

A variationist corpus analysis of the definite article with personal names across three varieties of Spanish (Chilean, Mexican, Andalusian)

Robin Tieperman
Texas Tech University

Brendan Regan
Texas Tech University
brendan.regan@ttu.edu



Received: 16-01-2023
Accepted: 03-04-2023
Published: 12-08-2023

How to cite: Tieperman, Robin & Brendan Regan. 2023. A variationist corpus analysis of the definite article with personal names across three varieties of Spanish (Chilean, Mexican, Andalusian). *Isogloss. Open Journal of Romance Linguistics*. 9(1)/10, 1-27. DOI: <https://doi.org/10.5565/rev/isogloss.300>

Abstract

It has been shown that the presence of a definite article (DA) prior to a personal name (PN) varies not only across languages, but also across varieties of the same language. Spanish is an example of a language that demonstrates DA+PN variation across varieties. However, few studies have quantitatively analyzed this DA+PN variation. The current study examined which linguistic and social factors govern the variation of DA+PN in the speech (sociolinguistic interviews) of 54 speakers from Santiago (Chile), Mexico City, and Granada (Spain) from the online PRESEEA Corpus. Statistical analyses found that the DA+PN realizations are most favored by speakers from Santiago across all educational levels. Speakers from Santiago also favor DA+PN realizations more with female names than male names, demonstrating a gendering of

language. Speakers from Mexico City and Granada demonstrated few overall DA+PN realizations, with the least realizations among those with more educational attainment, possibly indicating the effect of a standard language ideology for these communities with regard to DA+PN. Thus, not only do these varieties differ in overall frequency in which DA+PN appears to be the norm in Santiago and much less common in Granada and Mexico City, they also differ in terms of which social factors govern DA+PN realizations in each variety.

Keywords: variationist sociolinguistics, variational pragmatics, definite article, personal name, corpus linguistics, PRESEEA.

1. Introduction

It has been shown that the presence of definite article before a personal proper name varies widely across languages and dialects. In some languages the combination of a definite article (henceforth DA) before a personal name (henceforth PN) is viewed as ungrammatical while in other languages it is a required morphosyntactic feature (Lyons 1999). Additionally, there are languages that require DA+PN only in certain syntactic positions as well as languages in which DA+PN is variable across dialects (Matushansky 2006). Spanish is an example of a language that presents variable DA+PN realizations across dialects. For example, one could say either *¿Cómo está Mari?* ‘How is Mari?’ or *¿Cómo está la Mari?* ‘How is [the] Mari?’. While DA+PN realizations have been shown to be present across several varieties of Spanish (De Mello, 1992; Hunt, 2021), it has been found to be frequently used only in Chilean Spanish (De Mello 1992; Hunt 2021; Tieperman 2018, 2020). Nevertheless, there is a lack of quantitative studies analyzing the variable phenomenon of DA+PN across different varieties of Spanish with the main exception being Hunt (2021) who examined DA+PN in written speech. However, as DA+PN has been observed to be more common in informal spoken speech (Calderón Campos 2015, 2018), the current study comparatively analyzes the factors that govern DA+PN use in spoken speech in three geographically diverse regional varieties of Spanish including Chilean, Mexican, and Andalusian Spanish. This study specifically builds on De Mello’s (1992) small scale corpus study to provide an overall descriptive statistical analysis of the frequency of DA+PN across three varieties as well as an inferential statistical analysis to demonstrate which linguistic and social factors favor DA+PN realizations across these three varieties. Here only PNs that are in a referential context, that is, referring to a third person, are examined as part of the study.

This paper begins with Section 2 that provides the relevant background information, with a focus on sociolinguistic and pragmatic variation, previous studies of DA+PN, and the research questions. Then Section 3 reviews the methods. Section 4 presents the results. Finally, Section 5 discusses the results in connection to the previous literature and research questions.

2. Background

2.1 Variationist sociolinguistics & pragmatic variation

The study of pragmatic variation currently has two main directions which include VARIATIONIST PRAGMATICS (Cameron & Schwenker 2011) and VARIATIONAL PRAGMATICS (Barron 2017; Schneider & Barron 2008; Barron & Schneider 2009; Schneider 2010, 2021). A variationist approach to pragmatic variation utilizes the quantitative approaches of variationist sociolinguistics, which involves the five-step process of identifying a pragmatic variable and its variants, determining the context in which the variable appears, coding, conducting statistical analysis, and interpreting the results (Cameron & Schwenker 2011: 466; Terkourafi 2012: 297). Following this approach, it is the last stage, that of the interpretation of the results, where pragmatic meaning first enters into the analysis (Terkourafi 2012: 297). Pragmatic meaning is commonly referred to as a speaker's intention (Grice 1969). Terkourafi (2012: 313-314) suggests that this approach is "methodologically robust", but allows "only a limited (explanatory) role for pragmatics" and is "more interested in linguistic rather than social constraints on pragmatic variation."

Variational pragmatics, on the other hand, is an interdisciplinary field that exists at the intersection of pragmatics, sociolinguistics, and dialectology (Barron & Schneider 2009; Barron 2017; Schneider & Barron 2008; Schneider 2010, 2021). Unlike many traditional pragmatic studies, the field of variational pragmatics is interested in examining macro-social factors (i.e., gender, age, social class, region, ethnicity) and the role of these factors in speech acts at the pragmatic level (Schneider & Barron 2008; Barron 2017; Schneider 2010, 2021). There is a strong dialectological element of comparative regional variation (Schneider & Barron 2008; Barron & Schneider 2009). Consequently, one of the main purposes of variational pragmatics is "to problematize the underlying assumption that populations of native speakers are homogeneous with respect to their pragmatic choices" (Terkourafi 2011: 356). In this way, it is similar to the variationist comparative method (Poplack & Tagliamonte 2001), but instead of comparing varieties to examine how a particular sociolinguistic variable differs between communities based on linguistic factors, variational pragmatics compares speech communities of different dialects to see how a particular pragmatic variable varies based on macro-social factors across dialects. Due to its intra-lingual nature, a common sub-field within variational pragmatics is known as REGIONAL PRAGMATIC VARIATION (Placencia 2012). Thus, the intersectionality of this subdiscipline of variational pragmatics allows for a variety of new perspectives to be considered in what traditionally would have been studied in three distinct areas.

