

*Coloquio sobre el Diccionario de colombianismos,
Embajada de Colombia en Tokio, 29 de junio, 2021.*

La forma colombiana de pregunta confirmativa «¿cierto?» en corpus, mapa y dialectometría

Hiroto Ueda, Universidad de Tokio

1. Introducción

PRESEEA

«*Proyecto para el Estudio Sociolingüístico del Español de España y de América*», Moreno Fernández 2021, Moreno Fernández y Cestero 2020.

<http://preseea.linguas.net/>

LYNEAL «*Letras y Números en Análisis Lingüísticos*», con los datos concretos de PRESEEA.

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/preseea.htm>

Cuestiones de investigación:

- ¿Cuál es la frecuencia de las preguntas confirmativas empleadas en el corpus?
- ¿Cuál es la distribución geolectal de las preguntas confirmativas, en atención a su frecuencia de empleo?

2. Análisis

2. 1. Distribución de frecuencia

Pereira (Colombia):

- * lo primero que hago es como darle gracias a Dios *¿cierto?* / eeh lo segundo es como jugar con mi hija *
- * / ya que es una fecha tan especial en familia *¿verdad?* \$ si puede ser // pero igual hay períodos de / *
- * y al cabo un es como un estado de / de descanso *¿no?* / por ejemplo la semana santa / ahorita a mitad *
- * de salsa? / podría ser como una salsa de piña *¿cierto?* \$ si algo así / una salsa rosadas / cosas así
- * y las malas noticias que se oyen en la radio *¿verdad?* / ¿o usted no tiene? \$ no yo para eso no tengo *

Alcalá de Henares (España):

- * ha ¿me comprende lo que le quiero decir? / **¿eh?** / \$ sí sí sí
- ¿y tenía cáscaras? \$ **¿eh?** \$ ¿tenía cáscaras? /
- * un poquito la de la para la la selectividad **¿no?** de de de \$ para gen... no para gente de *
- sí sí era un \$ era selectivo **¿no?** \$ a estilo de cafetería / que allí / yo si pas*
- * mierda para ellos! \$ una mierda para ellos **¿verdad?** una mierda para ellos // \$ aquí en Alcalá no *

Ueda, 2017b, 9 ciudades:

¿no? (frecuencia: 4608), *¿sí?* (874), *¿eh?* (467), *¿cierto?* (341), *¿verdad?* (235), *¿cachái(s)?* (173), *¿sabes?* (158), *¿viste?* (116), *¿hm?* (101), *¿ah?* (57), *¿entiendes?* (27), *¿ves?* (16), *¿as?* (14), *¿te acuerdas?* (9), *¿comprendes?* (9), *¿te parece?* (8), *¿oíste?* (7), *¿vale?* (5), *¿no es cierto?* (5), *¿no es verdad?* (5), *¿no le parece?* (5), *¿ta?* (5), *¿correcto?* (4), *¿te das cuenta?* (4), *¿de acuerdo?* (3), *¿estamos?* (2), *¿no cree?* (2), *¿mirá?* (1), *¿ok?* (1).

En el estudio actual, 22 comunidades:

Nueve formas más frecuentes: *¿no?*, *¿sí?*, *¿eh?*, *¿cierto?*, *¿verdad?*, *¿cachái(s)?*, *¿sabes?*, *¿viste?*, *¿hm?*

N	FA	<i>¿cacháis?</i>	<i>¿cierto?</i>	<i>¿eh?</i>	<i>¿hm?</i>	<i>¿no?</i>	<i>¿sabes?</i>	<i>¿sí?</i>	<i>¿verdad?</i>	<i>¿viste?</i>
1	CU-La Habana	0	0	5	0	308	1	5	8	0
2	MX-Mexicali	0	0	16	0	326	1	129	65	1
3	MX-Monterrey	0	0	41	0	249	1	139	162	0
4	MX-Guadalajara	0	0	11	0	402	7	118	162	0
5	MX-México, D. F.	0	0	92	0	2085	0	100	74	0
6	GU-Guatemala	0	0	0	0	124	0	6	986	0
7	CO-Barranquilla	0	22	3	7	176	4	83	113	8
8	CO-Medellín	0	317	0	0	21	0	33	0	0
9	CO-Pereira	0	61	0	0	36	0	26	19	0
10	CO-Cali	0	30	0	0	104	0	8	0	0
11	VE-Caracas	0	1	0	0	308	34	38	39	0
12	PE-Lima	0	0	3	0	1922	0	15	1	0
13	CH-Santiago de Chile	164	24	2	0	59	0	55	1	7
14	UR-Montevideo	0	0	24	0	413	2	159	18	110
15	ES-Santander	0	0	240	0	549	60	79	92	0
16	ES-Santiago	0	0	137	24	724	513	97	12	0
17	ES-Alcalá	0	0	94	0	342	48	142	12	0
18	ES-Madrid	0	0	164	61	1073	91	354	30	0
19	ES-Valencia	0	0	121	0	244	9	23	3	0
20	ES-Sevilla	0	0	106	0	1000	37	110	15	0
21	ES-Granada	0	0	29	0	363	5	70	7	0
22	ES-Málaga	0	0	117	0	1106	29	186	11	0

