


February 2024

Discursive Variation in Texas Spanish: A Dialectometric Approach

James Ramsburg

University of Minnesota - Twin Cities, ramsburg@umn.edu

Follow this and additional works at: <https://engagedscholarship.csuohio.edu/ijlasw>

 Part of the [Language Description and Documentation Commons](#), and the [Spanish Linguistics Commons](#)

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Ramsburg, James (2024) "Discursive Variation in Texas Spanish: A Dialectometric Approach," *International Journal of the Linguistic Association of the Southwest*: Vol. 42: No. 1, Article 6. Available at: <https://engagedscholarship.csuohio.edu/ijlasw/vol42/iss1/6>

This Research Paper is brought to you for free and open access by EngagedScholarship@CSU. It has been accepted for inclusion in International Journal of the Linguistic Association of the Southwest by an authorized editor of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

DISCURSIVE VARIATION IN TEXAS SPANISH: A DIALECTOMETRIC APPROACH

James Ramsburg
University of Minnesota

ABSTRACT. This study presents a comprehensive analysis of Spanish discourse in Texas, focusing on the variation of discourse markers and fillers across different regions. Utilizing a dialectometric approach, the study combines quantitative and corpus linguistics methods to analyze speech samples from nine cities across the state. The investigation identifies significant geolinguistic variation, revealing six distinct dialect clusters and larger dialect cluster divisions within Texas Spanish discourse. Moreover, the study examines the use of English discourse markers in Texas Spanish, finding notable variations, particularly near the Mexican border. This research contributes to understanding the linguistic diversity of Texas Spanish and demonstrates the utility of dialectometric methods in analyzing language variation and change.

1. INTRODUCTION. Texas Spanish has been characterized from a variety of perspectives, whether that be descriptively (Cardenas 1970), historically (Lipski 1988), or sociolinguistically (Chaston 1996, Martínez 2003, Bayley et al. 2012, Carter and Wolford 2016). However, to date, no study has sought to produce a dialectological account of Spanish discourse in Texas. By combining methods of quantitative and corpus linguistics, the present study seeks to investigate the dynamics of how Spanish discourse features, namely discourse markers and fillers, vary across multiple regions in Texas. To do so, this study utilizes a corpus of contemporary Texas Spanish and evaluates it via a dialectometric approach (Speelman and Geeraerts 2009), applying various computational methods to analyze this variation from a geolinguistic perspective.

2. Previous research and theoretical framework

2.1 DISCOURSE MARKERS AND FILLERS. Discourse markers are a functional class of linguistic items that indicate a speaker's prospective communicative intents. They are also known as discourse particles or pragmatic markers (Fraser 1996). Despite the fact that they typically do not alter a sentence's propositional content, they are crucial for discourse organization because they

reveal the speaker's perspectives on the proposition being made and facilitate the formulation of pragmatic inferences (Furko 2017: 2). Although these markers have been extensively examined in the field since the late 20th century, there is no unified typology of them. In the present study, Spanish discourse markers will be categorized using Robles and Bertomeu's (2017) modified version of Portolés' (2007) five-type, role-based taxonomy: connectors (those that link two units of speech semantically and pragmatically); exclamatory markers (those that convey surprise or a strong emotional response to the interlocutor); acceptance markers (those that convey agreement with the interlocutor); alterity focusers (those that serve to focus the interlocutor's attention to the speaker's argument); and fillers (those that fill pauses between units of discourse):

Connectors	<i>Entonces, eh, pienso yo que esas cosas son importantes y para mí [...]</i> (AF133_1975_CT_SP201_3_CI) 'So, eh, I think that those things are important and for me [...]'
Exclamation	<i>Y... y no le gusta leer. Hijole. Le gusta jugar.</i> (AF054_1955_SA_SU201_2_JB) 'And... and she doesn't like to read. Gosh. She likes to play.'
Acceptance	<i>Aha. En un año nuevo, en un treinta y uno de diciembre.</i> (MF110_1958_GV_SU2012_AW) 'Uhuh. On a New Year's, on a 31st of December.'
Alterity Focusers	<i>Oye, me podrías hacer el favor de mandarme este reporte [...]</i> (AM059_1951_SA_SU2012_JB) 'Listen, could you do me the favor of sending me this report [...]'
Fillers	<i>Ehm... no recuerdo ahorita.</i> (MF110_1958_GV_SU2012_AW) 'Ehm... I don't remember right now.'

