

Proposta de análise dialectométrica da prosodia galega

Elisa Fernández Rei e Adela Martínez Calvo

Instituto da Lingua Galega (ILG) – Universidade de Santiago de Compostela (USC)

elisa.fernandez@usc.es adela.martinez@usc.es

Abstract:

In the last decade, an important effort was made to describe the Galician intonation and its dialectal variation. This description, which was fundamentally acoustic, was developed using a corpus from 15 different points distributed all over Galicia. The diatopic study aroused interest in delimiting the prosodic dialectal areas of Galician language. In order to answer this question, a dialectometrical analysis was developed to measure the distances among the different Galician dialects and to define the main areas of the Galician prosody. The most significant results of this analysis are presented in this work.

Keywords/Palavras-chave: dialectometry, Galician, intonation, prosodic areas / dialectometria, galego, entonação, áreas prosódicas.

1. Introducción

Na última década fíxose un importante esforzo para describir a entoación do galego e a súa variación dialectal. Esta descrición, fundamentalmente acústica, realizouse a partir dun corpus recollido en 15 puntos de enquisa repartidos ao longo de todo o territorio galego. O material recompilado e analizado foi, e é, de gran utilidade para desenvolver traballos no marco, por exemplo, da fonoloxía. Agora ben, desde o punto de vista da dialectoloxía, esta recolla de datos prosódicos suscita, obviamente, a pregunta de se é posible establecer ou delimitar as áreas prosódicas dialectais do galego.

Para contribuír a responder a esta pregunta, por un lado, levouse a cabo unha análise dialectométrica que permitise medir as distancias entre as distintas variedades; e, por outro lado, desenvolveuse un estudo perceptivo co obxectivo de medir a distancia que os falantes percibían entre as distintas entoacións dialectais. No presente traballo presentamos os resultados desa primeira parte, a análise dialectométrica.

O método dialectométrico utilizado consiste no cálculo, para cada par de curvas entoativas coa mesma estrutura e modalidade, dunha medida de correlación ponderada que permite avaliar de forma cuantitativa a similitude acústica entre elas. A aplicación desta medida de correlación a pares de curvas orixinadas en localizacións distintas permite definir unha distancia prosódica entre cada par de puntos e agrupalos segundo a súa proximidade mediante un algoritmo de agrupamento xerárquico. Finalmente, este agrupamento xerárquico permite xerar un primeiro mapa das variedades prosódicas galegas.

Todos os cálculos executados para a obtención dos resultados e as gráficas incluídas neste traballo realizáronse co software estatístico R (R Core Team, 2013).

Textos Seleccionados, XXIX Encontro Nacional da Associação Portuguesa de Linguística, Porto, APL, 2014, pp. 249-254, ISBN 978-989-97440-3-5

2. Descrición dos datos

Neste traballo propónse unha análise dialectométrica da prosodia galega co obxectivo de determinar e delimitar as áreas prosódicas do galego. Para este fin, analizouse un corpus de máis de 7.000 frases enunciativas e interrogativas correspondentes a 63 estruturas, con esquemas acentuais e lonxitude diferentes, provenientes de 15 puntos de enquisa de Galicia e un total de 28 informantes masculinos e femininos. Para cada unha das frases analizadas dispónse de medicións precisas de frecuencia fundamental (F0), da duración e da intensidade de cada unha das vogais do enunciado: no caso da intensidade recóllese un único valor, pero no caso de F0 tómase o valor inicial, medio e final en cada vogal.

A distribución xeográfica dos puntos de enquisa incluídos no corpus pode verse na Figura 1. Este corpus recolleuse dentro do proxecto AMPER-Galicia (<http://ilg.usc.es/amper/>), sección pertencente ao proxecto internacional AMPER (<http://w3.u-grenoble3.fr/dialecto/AMPER/amper.htm>), co fin de contribuír ao obxectivo común de estudar e comparar os trazos prosódicos das distintas variedades lingüísticas no ámbito da Romania.

