

A San Diegan debate: Take 8 or Take the 8 Investigating the use of the determiner “the” before numeric freeways in San Diego

Brianna O’Boyle*

Abstract. The determiner “the” before numeric roadways in Southern California has become salient enough to have reached mainstream media. However there has been little formal research done on this linguistic variant, and especially in subregions of California such as San Diego County. This study examines the use of *the* by ten speakers through sociolinguistic interviews and a map task. It was found that people from San Diego had the highest usage of *the*, long term transplants also favored *the* but to a lesser degree, and recent transplants disfavored *the*. Additionally, it was found when recent transplants used this variant they did not apply it to unknown roads, though long term transplants used *the* not only for roads in San Diego but also for unknown roads. These differences between transplants may inform us about factors influencing second dialect acquisition. Furthermore, a mixed methods analysis revealed that *the* might have diffused from Los Angeles to San Diego in the 1980s.

Keywords. sociolinguistics; second dialect acquisition; San Diego; Southern Californian English; numeric roadways; determiner; *the*

1. Introduction. California has gained attention as a notable dialect region, with recent research on the Californian Vowel Shift exemplified by Podesva et al. (2015) and D’Onofrio et al. (2019). However there are still linguistic features and smaller regions within California that remain relatively unexplored. This paper aims to shed light on one of those features in San Diego: the usage of the determiner *the* before numeric freeways (e.g., referring to Interstate 5 as *the* 5). This variable will be referred to as (the) and the specific variant examined will be referred to as *the*, some other variants encountered were *I* as in I-8, *Interstate* as in Interstate-10, *Route* as in Route 1, *the I* as in the I-5, *Highway* as in Highway 101, among others. It was hypothesized for (the) that Native San Diegans would have high usage of *the* because presumably this is a linguistic feature of San Diegan English. It was also hypothesized that long term transplants would have some usage of *the* but primarily for roads located in San Diego as there may be associations with the feature and the region. Finally recent transplants would have low usage of *the* as they have not yet had adequate time to adopt it into their language.

2. Literature Review. This section introduces previous work on second dialect acquisition (Section 2.1) and the status of *the* in Southern California (Section 2.2).

2.1. SECOND DIALECT ACQUISITION. There has been some substantial research and claims made of second dialect acquisition. Chambers (1992) outlined eight principles of dialect acquisition based on findings from six young Canadians who transplanted to Southern England. Some of these findings include how lexical replacements are acquired faster than pronunciation or phono-

* Thank you to Professor Aaron Dinkin who served as an invaluable mentor on this project. I would also like to thank my classmate Laura Fitzgerald for acting as a consult on classifying the word classes of the words preceding (the). And most of all I would like to thank the interviewees for donating their time and sharing their stories, without them this paper would not have been possible. Author: Brianna O’Boyle, San Diego State University (brianna.oboyle@outlook.com).

logical variants, and simpler phonological rules are acquired faster than more complex ones, among other topics. There also have been other studies such as Tagliamonte and Molfenter (2007) which build upon Chambers' findings – again examining how Canadian children acquire British English looking into the phonological variation of T-voicing. Though a lot of the research on second dialect acquisition has been focused on children still in the process of acquiring their language.

There has been some research investigating second dialect acquisition in adults as explored by Rampton (2013) and Nycz (2015). Rampton (2013) was limited though in the focus of the research was based on a single migrant from India who moved to London and learned English as his L2 later in life, but then had learned to style shift, akin to using multiple dialects in one language, after immigrating to England. Nycz's dissertation (2011) also investigated this, examining how 17 Canadians who moved to New York and New Jersey retained much of their first dialect in phonological perception and production, though some influences of the second dialect can be observed. Is it also of interest to note that of these studies, the majority of them are focused on speakers moving between countries, and there still seems to be much left to explore in how dialect acquisition may occur within different dialect regions within the same country. Speakers who move between countries may have more identity values tied to being from a different nationality that may influence their dialect acquisition.