TABLE 1: SPANISH DISCOURSE MARKERS AND FILLERS

2.2 DISCURSIVE VARIATION IN TEXAS AND THE SPANISH-SPEAKING WORLD. In his descriptive account of Spanish dialects in the United States, Cárdenas (1970) characterizes Texas Spanish in a few general terms: by

the absence of /θ/ and /k/, the absence of the second person plural pronoun and the corresponding verb form, and an absence of the possessive *nuestro* (our). Additionally, he also describes the dialect as uniquely Mexican due to the “overwhelming” migration from the country into Texas (Cárdenas 1970: 21). While this conclusion is logical, and the dialect does share many similarities with Mexican Spanish, the Spanish spoken in Texas is by no means monolithic. For example, in a small-scale sociolinguistic study of San Antonio, Bayley et al. (2012) found that Texans of Puerto Rican origin largely maintained their distinctive rates of subject personal pronoun use despite only comprising a small percentage of the local Spanish-speaking population; in sum, there appears to be more variation in Texas Spanish than is readily apparent. However, despite the numerous linguistic studies in this region, Texas Spanish discourse remains uncharacterized. Curiously, discourse marker usage has been studied in the lesser known contact variety of Texas German. In a corpus-based analysis of German speakers in the state, Boas and Weilbacher (2007) identified that the English discourse marker *you know* co-occurs and shares the same pragmatic functions as its German counterpart *weisst du/weisst(e)*. Furthermore, *you know* has almost entirely replaced *weisst du/weisst(e)* in Texas, indicating more extensive anglicization of this dialect compared to others such as Pennsylvania German (Boas and Weilbacher 2007: 55).

Outside of the Texan context, however, there has been considerable research about the characteristics, functions, and variation of discourse markers/fillers in the Spanish-speaking world. Regarding situations of language contact, Brody (1987) found that Spanish discourse markers were borrowed substantially into Mayan languages, disproving the popular notion at the time that discursive features were unlikely to transfer between languages. As for sociolinguistic variation, previous studies generally concur that the use of discourse markers in Spanish is highly variable. For example, Roggia’s (2012) multivariate analysis of Dominican Spanish encountered that the multifunctional discourse marker *eh* is more frequent among older, higher class, and male speakers. Similarly, Torres and Potowski (2008) examined the use of Spanish and English discourse markers (*so* and *entonces*) across different generations of Spanish speakers in Chicago, Illinois. Although there were individual differences between dialect groups, *so* was an established feature among all speakers and use of this marker correlated with weaker Spanish proficiency. Notably, the Mexican Spanish speakers utilized *so* with the lowest frequency relative to Puerto Rican and MexiRican speakers. Variation has even been discovered among pairings between clicks and discourse markers in Peninsular Spanish (Pinto and Vigil 2020). However, it is worth noting that discourse markers use is not always so varied. In their study of bilingual Spanish-speaking youth in Florida, Said-Mohand (2008)

observed that all speakers regardless of generation utilized the discourse marker *como* in favor of the English equivalent *like*.

2.3 DISCOURSE IN DIALECTOMETRY. Although discourse markers have traditionally been assessed through conversation analysis and corpus-based approaches, some researchers have employed dialectometry to study discursive features. Most relevantly, Bloem (2017) used various statistical methods to create a new computational method by which potential discourse markers can be identified and extracted from corpus data. Although different in scope to the present study, the Bloem investigation validates that dialectometric methods can be successfully applied to discourse phenomena.

3. METHODOLOGY

3.1 RESEARCH QUESTIONS. The present study seeks to further characterize Texas Spanish through the pursuit of two questions:

RQ1: Is there geolinguistic variation among the use of discourse markers and fillers across different locations/regions in Texas?

RQ2: Does the use of loaned and code-switched English discourse markers vary across different locations/regions in Texas?

3.2 DATA. This investigation makes use of the *Spanish in Texas Corpus* (Bullock and Toribio 2013), a diverse collection of Spanish and bilingual Spanish-English speech samples from the state. Compiled from 2011-2013, the corpus consists of 97 interviews (around 500,000 total words) from nine different cities: Austin, Brownsville, Edinburg, El Paso, Houston, Irving, Pearland, San Antonio, and Weslaco. It is also the largest and most geographically diverse corpus of Texas Spanish to date. The conversations were structured as traditional sociolinguistic interviews, and the participants discussed their personal experiences as well as their linguistic habits.



FIGURE 1: LOCATIONS INCLUDED IN THE STUDY

For the purposes of the present study, six interviews¹ were included from seven different locations. Due to a lack of data at some sites, geographically adjacent cities (Edinburg/Weslaco and Houston/Pearland) were combined to make possible the inclusion of every city without sacrificing a balanced dataset. Similarly, sex (three males and three females from each location) and age (two participants <30, two 30-50, and two 50> at each location) were controlled for to ensure a representative sample.² While birthplace was not a factor for participant selection, the speakers in the resulting sample were born in the United States, Mexico, El Salvador, Colombia, Venezuela, and Argentina. As mentioned above, all participants had some degree of Spanish-English bilingual competency. Overall, 41 participants were selected (42% of the total corpus):

¹ There were only five interviews performed in Brownsville.