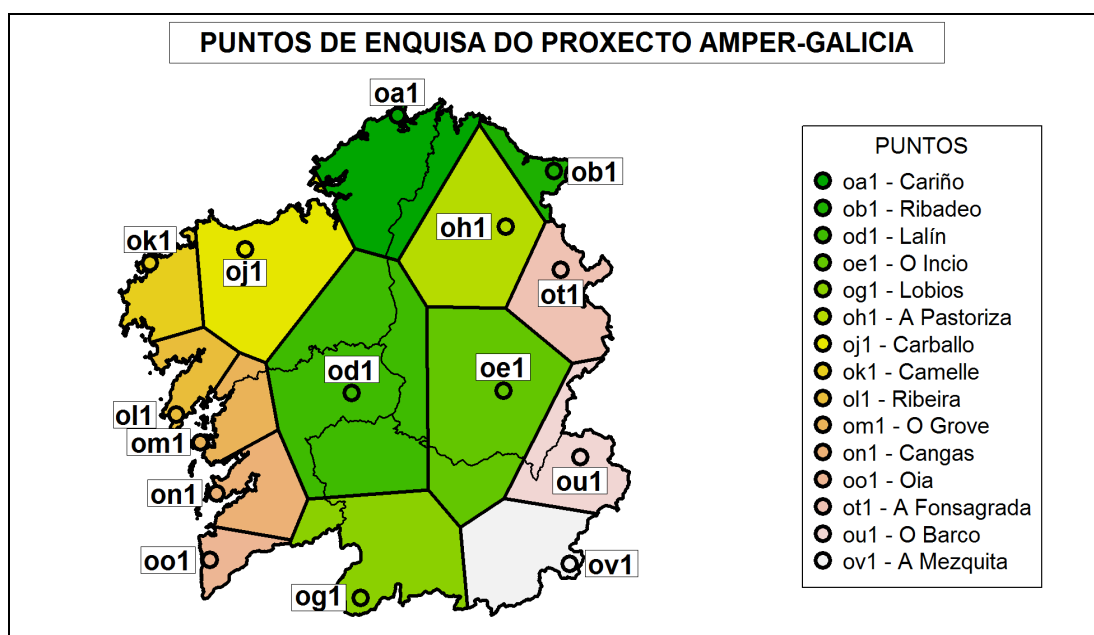


Figura 1: Codificación e localización xeográfica dos puntos de enquisa do corpus

Como xa se ten referido na literatura, é habitualmente na entoación dos enunciados interrogativos onde aparecen os trazos rexionais máis marcados. Este feito tamén se observou no corpus analizado neste traballo. Por este motivo, se ben se levou a cabo a análise completa das enunciativas e interrogativas do corpus, preséntanse a continuación só os resultados destas últimas, pois é nelas onde se manifesta de maneira notoria a variación diatópica, como pode verse nos exemplos representados na Figura 2.