Tabla 1. Preguntas confirmativas en 22 muestras del corpus. Frecuencia absoluta.

N	FN.Pl.:100000	<i>¿cacháis?</i>	<i>¿cierto?</i>	<i>¿eh?</i>	<i>¿hm?</i>	<i>¿no?</i>	<i>¿sabes?</i>	<i>¿sí?</i>	<i>¿verdad?</i>	<i>¿viste?</i>
1	CU-La Habana	0	0	3.8	0	231.8	0.8	3.8	6	0
2	MX-Mexicali	0	0	14.5	0	296.3	0.9	117.2	59.1	0.9
3	MX-Monterrey	0	0	19.2	0	116.7	0.5	65.2	76	0
4	MX-Guadalajara	0	0	7.6	0	277.4	4.8	81.4	111.8	0
5	MX-México, D. F.	0	0	43.8	0	992.6	0	47.6	35.2	0
6	GU-Guatemala	0	0	0	0	101.3	0	4.9	805.1	0
7	CO-Barranquilla	0	13.8	1.9	4.4	110.3	2.5	52	70.8	5
8	CO-Medellín	0	255.1	0	0	16.9	0	26.6	0	0
9	CO-Pereira	0	57.7	0	0	34.1	0	24.6	18	0
10	CO-Cali	0	29.5	0	0	102.2	0	7.9	0	0
11	VE-Caracas	0	0.6	0	0	187.2	20.7	23.1	23.7	0
12	PE-Lima	0	0	2.3	0	1502.6	0	11.7	0.8	0
13	CH-Santiago de Chile	94.8	13.9	1.2	0	34.1	0	31.8	0.6	4
14	UR-Montevideo	0	0	18.8	0	323.9	1.6	124.7	14.1	86.3
15	ES-Santander	0	0	146	0	334	36.5	48.1	56	0
16	ES-Santiago	0	0	63.5	11.1	335.5	237.7	45	5.6	0
17	ES-Alcalá	0	0	73.6	0	267.6	37.6	111.1	9.4	0
18	ES-Madrid	0	0	100.9	37.5	660.3	56	217.8	18.5	0
19	ES-Valencia	0	0	111.3	0	224.5	8.3	21.2	2.8	0
20	ES-Sevilla	0	0	75.5	0	712.4	26.4	78.4	10.7	0
21	ES-Granada	0	0	24.1	0	302.2	4.2	58.3	5.8	0
22	ES-Málaga	0	0	70.2	0	663.6	17.4	111.6	6.6	0

Tabla 2. Preguntas confirmativas en 22 muestras del corpus. Frecuencia normalizada por cien mil palabras.