² There were a few exceptions to this rule (see Edinburg/Weslaco and Houston/Pearland) due to unequal representation of males and females in the corpus.

	Sex		Average Age	Word Count
	Male	Female		
Austin	3	3	35	32,043
El Paso	3	3	49	31,551
Edinburg/Weslaco	4	2	37	27,824
San Antonio	3	3	47	28,187
Brownsville	2	3	32	23,308
Houston/Pearland	2	4	39	29,720
Irving	3	3	52	42,785
Total	20	21	42	215,418

TABLE 2: SAMPLE DEMOGRAPHICS

In terms of the study's tokens, all discourse markers and fillers that appeared in the sample were included. However, it should be noted that some discourse markers appeared in the data (e.g. counter argumentative, sequential, additive) very infrequently and therefore needed to be removed so as to not jeopardize the validity and functionality of the statistical models. Therefore, the final analysis comprised 30 different discourse markers and fillers (5095 tokens in total) belonging to five different categories:

Connector	Exclamation	Acceptance	Alterity Focuser	Filler
<i>después</i>	<i>ah</i>	<i>aha</i>	<i>Fijate</i>	<i>Ahh</i>
<i>entonces</i>	<i>ay</i>	<i>okay</i>	<i>mira</i>	<i>bueno</i>
<i>luego</i>	<i>hijole</i>	<i>vale</i>	<i>¿no?</i>	<i>ehh</i>
<i>so</i>	<i>oh</i>	<i>yeah</i>	<i>oye</i>	<i>ehm</i>
	<i>wow</i>		<i>¿verdad?</i>	<i>este</i>
	<i>órale</i>			<i>I guess</i>
				<i>like</i>
				<i>mmm</i>
				<i>pues</i>
				<i>uh</i>
				<i>uhm</i>

TABLE 3: TOKENS

3.3 PROCEDURE. After coding for each token in each corresponding location, I utilized the dialect analysis software Gabmap (Nerbonne et al. 2011) to assess the geolinguistic variation among the data. This application allows researchers to physically map linguistic data by measuring both geographic and linguistic distances between languages/dialects. Due to the large range of frequencies among the tokens, I processed the data using the Gewichteter Identitätswert method (Goebl 1984). This approach assigns more value to infrequent words rather than frequent ones, ensuring that the sample's less frequent tokens were not treated as noise.³ The data were analyzed using multidimensional scaling (MDS) and cluster analysis. MDS is a statistical technique that provides a visual representation of the distances between a set of objects. When applied to linguistic difference data, this method plots the given geographic locations onto a two-dimensional axis. Dialect areas are then identified by grouping the locations which appear closest to each other on the axis. Dialectometric cluster analysis also attempts to identify dialect areas by placing locations into groups based on their linguistic similarity. However, discrete clustering has been shown to be unreliable (Kleiweg et al. 2004, Nerbonne et al. 2008) and should be corroborated through other means. Therefore, upon completion of my analysis, I verified my results using cluster validation (i.e. comparison to the MDS plot) and fuzzy clustering (where small amounts of random noise test the resiliency of the discrete cluster model). Details of the analyses are presented in the next section.

4. Results

4.1 AGGREGATE LINGUISTIC DIFFERENCES AND MODEL METRICS. The general linguistic difference analysis yielded a reasonably reliable model with a local incoherence⁴ value of .35 and a Cronbach's alpha⁵ coefficient of .46. Table 4 presents the difference matrix for each of the seven locations. Overall, Irving and San Antonio differed the most (.735) and Brownsville and San Antonio were the most similar (.474). The average linguistic difference was .626.

³ Additionally, during preliminary analysis the Gewichteter Identitätswert method created a more reliable model (lower local incoherence and higher Cronbach's alpha) than the binary comparison method.

⁴ Local incoherence refers to a lack of coherence in the data. Lower values indicate better measurements (Nerbonne and Kleiweg 2007).

⁵ Cronbach's alpha is a measure of consistency. Higher values are preferred, and .70 is generally accepted as a threshold for consistency (Heeringa 2004).

