en base a dous dialectos dunha única *lingua distanciada*, a galaicoportuguesa: "si plus d'un dialecte d'une LD [*langue par distanciation*] a été élaboré avec succès, je parle des *langues bi-* ou *polyélaborées*, cfr. le galaïoportugais > le portugais + le galicien". Cintra (1970) e Hilty (1994) tamén expresan a súa convicción de que estes dous conxuntos de variedades forman unha unidade tanto dende o punto de vista histórico como dende o punto de vista dialectal, de maneira que, tal como afirma Coseriu (1987),

a pesar de que durante siglos no ha habido contactos efectivos, el gallego y el portugués siguen perteneciendo al mismo conjunto, al mismo *continuum* lingüístico. La separación puede, sin duda, establecerse en el plano de la lengua común (que para el gallego, en parte, se está todavía elaborando), pero no a nivel popular o dialectal, de suerte que, para la lingüística histórica, sigue teniendo plena vigencia la denominación compuesta 'gallego-portugués' (Coseriu 1987: 132).

A derradeira pasaxe, tirada de Fernández Rei (1988: 100), amosa como este punto de vista foi asumido, tamén, polo menos por algúns dos principais responsábeis do proceso de codificación recente do galego:

Desde unha perspectiva estritamente lingüística, como antes dixen, ó N e ó S do Miño fálase o mesmo *dialecto* latino (ou a mesma *lingua* latina) porque non existen diferenzas sustanciais entre os falares galegos e os falares interamnenses e trasmontanos portugueses; pero no plano da lingua común, que no caso do galego aínda está a elaborarse, si que existe unha lingua galega e unha lingua portuguesa, con diferenzas fonéticas, morfosintácticas e léxicas, diferenzas que mesmo poden non impedi-la intercomprensión [sic].

4.2. Novas perspectivas baseadas en técnicas de análise cuantitativas

Aínda que a tendencia imperante no ámbito da romanística ao longo do século XX foi considerar o galego como unha variedade do portugués setentrional, algúns estudosos tentaron analizar a relación do galego co portugués e o castelán non con base en criterios diacrónicos e cualitativos, senón con base en criterios de tipo sincrónico e técnicas de análise cuantitativas. Quizais o intento máis acertado de demostrar a individualidade do galego respecto das linguas veciñas foi o de Feldman (1974). Este autor chega á conclusión de que tanto os feixes de isoglosas como os tests sobre o grao de intelixibilidade mutua entre estas variedades bastan para establecer unha fronteira clara entre o galego e o castelán, pero son insuficientes para determinar se o galego forma ou non unha unidade coas variedades dialectais da lingua portuguesa. Por esta razón, este autor opta por recorrer a un terceiro sistema de análise da variación lingüística: a cuantificación de diferenzas estruturais entre variedades. Feldman (1974: 46) afirma que, despois de se analizaren os inventarios estruturais de máis de 300 linguas, se chegou á conclusión de que o teito de diferenzas estruturais a partir do cal dúas variedades deben considerarse linguas diferentes é do 75%. Os resultados do seu estudo, no que compara uns oitenta puntos estruturais do galego, o castelán e o portugués, demostran que “el gallego y el español son lingüísticamente distintos como dos lenguas independientes”, xa que manteñen un grao de semellanza estrutural do 45%. No que fai ao galego e o portugués, malia que o grao de semellanza estrutural chega ao 69%, Feldman (1974: 48) observa que esta porcentaxe aínda é inferior ao 75%, de xeito que, segundo o seu parecer, “el gallego y el portugués [son] hoy dos idiomas independientes”. Esta metodoloxía válelle ao autor para afirmar que

aunque en ciertas épocas anteriores eran más estrechos los vínculos entre el gallego, el español y el portugués, hoy parece cada vez más claro que el efecto principal del cambio lingüístico ha sido la demarcación de los límites entre los tres idiomas. [...] Hoy, por lo menos, podemos poner manos a la obra convencidos de que, al hablar de *lingua* gallega, estamos en lo científicamente cierto (Feldman 1974: 48).

Todos os traballos aos que nos referimos obvian, porén, unha cuestión crucial no debate actual sobre a individuación das linguas, e é que hoxe en día é un feito aceptado que a asignación do status de lingua autónoma a unha determinada variedade ou conxunto de variedades é unha cuestión basicamente política. Kremnitz (2008) demostra —dende un chauvinismo que non compartimos en absoluto— que nas últimas décadas o xurdimento ou a consolidación de movementos nacionalistas no marco dos estados-nación europeos non só comportou unha crítica acérrima ao concepto de *lingua nacional* entendida como a lingua do estado-nación, senón que tamén favoreceu un aumento exponencial das comunidades que loitan por conseguir elevar as súas variedades á categoría de linguas autónomas.

Neste traballo non se pretende, pois, ofrecer unha información *definitiva* sobre o grao de individuación da lingua galega. O noso obxectivo é unicamente achegar un novo punto de vista que lles permita aos lingüistas —especialmente a aqueles encargados da codificación da lingua— coñecer con máis exactitude cal é a posición das variedades galegas —das variedades xeolectais, non do estándar— en relación co resto de variedades románicas. Trátase, en definitiva, de proporcionar unha nova información que poida ser tida en conta durante a planificación do corpus coa intención de que repercuta positivamente tamén na planificación do status e na matriz de trazos valorativos que os galegofalantes asocian á súa lingua.

5. RESULTADOS

A figura 2 mostra un mapa da Romanía no que se proxectaron os doce clústers principais das variedades románicas —que é o número máximo de agrupamentos que permite mostrar a interface *Gabmap*. Como se dixo no apartado 3, utilizamos esta interface para calcular a distancia de Levenshtein entre as variedades románicas a partir dos 36.808 ítems do corpus. Unha vez obtidas as interdistancias entre as 1.010 variedades analizadas, aplicouse un algoritmo de clusterización —o método de Ward— á matriz de distancia fonética. O resultado deste procedemento é o clúster da figura 3, do que se proxectaron os doce agrupamentos principais á figura 2 para facilitar unha primeira aproximación aos resultados.