2. 2. Mapa lingüístico

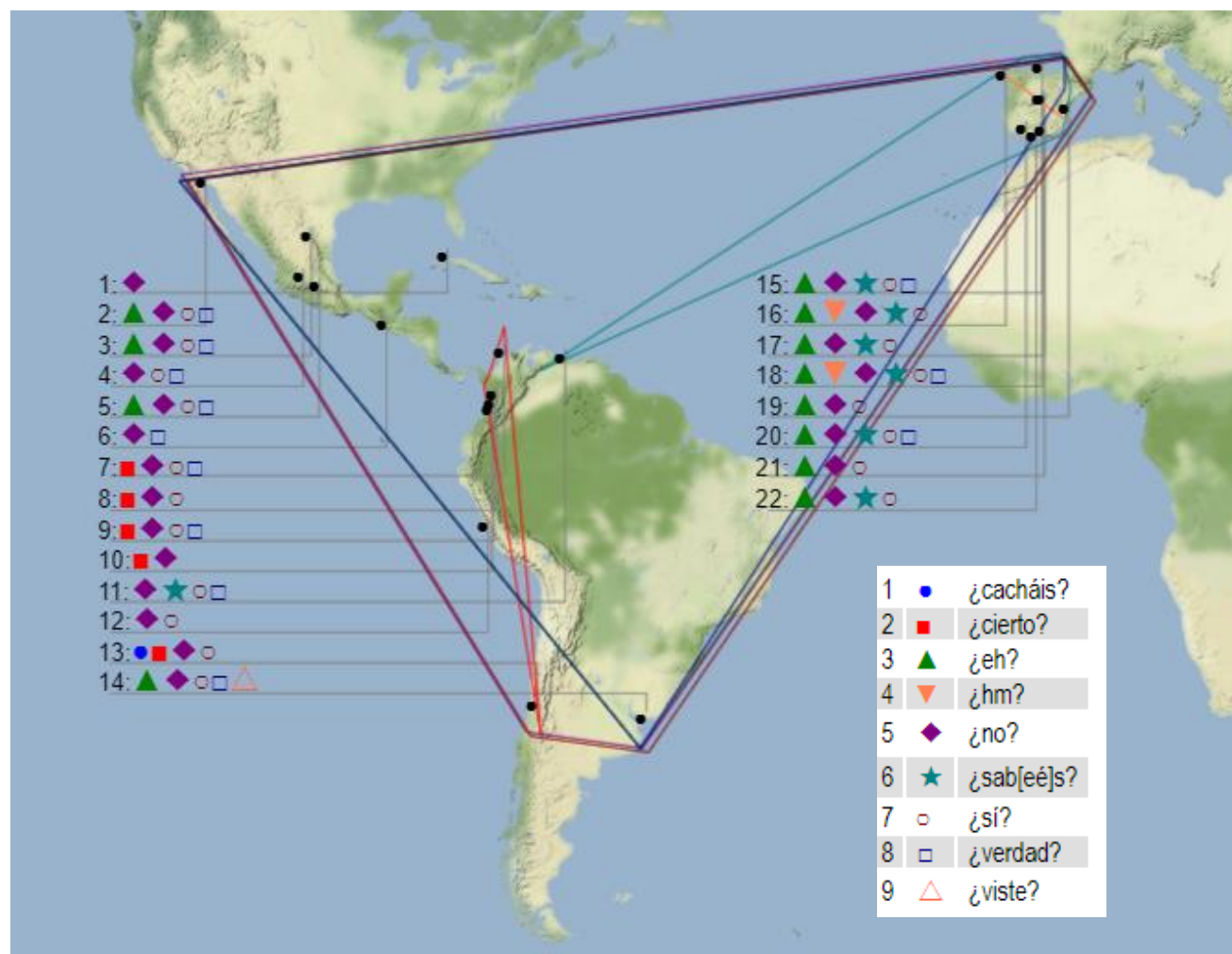


Fig. 1. Preguntas confirmativas en 22 muestras del corpus. Frecuencia mayor de 10.

2. 3. Análisis de correspondencia

FN.Pl.:100000	¿cierto?	¿eh?	¿no?	¿sabes?	¿sí?	¿verdad?
CU-La Habana		3.8	231.8	.8	3.8	6.0
MX-Mexicali		14.5	296.3	.9	117.2	59.1
MX-Monterrey		19.2	116.7	.5	65.2	76.0
MX-Guadalajara		7.6	277.4	4.8	81.4	111.8
MX-México, D. F.		43.8	992.6		47.6	35.2
GU-Guatemala			101.3		4.9	805.1
CO-Barranquilla	13.8	1.9	110.3	2.5	52.0	70.8
CO-Medellín	255.1		16.9		26.6	
CO-Pereira	57.7		34.1		24.6	18.0
CO-Cali	29.5		102.2		7.9	
VE-Caracas	.6		187.2	20.7	23.1	23.7
PE-Lima		2.3	1502.6		11.7	.8
CH-Santiago de Chile	13.9	1.2	34.1		31.8	.6
UR-Montevideo		18.8	323.9	1.6	124.7	14.1
ES-Santander		146.0	334.0	36.5	48.1	56.0
ES-Santiago		63.5	335.5	237.7	45.0	5.6
ES-Alcalá		73.6	267.6	37.6	111.1	9.4
ES-Madrid		100.9	660.3	56.0	217.8	18.5
ES-Valencia		111.3	224.5	8.3	21.2	2.8
ES-Sevilla		75.5	712.4	26.4	78.4	10.7
ES-Granada		24.1	302.2	4.2	58.3	5.8
ES-Málaga		70.2	663.6	17.4	111.6	6.6