	Austin	El Paso	Edinburg/ Weslaco	San Antonio	Brownsville	Houston/ Pearland	Irving
Austin	0	.6501	.6193	.6472	.6608	.5215	.6858
El Paso	.6501	0	.5037	.6022	.6449	.6375	.6778
Edinburg/ Weslaco	.6193	.5037	0	.6632	.7072	.6404	.7353
San Antonio	.6472	.6022	.6632	0	.4736	.6070	.5547
Brownsville	.6608	.6449	.7072	.4736	0	.6517	.6122
Houston/ Pearland	.5215	.6375	.6404	.6070	.6517	0	.6486
Irving	.6858	.6778	.7353	.5547	.6122	.6486	0

TABLE 4: AGGREGATE LINGUISTIC DIFFERENCES OF TEXAS SPANISH DISCOURSE

4.2 MULTIDIMENSIONAL SCALING. Figure 2 plots the results of the multidimensional scaling of the linguistic difference data:

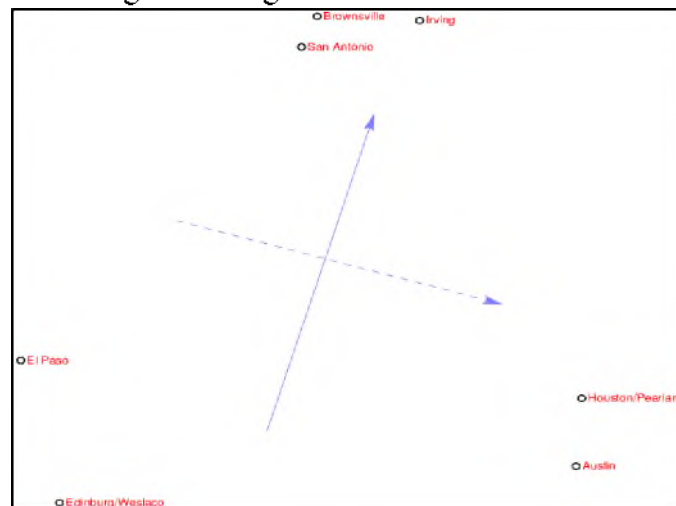


FIGURE 2. MULTIDIMENSIONAL SCALING PLOT

The plot details three distinct groups of locations: Brownsville, San Antonio, and Irving; Houston/Pearland and Austin; and El Paso and Edinburg/Weslaco. The physical distance between the plot points is significant as well; the high degree of similarity between the speech of Brownsville and San Antonio is evident, while the greater distances between the other groups at the bottom of the axis indicate less cohesion between those dialect areas.

4.3 DIALECT CLUSTERING AND VALIDATION. The results of the cluster analysis reflect those of the MDS plot, albeit with more specificity. As the map indicates,⁶ there are six major dialect clusters: (1) Austin; (2) Houston/Pearland; (3) Edinburg/Weslaco; (4) El Paso; (5) Brownsville and San Antonio; and (6) Irving.

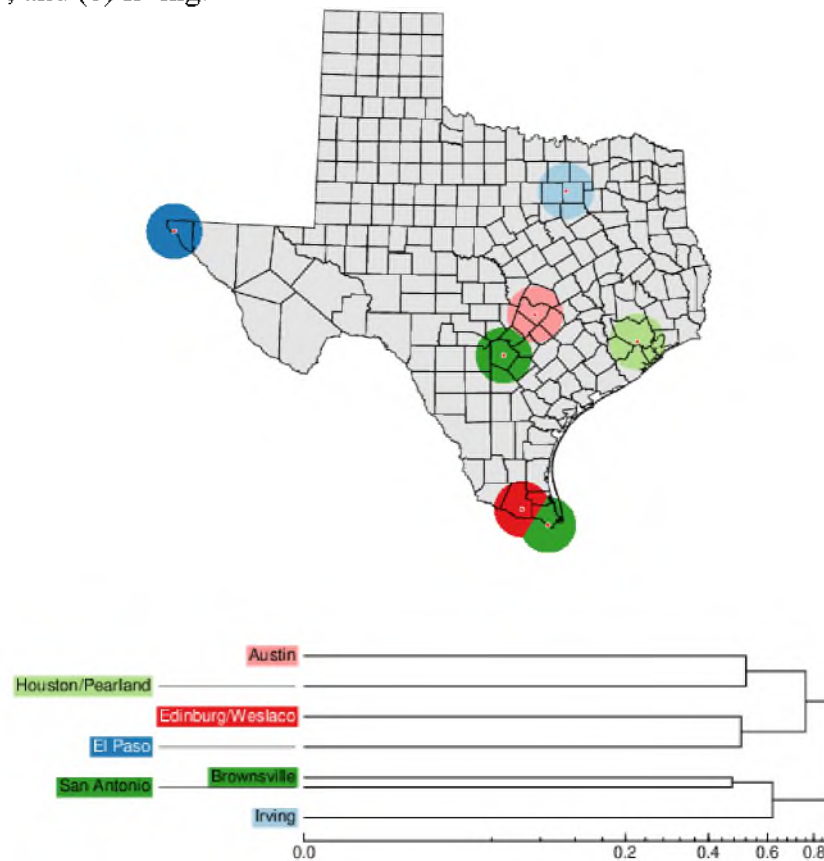


FIGURE 3: DISCRETE CLUSTER MAP AND DENDROGRAM OF TEXAS SPANISH DISCOURSE

⁶ For clarification, different shades of similar colors do not imply a relationship. Instead, areas that show the exact same color are related.