En primeiro lugar, a figura 2 demostra que un subcorpus relativamente pequeno —de só 48 conceptos, e mesmo menos na maioría de variedades— abonda para que emerxan as principais unidades lingüísticas das que se ocupou a romanística; en concreto, identifícanse claramente os clústers —ou grupos máis ou menos homoxéneos de variedades— correspondentes ás falas portuguesas, galegas, castelás, catalás, occitanas, francoprovenzais, de oïl, italianas setentrionais, italianas centrais, italianas meridionais e romanesas, e un último grupo de variedades que integra, grosso modo, as falas sardas, corsas, friulanas, ladinas, réticas e os enclaves lingüísticos de Italia. A análise deste derradeiro grupo de variedades, tan heteroxéneo, abordarase a partir do dendrograma probabilístico da figura 11.

A proxección destes doce agrupamentos no mapa permite unha primeira aproximación á cuestión de como se organiza o espazo lingüístico románico dende un punto de vista xeográfico, pero ten o inconveniente de que non reflicte as relacións internas das variedades románicas en termos de distancia lingüística. A partir da observación deste mapa, por exemplo, un podería pensar que entre as variedades de oc e de oïl, por unha banda, e as variedades italianas centrais e meridionais, pola outra, debe de haber unha distancia lingüística similar, posto que se trata en ambos os dous casos de variedades contiguas; ou un podería pensar, dado que a distancia xeográfica que as separa é similar, que a distancia lingüística entre o romanés e o italiano setentrional é semellante á distancia lingüística entre este último grupo de variedades e o castelán. Como se desprende da figura 3, porén, estas suposicións están moi afastadas da realidade.

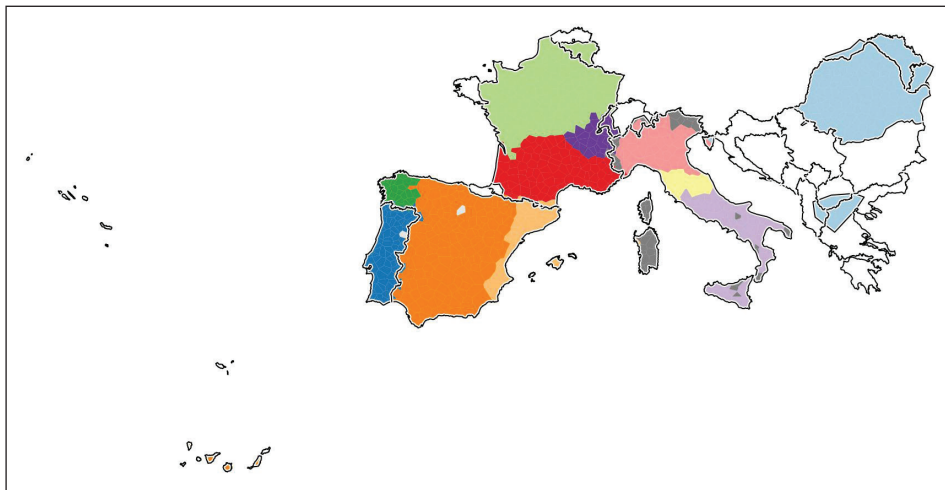


Figura 2. Os 12 clústers principais das variedades románicas obtidos mediante o método de Ward⁷.

A figura 3 mostra o dendrograma xerárquico aglomerativo obtido mediante a aplicación do método de Ward á matriz de distancias. Este dendrograma permite visualizar, por unha banda, como se agrupan entre si as diferentes variedades analizadas; e, pola outra, a que distancia o

⁷ Esta figura correspóndese co dendrograma da figura 3.

fan, xa que a lonxitude das liñas horizontais é directamente proporcional á distancia cofenética media que separa entre si dous grupos de variedades. Así pois, canto máis á dereita se xuntan dous clústers, maior é a distancia que os separa. E, pola contra, canto máis á esquerda se xuntan dous clústers —ou, á esquerda de todo, dúas variedades— menor é a distancia que os separa.

A figura 3 reflicte claramente que as variedades occitanas e francoprovenzais e, a unha distancia un chisco maior, as catalás, forman un primeiro grupo de variedades próximas. Este primeiro grupo únese a un segundo grupo formado polas variedades do italiano —tanto centrais como meridionais e setentrionais— e polas variedades sardas, corsas, ladinas, friulanas e réticas. Da figura 3 despréndese, polo tanto, que a área mediterránea deu lugar ao maior contínuum da *Romania* —polo menos en termos de distancia fonética global entre variedades—; este contínuum parece estenderse do sur de Alacant ao sur da Península Itálica por todo o arco mediterráneo e constitúe —incluídas as Baleares, as Pitiusas, Córsega, Sardeña e Sicilia— unha grande área lingüística caracterizada pola presenza de variedades de transición que favorecen un tránsito suave entre os distintos grupos lingüísticos.

A análise do resto de agrupamentos da figura 3 permite constatar tres feitos relevantes: en primeiro lugar, obsérvase que o conxunto de dialectos que mantén máis similitudes coas variedades do contínuum mediterráneo é o das linguas de oïl —aínda que a agrupación se fai a unha gran distancia, cousa que indica a marcada personalidade destas últimas variedades. En segundo lugar, constátase que as relacións entre as falas galegas e as falas castelás son semellantes, en termos de distancia lingüística, ás que manteñen entre si as falas francoprovenzais e as occitanas ou as italianas setentrionais e o resto de falas itálicas. E, en terceiro lugar, obsérvase que as áreas laterais do ámbito lingüístico románico se atopan a unha gran distancia do resto de variedades estudadas: así pois, aínda que o portugués se agrupa coas variedades galegas e castelás nun conxunto de variedades iberorrománicas que non incluíría o catalán, faíno a unha distancia notábel; e, finalmente, as variedades romanas non aparecen emparentadas con ningún outro grupo de variedades, feito que demostra unha vez máis a especificidade deste conxunto de falas no seo da *Romania*.