Both approaches differ methodologically in their conceptualization of the pragmatic variable. As Cameron & Schwenker (2011) note, variationists pursuing this agenda may follow Walker's (2010) distinction between "form-based" and "function-based" approaches to define the variable context (Cameron & Schwenker 2011: 467). The "form-based" approach is more common in studies of phonetic and phonological variation as traditionally the sociolinguistic variable has been defined as "a class of variants which are ordered along a continuous dimension and whose position is determined by an independent linguistic or extralinguistic variable" (Labov 1966: 15). That is, the linguistic variable is defined as an "alternate ways of saying 'the same' thing" (Labov 1972: 188, 1978: 2). This has been debated since Lavandera's (1978)

critique in which such a definition functions best for phonetic and phonological variables, but has proven more complicated for morphosyntactic or discursive-pragmatic variables, as two different variants do in fact have semantic differences. Variational pragmatics may provide a practical solution to this dilemma as Schneider (2010: 251) points out that the term *variational* as opposed to *variationist* in the name is not an accidental one and allows for less strict definitions of “variables” and “variants.” Specifically, the pragmatic variable is simply “‘that which varies’ on the pragmatic level across varieties of the same language, and the respective variants are the options available of which different social groups of speakers make use.” Of note is that Terkourafi (2011: 357) indicates that a pragmatic variable should be defined in two ways: “(i) variation in the range of pragmatic meanings expressed by a single form, and (ii) variation in the range of forms that express a single pragmatic meaning.” Thus, following Terkourafi’s definition, the current study would fall more into *variationist* pragmatics (Cameron & Schwenker 2010). However, given the focus on dialect comparisons, this study is at the intersection of *variationist* and *variational* pragmatics.

Rather than ascribe to one methodological approach or the other, the current study blends these approaches in using the robust coding and statistical methodologies of *variationist* pragmatics, but also incorporating the role of social factors and dialect comparison, some of the principle foundations of *variational* pragmatics. This complementary method has the advantage of providing insights into which social and linguistic factors govern the use of DA+PN and how this varies between dialects. This combination of methods is particularly useful for examining DA+PN as several studies (Caro Reina 2022; Caro Reina & Helmbrecht 2022; De Mello 1992; Hunt 2021; Tieperman 2018) have indicated that social, linguistic, and pragmatic factors may govern DA+PN use across different varieties. However, the limitation here is that, as Terkourafi (2012) notes, we will not be able to go beyond mere inferences of the pragmatic meaning of the variation, which would result in the fifth step. Consequently, this study is seen as a first step in examining the role of macro-social factors in DA+PN realization across dialects (*variational* pragmatics) by using the statistically robust methods of *variationist* sociolinguistics. A *variational* pragmatic perspective has been used frequently to examine pragmatic differences across dialects of Spanish, in the context of service encounters (Bataller 2015; Escalona Torres 2019; Félix-Brasdefer 2015; Félix-Brasdefer & Yates 2019; Placencia 2008; Placencia & García 2019), invitations (García 2008), and discourse markers (Fuentes-Rodríguez et al. 2016). While few studies have examined *variational* pragmatics using a corpus approach, there are several scholars who have called for more collaboration between corpus linguistics and *variational* pragmatics (Clancy 2011; Murphy 2021). To the best of the knowledge of the authors, this is the first study of DA+PN that analyzes a corpus of spoken language with a *variational* and *variationist* pragmatics perspective.

2.2 Sociolinguistic/pragmatic variable: Definite article with personal names

Proper names are thought to be inherently definite¹ (Caro Reina & Helmbrecht 2022; Hawkins 1978; Helmbrecht 2022; Löbner 1985; Lyons 1999; RAE/ASALE 2009: 836;

¹ While it is acknowledged that proper personal names often behave in a definite manner, scholars in linguistics and philosophy differ with regard to the specific syntactic and

Wotjak 1985; Van Langendonck 2007). That is, they refer to a singular and unique reference, a person or an object, in the world (Helmbrecht 2022: 124; Caro Reina & Helmbrecht 2022: 1). Given their definiteness, one would assume that proper names “are not compatible with definiteness marking- that is, with definite articles” (Caro Reina & Helmbrecht (2022: 6). However, many languages do allow for the definite article with proper names.

The construction of DA+PN is not universally used in the same way across languages, even among languages which do utilize DAs as definiteness markers. For example, Helmbrecht (2022: 138) divides languages into three categories with regards to DA+PN realizations: (i) those that do not allow the DA with PN; (ii) those that require the DA with PN; and (iii) those that allow (but do not require) the DA with PN depending upon “syntactic, semantic, pragmatic or sociolinguistic conditions.” The first language group, languages which do not allow the DA with the PN include English (Helmbrecht 2022). The second language group, in which the DA use is required with the PN include languages like Modern Greek and Albanian (Lyons 1999: 121). The third language group, in which DA with PN varies based on dialect variety and other factors is the focus of the current paper. It has been shown that DA+PN varies across varieties of Portuguese, Spanish, Italian, and German (to name a few). Regarding Portuguese, in European and Brazilian² Portuguese the DA with PN is more common in a more informal register (Caro Reina 2022; Cunha & Cintra 2017). Spanish also demonstrates differences in DA+PN realizations across dialects³ (Caro Reina 2022; De Mello 1992; Hunt 2021). While in “Standard” German the DA with the PN is considered ungrammatical, in other dialects of German such as in Southern Germany the DA is more common than not (Helmbrecht 2022; Werth 2020). Similarly, “Standard” Italian does not allow the DA with PNs, but in other varieties, such as in Northern Italy, the DA is optional (Helmbrecht 2022). Furthermore, it has been shown that the DA is required in certain syntactic positions but not others for languages such as “colloquial Icelandic, Northern Norwegian and Northern Swedish, Catalan, Tagalog, the Uto-Aztecan language Pima, Albanian” (Matushansky 2006: 287). While there have been several cross-linguistic studies (Caro Reina 2022; Handschuh 2017; Helmbrecht 2022; Stolz & Levkovych 2022) to examine differences between languages, there have also been a few studies that have examined several pragmatic, lexical, morphosyntactic, and social factors that govern the use of DA with PN. These factors include: famous name (common versus ordinary name) (De Mello 1992), personal name type (first name, last name, nickname, etc.) (Caro Reina 2022; Caro Reina & Helmbrecht 2022; De Mello 1992) as well as dialect variety (Caro Reina 2022; Caro Reina & Helmbrecht 2022).