Another important point to make is that there is some evidence that when one first enters a speech community, this acts like the "critical period" in which one will be able to acquire a linguistic feature more quickly. This is mentioned in Danescu-Niculescu-Mizil et al. (2013) in which members of an online community would be more likely to adopt new terms or language when they first entered the community. This appears to be in alignment with Chamber's second assertion of dialect change in which lexical replacements occur more quickly in the first stage of dialect acquisition. However there is still opacity regarding how long this process may take and what would influences be on this acquisition process beyond simple exposure.

2.2. ORIGINS OF *THE* AND ITS STATUS IN SOUTHERN CALIFORNIA. It is not very clear when or where the variant *the* originated. Geyer (2001) recognizes *the* as a feature of Southern Californian English and argues that it came from the tendency to refer to roads by their names in the 1970s, which oftentimes required the usage of the word "the." One example being Route 99 was also known as the Golden State Freeway. He then extends this point by asserting with the popularization of freeways it seemed natural to carry over the "the" from roadway names to their numbers as well resulting in people saying *the* 405 instead of the San Diego Freeway. Geyer (2001) also mentions that *the* became widely used in 1975 in LA when several minor interstates and bypasses formed, but it seems difficult to pinpoint this variant to a specific year. He also argues this did not become a feature of the Bay Area, because when they began to build more roadways with numeric names they were not as entrenched. He also brings up a point of car culture, and mentions this as an archetype of Southern California, which may have influenced usage of this variant. It is uncertain exactly when *the* started to become used and become recognized as a feature of Southern Californian English. It is also unclear if this feature functions the same in all of the subregions of Southern California.

There seems to be some awareness of *the* as a feature of Southern California by sociolinguistic researchers. In a perceptual dialectology study Bucholtz et al. (2007) found *the* to be mentioned as a feature of Southern Californian English. Their paper investigated opinions of Californians regarding their perceptions of California as a dialect region, especially focusing on differences between Northern and Southern California. But, most of the lexical items investigat-

ed were related to slang terms, especially through the use of intensifiers like *hella* and *totally*. This was an extension of Fought (2002) in which she looked into Californians' language attitudes of the US as a whole. However, in Villareal's (2016) study on people's perceptions and attitudes of Californian English, *the* was not mentioned or investigated at all. To be fair Villareal's study was more focused on assessing people's language attitudes and focused on the California Vowel Shift rather than lexical items. Outside of research circles, *the* is salient enough to have reached mainstream media. In Saturday Night Live's recurring skit "The Californians," actors frequently mention freeway names with *the* (Pratt et al 2017). While *the* has obviously become associated as a feature of Southern California English, there has been little done in documenting its use, especially in the region of San Diego.

3. Methodology. For this study to see how this linguistic feature manifests it was decided to have three population groups: native San Diegans, long term transplants (20+ years in San Diego), and recent transplants (less than 1 year in San Diego). There also ended up being an unexpected fourth category of a San Diegan native who recently transplanted to the East Coast, though unless specified this participant is grouped with the San Diegan natives to avoid comparing an individual to a group. For this study there were a total of 10 participants: two recent San Diegan transplants, four long term transplants, and four San Diegan natives. Because of time and logistical constraints, it was not able to achieve an even distribution of gender, age, socioeconomic status, ethnicity, or other factors. Despite having a small sample size the results were remarkably clear.

For transplants, because usage of *the* seems to be a feature of not just San Diegan English, but of Southern California English it was important to seek transplants from outside of Southern California. Additionally since there is some debate where exactly the boundary between Northern and Southern California is, and if there are linguistic boundaries they would likely be ambiguous as well, it was decided to recruit transplants from outside of California entirely. To the knowledge of the author there are no other regions within the US that use *the* before numeric freeways, though this may be a feature in Toronto (Aaron Dinkin, p.c.). One of the transplants was from Florida, two of the transplants were from Wisconsin, and the remaining three were from New York.