Tabla 3. Preguntas confirmativas principales en 22 ciudades. Frecuencia normalizada.

Correspondencia	<i>¿sabes?</i>	<i>¿eh?</i>	<i>¿no?</i>	<i>¿verdad?</i>	<i>¿sí?</i>	<i>¿cierto?</i>
ES-Santiago	237.7	63.5	335.5	5.6	45.0	
ES-Valencia	8.3	111.3	224.5	2.8	21.2	
ES-Santander	36.5	146.0	334.0	56.0	48.1	
PE-Lima		2.3	1502.6	.8	11.7	
CU-La Habana	.8	3.8	231.8	6.0	3.8	
MX-México, D. F.		43.8	992.6	35.2	47.6	
ES-Sevilla	26.4	75.5	712.4	10.7	78.4	
GU-Guatemala			101.3	805.1	4.9	
ES-Málaga	17.4	70.2	663.6	6.6	111.6	
VE-Caracas	20.7		187.2	23.7	23.1	.6
ES-Granada	4.2	24.1	302.2	5.8	58.3	
ES-Alcalá	37.6	73.6	267.6	9.4	111.1	
ES-Madrid	56.0	100.9	660.3	18.5	217.8	
MX-Guadalajara	4.8	7.6	277.4	111.8	81.4	
MX-Monterrey	.5	19.2	116.7	76.0	65.2	
MX-Mexicali	.9	14.5	296.3	59.1	117.2	
UR-Montevideo	1.6	18.8	323.9	14.1	124.7	
CO-Barranquilla	2.5	1.9	110.3	70.8	52.0	13.8
CH-Santiago de Chile		1.2	34.1	.6	31.8	13.9
CO-Cali			102.2		7.9	29.5
CO-Pereira			34.1	18.0	24.6	57.7
CO-Medellín			16.9		26.6	255.1

Tabla 4. Análisis de correspondencia. Distribución diagonalizada.

2. 4. Análisis de conglomerados

(A) Formas

Pregunta confirmativa	<i>¿sabes?</i>	<i>¿eh?</i>	<i>¿no?</i>	<i>¿verdad?</i>	<i>¿sí?</i>	<i>¿cierto?</i>
Valor de peso	-.239	-.221	-.162	-.150	.034	4.600

Tabla 5. Preguntas confirmativas. Valor de peso.

Distancia (*¿sabes?*, *¿eh?*):

$$|(-.239) - (-.221)| = .018$$

Distancia	<i>¿sabes?</i>	<i>¿eh?</i>	<i>¿no?</i>	<i>¿verdad?</i>	<i>¿sí?</i>	<i>¿cierto?</i>
<i>¿sabes?</i>	-	.018	.077	.089	.274	4.839
<i>¿eh?</i>	.018	-	.059	.071	.255	4.821
<i>¿no?</i>	.077	.059	-	.012	.196	4.762
<i>¿verdad?</i>	.089	.071	.012	-	.184	4.750
<i>¿sí?</i>	.274	.255	.196	.184	-	4.566
<i>¿cierto?</i>	4.839	4.821	4.762	4.750	4.566	-

Tabla 6. Preguntas confirmativas. Matriz de distancia.