Regarding the validation of these groupings, the results of the fuzzy cluster analysis confirm the trends established through multidimensional scaling and discrete clustering:

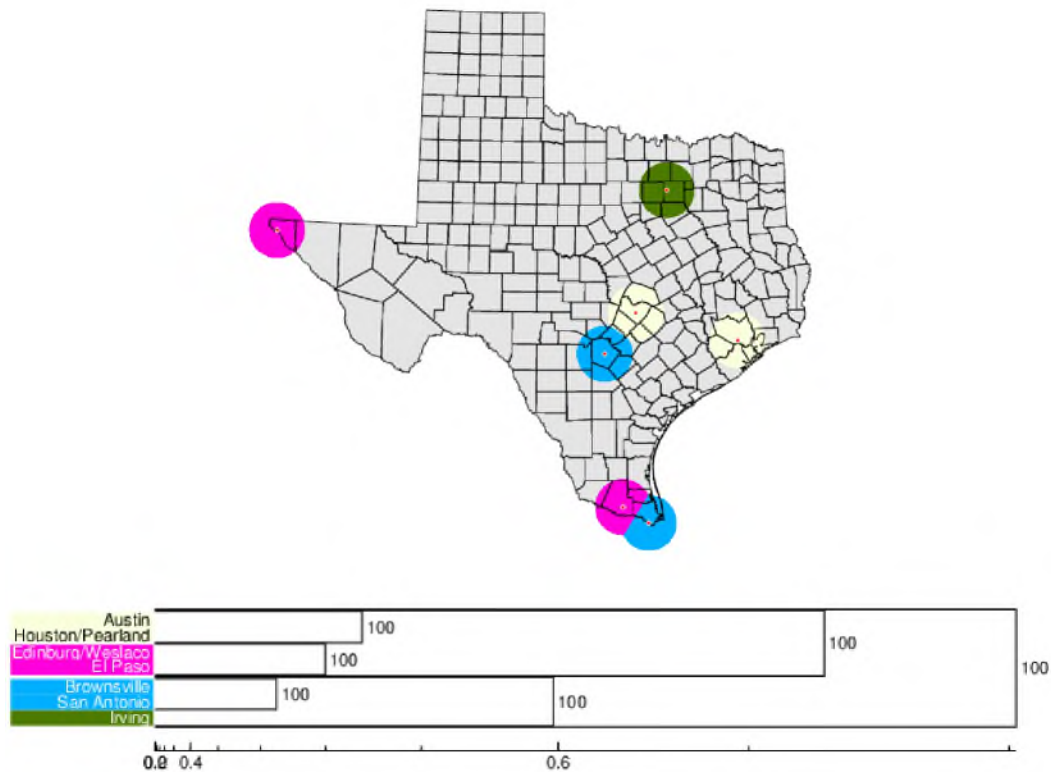


FIGURE 4: FUZZY CLUSTER MAP AND PROBABILISTIC DENDROGRAM OF TEXAS SPANISH DISCOURSE

In contrast to the discrete model, the results do not differentiate between the proposed individual location clusters, instead grouping them into larger clusters comprising two locations: (1) Austin and Houston/Pearland; (2) Edinburg/Weslaco and El Paso; and (3) Brownsville and San Antonio. As before, however, (4) Irving retains its dialectal independence. Overall, as the uniform results of the probabilistic dendrogram indicate, the division between the four larger clusters is exceedingly stable.

4.4 THE DISTRIBUTION OF ENGLISH MARKERS. Figure 5 illustrates the distribution of four English discourse markers found in the data.⁷ The maps are coded on a white to blue spectrum, and darker colors indicate higher frequency.