To observe the usage of (the) tokens were collected from the interviewees during sociolinguistic interviews that had a duration of no less than 1 hour, in hopes of eliciting more vernacular speech. Additionally interview module questions were designed around asking about transportation, traffic, and driving to encourage more tokens of the desired variable. Furthermore, a map task was designed in which participants were shown a map of San Diego county and asked to identify the major roads on the map and to describe how to go to and from different places. To further test the dialectal grammar, and see whether or not there was a tendency to only use *the* for roadways in Southern California participants were also shown a map of an unknown region that also had roads labeled with numbers and asked the same questions, to identify the major roads and describe how to get to and from different places. These strategies were met with relative success as each speaker produced an average of 76 tokens of the (the) variable.

There were a total of nine interviews, most of them being one-on-one with the exception of Interview #6 being a group interview with two interviewees for ten speakers in total. All interviews lasted between 1 hour and 1 hour and 45 minutes. Two different high-quality recorders were used: Zoom H2next Handy Recorder and Zoom H5 Handy Recorder. The Zoom H2next is a stand up microphone and its internal mic was used for Interviews #1, #2, #3, and #6, and the Zoom H5 with a clip-on Comica XLR Lavalier microphone was used for Interviews #4, #5, #8,

and #9. Interview #7 was conducted over Zoom and it was decided to use the built in microphone on Zoom.

Since there has been little research done on this variable it was necessary to consider what would be the envelope of variation. The envelope of variation for this variable was whenever a numeric highway was mentioned, excluding instances in which it was a media reference such as the name of a show. Any instances in which it was not clear if the speaker was using the number as the name of a freeway, as well as those that were difficult to clearly distinguish were excluded. Additionally any instances in which participants were obviously aware of the feature and were attempting to perform *the* or lack of *the* these tokens were also excluded for potentially not representing the speaker's actual speech. A total of 759 tokens were collected.

	SD Natives		Long Term Transplants		Recent Transplants		Totals	
	F	M	F	M	F	M	F	M
Ages 20-50	71	47	0	107	54	70	125	224
Ages 51-80	0	188	105	117	0	0	105	305
Total	306		329		124		759	

Table 1. Number of Tokens (F=tokens from women, M=tokens from men)

The data was analyzed using Rbrul to build a mixed-effects step down multiple logistic regression model (Johnson 2009). For the analysis the model was built with a step-down approach, the dependent variable examined was the target variable (*the*). The variable was grouped into two categories with the first being *the* and the second being all other variants. Some of the other variants were *0*-nothing such as “take 8,” *I/Interstate* such as “take I/Interstate 8,” *Route*, such as “take Route 8,” and *SR/State Route*, such as “take State Route 8,” among other variants. It is important to note that if there was a token that had *the* combined with another variant, this was grouped to be with *the* (example “the I-8”). The logic behind this grouping is if a speaker was using a “the” this should be grouped with the other *the* tokens since it is implementing a similar dialectal grammar structure.

The independent variables used as candidates for the model were: age, gender, ethnicity, location of the road, word class of the word preceding the variable, map presence, and speaker status (if they were a SD native, long term transplant or recent transplant). It was decided not to consider education as a factor since the educational status of all the interviewees was relatively similar with all ten speakers having at least some education beyond high school, and eight out of ten having at least a BA. Additionally the actual name of the roads were not considered since there were over 50 different roadways mentioned, with more than half being mentioned less than 10 times.

4. Results. Three variables made it into the model for being statistically significant $p < 0.05$, these were found to be word class of the preceding word, speaker status, and age. In the mixed effects regression model, speaker was used as a random factor, $N=665$, intercept 1.10 positive coefficients indicate increased use of *the*.

	p value		Coefficient
Word Class of the preceding word	p<0.001	Preposition	0.90
		Verb	0.72
		Conjunction	0.06
		Adverb	-0.29
		Noun	-0.46
		0 (nothing)	-0.92
Speaker Status	p<0.001	Native	2.13
		Long Term Transplant	0.57
		Recent Transplant	-2.70
Age	p<0.01	+1 year	-0.04

Table 2. Predictors that were statistically significant

Word class of the word preceding the variable was found to be statistically significant. Use of *the* was favored when preceded by a preposition or a verb, but disfavored by preceded by a noun or nothing. This might be explained by the fact that use of this variable could be perceived to be more common when used in a prepositional phrase or verb phrase. Speaker status was perhaps unsurprisingly statistically significant. As expected San Diego natives were most likely to favor *the*, the long term transplants also favored *the* though to a lesser degree, and the recent transplants disfavored *the*.