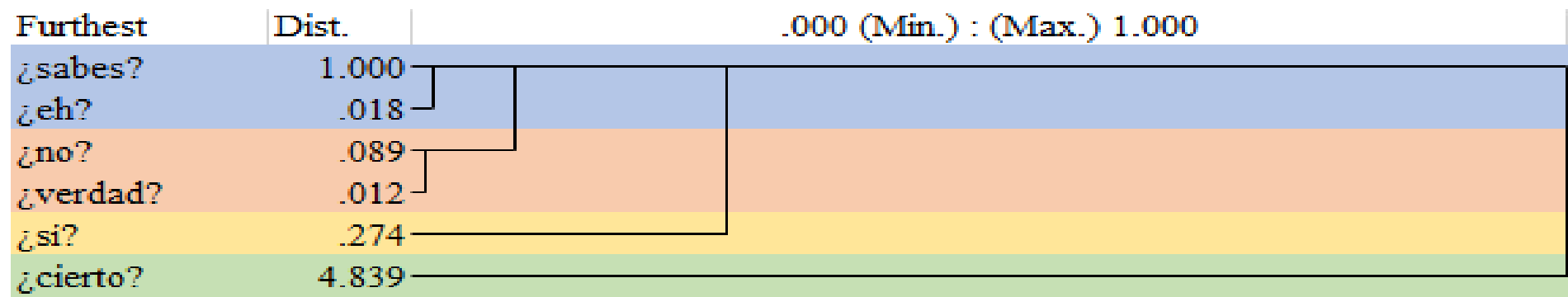


Fig. 2. Preguntas confirmativas. Análisis de conglomerados.

(A) Ciudades

Row	Xn
ES-Santiago	-.221
ES-Valencia	-.207
ES-Santander	-.200
PE-Lima	-.196
CU-La Habana	-.195
MX-México, D. F.	-.190
ES-Sevilla	-.185
GU-Guatemala	-.183
ES-Málaga	-.174
VE-Caracas	-.168
ES-Granada	-.167
ES-Alcalá	-.162
ES-Madrid	-.160
MX-Guadalajara	-.156
MX-Monterrey	-.142
MX-Mexicali	-.140
UR-Montevideo	-.138
CO-Barranquilla	.173
CH-Santiago de Chile	.883
CO-Cali	1.042
CO-Perera	2.339
CO-Medellin	4.780

Tabla 7. Valor de peso de ciudades.