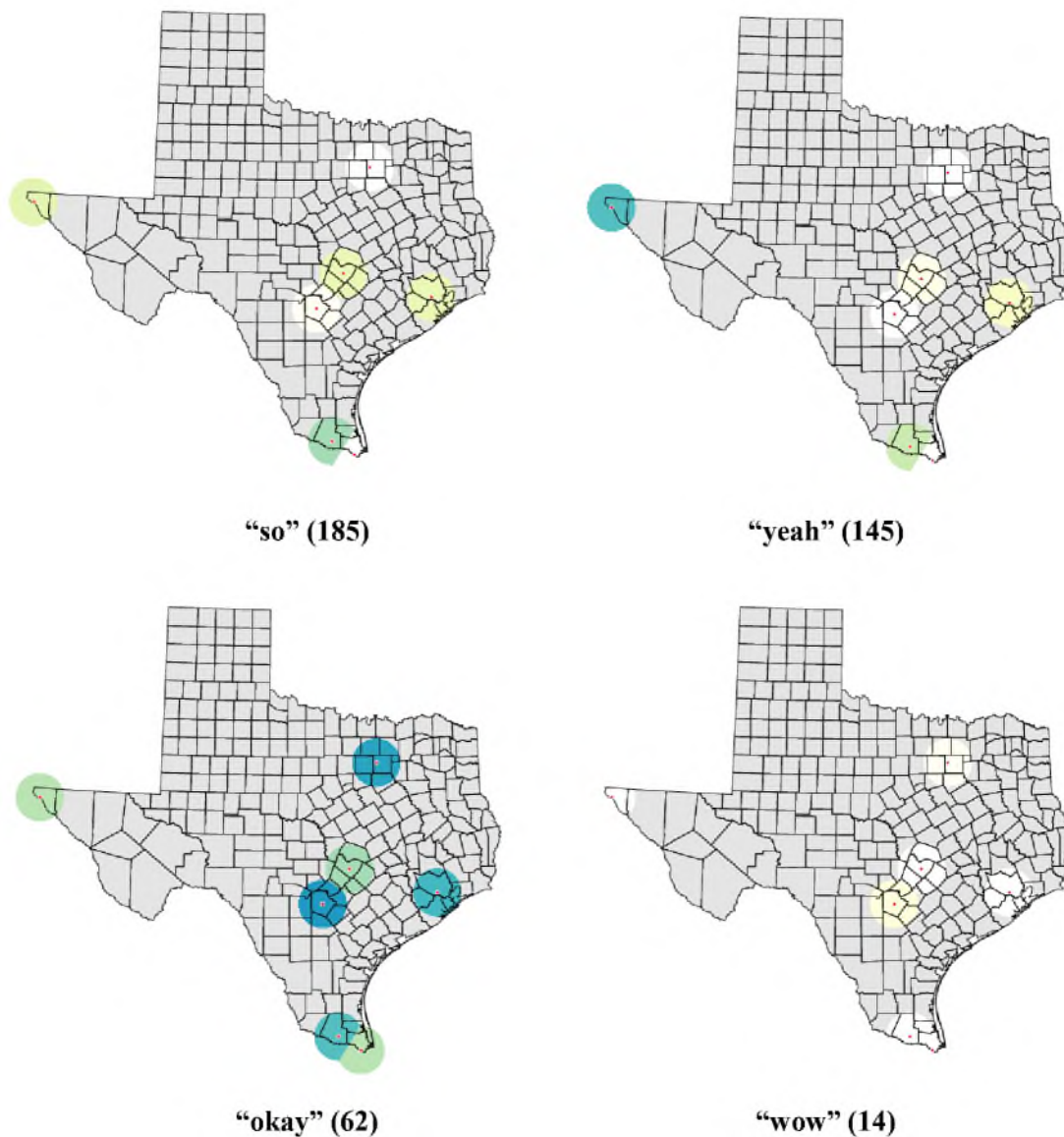


FIGURE 5: DISTRIBUTION MAPS OF ENGLISH DISCOURSE MARKERS

⁷ The English markers “I guess” and “like” were also present in the data, but their low frequency yielded distribution maps with no discernable variation.

Okay exhibits the most balanced distribution, having been used frequently across all locations in the analysis. Contrarily, both *so* and *yeah* were more frequently used in El Paso and Edinburg/Weslaco, and *wow* was almost exclusive to San Antonio. These maps also reveal that speakers in Brownsville, in sharp contrast to the other border locations, hardly used any English markers.