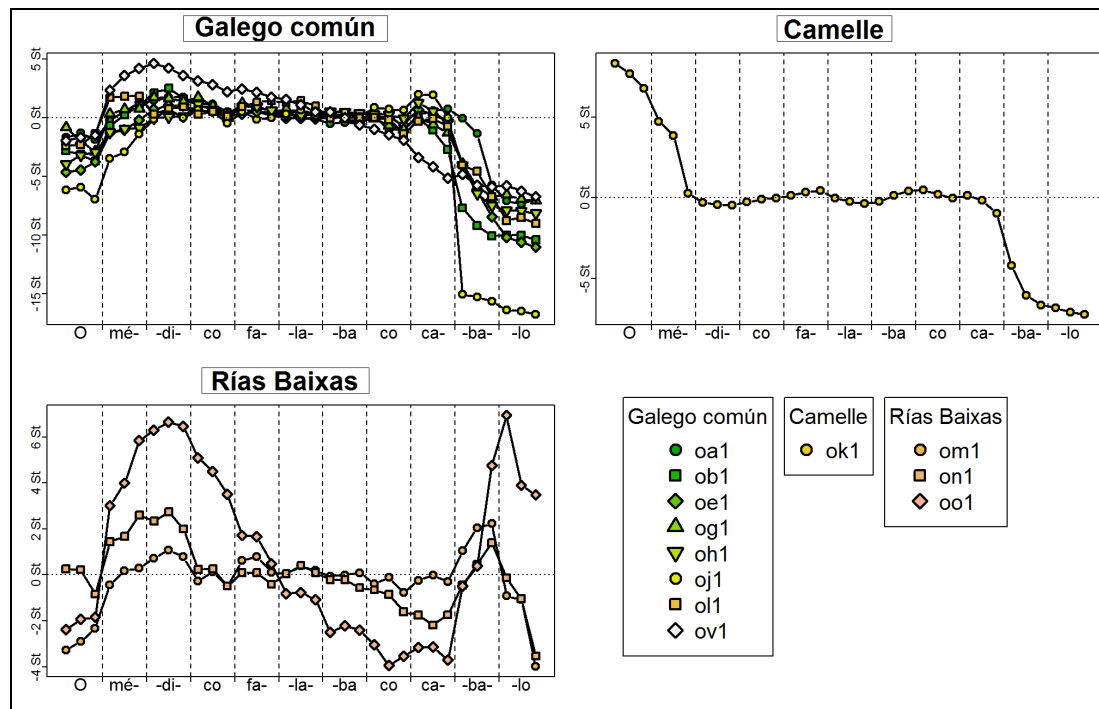


Figura 2: Media para tres repeticións da frecuencia fundamental (medida en semitóns) das mulleres do corpus para a frase interrogativa “o médico falaba co cabalo?”

3. Metodoloxía estatística

Para medir a proximidade entre dúas curvas entoativas, coa mesma estrutura e modalidade, calculouse a medida de correlación ponderada pola enerxía do sinal entre curvas de frecuencia fundamental definida por:

$$corr(f_1, f_2) = \frac{\sum_i w(i)(f_1(i) - m_1)(f_2(i) - m_2)}{\sqrt{\sum_i w(i)(f_1(i) - m_1)^2 \sum_i w(i)(f_2(i) - m_2)^2}}$$

onde f_1 e f_2 son as curvas da F0 (medidas en semitóns) para cada unha das frases, m_1 e m_2 son os valores medios da F0 para f_1 e f_2 respectivamente (medidos en semitóns), e w é o promedio dos valores de enerxía das dúas curvas f_1 e f_2 (medido en decibelios). Esta fórmula deriva dos traballos de Hermes (1998), foi retomada por d’Alessandro *et al.* (2011), e foi adaptada en Moutinho *et al.* (2011) para medir a distancia entre as variedades do portugués europeo.

A aplicación desta medida de correlación a pares de curvas orixinadas en localizacións distintas permite avaliar de forma cuantitativa a similitude entre elas. En particular, calculando a mediana das correlacións ponderadas entre cada par de curvas interrogativas coa mesma estrutura e de puntos distintos, obtívose unha medida da proximidade prosódica entre as diferentes localizacións. A Figura 3¹ representa graficamente esta medida de proximidade: se dous puntos teñen unha correlación moi alta (próxima a 1 ou a -1), o cadrado da Figura 3 que teñan asociado será de cor vermella escura (se a correlación é positiva) ou azul escura (se a correlación é negativa). Por exemplo, o cadrado que se atopa na intersección da fila sinalada como ob1 no eixe vertical e a columna sinalada como ov1 no eixe horizontal representa a mediana das correlacións ponderadas entre frases coa mesma estrutura de Ribadeo e A Mezquita. A cor do cadrado indica que a correlación entre os dous puntos é positiva e moi alta.