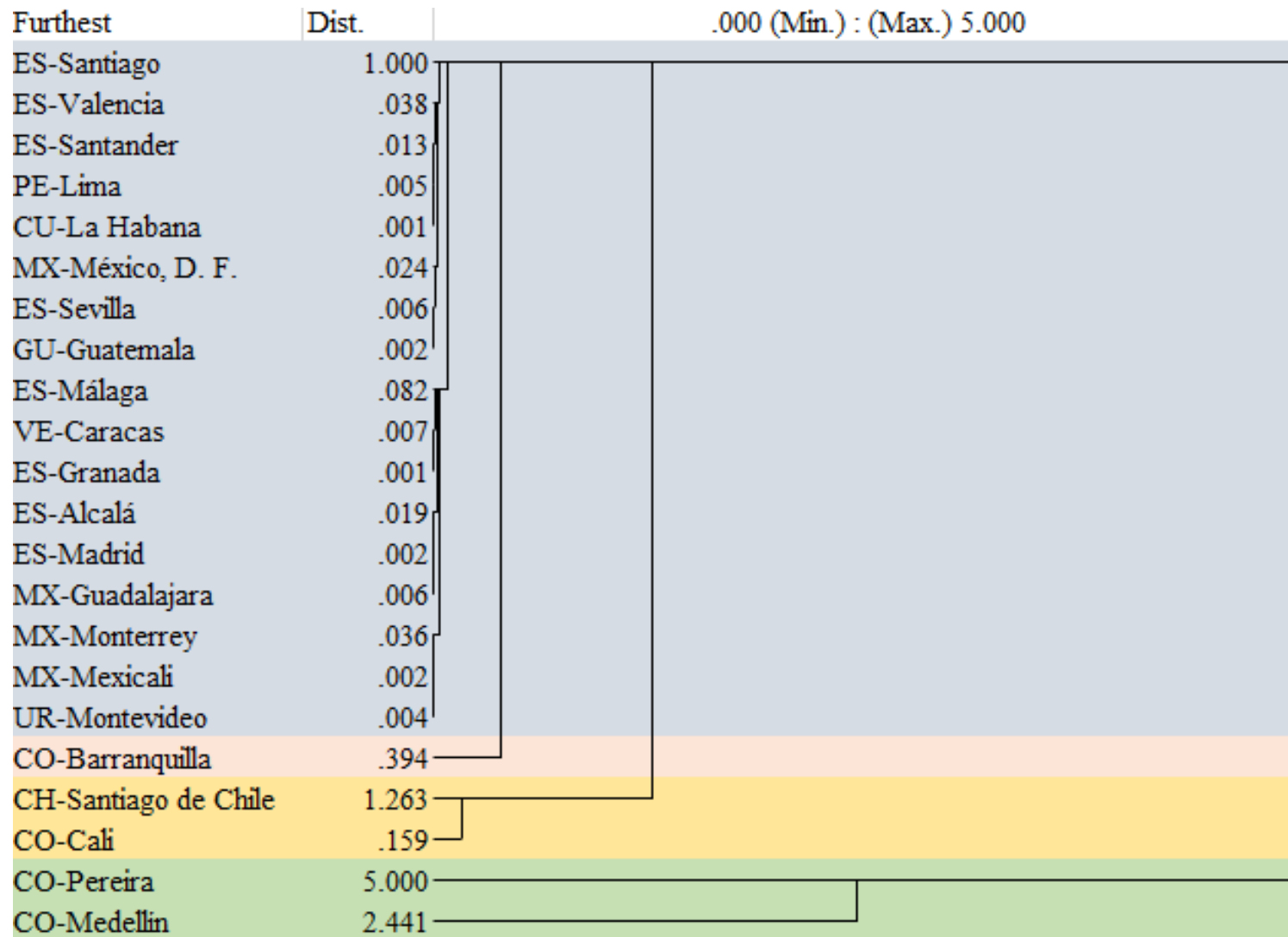


Fig. 3. Análisis de conglomerados. Ciudades.

2. 5. Distribución diagonal clasificada

Correspondence	¿sabes?	¿eh?	¿no?	¿verdad?	¿sí?	¿cierto?
ES-Santiago	237.7	63.5	335.5	5.6	45.0	
ES-Valencia	8.3	111.3	224.5	2.8	21.2	
ES-Santander	36.5	146.0	334.0	56.0	48.1	
PE-Lima		2.3	1502.6	.8	11.7	
CU-La Habana	.8	3.8	231.8	6.0	3.8	
MX-México, D. F.		43.8	992.6	35.2	47.6	
ES-Sevilla	26.4	75.5	712.4	10.7	78.4	
GU-Guatemala			101.3	805.1	4.9	
ES-Málaga	17.4	70.2	663.6	6.6	111.6	
VE-Caracas	20.7		187.2	23.7	23.1	.6
ES-Granada	4.2	24.1	302.2	5.8	58.3	
ES-Alcalá	37.6	73.6	267.6	9.4	111.1	
ES-Madrid	56.0	100.9	660.3	18.5	217.8	
MX-Guadalajara	4.8	7.6	277.4	111.8	81.4	
MX-Monterrey	.5	19.2	116.7	76.0	65.2	
MX-Mexicali	.9	14.5	296.3	59.1	117.2	
UR-Montevideo	1.6	18.8	323.9	14.1	124.7	
CO-Barranquilla	2.5	1.9	110.3	70.8	52.0	13.8
CH-Santiago de Chile		1.2	34.1	.6	31.8	13.9
CO-Cali			102.2		7.9	29.5
CO-Pereira			34.1	18.0	24.6	57.7
CO-Medellin			16.9		26.6	255.1

Tabla 8. Distribución diagonal clasificada de frecuencia.

3. Final

- En este breve estudio sobre distribución geolectal de las preguntas confirmativas documentadas en el corpus de muestras de habla PRESEEA, de 22 comunidades de España y América, espero haber demostrado que existe una continuidad en la distribución de frecuencia de su uso en la amplia geografía de la lengua española.
- Precisamente al analizar la distribución geográfica, hemos observado la peculiaridad relevante de *¿cierto?*, cuyo uso se concentra en las ciudades colombianas, a las que se suman Santiago de Chile y, en mucha menor medida, Caracas (Venezuela).

- Para realizar los análisis estadísticos, se ha utilizado el *análisis de correspondencia* y el *análisis de conglomerados*, métodos tradicionales de la estadística multivariante.
- He propuesto el uso de la matriz de distancia a partir del vector de valores de peso de formas lingüísticas y ciudades, que se obtienen en el *análisis de correspondencia*, para conseguir la distribución diagonal, ahora clasificada por el *análisis de conglomerados*.

Referencia

Asociación de Academias de la Lengua Española. (2010) *Diccionario de americanismos*. Madrid: Santillana Ediciones.