semantic functions that proper personal names serve (Burge 1973; Allerton 1987; Geurts 1997, 2002; Abbot 2002)

² Christodulelis (2014) examined which factors favor DA+PN in Brazilian Portuguese. Based on an acceptability judgment task to determine the felicity of different situations, the results indicated that only “weak familiarity” (Roberts 2003) was needed to favor the DA.

³ For a summary of difference across European and Latin American varieties of Spanish see Caro Reina (2022: 75-76).

2.3 Grammar accounts of DA+PN in Spanish

While grammars indicate that the prescriptive norm is the absence of the DA with PNs, descriptively, they also recognize its existence across varieties of Spanish as well as several social connotations association with DA+PN. It is said that “*Los nombres propios rechazan, en principio, la presencia del artículo, ya que constituyen de por sí SSNN definidos*” ‘Proper names reject, in principle, the presence of an article, as they already constitute a definite nominal phrase themselves’ (Leonetti 1999: 811). Or, said less categorically, “*los nombres propios, al constituir por sí solos expresiones referenciales, no van normalmente acompañados del artículo*” ‘proper names, in constituting referential expressions by themselves, are not usually accompanied by the article’ (Laca 1999: 924). However, descriptive grammars acknowledge that DA+PN variation occurs across dialects (RAE/ASALE 2009: 840). It has been noted that DA+PN realizations have particular social connotations of ‘familiar or colloquial’ (Fernández-Leborans 1999: 112-113), “*connotaciones populares y familiares*” ‘familiar and popular connotations’ (Laca 1999: 924), or “*lengua popular*” ‘informal language’ (RAE/ASALE 2009: 840) and are not as common with “*la lengua estándar*” ‘standard language’ (RAE/ASALE 2009: 840). It is also said that the use of DA+PN is associated with a specific socioeconomic status. For example, the *Diccionario panhispánico de dudas* (RAE) states, “*En la lengua culta, los nombres propios de persona se emplean normalmente sin artículo [...] La anteposición del artículo, en estos casos, suele ser propia del habla popular*” ‘In the educated/cultured speech, proper personal names are employed normally without article [...] The preposed article, in those cases, tends to be from slang/folk speech’ (RAE 2005: 249). Laca (1999: 924) similarly states that DA use with PN “*pertenece casi exclusivamente a la lengua hablada de nivel socio-cultural no alto*” ‘pertains almost exclusively to the spoken language of a non-high socioeconomic level’. While prescriptive grammar associates DA+PN with lower socioeconomic groups, it also acknowledges that DA+PN is common in “*el habla culta*” in Chile (RAE 2005: 249). Finally, while it is said to be common with first names, it is even more common for *los apodos* ‘nicknames’ (RAE/ASALE 2009: 845). While these grammar accounts provide insights that DA+PN realizations across dialects and may have social connotations, it is important to provide empirical evidence regarding these claims of DA+PN variation.

2.4 Sociolinguistic studies of DA+PN variation in Spanish

In spite of the dialect variation that exists in Spanish regarding DA+PN realizations, surprisingly few studies have quantitatively analyzed this feature. De Mello (1992) was the first study to quantitatively examine this variable phenomenon in Spanish with a descriptive statistical analysis of the DA+PN phenomenon using a corpus of “*el habla culta*” ‘educated speech’ of ten Spanish-speaking cities (Bogotá, Buenos Aires, Caracas, La Habana, Lima, Madrid, Mexico City, San Juan, Santiago [Chile], Sevilla). De Mello found that while this structure appears in various instances throughout the Spanish-speaking world, it only frequently occurs in Santiago, Chile (98 of the total 135 DA+PN found in the study). Some cities, such as La Habana, Lima, Mexico City, and Sevilla, demonstrated no DA+PN realizations. De Mello (1992: 222-223) also found that DAs were more common with female names than male names, particularly for Santiago (90 of the 98 Santiago tokens were female names). However, it should be

noted that if one takes out the tokens from Santiago, there are slightly more tokens for male referents ($n = 23$) than female referents ($n = 14$) for the remaining tokens across cities, which could indicate that only in Santiago is a definite article more common with a female referent name. De Mello (1992) also examined famous versus common names without a clear pattern as well as name type (first name, diminutive, nickname, last name, full name) with first name being the most frequent followed by nicknames⁴.

Of recent, there have been several additional quantitative studies building on De Mello's (1992) DA+PN findings. In Santiago, Chile, Dixon (2011) conducted an anonymous questionnaire balancing participants across social class, age, and gender. Participants were given a questionnaire with several different situations varying in formality and were asked whether or not they agreed on the use of "el" or "la" with a PN in each situation. Overall, most participants disagreed with the use of DA+PN in more formal contexts, indicating that DA+PN may be more common in informal contexts.