¹ Na Figura 3 os puntos de enquisa aparecen ordenados segundo os agrupamentos que se obtiveron na análise de conglomerados que se describe máis adiante. Empregouse esta reordenación na gráfica para resaltar as diferenzas que hai en termos de correlación entre os clústers que se obtíñen e as semellanzas que se observan, tamén en termos de correlación, entre os puntos que pertencen a un mesmo clúster.

A partir da matriz de correlacións representada na Figura 3, construíuse unha matriz de distancias prosódicas entre os puntos. Esta matriz de distancias obtívose calculando as distancias euclidianas entre cada par de filas da matriz de correlacións, e represéntase graficamente na Figura 4². Continuando co exemplo anterior, a distancia entre Ribadeo e A Mezquita obtense ao calcular a distancia euclidiana entre as filas de ob1 e ov1 na matriz de correlacións. Ese valor, representado na Figura 4 no cadro que resulta da intersección da fila ob1 e a columna ov1, toma un valor próximo a cero por tratarse de dous puntos próximos en termos de distancia prosódica.

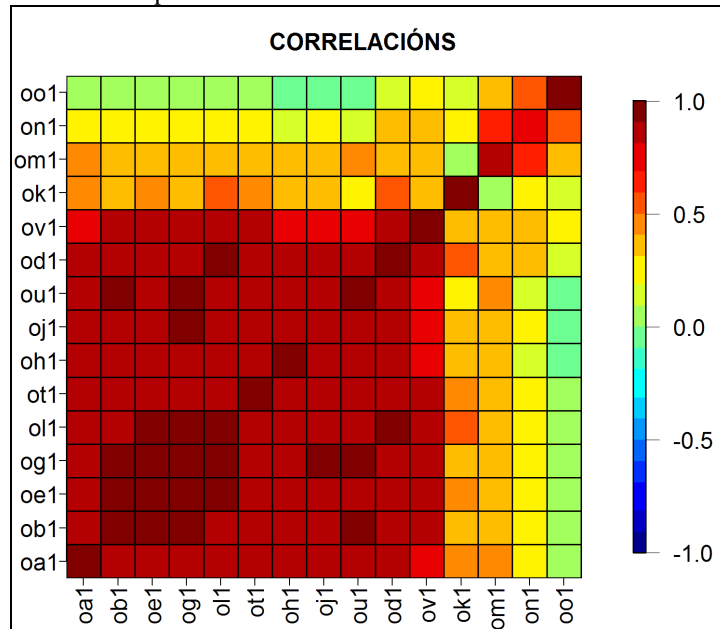


Figura 3: Medianas das correlacións ponderadas para cada par de frases interrogativas coa mesma estrutura

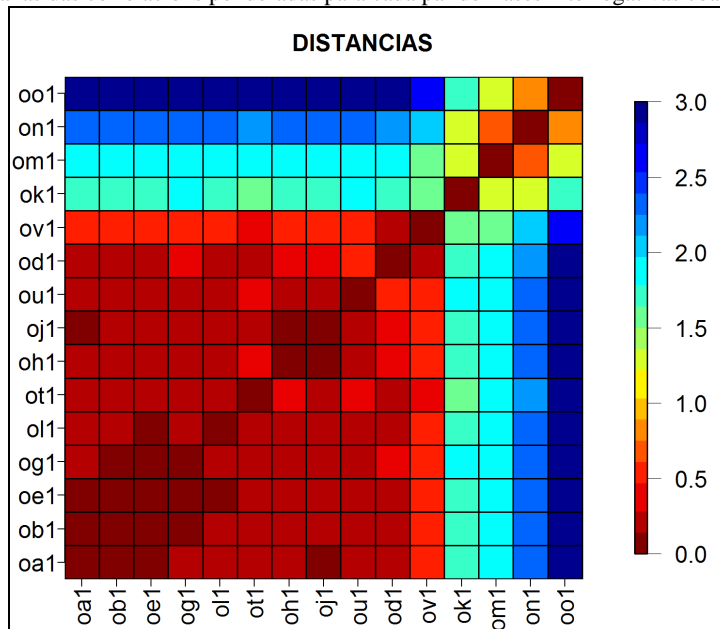


Figura 4: Distancias para cada par de frases interrogativas coa mesma estrutura

² Como acontecía na representación gráfica das correlacións, na Figura 4 os puntos de enquisa aparecen ordenados segundo os agrupamentos xerados pola análise de conglomerados para destacar as diferenzas entre os clústers e as semellanzas entre os puntos que pertencen a un mesmo clúster, neste caso en termos de distancia.