This can likely be explained by the fact that this is a feature of San Diegan English and natives would have greater usage of this variant. It appears long term transplants have begun to acquire this feature which can explain why long term transplants favor it, though their usage is not as high as the San Diegan natives. However short term transplants disfavor it, possibly because they have not yet had the time or exposure to acquire the feature.

In regard to age it was found that with increased age there is a slight tendency to disfavor *the*. This can likely be explained by the fact that there appears to be generational change. As reflected by a couple of the anecdotes from some of the interviewees the determiner may be a linguistic feature from LA that diffused to SD in the 1980s:

- (1) Lindsay: When I moved here... I met this guy from San Diego, he was from La Mesa and he said people fr-he never called it the 8, and he said that's LA. He hated LA, San Diegans hate LA or that's how it was. It was like LA, San Diego, he said they say the 5 the-he's like you don't say, it's 8 take 8 east... after a while it's like you just say that and now there's so many people from LA coming they-they call it the 5... but yeah that's pretty much not just LA people say that now...you hear it on the news and everything.
- (2) Manny: The only specific little regional thing I can recall uh that has gone away to a large extent is the names of our freeways. Everyone I grew up with talked about uh taking taking 94 to 5 [mm] and folks from Los Angeles said the 94 the 5 uh now we've adopted Los Angeles' language [So you-that's something that came from LA?] Yes [ok] the 8, 8's go east...I haven't heard it being a controversy for a long time [okay] I know it used to slightly offend me, going no you're in San Diego now, it's-it's 94, it's not the 94 and now I've-I've adopted it myself and it doesn't, and it just seems totally neutral I don't think of it at all.

This could explain why among the native San Diegans the younger speakers have higher percentages of usage of the determiner compared to their older counterparts since this is an in-

coming linguistic variant that would have been present in the younger San Diegan natives dialect from the time they were born, versus for the older San Diegans they would not have had the feature when they were younger but then began to acquire it as the feature diffused to San Diego.

Younger SD Natives	Percentage	Older SD Natives	Percentage
Ryan	79%	Bob	49%
Mary Ann	93%	Manny	69%

Table 3. Percentage of *the* tokens by San Diegan natives

4.1. INTERACTION OF ROAD LOCATION AND SPEAKER STATUS. While it was not surprising to see that recent transplants disfavored *the*, there were some instances in the interviews of limited use of *the* when referring to roadways in Southern California and San Diego.

- (1) Greg: if I had to go from Julian I would definitely...mmm you could obviously just g-take uh from here you could get on 94 east and then from there go on *the* 125 no-and then 94 yeah 94 east right onto I-8 and then just follow 8, take 79 up north.
- (2) Poppy: I go on the freeway I think I take *the* 5...Mkay 5, 54 to *the* 8, to 79 to 78.

This led to a discovery of an interaction between location of the road and speaker status. In the limited instances recent transplants used *the*, they favored it for roadways in San Diego and disfavored it for roadways that were from the map of the region they were unfamiliar with. However what is surprising is that for long term transplants they did not prefer to use *the* with roadways in San Diego, disproving the hypothesis that there would be a greater association with *the* for roadways in San Diego. This indicates on some level that the dialectal grammar of using *the* by the long term transplants seems to be more similar to San Diegan natives, which may indicate a more complete second dialect acquisition. Also there was the unexpected finding that for San Diegan natives they favor *the* with unknown roadways. This may be explained by the fact all of the instances they were referencing unknown roadways was during the map task which may have prompted more careful speech. The San Diego roadways may have come up naturally in conversation and may have had less careful speech and less use of *the*.

This mixed effects regression model also used speaker as a random factor N=708, intercept 6.86, $p < 0.001$ positive coefficients indicate use of *the*.