From a diachronic perspective, Calderón Campos (2015) quantitatively analyzed DA+PN realizations utilizing the *Corpus diacrónico del español* (CORDE) in examining texts from the 13th through the 19th century. He found that the pragmatic meaning of DA+PN varied based on the type of textual genre that it appeared in. In the chronicles and testimonies DA+PN served a more anaphoric textual value, while in "humanistic/didactic" texts, it served to indicate a famous person, "the known by all" (2015:90). Finally, in literary texts that imitated spoken speech, DA+PN demonstrated several values including (i) intimacy or closeness towards the person; (ii) either a subtle hint of a slight or of a praise; and, (iii) an indication of a person from a lower socioeconomic status. Calderón Campos (2018: 95) indicates that today in spoken speech, the only use of DA+PN that still persists is "the informal use in which the article may help activate inferences in the addressee's mind" (see also Carranza Brito 2008).

Christodulelis (2017) conducted an online survey to determine acceptability of DA+PN use across all varieties of Spanish. It was found that DA presence or absence was the greatest predictor of acceptability, even more so than other pragmatic factors commonly reported anecdotally, such as in "emotionally charged contexts" (2017).

Tieperman (2018) analyzed the constraints governing DA+PN in Chilean Spanish by examining episodes of the Chilean TV series *Casado con hijos* (Roos Film 2006), a Chilean spinoff of the American show *Married...with Children*. Of all PN produced in the show, 54.6% were accompanied with a DA. A mixed-effects logistic regression found that referent gender was the most significant predictor of DA+PN realizations in which female referents (i.e., *la Laura*) favored DA+PN realizations more than male referents (i.e., *el Pablo*).

Aliaga Rovira et al. (2020) examined 5 Chilean newspapers (*La Tercera*, *La Cuarta*, *Las Últimas Noticias*, *Emol.cl*, & *La Nación*) over 28 consecutive days in spring 2018. They examined the presence, or lack thereof, of the DA with the PN between masculine and feminine names. Overall, they found that the DA occurred 65.9% of the time with female names and only 4.5% with male names. Given the

⁴ While an overall frequency count is helpful, it should be noted that overall patterns are difficult to ascertain as these frequency counts do not include the realization of PNs without the DA.

formal context of a newspaper, they suggest that such gendered differences could be evidence of social prejudice towards women.

Most recently, Hunt (2021) conducted a large-scale corpus study that examined the use of DA+PN across eight varieties of Spanish (Argentina, Chile, Colombia, Cuba, Mexico, Paraguay, Peru, Spain) using the *Corpus del Español: Web/Dialects* (Davies 2016). Overall, DA+PN realizations were quite low. While Chile presented significantly more DAs than any other dialects, it was only present for 1.49% of all PNs. However, as Hunt (2021: 19, 39) indicates, given that this corpus uses written instead of spoken speech, while overall trends of use of DA+PN should be fairly representative of dialect differences, the overall frequencies are most likely lower than a spoken corpus. Of note is that referent gender was not a significant predictor for any of the dialect varieties. What is also noteworthy in this study is that DA+PN realizations were present in all eight varieties of Spanish. As Hunt (2021: 39) indicates, while this corpus study provides an incredible amount of data points for the analysis, another limitation is that it does not separate tokens based on city but only country. Consequently, as Hunt mentions, it could be that Santiago, Chile produces more DA+PN than other regions of the country, but this would be masked by grouping all of Chile together. Thus, in building on these studies, the current study aims to examine three specific cities using the online PRESEEA corpus of spoken speech of a socially stratified sample (PRESEEA 2014-).

2.5 Research questions

This study was carried out to provide a quantitative analysis of DA+PN variation in three different varieties of the same language including: the Spanish of Andalucía, Spain, the Spanish of central Mexico, and the Spanish of central Chile. The following two research questions guided the study.

RQ1: What are the overall rates of DA+PN in the respective corpora of Granada, Spain, Mexico City, Mexico, and Santiago, Chile?

RQ1_H: It was hypothesized that DA+PN realizations would prove to be the most frequent in the corpus of Santiago, Chile, based on previous comparative studies (De Mello 1992; Hunt 2021). De Mello (1992) did not find any tokens of DA+PN for either Mexico City or Sevilla (Andalusian Spanish), however De Mello's corpus was based on *el habla culta* 'educated speech', while PRESEEA includes speech for all educational levels. Additionally, Hunt (2021) did find DA+PN in Spain. Based on the authors' participant observations throughout Andalucía, it was hypothesized that Granada would have higher overall rates of DA+PN than that Mexico City.

RQ2: What are the linguistic and social (extra-linguistic) constraints that govern the variable use of DA+PN in the respective corpora of Granada, Mexico City, and Santiago?

RQ2_H: As previous studies (De Mello 1992; Aliaga Rovira et al. 2020; Tieperman 2018) found that DA+PN realizations were more favored with a female referent names as opposed to male referent names, it was hypothesized that referent gender would prove to be a significant social factor in female referents would favor DA+PN realizations. Additionally, based on qualitative prescriptive comments (RAE 2005), it was hypothesized that those with more educational attainment may disfavor DA+PN realizations more than those with less educational attainment.

3. Methodology

3.1 Participants

Using the available online data from the PRESEEA corpus⁵, we analyzed the transcribed interviews from the three cities of Granada, Spain, Mexico City, Mexico, and Santiago, Chile. The rationale for selecting these three cities in particular was that the three varieties represent three distinct geographic varieties of Spanish in which De Mello (1992) had observed differences in a small-scale study. Previous research (Aliaga Rovira et al. 2020; De Mello 1992; Hunt 2021; Tieperman 2018) indicated that indicated that Santiago would demonstrate highly frequent DA+PN realizations. Given the lack of quantitative research on other speech communities, both Mexican and Andalusian⁶ Spanish were selected based on the authors' participant observation and in conversations with speakers from these three varieties in which DA+PN was perhaps relatively infrequent (but definite present) in Andalusian Spanish and almost non-existent in Mexican Spanish. Thus, it was hypothesized that each of these varieties would show a different amount of DA+PN realizations and as such would potentially reveal differences between speech communities in which social and linguistic factors (dis)favor DA+PN realizations.