Finalmente, a matriz de distancias empregouse para aplicar un método de agrupamento xerarquizado divisivo que agrupou os puntos do corpus en clústers segundo a súa proximidades en termos de distancia prosódica. O agrupamento obtido tras a aplicación desta metodoloxía pode verse no dendrograma da Figura 5. Para os tres primeiros clústers que se obtiveron coa análise de conglomerados, xerouse un novo mapa xeográfico dos puntos de enquisa, coloreando cada un deles segundo o clúster ao que pertence (ver Figura 6).

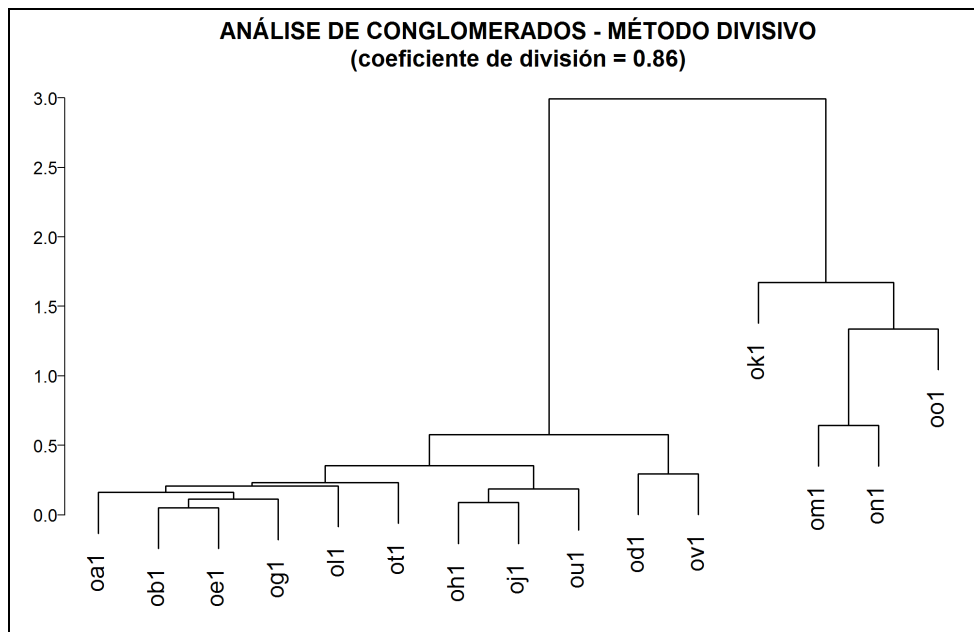


Figura 5: Dendrograma derivado da análise clúster divisiva para as frases interrogativas

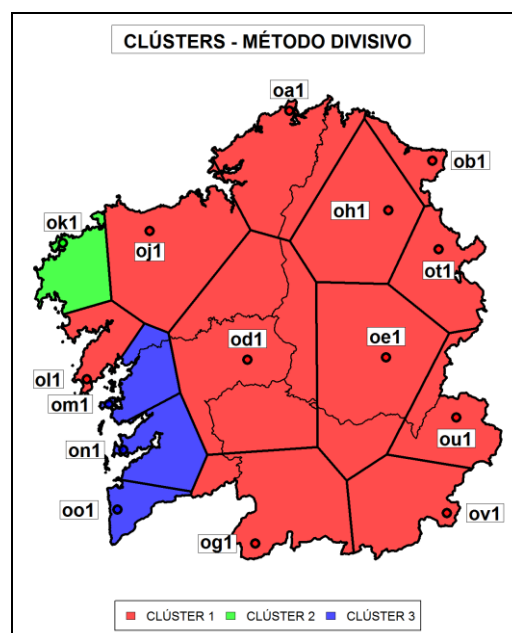


Figura 6: Localización xeográfica dos tres primeiros clústers obtidos coa análise clúster divisiva para as frases interrogativas