		Coefficient
Location of the Road	San Diego	2.49
	Unknown to the Speaker	-2.49
Speaker Status	Native San Diegan	4.17
	Long Term Transplant	2.61
	Recent Transplant	-6.78
Location: Status	N	Coefficient
SD: Recent	56	5.28
Unknown: Native	57	2.79
Unknown: Long Term	29	2.50
SD: Long Term	116	-2.50
SD: Native	228	-2.79
Unknown: Recent	43	-5.28

Table 4. Interaction between location of the road and speaker status

5. Discussion. This section is broken up into two parts, one about second dialect acquisition and one about implications of the diffusion of *the*.

5.1. SECOND DIALECT ACQUISITION. It may also be interesting to consider how this acquisition of the grammar of San Diegan English may influence long term transplants. It seems obvious that there has been some second dialect acquisition since the long term transplants favored *the*, and through the interaction of location of the road and speaker status their dialectal grammar appears to be more similar to San Diegan Natives than recent transplants. It could be interesting though to see how long term and recent transplants refer to roads in their hometowns. It would be assumed that it is likely these roadways would have a different variant of (the).

Perhaps this could be further investigated by examining how transplants refer to roads in San Diego, roads in unknown regions and roads in the regions they grew up in. Would transplants use *the* with roadways they may have grown up with or would they use another variant of (the) with these roads? This could give great insight into aspects of second dialect acquisition by revealing how engrained some of these dialectal grammars are and how or when one's first dialect may interfere with one's second dialect as this would present a case in which the two of them would be in conflict. It could also be interesting to play the idea of audience design and prompt interviewees to imagine they are speaking to someone from their hometown, or someone from Southern California.

Further work may also want to consider an effect on the name of the road, and if there are prosodic influences on the variant such as syllable length. This may be interesting to investigate for transplants, as they may be trying to develop intuitions on how to use this variant and if there is a phonological component. Because of the constraints of this data, there are limited analyses that can be conducted on this matter without compromising the results with overfitting. Duration of time as a transplant would also potentially be interesting to investigate, as this study was limited in only looking at recent transplants and transplants that had been in San Diego for over 20 years. Recruiting participants with a wider variety of durations of time as a transplant may allow for finer gradients of second dialect acquisition to be observed.

Beyond time as a transplant, the long term goals of transplants may have an effect on their acquisition of variants. If someone is only planning on staying in San Diego for a year or two, perhaps they would not have as much motivation to adjust their language. However if their intention is to move here permanently then they may seek to adapt to variants more quickly. Additionally ties to their old community may also have an impact, similar to ethnic orientation, as someone who moves to San Diego with the desire to become a San Diegan may acquire this variant faster than someone who has moved to San Diego, but maintains a regional identity as an outsider. Motivations for moving to San Diego should be considered as well, have they married someone from this region, or did they move here with their partner who is also an outsider? There are still many factors that should be considered as possibly having an influence on second dialect acquisition.

5.2. IMPLICATIONS OF DIFFUSION OF *THE*. There were some preliminary syntactic analysis conducted in this work, through the effect of word class on the word preceding the variable. In this data *the* was favored before prepositions and verbs, but disfavored before nouns and nothing. In Tagliamonte and Denis (2014) it was seen that lexical items that have diffused from different cities may present different syntactic properties. If it is true that feature diffused from Los Angeles

to San Diego in the 1980s, it would be interesting to see if there are similar syntactic patterns of usage of *the* in Los Angeles or if there are different patterns of usage. Nonetheless it seems likely that *the* diffused from Los Angeles to San Diego as the two cities are in close proximity, even though there are still mysteries on the exact details of the diffusion and how this variant manifests.

6. Conclusion. In the end it may not have been particularly surprising to discover that indeed San Diegan natives favored *the* the most as it is a feature of San Diegan English. Long-term transplants had greater usage of *the* than their recent transplant counterparts who disfavored *the*. Short-term transplants had limited usage of *the* and seemed to exclusively use it for roadways they associate with San Diego. However, for long-term transplants there was no special region association with this variant as originally hypothesized as they applied *the* to roadways in San Diego and an unknown region, and actually favored it with unknown roads. This could lead us to believe that usage of *the* has now become at least in part engrained into the dialectal grammar of the long term transplants as *the* is not only reserved for roadways in Southern California. This is interesting to consider how this may influence theories of second dialect acquisition and what influences have an effect on people's usage of variants that manifest differently in different dialects.