Cómpre ter en conta, porén, que a análise de conglomerados, malia ser un sistema de clasificación moi empregado nos estudos de tipo dialectométrico,

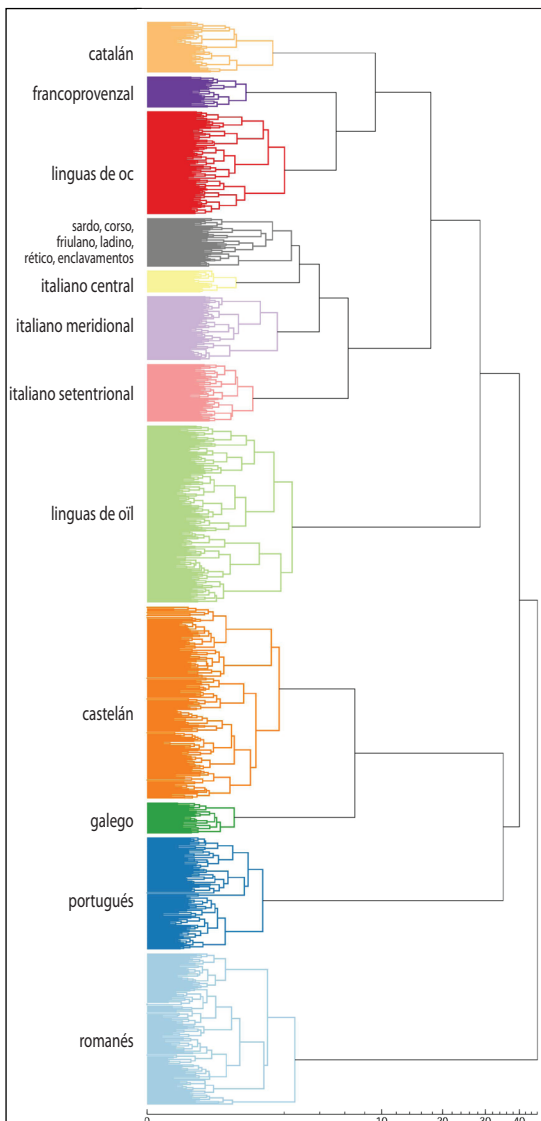


Figura 3. Dendrograma cos 12 clústers principais das variedades románicas obtido mediante o método de Ward⁸.

⁸ A figura 2 mostra a proxección destes clústers no mapa.

é relativamente inestável por mor do seu carácter binario. Esta inestabilidade concretase en que, cando hai varios pares de elementos que se atopan a unha distancia similar na matriz, pequenas diferenzas nos inputs poden dar lugar a clasificacións dendrográficas diferentes (Prokić / Nerbonne 2010). Para superar esta inestabilidade, neste traballo recorreremos aos sistemas de análise multidimensional, que presentan varias vantaxes respecto dos sistemas de clusterización tradicionais: en primeiro lugar, son estatisticamente estábeis; e, en segundo lugar, non agrupan as variedades entre si segundo un procedemento binario e xerárquico, senón que simplemente reflicten no plano as distancias que manteñen entre elas todas as variedades examinadas; isto favorece unha interpretación da realidade lingüística en termos de contínuum, máis que en termos de grupos excluíntes de variedades.

A figura 4 amosa os resultados de realizar unha análise multidimensional á matriz de distancia obtida mediante a aplicación do algoritmo de Levenshtein. Os doce clústers principais que emerxían na figura 3 coloreáronse coas mesmas cores no gráfico para comprobar que se trata realmente de grupos lingüísticos estábeis e independentes e non de clústers emerxidos polo carácter binario da análise de conglomerados. Unha primeira ollada á figura 4 só permite chegar con seguridade a unha dobre conclusión: obsérvase, en primeiro lugar, que as variedades daco-romanesas, aromanesas, megleno-romanesas e máis istrio-romanesas forman un clúster que se sitúa a gran distancia do resto de variedades románicas; en segundo lugar, obsérvase que o resto de variedades debuxan un contínuum que se estende das variedades iberorrománicas, na parte inferior do plano, ás variedades galorrománicas, na parte superior do plano. O catalán sitúase nunha posición de transición entre estes dous conxuntos de variedades e as variedades italo-románicas. A partir da información que proporciona a figura 4 faise imposible, en cambio, determinar os once clústers de variedades románicas —excluído o romanés— que identificabamos na figura 3.

Para podermos analizar a composición deste gran contínuum románico que se estende pola metade esquerda do plano da figura 4 debemos aumentar o zoom sobre a zona de interese, e isto é posible realizando unha nova análise multidimensional que exclúa as variedades lingüisticamente máis distantes, é dicir, as romanesas. Os resultados desta nova análise —v. a figura 5— demostran un feito que xa se detectara no dendrograma da figura 3: que o portugués, por ser unha área lateral do ámbito lingüístico, sitúase a unha distancia bastante importante do resto de variedades románicas, incluídas as que máis se lle achegan, as da lingua galega.

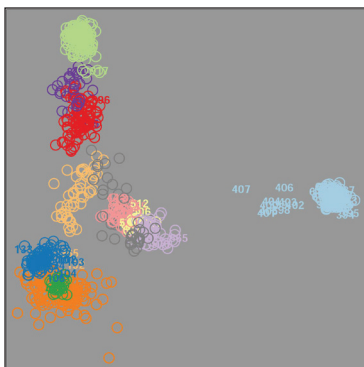


Figura 4. Resultados da análise multidimensional cos 12 clústers principais das variedades románicas. $R = 0,70$.

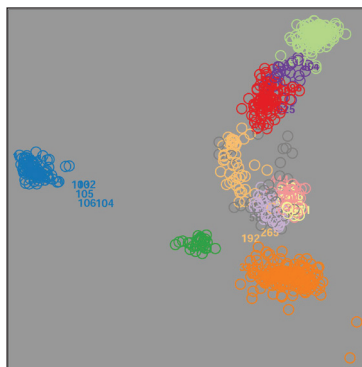


Figura 5. Resultados da análise multidimensional das variedades románicas sen o romanés. $R = 0,74$.