There were 54 participants included in this study ($n = 18$ per speech community⁷), whose data was collected from the *Proyecto para el estudio sociolingüístico del español de España y de América* 'Project for the sociolinguistic study of Spanish in Spain and America' corpus (PRESEEA 2014-). The PRESEEA corpus is a searchable online database comprised of 18 sociolinguistic (semi-structured) interviews per city from across the Spanish-speaking world. These sociolinguistic interviews have been transcribed and uploaded to the corpus. These interviews follow eight thematic modules including: greetings, weather, place of residence, family and friends, customs, danger of death, important anecdotes, and desire of economic improvement. The corpus lends itself to analyzing morphosyntactic and discursive variation across varieties of Spanish (Moreno Fernández 2005).

In each city, the 18 informants were evenly divided by gender, age group, and educational level, such that in each city, there is one representative informant for every possible combination of these three social groups as seen in Table 1. The only exception to this was in Santiago, in which case although there were still 18 informants, there were no men between the ages of 35-54 who were classified as being of a lower educational level. Instead there were two men between the ages of 35-54 who were classified as being of a medium educational level. Unfortunately, there were

⁵ One reviewer questioned whether the PRESEEA Corpus is well suited for the study of PNs with the use of the DA. This is a valid question as the overall low tokens of PNs would indicate that PNs in general are not common in spontaneous spoken speech. However, we believe that the PRESEEA corpus allows a great first step to provide a quantitative analysis of DA+PN in informal spoken speech as most documentation of this feature has been based on anecdotes, traditional dialectology, qualitative ethnographic observations, and/or grammar textbooks.

⁶ Granada was selected as at the time of coding; the data was not available for Sevilla.

⁷ As one reviewer pointed out, only 18 speakers are accessible per community via the online PRESEEA corpus as opposed to the 54 recordings that exist for each community.

several speakers ($n = 10$) without any tokens of PNs and therefore were not included in the final analysis. Thus, although the speech of 54 speakers was analyzed, only the data from 44 speakers was included in the statistical analysis ($n = 13$ Granada, $n = 17$ Mexico City, $n = 14$ Santiago).

Table 1. Distribution of participants, divided according to gender, age, and educational level for each speech community following the PRESEEA norms.

Age / Education	Male			Female		
	Lower	Middle	Upper	Lower	Middle	Upper
20-34	1	1	1	1	1	1
35-54	1	1	1	1	1	1
55+	1	1	1	1	1	1

3.2 Dependent variable, data collection, & data coding

The dependent variable in this study was the binary realization of the DA, or lack of DA, with a PN coded by the first author. Every PN mentioned in a referential context, that is, referring to a third person as opposed to the speaker/interviewee or the listener/interviewer, was included as seen in each transcript. It was coded for as *yes* when the PN was realized with the DA, as shown below in Examples (1)-(5) from PRESEEA.

- (1) “*uno de mis mejores amigos, el L, él es súper buen bailarín también*”
‘one of my best friends, [the] L, is a super good dancer also’
(female, SCHI_M13_079)
- (2) “*...la S. sí me puede traer algo...*”
‘... [the] S. can bring me something...’
(male, Santiago, SCHI_H23_085)
- (3) “*...y la R nació el ochenta y nueve.*”
‘...and [the] R was born in eighty-nine.’
(female, Santiago, SCHI_M22_055)
- (4) “*andábamos tres- cuatro, el L, el V, el P y el C, cinco éramos.*”
‘three-four of us went, [the] L, [the] V, [the] P, and [the] C, we were five.’
(male, Santiago, SCHI_H32_061)
- (5) “*¿Este es el D. de Barcelona?*”
‘This is [the] D. from Barcelona?’
(male, Granada, GRAN_H22_026)

Alternatively, when there was no DA realization with the PN, it was coded as *no*, as in the case of Examples (6)-(10) from PRESEEA.

- (6) “*...aunque ya se casó M.*”
‘...although M already got married.’
(male, Mexico City, MEXI_H32_066)
- (7) “*Llévate a L, yo no voy.*”
‘Take L, I’m not going.’
(female, Mexico City, MEXI_M21_096)
- (8) “*yo me acuerdo que merendando, cuando iba a buscar a E...*”
‘I remember while afternoon snacking, when I was going to look for E...’
(male, Granada, GRAN_H23_007)

- (9) “E. me miraba con cara de ‘¿qué dice este tío?’”
 ‘E. looked at me with a face of ‘what is this dude saying?’’
 (male, Granada, GRAN_H23_007)
- (10) “*hay un hombre allí que es D. que es molinero.*”
 ‘there is a guy over there that is D. who is a miller.’
 (male, Granada, GRAN_H21_043)

It is important to note that, as there were different teams in charge of conducting and transcribing the sociolinguistic interviews for the PRESEEA corpus, different transcription techniques were implemented. For example, in the corpus of Granada, the personal names in referential use were left in their complete forms⁸. In the corpora of Mexico City and Santiago, however, in general, names were reduced to a capitalized first letter of the name such as “*Vi a S*” ‘I saw S’ or “*Vi a la S*” ‘I saw [the] S’ with the initial S used to replace the PN for confidentiality reasons. Because of these differences, it was necessary to rely on contextual clues in the Mexico City and Santiago corpora. For example, if a person referred to in the transcript by the initial L was described as *bonita* ‘pretty’ with the feminine form of the adjective, it was inferred that “L” was female. However, if there were not sufficient contextual clues surround the personal names with initials to determine the gender of the referent, the referent gender was labeled unknown. This led to 126 tokens being excluded from the analyses as the referent gender was unknown. Originally type of personal name was coded (first name, last name, nickname), but given two of the cities only used first letters, this independent variable could not be reliably included in the statistical analysis. This is a limitation as previous scholars have suggested that nicknames in Spanish may favor the definite article (Laca 1999: 924; RAE/ASALE 2009: 845).