Beinhauer, W. (1978) *El español coloquial*. Madrid. Gredos.

Briz Gómez, A. (1998) *El español coloquial en la conversación. Esbozo de pragmática*. Barcelona: Ariel.

_____. / Grupo Val.Es.Co. Universidad de Valencia (2000) *Diccionario de partículas discursivas del español*. <http://www.dpde.es/>

Casado Velarde, M. (1998) "Lingüística del texto y marcadores del discurso", en Martín Zorraquino, M. A. / Montolío Durán, E. *Los mardadores del discurso. Teoría y análisis*. Madrid Arco / Libros.

Cestero Mancera, A. M. (2000) "La función fática del lenguaje en el

discurso y en la conversación", *Actas del IV Congreso de Lingüística General*, Cádiz, vol. II, pp. 617-629.

_____. (2003) "El funcionamiento de los apéndices interrogativos en la conversación y en el discurso académico", Castillo Martínez, Cristina / Lucía Megías, J. M. (eds.) *Decíamos ayer... Estudios de alumnos en honor a María Cruz García de Enterría*. Universidad de Alcalá. pp. 83-127.

Cortes Rodríguez, L. / Camacho Adarve, M. M. (2005) *Unidades de segmentación y marcadores del discurso: Elementos esenciales en el procesamiento discursivo oral*. Madrid. Arco / Libros.

Fuentes Rodríguez, C. (1990) "Apéndices con valor apelativo", P. Carbonero Cano (coord.) y M. T. Palet Plaja (ed.) *Sociolingüística andaluza*, 5, *Habla de Sevilla y hablas americanas*. Sevilla:

Universidad de Sevilla, pp.171-196.

Goebel, H. (2013) “La dialectometrización del ALPI: rápida presentación de los resultados”, *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas*, E. Casanova Herrero, C. Calvo Rigual (eds.) Valencia 2010, De Gruyter: pp. 143-154.

Moreno Fernández, F. (2021) “Metodología del «Proyecto para el estudio sociolingüístico del español de España y de América»”. *Documentos PRESEEA de Investigación 1*. DOI: 10.37536/PRESEEA.2021.doc1

Moreno Fernández F. and Ueda, H. (2018) «Cohesion and Particularity in the Spanish Dialect Continuum», *Open Linguistics*, 4: 722–742.

<https://www.degruyter.com/view/journals/oplí4/1/article-p722.xml>

Moreno Fernández, F. y A. M. Cestero Mancera. 2020. El proyecto PRESEEA: desarrollos analíticos, *Verba*, anexo 80 (2020), editado por

Á. J. Gallego y F. Roca Urgell. *Dialectología digital del español*. pp. 119-138.

<https://dx.doi.org/10.15304/9788418445316>

Paredes, F. (2002) “Dialectometría en Venezuela”, *XIII Congreso Internacional de la Asociación de Lingüística y Filología*, San José: pp. 327-337.

Quilis, A. (1981) *Fonética acústica de la lengua española*. Madrid. Gredos. p.440-442.

_____ (1993) *Tratado de fonología y fonética españolas*. Madrid. Gredos. pp. 451-452.

Ueda, H. (1995). *Problemas y métodos de la lexicología de variación: Palabras y cosas de la vida urbana*. Tesis doctoral (Universidad de Alcalá)

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyuchiritesis/tesis-doctoral.pdf>

_____. (2017a) *Análisis de datos cuantitativos para estudios lingüísticos*.

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengó4-numeros/doc/numeros-es.pdf>

_____. (2017b) “Media proporcional y frecuencia probabilística de las preguntas confirmativas españolas. Macronálisis y micronálisis de los datos de PRESEEA en el sistema LYNEAL”, *XVIII Congreso Internacional de la Asociación de Lingüística y Filología de América Latina (ALFAL)*, Universidad Nacional de Colombia, 24 - 28 de julio de 2017.

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyuchiripregunta-confirmativa/pregunta-confirmativa.pdf>

_____. (2019) «Investigación de la variación léxica española. Método de la zonificación multivariante y visualización cartográfica», *I Congress of Geolinguistic Society of Japan*, Aoyama Gakuin University, October 6, 2019.

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/chiri/kukakuka/kukaku-ka-sp.pdf>

Muchas gracias.