Este artigo céntrase especialmente no debate sobre a posición do galego respecto das variedades portuguesas e castelás. Como se expuxo en 4.1, até o ano 1988 os principais manuais sobre romanística adoitaban considerar esta lingua como unha variedade dialectal do portugués ou do español, e non foi até entón que algúns expertos optaron por considerala unha lingua iberorrománica independente. Para determinar o grao de individuación do galego respecto do portugués e do castelán elaborouse a figura 6, que amosa os resultados da análise multidimensional nas variedades destas tres linguas. A conclusión é taxativa: as variedades do

galego forman un clúster tan homoxéneo e independente, como mínimo, como os clústers correspondentes ás variedades portuguesas e castelás⁹.

A independencia lingüística do galego en termos de distancia fonética é, xa que logo, un feito empiricamente contrastado. Agora ben, aínda que esta individualidade non se cuestiona, se se incorpora unha cuarta variedade á análise —a catalá, por ser tamén do ámbito ibérico—, vese que realmente a relación entre o galego e o castelán e entre o galego e o portugués non é de equidistancia, como suxería a figura 6. A figura 7, que plasma os resultados da análise multidimensional nas variedades da Península Ibérica, mostra que a posición do galego respecto ao castelán é de maior proximidade que respecto ao portugués¹⁰.

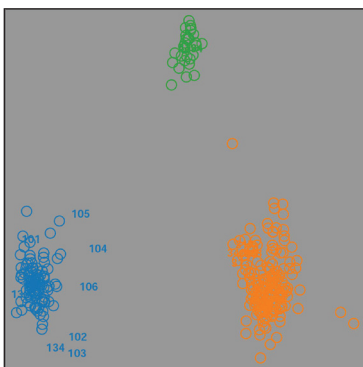


Figura 6. Resultados da análise multidimensional das variedades castelás, galegas e portuguesas. $R = 0,95$.

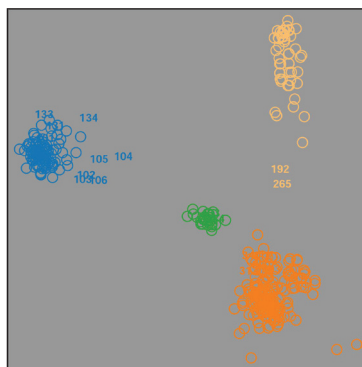


Figura 7. Resultados da análise multidimensional das variedades da Península Ibérica. $R = 0,94$.

As figuras 8 e 9 amosan os resultados de senllas análises multidimensionais levadas a cabo para clasificar as variedades galas e itálicas —incluídas as de Córsega, Sardeña e Sicilia—, respectivamente. Da figura 8 despréndese claramente que as variedades de oil, de oc e francoprovenzais constitúen tres grupos lingüísticos claramente independentes; o mesmo acontece —v. a figura 9— entre as variedades itálicas setentrionais, centrais e meridionais.

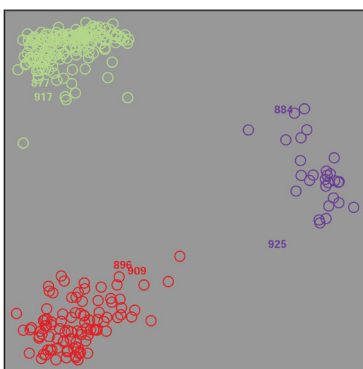


Figura 8. Resultados da análise multidimensional das variedades galas. $R = 0,71$.

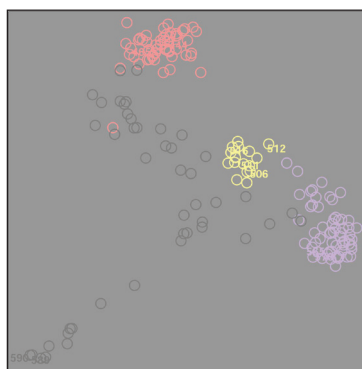


Figura 9. Resultados da análise multidimensional das variedades itálicas. $R = 0,71$.

⁹ A causa dos poucos datos dispoñíbeis sobre as falas asturleonesas e aragonesas no corpus, estas variedades non emerxen do clúster de falas castelás. A nosa hipótese, porén, é que cunha maior cantidade de datos estes dous conxuntos de variedades aparecerían como dous clústers identificábeis á marxe do da lingua castelá.

¹⁰ O feito de que as variedades patrimoniais da lingua galega presenten a miúdo un elevado número de interferencias léxicas procedentes do castelán ten que se ter en conta á hora de interpretar estes resultados, porque este traballo parte, como se viu, dun corpus exclusivamente léxico. De feito, a depuración das interferencias que todo proceso estandarizador leva asociada podería contrarrestar, ou cando menos mitigar, nos vindeiros anos, a tendencia á adxerencia cara á lingua castelá que caracterizou a evolución do galego nos últimos séculos.

A figura 10, finalmente, permite abordar outro asunto que ten xerado moita controversia entre os romanistas: a subagrupación románica do catalán¹¹. Co obxectivo de achegarse a esta cuestión realizouse unha análise multidimensional que incluíse unicamente as variedades que a romanística considerou claramente iberorrománicas —as portuguesas, galegas e castelás— e as que considerou claramente galorrománicas —as de oïl, occitanas e francoprovenzais. Os resultados mostran que, polo menos a comezos e mediados do século XX, cando se levaron a cabo as enquisas a partir das cales se elaborou este subcorpus de datos fonéticos, o agrupamento da lingua catalá con algún destes dous grupos de variedades xa non era viábel: a posición do catalán é claramente de transición entre as variedades galorrománicas meridionais —languedocianas e gascoas— e as variedades iberorrománicas setentrionais. De feito, o propio clúster de variedades iberorrománicas —que se identificaba claramente na figura 4— desdebúxase na figura 10, en consonancia co que vimos nas figuras 6 e 7.

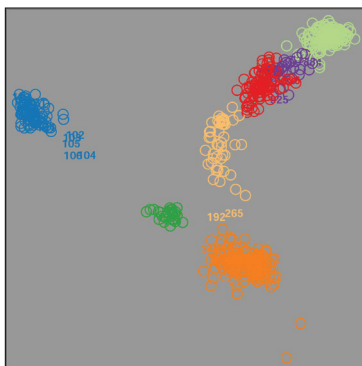


Figura 10. Resultados da análise multidimensional das variedades iberorrománicas e galorrománicas. $R = 0,86$.