There were also some tokens that were excluded from the analysis as they fell outside of the envelope of variation. For example, the analysis did not include the case of religious figures such as *la Virgen* ‘the Virgin [Mary]’. Last names referring to the family (i.e., *los Sánchez*) were also excluded as all varieties produce the DA with a last name in the plural (Laca 1999: 925). Instances of third-person referential PNs that were spoken by someone other than the participants themselves were also excluded. That is to say, in many sociolinguistic interviews, there were either family members or friends present during the interview and occasionally shared their opinions. These tokens were not considered as there was not enough information provided about these speakers. Tokens in which other determiners, such as “*mi*” ‘my’, occurred directly before the PN were also excluded (i.e. “*mi Cristina*”), because they already occupied the space that other determiners, such as *el* or *la*, might otherwise occupy. Names of TV series (when there was a PN included in the TV series) or when someone was asking how a name was spelled were also excluded, in addition to first names that followed a title such as *don* (i.e. “*el don Luis*”), or for any names that occurred within the context of a song or work title (i.e. “*Don Manolito*”). The reason that titles such as *don*, *señor*, *señora*, etc. were excluded from the data collection process was because of the fact that the DA is often used in front of these titles across varieties.

⁸ Since the original coding of this study, the online PRESEEA corpus has been updated and the Granada corpus now only presents PNs with the initial letter of each name.

3.3 Independent variables

The study examined five social (extra-linguistic) independent variables for each referential PN token, which included: (i) CITY (Granada, Mexico City, Santiago); (ii) SPEAKER GENDER (male, female); (iii) AGE GROUP (18-34, 35-54, >55); (iv) EDUCATIONAL ATTAINMENT (low, medium, high); and (v) REFERENT GENDER⁹ (male, female). For age group, the study followed the groupings providing by the PRESEEA corpus of Group 1 (18-34 years), Group 2 (35-54 years), and Group 3 (55+ years). Similarly, for educational attainment, the current study followed the PRESEEA corpus' grouping of low (0-5 years of schooling), middle (10-12 years of schooling), high (15+ years of schooling). Finally, referent gender refers to the gender of the individual being referred to as a third person within the context of the conversation.

For each referential PN token, two linguistic variables were also coded: (i) POSITION IN UTTERANCE (phrase-initial, phrase-internal); and (ii) PRECEDING PHONOLOGICAL CONTEXT (vowel, consonant, following a pause). Position in utterance refers to whether the DA realization or lack thereof occurred at the beginning of the phrase or in the middle of the phrase; therefore, it was accordingly coded as either phrase-initial or phrase-internal. Preceding phonological context refers to whether the phoneme preceding the variable DA realization was a vowel or a consonant and was therefore coded as either vowel, consonant, or following a pause. An example of each possible preceding phonological context would be (i) *Vi a la Laura* (vowel), (ii) *Estaba hablando con la Laura* (consonant), or (iii) *La Laura me vino a visitar el otro día* (following a pause). As noted above, differences in transcription between the three speech communities regarding names made it difficult to code for type of name, which is why this linguistic factor was not included. In the Granada corpus, since the PNs were written in their entirety in the transcripts at the time of analysis, type of name was coded. For example, it was recorded whether the name was a first name, a last name, a nickname, a full name (comprised of the first and last name), etc. Due to the transcription methodology used in the Mexico City and Santiago corpuses, it was not possible to include type of name for those regions, and therefore, this variable was excluded from further analyses. While one option could have been to eliminate all tokens from Granada that were not first names, we opted against this idea as the PNs from Mexico City and Santiago most likely had other types of names as well but given they were only first initials, we were unable to reliably code for type of name. Thus, we included all PNs from each community in the analysis recognizing this as a limitation. Type of name is worthwhile factor to examine in future studies as Laca (1999: 924) indicates DA usage may be more common with nicknames.

Three pragmatic variables were also coded including: (i) the closeness of the participant's relationship with the referent; (ii) the quality of the participant's relationship with the referent; and (iii) the quality of the conversational context. With regard to the closeness of the participant's relationship with the referent, it was coded as being either close or distant. The "quality of the conversational context" was coded as being positive or negative. The term "conversational context", referring to the

⁹ One reviewer suggested that REFERENT GENDER be listed as a linguistic factor instead of a social factor. While the authors recognize the grammatical gender component of referent gender, we have decided to place it under social factors as we believe there is still a social component to PNs within societal norms of *doing* gender (Eckert & McConnell-Ginet 2003: 10).

content of the conversation at the time of the utterance of the token, was also coded positive or negative. Given the nature of the corpus, it was determined that there were not always enough context clues to reliably and objectively code for each context and therefore these independent variables were not included in the statistical analysis.

3.4 Statistical analysis

First, a random forest was conducted in R (R Core Team 2022) using the *cforest* function from the *party* package (Hothorn et al. 2020) to determine the variable importance. This follows the suggestion by Tagliamonte and Baayen (2012) as then researchers are able to place each independent variable in order of most to least importance in the regression models. That is, because the order one places each independent variable in the regression model can impact the model. Once this was completed, a mixed effect logistic regression model was fitted for all of the data together using the *lmer* function (Bates et al. 2015) and *lmerTest* function (Kuznetsova et al. 2017). To provide a goodness-of-fit of the variation for each final model, the marginal R-squared (R^2_m) and conditional R-squared (R^2_c) values are listed (Nakagawa & Schielzeth 2013). All independent variables were tested within model construction and all possible interactions were tested as well. Several different regression models were compared using ANOVA testing. Post-hoc analyses were conducted with estimated marginal means (Lenth et al. 2018) for all significant main effects of discrete independent variables with more than two categorical levels. Finally, a conditional inference tree was conducted using the *cforest* function from the *party* package (Hothorn et al. 2020) to explore the relationship between significant main effects in the regression model. All figures were created with *ggplot2* (Wickham 2016).