4. Conclusión e liñas futuras

Os resultados da análise dialectométrica permiten definir clústers de variedades segundo a distancia que presentan as entoacións dos enunciados interrogativos. De aí, pódese xerar un primeiro mapa das variedades prosódicas do galego, no que se observa unha ampla área, que correspondería co denominado galego común (Fernández Rei *et al.*, 2005), unha área nas Rías Baixas e unha pequena zona da Costa da Morte. Confírmase, pois, as descrições feitas sobre a base das análises acústicas por Fernández Rei & Escourido (2008), onde, tomando só datos de puntos situados na costa galega, se establecían tres patróns de interrogativas para o galego, que se corresponden cos aquí sinalados. En consecuencia, constátase que o método de análise dialectométrica utilizado é válido para o tratamento de datos deste tipo e para o establecemento de áreas dialectais. Outra das súas virtudes é a súa capacidade para o procesamento dunha grande cantidade de datos prosódicos, de xeito que vai permitir no futuro o estudo de áreas meirandes, así como a inclusión dun maior número e variedade de estruturas e modalidades.

O presente estudo abre, polo tanto, novas liñas de traballo a desenvolver no futuro. A primeira delas baséase na ampliación do corpus, engadindo modalidades oracionais e estruturas agora non representadas e completando os datos nalgunhas localizacións; este novo corpus permitiría constatar se efectivamente se manteñen as isoglosas aquí presentadas ou se hai unha distribución dialectal dispar en función da estrutura e/ou modalidade analizada. En segundo lugar, abre a vía para a realización de tests perceptivos co obxectivo de comparar os seus resultados cos obtidos na presente análise: esta comparativa permitiría analizar se as distancias percibidas polos falantes se corresponden ou non coas distancias prosódicas calculadas na análise dialectométrica. Finalmente, sería de grande interese a exploración doutras técnicas estatísticas, como a análise de datos funcionais, no contexto presentado neste traballo.

5. Agradecementos

Este traballo foi financiado pola *Rede TECANDALI - Tecnoloxías e Análise dos Datos Lingüísticos* (CN 2012/179) da Consellería de Cultura, Educación e Ordenación Universitaria (Xunta de Galicia), e o proxecto *Estudio perceptivo de la variación prosódica dialectal del gallego* (FFI2009-12738) do Ministerio de Ciencia e Innovación (Goberno de España).

Referencias

- d'Alessandro, Christophe; Rilliard, Albert & Le Beux, Sylvain. (2011) Chirinomic stylization of intonation. *Journal of the Acoustical Society of America* 129 (3), pp. 1594-1604.
- Fernández Rei, Elisa & Escourido Pernas, Ana (2008) La entonación de las interrogativas totales a lo largo de la costa gallega. In. Adrian Turculet (ed) *La variation diatopique de l'intonation dans le domaine roumain et roman*. Iași: Editura Universităţii “Alexandru Ioan Cuza”, pp. 151-166.
- Fernández Rei, Elisa; González González; Manuel, Xuncal Pereira; Luís & Caamaño Varela, Maruxa (2005) Achega á entoación dunha fala do centro de Galicia. Contribución para o AMPER. *Géolinguistique*, Hors de Série 3, pp. 87-102.
- Hermes, Dik J. (1998) Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language and Hearing Research* 41, pp. 73-82.
- Moutinho, Lurdes de Castro; Coimbra, Rosa Lidia; Rilliard, Albert & Romano, Antonio (2011) Measure de la variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental* 20, pp. 33-55.
- R Core Team (2013) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.