Alén das análises multidimensionais, hai un segundo método para se achegar ás relacións entre grupos de variedades que non sofre as limitacións dos sistemas xerárquicos aglomerativos tradicionais e que ademais reflicte as relacións xerárquicas entre variedades e grupos de variedades. Trátase dos sistemas de clusterización estábel, como o *noisy clustering*. Este sistema consiste en engadir pequenas cantidades de ruído aleatorio —*random noise*— á matriz de distancias en sucesivas clusterizacións e dá como resultado un *dendrograma probabilístico* ou *de consenso* —en inglés, *consensus* ou *probabilistic dendrogram*. Este dendrograma proporciona dúas informacións relevantes: en primeiro lugar, os números asociados aos clústers indican a cantidade de veces que un conxunto de variedades constituíu un mesmo grupo nas distintas iteracións do proceso; en segundo lugar, a lonxitude horizontal das liñas reflicte a distancia cofenética media á que un conxunto de variedades se agrupou entre si no total de clusterizacións: canto máis longa é unha liña menos homoxéneas son as variedades que agrupa, por máis que estas sexan máis semellantes entre elas que respecto a calquera outra variedade do corpus.

Unha primeira ollada ao dendrograma probabilístico da figura 11¹² xa permite chegar a unha conclusión importante: que as relacións entre variedades románicas deben entenderse máis en termos de continuum que en termos de subagrupacións pechadas; isto é así porque, aínda que este sistema de clusterización detecta moi ben as principais unidades lingüísticas románicas,

¹¹ En Valls / González (no prelo) abórdase, tamén dende unha óptica dialectométrica e en base ao mesmo corpus do *ALiR*, a cuestión da subagrupación románica da lingua catalá. En certa maneira trátase, pois, dun artigo complementario a este traballo.

¹² Por falta de espazo non puidemos incluír esta figura no artigo, pero pódese visualizar e descargar no enderezo <http://www.usc.es/revistas/docs/3175-16053-1-SP.pdf> [consulta: 28/07/2016]. Para interpretar correctamente as abreviaturas coas que se designan as distintas variedades románicas, pódense consultar as táboas explicativas incluídas en Valls / González (2014: 341, 343-344).

estas unidades non se agrupan entre si de maneira estábel —isto é, en máis do 80% de clusterizacións— en case ningún caso.

De novo, este sistema de clasificación fai emerxer os clústers das variedades catalás, castelás, galegas, de oïl, francoprovenzais, occitanas, italianas setentrionais, italianas centrais, italianas meridionais, romanesas e mais portuguesas. Ademais, porén, a maior sofisticación do método permite que emerxan outros grupos homoxéneos de variedades que até o de agora non se detectaran de maneira nítida: en primeiro lugar, emerxe un grupo de variedades corsas, que forman un único clúster cunha certa variación interna. En segundo lugar, emerxe un grupo de variedades sardas que se agrupan entre si no 100% de clusterizacións, aínda que tamén presentan unha certa variación interna —que se identifica, como dixemos, pola lonxitude das liñas horizontais do dendrograma. En terceiro lugar, emerxe un grupo de variedades ladinas e friulanas —que á súa vez se agrupan coas variedades italianas setentrionais nun segundo nivel. A raíz disto pódese concluír que estas variedades —antigas integrantes dun contínuum lingüístico hoxe fragmentado— seguen mantendo unha relación de gran proximidade. Este non é o caso, en cambio, das variedades retorrománicas, as terceiras integrantes deste contínuum histórico, que non só aparecen desvinculadas das falas ladinas e friulanas, senón que —e isto podería deberse á influencia das variedades xermánicas de Suíza— non se agrupan con ningún outro conxunto de variedades analizadas. A figura 11 permite identificar, finalmente, un clúster bastante homoxéneo formado polas variedades aromanesas e, nun segundo nivel, megleno-romanesas. Estas variedades, xunto co istrio-romanés, a máis distancia, agrúpanse regularmente coas variedades daco-romanesas.

Un último aspecto interesante que se desprende da observación da figura 11 é a posición que ocupan as variedades galegas: no 96% dos casos aparecen agrupadas entre si, como un grupo de variedades moi homoxéneas no contexto románico, pero no marco do clúster de variedades da lingua castelá. Isto significa que, de todas as variedades románicas que hoxe reciben unanimemente a consideración de linguas independentes, as variedades da lingua galega son, polo que semella, as que presentan unha distancia lingüística menor respecto das variedades doutra lingua, neste caso a castelá. Neste sentido, a relación entre o galego e o castelán é, en termos de distancia fonética, similar á que manteñen entre si variedades como o gascón e o languedociano, que se adoitan considerar variantes dialectais da lingua occitana.

Os sistemas de clusterización estábeis, como o *noisy clustering*, tamén teñen a vantaxe de permitir proxectar subsets de datos nos chamados *mapas probabilísticos* —ou *fuzzy cluster maps*—, grazas aos cales é posíbel analizar a partir dun mapa a composición interna dos principais clústers detectados nun dendrograma de consenso. Tal como se explica no manual de uso de *Gabmap*, un mapa probabilístico

visualizes something between multidimensional scaling (MDS) and cluster analysis: main dialect groups are identified in the map, but continuous relationships are displayed for places which cannot be put in one group with high probability. The map is created by running MDS on the branch lengths of the dendrogram (so-called cophenetic distances) instead of on the original linguistic distances¹³.