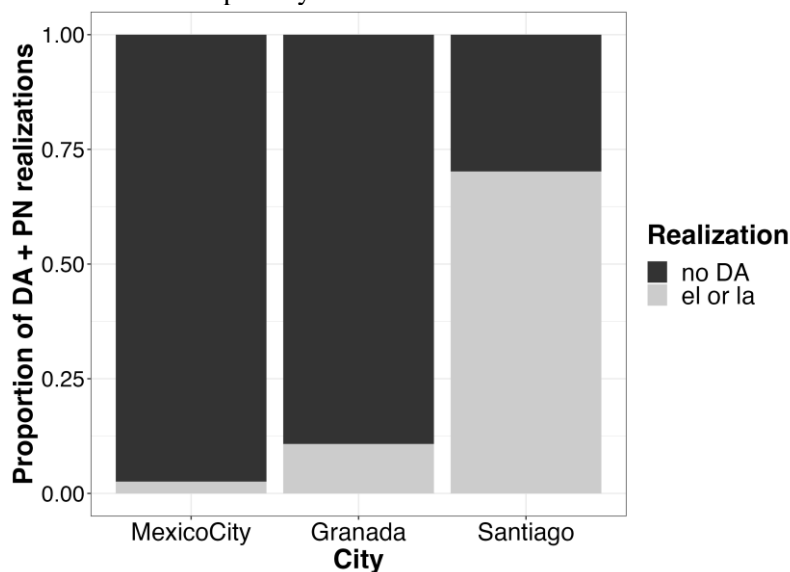
4. Results

4.1 Overall descriptive statistics

For all three cities together, of the 1,039 total tokens included in the analysis¹⁰, there were 273 DA+PN realizations (26.3%) and 766 PN realizations without a DA (73.7%). Specifically, for Mexico City, of the 585 total tokens there were 15 DA+PN realizations (2.6%) and 570 PN realizations without a DA (97.4%). Of the 102 total tokens for Granada, there were 11 DA+PN realizations (10.8%) and 91 PN realizations without a DA (89.2%). Of the 352 total tokens for Santiago, there were 247 DA+PN realizations (70.2%) and 105 PN realizations without a DA (29.8%). A Chi-square test revealed a significant difference in the realization of DA+PN between the three speech communities, $\chi^2(2) = 532.54$, $p < 0.0001$, as seen in Figure 1. Overall percentages should be taken with caution given the interspeaker variation within each speech community (see Appendix I). While these percentages reveal important trends, the inferential statistics with participant as a random factor in mixed effects models allows for us to more accurately observe which factors most favor DA+PN.

¹⁰ The 1,039 tokens in the analysis does not include the 126 unknown referent gender tokens. Those tokens were excluded prior to statistical analyses.

Figure 1. DA+PN realizations per city.



4.2 Inferential statistics

The results of the random forest suggest that for DA+PN variation, the most important predictor is participant (i.e., speaker), followed by city, education, referent gender, speaker age, position in utterance, preceding phonological context, and finally speaker gender (see Figure 2). Following this variable importance, the mixed effect logistic regressions were conducted. As the factors of age, position in utterance, preceding phonological context, and speaker gender were not statistically significant, they will not be discussed further.

Figure 2. Random forest of the DA+PN variation.

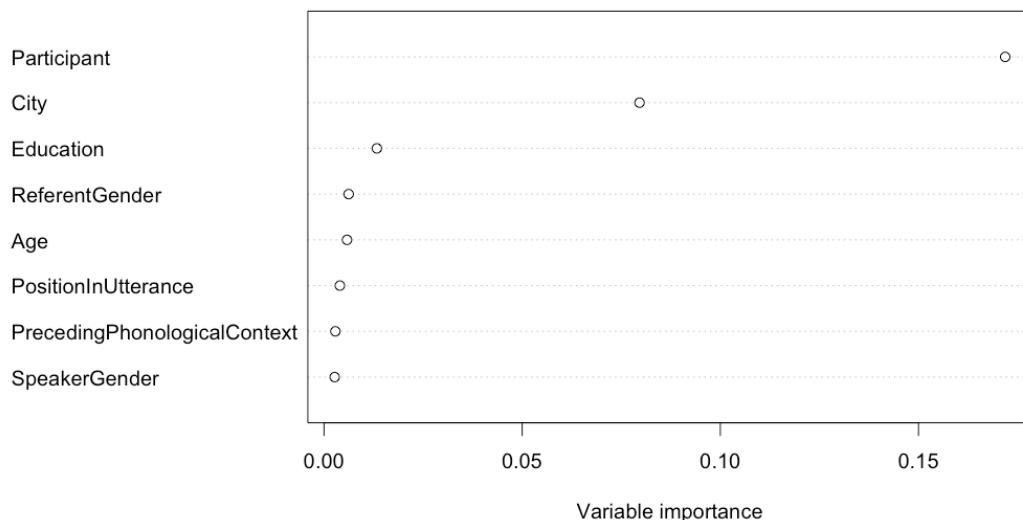


Table 2 presents the best fit mixed effects logistic regression, which displays the estimate, standard error (*SE*), t-value, percent of DA+PN per level, number of tokens per level, and finally *p*-value. Given the model is in reference to DA realizations (*el/la*), positive estimates indicate that the listed factor level favors DA more than the reference level and that negative estimates indicate that the listed level disfavors DA realizations more than the reference level.

Table 2. Summary of mixed effects logistic regression for DA+PN variation in reference to DA (*el/la*) realizations (vs. null realization), speaker as a random factor; $n = 1,039$ (R^2_m : 0.59, R^2_c : 0.67).