5. DISCUSSION. Returning to RQ1, the results from the various dialectometric analyses confirm the existence of geolinguistic variation among the usage of discourse markers and fillers in Texas Spanish. More precisely, there appears to be six dialect clusters and larger four dialect cluster divisions:

Dialect Clusters (discrete clustering approach)	Dialect Cluster Divisions (fuzzy clustering approach)
Austin	Austin and Houston/Pearland
Houston/Pearland	Edinburg/Weslaco and El Paso
Edinburg/Weslaco	Brownsville and San Antonio
El Paso	Irving
Brownsville and San Antonio	
Irving	

TABLE 5: DIALECT CLUSTERS AND DIVISIONS

Although the results of the fuzzy clustering differ slightly from the initial discrete clustering, I am still inclined to believe that the discrete analysis is valid for two reasons. First, there is simply too great of a linguistic difference between many locations in the data for them to reasonably be considered part of a single cluster. For example, there is considerable linguistic difference between each dialect (0.4-0.7), markedly more than in other studies where individual clusters can display aggregate differences of less than 0.1 (Snoek 2014). Additionally, and more importantly, the data used in the present study are not geographically representative enough to fully determine the true geolinguistic dynamics of this context. Therefore, it is more cautious and prudent to conclude that while four major cluster divisions do seem to exist in Texas Spanish discourse, most of the individual locations retain enough linguistic difference to merit the distinction between six dialect clusters. Nonetheless, the results of this study add further evidence of the

heterogeneity of Texas Spanish. This diversity is expected, as it is a common characteristic shared with other Mexican-origin Spanish dialects spoken in the United States: “[...] the Spanish of Mexican Americans is not monolithic, but covers a broad range of social and regional variants, reflective of the immense linguistic diversity of Mexico itself” (Lipski 2008: 75). This remarkable intra-dialectal diversity may call into question the validity/utility of the notion of “Texas Spanish” itself. Working in the context of Spanish in the United States, various authors (Erker 2017, Otheguy et al. 2015) have critiqued named languages as socio-political constructs that obfuscate the real drivers of language variation and change: the language users. While I agree that discrete, named language varieties can be a problematic notion, often rife with *a priori* categorizations inconsistent with linguistic realities, the findings of this study demonstrate that 1) variation can be tied to a specific geographical region and 2) geography may play a significant role in the dynamics of that variation. In other words, while it is imperative that we focalize speakers in analyses of language use, we must also not ignore the role of the places where those speakers find themselves at the time of analysis. So, although “Texas Spanish” may be an increasingly dubious concept, I argue that the following statement is still linguistically founded and useful: the Spanish spoken in Texas presents considerable geolinguistic variation at the level of discourse markers and fillers.

In response to RQ2, the use of loaned and code-switched English discourse markers does vary across different locations/regions in Texas. While these markers appeared across all locations, the majority of them were utilized in locations near the Mexican border (El Paso and Edinburg/Weslaco). Therefore, although somewhat curious, proximity to Mexico does not seem to correlate with less English transfer. While the present study is primarily descriptive in scope, this finding could suggest (among other explanations) that markers such as *so*, *okay*, and *yeah* are also present in the Spanish used on the Mexican side of the border. Overall, these results demonstrate the utility of dialectometric methods in sociolinguistic, dialectological, and/or language contact research. In contrast to traditional, subjective dialectological studies, which are often plagued by inaccuracies in dialect classification, or variationist sociolinguistic studies that tend to treat geographical location as just one of numerous statistically (in)significant factors, dialectometry offers a precise and rigorous approach to parse out the impact of geography on language variation, change, and use. However, it is important to acknowledge that every method has its limitations, and dialectometry is most effectively employed in conjunction with other approaches that can more comprehensively address social and linguistic variables.

6. CONCLUSIONS AND FURTHER RESEARCH. All in all, this investigation offers a novel approach to the study of Spanish discourse. A battery of dialectometric analyses reveal considerable subdialectal variation among the usage of discourse markers and fillers. More specifically, the results suggest the existence of six dialect clusters corresponding to four larger dialect cluster divisions. However, these conclusions should be considered preliminary until more data can be collected. Texas is a massive geographic area comprising millions of speakers; the analysis of nine cities is insufficient in accurately characterizing such a large territory. Future studies should prioritize the collection and processing of such data. Nonetheless, I do believe that the findings of this study are valid, albeit partial. The various computational methods provided consistent outcomes and validation tests indicate confidence in the given clustering models. In summary, the present study offers only a snapshot of the true nature of discursive variation in this context, and more data is necessary to fully understand geolinguistic realities of the Spanish spoken in Texas.