As figuras 12, 13, 14 e mais 15 son mapas probabilísticos correspondentes aos seguintes subsets de datos: as variedades de oïl, as de oc e francoprovenzais, as italo-románicas e as romanesas. Grazas a estes mapas pódense identificar as principais variedades que compoñen cada un destes clústers principais. Na figura 12, por exemplo, identifícanse as seguintes variedades de oïl —do suroeste ao nordés—: o poitevino-saintongeais, o grupo formado polo galó e o angevin, o normando, o picardo, o valón, o lorenés, o franco-contés e un gran *plateau* de falas centrais que comprendería as falas borgoñoas, o champenois, o berrichon e mais o francés, todas elas practicamente asimiladas a esta última variedade. No que fai á figura 13, correspondente ás variedades de oc e francoprovenzais, identifícanse claramente as falas gascoas, languedocianas, lemosinas, auvernesas, vívaro alpinas e provenzais, por unha banda, e un grupo de falas franco-provenzais bastante homoxéneas, pola outra.

¹³ Esta pasaxe foi tirada do manual en liña de *Gabmap*, de acceso aberto no enderezo <http://www.gabmap.nl/~app/doc/manual/clustering-fuzzy.html> [consulta: 27/01/2016].

A figura 14, pola súa banda, identifica claramente as variedades italianas setentrionais, centrais e meridionais, como a maioría de técnicas que viñemos utilizando até agora. Porén, tamén identifica moi claramente as falas ladinas e friulanas, o occitano falado nas *valadas* e o franco-provenzal do Val de Aosta, as falas sardas e, finalmente, as falas corsas, que teñen continuidade no sasará e o galurés do norte de Sardeña. A figura 15, finalmente, salienta catro grandes variedades dialectais da lingua romanese: por unha banda, as falas aromanesas e megleno-romanesas, nos Balcáns; pola outra, as tres variedades principais de daco-romanés: o valaco, no centro e sur da actual Romanía; o moldavo, no nordés; e as falas de Banat, no noroeste. No medio destas variedades, os tons verdosos e azulados, máis claros, sinalan a área de transición formada polos dialectos transilvánicos.

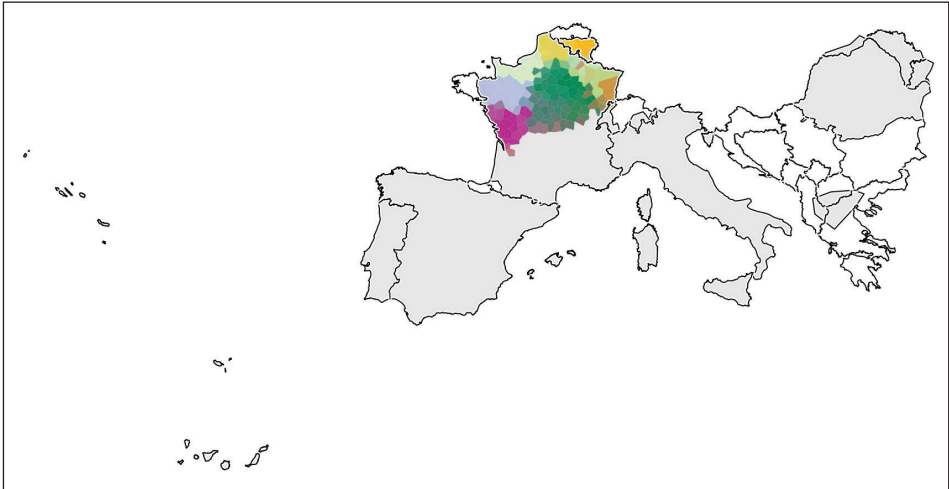


Figura 12. Mapa probabilístico das variedades de oil.

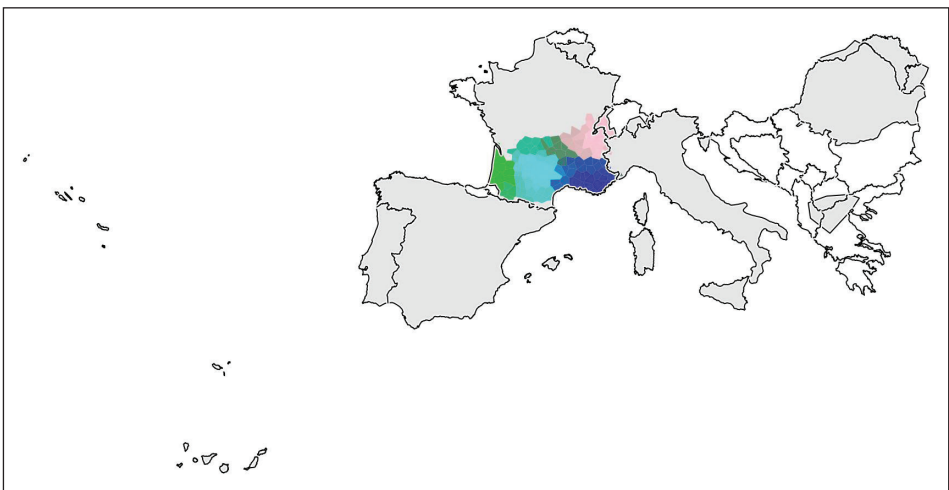


Figura 13. Mapa probabilístico das variedades de oc e franco-provenzais.

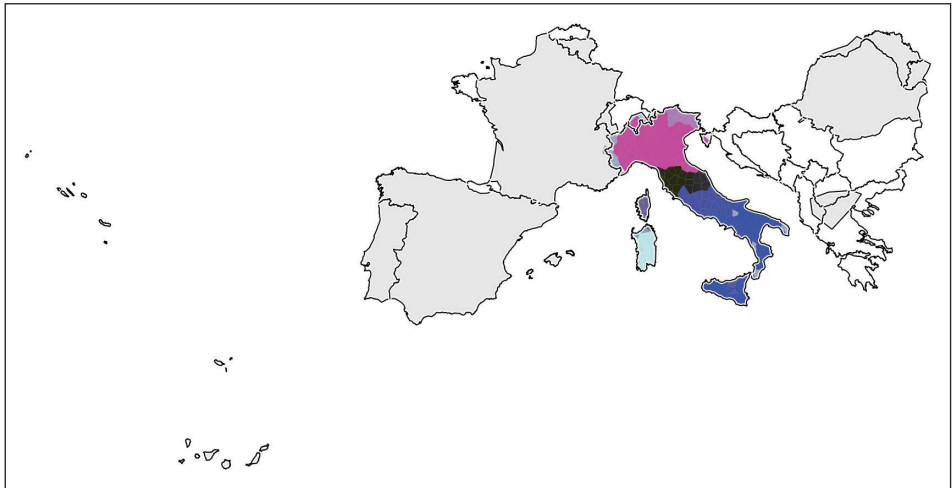


Figura 14. Mapa probabilístico das variedades italo-románicas.