Perhaps the most fascinating findings of this research relate to how a couple interviewees cite *the* as something that did not originally exist in San Diego, and that they believe it is a language feature that came from Los Angeles to San Diego, possibly in the 1980s. These anecdotes coupled with the comparison of usage of *the* between younger San Diego natives with older San Diegan natives may lead one to conclude that there may be change in apparent time motivated by generational change. This is also supposed by the fact that with increased age there is a slight disfavoring of *the*. Furthermore, this somewhat fits in with the timeline Geyer (2001) proposed of this feature becoming popularized in Los Angeles in the 1970s. It would be extremely interesting to investigate into this matter further and determine more about how and when the determiner diffused, as well as establish more of a baseline with more native San Diegans from a variety of ages. This paper hoped to bring attention to the understudied dialect region of San Diego as well as the underexplored variable of (the).

References

- Bucholtz, Mary, Nancy Bermudez, Victor Fung, Lisa Edwards & Rosalva Vargas. 2007. Hella Nor Cal or Totally So Cal? *Journal of English Linguistics* 35(4). 325–352. <https://doi.org/10.1177/0075424207307780>.
- Chambers, J. K. 1992. Dialect acquisition. *Language* 68(4). 673–705. <https://doi.org/10.2307/416850>.
- D'Onofrio, Annette, Teresa Pratt & Janneke Van Hofwegen. 2019. Compression in the California Vowel Shift: Tracking generational sound change in California's Central Valley. *Language Variation and Change* 31.2:193–217. <https://doi.org/10.1017/S0954394519000085>.
- Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec & Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In (n.eds.), *Proceedings of the 22nd International Conference on World Wide Web*, 307–318. New York: Association for Computing Machinery. <https://doi.org/10.1145/2488388.2488416>.
- Fought, Carmen. 2002. California students' perceptions of, you know, regions and dialects? In Dennis R. Preston & Daniel Long (eds.), *Handbook of perceptual dialectology* (Vol. 2), 113–134. Amsterdam: John Benjamins.

- Geyer, Grant. 2001. “The” freeway in Southern California. *American Speech*, 76(2), 221–224. <https://doi.org/10.1215/00031283-76-2-221>.
- Johnson, Daniel. E. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359–383. <https://doi.org/10.1111/j.1749-818x.2008.00108.x>.
- Nycz, Jennifer. 2015. Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass* 9(11). 469–482. <https://doi.org/10.1111/lnc3.12163>.
- Nycz, Jennifer. 2011. *Second dialect acquisition: Implications for theories of phonological representation*. New York: NYU dissertation.
- Podesva, Robert, Annette D’Onofrio, Janneke Van Hofwegen & Seung Kyung Kim. 2015. Country ideology and the California Vowel Shift. *Language Variation and Change* 27(2). 157–186. <https://doi.org/10.1017/S095439451500006X>.
- Pratt, Teresa & Annette D’Onofrio. 2017. Jaw setting and the California Vowel Shift in parodic performance. *Language in Society* 46(3). 283–312. <https://doi.org/mr6m>.
- Rampton, Ben. 2013. Styling in a language learned later in life. *The Modern Language Journal* 97(2). 360–382. <https://doi.org/10.1111/j.1540-4781.2013.12010.x>
- Tagliamonte, Sali A. & Derek Denis. 2014. Expanding the transmission/diffusion dichotomy: Evidence from Canada. *Language*, 90(1), 90–136. <https://doi.org/10.1353/lan.2014.0016>
- Tagliamonte, Sali A. & Sonja Molfenter. 2007. How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society* 36(05). 649–675. <https://doi.org/10.1017/s0047404507070911>.
- Villarreal, Dan. 2016. 4. “Do I sound like a valley girl to you?” Perceptual dialectology and language attitudes in California. *The Publication of the American Dialect Society* 101(1). 55–75. <https://doi.org/10.1215/00031283-3772901>.