REFERENCES

- AIJMER, KARIN. 2002. *English discourse particles: Evidence from a corpus*. Philadelphia, PA: John Benjamins Pub. Co.
- BAYLEY, ROBERT, NORMA L. CÁRDENAS, BELINDA TREVIÑO SCHOUTEN, AND CARLOS MARTIN VÉLEZ SALAS. 2012. Spanish dialect contact in San Antonio, Texas: An exploratory study. In *Selected proceedings of the 14th Hispanic linguistics symposium*. 48–60. Somerville, MA: Cascadilla Proceedings Project.
- BLOEM, JELKE. 2017. Empirical Evidence for Discourse Markers at the Lexical Level. In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (Eds.) *From Semantics to Dialectometry*. London: College Publications.
- BOAS, HANS C. & HUNTER WEILBACHER. 2007. How universal is the pragmatic detachability scale? Evidence from Texas German discourse markers. *Texas Linguistic Society IX: The Morphosyntax of Underrepresented Languages*. 33.
- BORDERÍA, SALVADOR PONS. 2006. A functional approach to the study of discourse markers. *Approaches to Discourse Particles*. 77–99.
- BRODY, JILL. 1987. Particles borrowed from Spanish as discourse markers in Mayan languages. *Anthropological linguistics*. 507–521.
- CÁRDENAS, DANIEL NEGRETE. 1970. *Dominant Spanish Dialects Spoken in the United States*. Washington D.C: Center for Applied Linguistics.
- CARTER, PHILLIP M. & TONYA WOLFORD. 2016. Cross-generational prosodic convergence in South Texas Spanish. *Spanish in Context* 13(1). 29–52.

- CHASTON, JOHN M. 1996. Sociolinguistic analysis of gender agreement in article/noun combinations in Mexican American Spanish in Texas. *Bilingual Review/La Revista Bilingüe* 21(3). 195–202.
- BULLOCK, BARBARA E., AND A. JACQUELINE TORIBIO. 2013. The Spanish in Texas corpus project. *Center for open education resources and language learning (COERLL)*.
- ERKER, DANIEL. 2017. The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass*, 11(1), 1–20.
- ERKER, DANIEL, AND JOANNA BRUSO. 2017. Uh, bueno, em...: Filled pauses as a site of contact-induced change in Boston Spanish. *Language variation and change* 29(2). 205–244.
- FRASER, BRUCE. 1996. Pragmatic markers. *Pragmatics* 6(2). 167–190.
- FURKÓ, BÁLINT PÉTER. 2017. Manipulative uses of pragmatic markers in political discourse. *Palgrave Communications* 3(1). 1–18.
- GOEBL, HANS. 1984. Dialektometrische Studien: Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. *Unter Mitarbeit von Siegfried Selberherr, Wolf-Dieter Rase und Hilmar Pudlatz*. 3.
- HEERINGA, WILBERT. 2004. Measuring dialect pronunciation differences using Levenshtein distance. dissertation.
- KLEIWEG, PETER, JOHN NERBONNE & LEONIE BOSVELD. 2004. Geographic projection of Cluster Composites. *Diagrammatic Representation and Inference*. 392–394.
- LAM, PHOENIX W. 2009. Discourse particles in corpus data and textbooks: The case of well. *Applied Linguistics* 31(2). 260–281.

- LIPSKI, JOHN. 1988 Sabine River Spanish: vestigial 18th century Mexican Spanish in Texas and Louisiana. *Southwest Journal of Linguistics* 8. 5-24.
- LIPSKI, JOHN. 2008. *Varieties of Spanish in the United States*. Georgetown University Press.
- MARTÍNEZ, GLENN A. 2003. Perceptions of dialect in a changing society: Folk linguistics along the Texas-mexico border. *Journal of Sociolinguistics* 7(1). 38-49.
- NERBONNE, JOHN, RINKE COLEN, CHARLOTTE GOOSKENS, PETER KLEIWEG, AND THERESE LEINONEN. 2011. Gabmap-a web application for dialectology. *Dialectologia: revista electrònica*. 65-89.
- NERBONNE, JOHN, PETER KLEIWEG, WILBERT HEERINGA & FRANZ MANNI. 2008. Projecting dialect distances to geography: Bootstrap clustering vs. Noisy Clustering. *Data Analysis, Machine Learning and Applications*. 647-654.
- NERBONNE, JOHN & PETER KLEIWEG. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2-3). 148-166.
- OTHEGUY, RICARDO, OFELIA GARCÍA, & WALLIS REID. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6 (3). 281-307.
- PORTOLÉS, JOSÉ. 2010. Los marcadores del discurso y la estructura informativa. In Óscar Loureda y Esperanza Acín (eds.), *La investigación sobre marcadores del discurso en español, hoy*. Madrid: Arco/Libros. 281-352.
- ROBLES I SABATER, FERRAN & PAU BERTOMEU PI. 2017. Els marcadors del discurs: caracterització general. *Zeitschrift für Katalanistik: Revista d'Estudis Catalans* 30. 3-18.

SNOEK, CONOR. 2014. Review of Gabmap: doing dialect analysis on the web. *Language documentation & conservation* 8. 192-208.

SPEELMAN, DIRK & DIRK GEERAERTS. 2009. The role of concept characteristics in lexical dialectometry. *Computing and Language Variation*. 221–242.