Figura 15. Mapa probabilístico das variedades romanesas.

6. CONCLUSIÓNS

Neste traballo presentáronse os resultados de analizar por primeira vez a variación e a distancia lingüística entre as variedades da Romania Antiqua dende unha óptica dialectométrica. A medida de distancia fonética utilizada foi a distancia de Levenshtein, que se aplicou a un corpus de 36.808 ítems repartidos en 1.010 localidades do ámbito lingüístico románico. Este corpus deseñouse no Centro Ramón Piñeiro para a Investigación en Humanidades e actualmente está en proceso de ampliación. Tanto o cálculo da matriz de interdistancias como os distintos sistemas taxonómicos e de visualización empregados aplicáronse mediante a interface *Gabmap*.

A primeira conclusión á que se pode chegar é que un corpus de datos relativamente reducido —de só 48 conceptos, e aínda menos na maioría de variedades— basta para que emer-

xan as principais unidades lingüísticas das que se ocupou a romanística. A análise cuantitativa deste corpus debuxa unha paisaxe xeolectal do ámbito románico que se pode resumir nos seguintes puntos:

- a. En primeiro lugar, parece —polo que se desprende da análise xerárquica aglomerativa— que a área mediterránea deu lugar ao maior contínuum de toda a Romanía; este contínuum esténdese do sur de Alacant ao sur da Península Itálica pasando por todo o arco mediterráneo e constitúe —incluídas as Baleares, as Pitiusas, Córsega, Sardeña e Sicilia— unha grande área lingüística caracterizada pola presenza de variedades de transición que favorecen un tránsito suave entre os distintos grupos lingüísticos.
- b. En segundo lugar, obsérvase —a partir das análises multidimensionais— que as variedades daco-romanesas, aromanesas, megleno-romanesas e istrio-romanesas forman un clúster que se sitúa a unha gran distancia do resto de variedades románicas. En contraste coas variedades do romanés, as variedades iberorrománicas, italo-románicas e galorrománicas forman un contínuum de falas separadas por unha distancia fonética moito máis reducida.
- c. En terceiro lugar, obsérvase que o portugués, pola súa situación lateral no ámbito lingüístico, sitúase a unha distancia notábel respecto ao resto de variedades románicas. Esta distancia é importante mesmo con respecto ás variedades que máis se lle achegan, as da lingua galega.
- d. En cuarto lugar, a análise do dendrograma probabilístico permitiu constatar que as relacións entre variedades románicas deben entenderse máis en termos de contínuum ca en termos de subagrupacións pechadas. Este dendrograma tamén permitiu observar que, a día de hoxe, as variedades ladinas e friulanas —antigas integrantes dun contínuum lingüístico hoxe fragmentado— seguen mantendo unha relación de gran proximidade. Isto contrasta coas variedades retorrománicas —as terceiras integrantes deste contínuum histórico—, que non só aparecen desvinculadas das falas ladinas e friulanas, senón que —e isto podería deberse á influencia das variedades xermánicas de Suíza— non se agrupan con ningún outro conxunto de variedades analizadas.
- e. No que fai á subagrupación románica da lingua catalá, as análises multidimensionais amosan que as variedades desta lingua se ubican claramente nunha posición de transición entre as variedades galorrománicas meridionais —languedocianas e gascoas— e as variedades iberorrománicas setentrionais.
- f. E, finalmente, no que fai ao grao de individuación da lingua galega, dende unha perspectiva cuantitativa global é posíbel chegar a unha dobre conclusión: por unha banda, as variedades do galego forman un clúster tan homoxéneo e independente, polo menos, como os clústers correspondentes ás variedades portuguesas e castelás; por outra banda, porén, a relación entre o galego e o castelán e entre o galego e o portugués non é de equidistancia, posto que nas análises multidimensionais se observa unha maior proximidade desta lingua coas variedades castelás e no dendrograma probabilístico mesmo se agrupa con elas nun 96% das clusterizacións. Isto parece indicar que o proceso de distanciamento do galego respecto do portugués que tivo lugar ao longo dos últimos séculos comportou, paralelamente, un proceso de adverxencia ao castelán. A raíz desta evolución, de todas as variedades románicas que hoxe reciben por consenso a consideración de linguas independentes, as da lingua galega son, polo que parece, as que presentan unha distancia lingüística menor respecto ás variedades doutra lingua — neste caso a castelá.

Segundo o noso parecer, estes resultados suxiren que a potenciación dos elementos converxentes co portugués durante a planificación do corpus podería repercutir moi positivamente na matriz de trazos valorativos asociada a este idioma en dous sentidos: por unha banda, favorecería que os falantes asumisen con máis naturalidade que o que falan é unha lingua claramente autónoma —e non un *castropo*, como aínda afirman numerosos falantes de todas as idades influenciados por vellos prexuízos lingüísticos—; e, pola outra, facilitaríala o acceso dos galegofa-

lantes á lusofonía, ben a través da súa propia lingua —unha lingua que tería unha maior presenza dos elementos comúns co portugués—, ben a través do portugués estándar, unha variedade cuxa aprendizaxe sería ben máis fácil nestas novas circunstancias. Para nós, é indubidábel que estas dúas consecuencias da nova etapa planificadora contribuirían significativamente ao aumento do prestixio da lingua e ao incremento do capital lingüístico dos galegofalantes.

É certo que no momento actual, cunha lingua galega inmersa dende hai xa demasiados anos nunha situación de *contraplanificación lingüística* —v. Pradilla (2003)—, é difícil imaxinar que haxa a vontade política necesaria para emprender con medios e convicción un cometido coma este. Pero vai chegar un día no que se vai retomar con forza a planificación do status da lingua galega, e ese día cómpre termos claro cara a onde queremos ir e cómpre termos traballado moito nesa dirección. A posición do galego respecto do castelán e do portugués aconsella reforzar a presenza e a visibilidade dos elementos comúns co resto de variedades da lusofonía —elementos, non o esquezamos, patrimoniais da lingua galega—: as desvantaxes de non facelo son demasiadas para non ser tidas en conta. Mans á obra, pois.

REFERENCIAS BIBLIOGRÁFICAS

- Alonso, Amado (1943): «Partición de las lenguas románicas de Occidente», en Joan Corominas (ed.), *Miscel-lània Fabra*. Buenos Aires: Coni, 81-101.
- Contini, Michel (1992): «L'Atlas Linguistique Roman», *IKER* 7, 339-356.
- Coseriu, Eugenio (1987): «El gallego y sus problemas. Reflexiones frías sobre un tema candente», *Lingüística española actual* 9, 127-138.
- Diez, Friedrich C. (1836): *Grammatik der romanischen sprachen*. Bonn: E. Weber.
- Elcock, William D. (1960): *The Romance Languages*. Londres: Faber & Faber.
- Feldman, David (1974): «Consideraciones sobre la identificación lingüística de la lengua gallega», *Grial* 43, 38-48.
- Fernández Rei, Francisco (1988): «Posición do Galego entre as linguas románicas», *Verba* 15, 79-107.
- Goebel, Hans (1991): «Dialectometry: A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data», en Reinhard Köhler / Burghard B. Rieger (eds.), *Contributions to quantitative linguistics. Proceedings of the First International Conference on Quantitative Linguistics*. Dordrecht: Kluwer, 277-315.
- Goebel, Hans (2003): «Regards dialectométriques sur les données de l'Atlas Linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics* XXV, 59-121.
- Goebel, Hans (2006): «Recent Advances in Salzburg Dialectometry», *Literary and Linguistic Computing* 21, 411-435. <http://dx.doi.org/10.1093/lc/fql042>.
- Goebel, Hans (2008): «La dialettometrizzazione integrale dell'ALS. Presentazione dei primi risultati», *Revue de linguistique romane* 72 (285-286), 25-73.
- Gröber, Gustav (ed.) (1888): *Grundriss der romanischen philologie*. Estrasbourg: K. J. Trübner.
- Hall Jr., Robert A. (1974): *External History of the Romance languages*. Nova York: American Elsevier.
- Harris, Martin / Nigel Vincent (eds.) (1988): *The romance languages*. Londres: Routledge.
- Hilty, Gerold (1994): «La place du galicien parmi les langues romanes», en Ramón Lorenzo (ed.), *Actas do XIX Congreso Internacional de Lingüística e Filoloxía Románicas*. A Coruña: Fundación Pedro Barrié de la Maza, 457-472.
- Holtus, Günter / Michael Metzeltin / Christian Schmitt (eds.) (1988-2005): *Lexikon der romanischen Linguistik*. Tübingen: Niemeyer.
- Kloss, Heinz (1967): «Abstand languages and Ausbau languages», *Anthropological Linguistics*, 9 (7), 29-41.
- Kremnitz, Georg (2008): «Sur la délimitation et l'indivision des langues. Avec des exemples pris principalement dans le domaine romain», *Estudis Romànics* 30, 7-38.
- Lausberg, Heinrich (1963): *Romanische Sprachwissenschaft I. Einleitung und Vokalismus*. Berlín: Walter de Gruyter.
- Meyer-Lübke, Wilhelm (1901): *Einführung in das Studium der romanischen Sprachwissenschaft*. Heidelberg: Winter.

- Muljačić, Žarco (1986): "L'enseignement de Heinz Kloss (modifications, implications, perspectives", *Langages* 83, 53-63. <http://dx.doi.org/10.3406/lgge.1986.2495>.
- Nerbonne, John *et al.* (2011): "Gabmap: A Web Application for Dialectology", *Dialectologia. Special issue 2011*.
- Pradilla, Miquel-Àngel (2003): "El País Valencià: un cas de contraplanificació lingüística", en *Actes del 2n Congrés Europeu sobre Planificació Lingüística*. Barcelona: Generalitat de Catalunya, 112-128.
- Prokić, Jelena / John Nerbonne (2010): "Recognizing Groups among dialects", en John Nerbonne *et al.* (eds.), *International Journal of Humanities and Arts Computing. Special issue on Language Variation*. Edimburg: Edinburgh University Press, 153-172.
- Renzi, Lorenzo (1985): *Nuova introduzione alla filologia romanza*. Bolonya: Il Mulino.
- Tagliavini, Carlo (1949): *Le origini delle lingue neolatine. Introduzione alla filologia romanza*. Bolonya: Pàtron.
- Tuaillon, Gaston / Michel Contini (eds.) (1996): *Atlas Linguistique Roman (ALiR). Volume I. Présentation. Cartes. Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.
- Valls, Esteve (2013): *Direccionalitat, ritme, abast i naturalesa del canvi lingüístic en català nord-occidental. De l'anàlisi dialectomètrica a la reflexió sociolingüística*. Universitat de Barcelona. Tese de doctorament inèdita.
- Valls, Esteve / Manuel González (2014): "Unha nova base de datos para a romanística a partir do *Atlas Linguistique Roman*", *Verba* 41, 337-363.
- Valls, Esteve / Manuel González (no prelo): "La subagrupació romànica del català: una aproximació dialectomètrica de base fonètica a l'*Atlas Linguistique Roman*", *Studia Romanica Posnaniensia*.
- Veny, Joan / Michel Contini (eds.) (2001): *Atlas Linguistique Roman (ALiR). Volume IIa. Cartes. Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.
- Veny, Joan / Michel Contini (eds.) (2009). *Atlas Linguistique Roman (ALiR). Volume IIb. Cartes. Commentaires*. Roma: Istituto Poligrafico e Zecca Dello Stato.
- Vidos, Benedek E. (1956): *Handboek tot de Romaanse taalkunde*. 'S-Hertogenbosch: Malmberg.