Predictors	Estimate	SE	z-value	% DA	n	p-value
(Intercept)	1.81	0.43	4.16	---	---	***
City (Ref = Santiago)	---	---	---	70.2	352	---
Granada	-3.62	0.73	-4.96	10.8	102	***
MexicoCity	-5.46	0.71	-7.71	2.6	585	***
Education (Ref = Medium)	---	---	---	42.2	436	---
Lower	-0.19	0.56	-0.34	41.0	134	0.731
Upper	-1.28	0.50	-2.53	7.2	469	*
Referent Gender (Ref = Female)	---	---	---	29.9	412	---
Male	-1.12	0.30	-3.77	23.9	627	***
City:ReferentGender (Sant:Female)	---	---	---	79.6	142	---
Granada:Male	0.63	0.76	0.83	9.1	55	0.408
MexicoCity:Male	1.70	0.68	2.49	3.0	362	*

Note: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

The main effect of city demonstrates that speakers from Santiago favor DA+PN realizations more than those of Granada and Mexico (Figure 1). A post-hoc analysis of the estimated marginal means shows that Santiago speakers produce significantly more DA+PN than Granada speakers ($p < 0.001$) and Mexico City speakers ($p < 0.0001$), but that there were no significant differences between speakers from Granada and Mexico City ($p = 0.12$). The main effect of education demonstrates that those with the highest amount of educational attainment produce the least amount of DA+PN (Figure 3A). A post-hoc analysis of the estimated marginal means shows that those from the middle educational level produce more DA+PN than those of upper educational attainment ($p < 0.5$), but no more than those from the lowest educational levels ($p = 0.94$). There were no differences in DA+PN production between those with the least and most educational attainment ($p = 0.18$). The main effect of referent gender demonstrates that DA+PN realizations were significantly more favored with female names than with male names (Figure 3B). Finally, a city by referent gender interaction indicates that Mexico City differs from Santiago in the gendered use of DA+PN (Figure 4). A post-hoc analysis of the estimated marginal means shows that speakers from Santiago produce more DA+PN for female than male referents ($p < 0.001$) while Granada ($p = 0.48$) and Mexico City ($p = 0.35$) speakers do not demonstrate differences in DA+PN realizations per referent gender. Regarding female referents, Santiago speakers produced more DA+PN realizations than those from Granada ($p < 0.0001$) and Mexico City ($p < 0.0001$), while there were no significant differences between Granada and Mexico City ($p = 0.078$). Regarding male voices, speakers from Santiago produced more DA+PN than those from Granada ($p < 0.001$) and Mexico City ($p < 0.0001$), while there were no significant differences between Granada and Mexico City ($p = 0.572$). The token count of city by referent gender for DA+PN realizations was: Granada (female names: 6/47, male names: 5/55); Mexico City

(female names: 4/223, male names: 11/362); Santiago (female names: 113/142, male names: 134/210).

Figure 3. Main effects of education (A) and referent gender (B) for DA+PN variation.

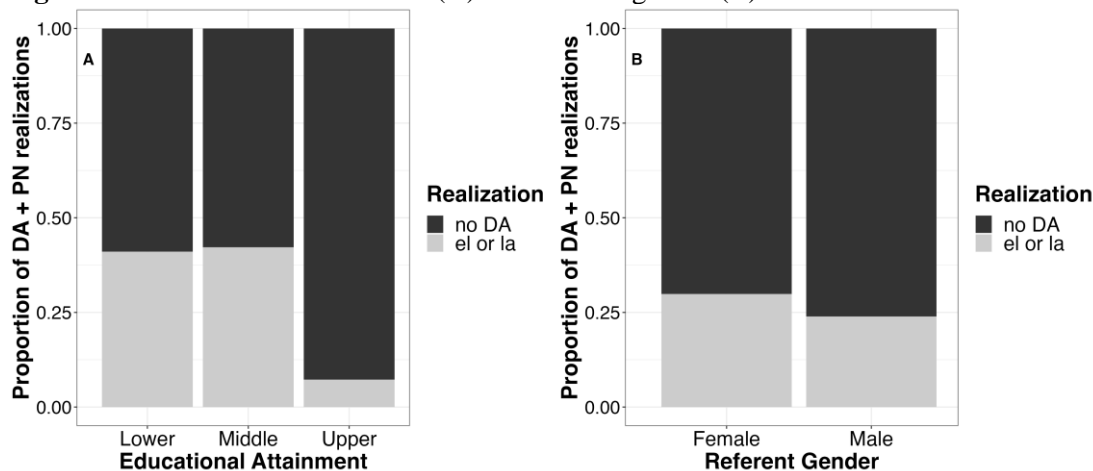
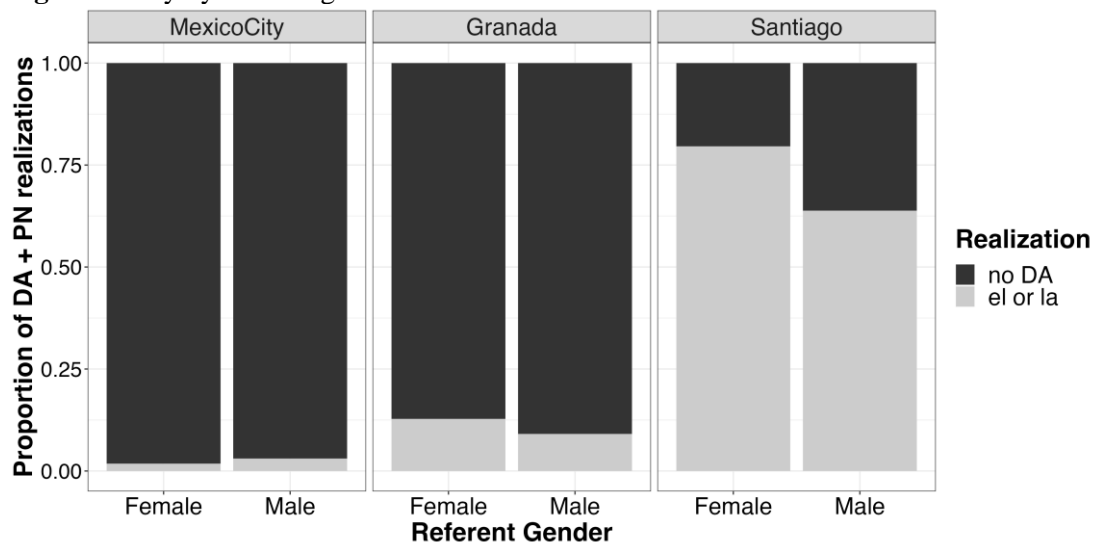
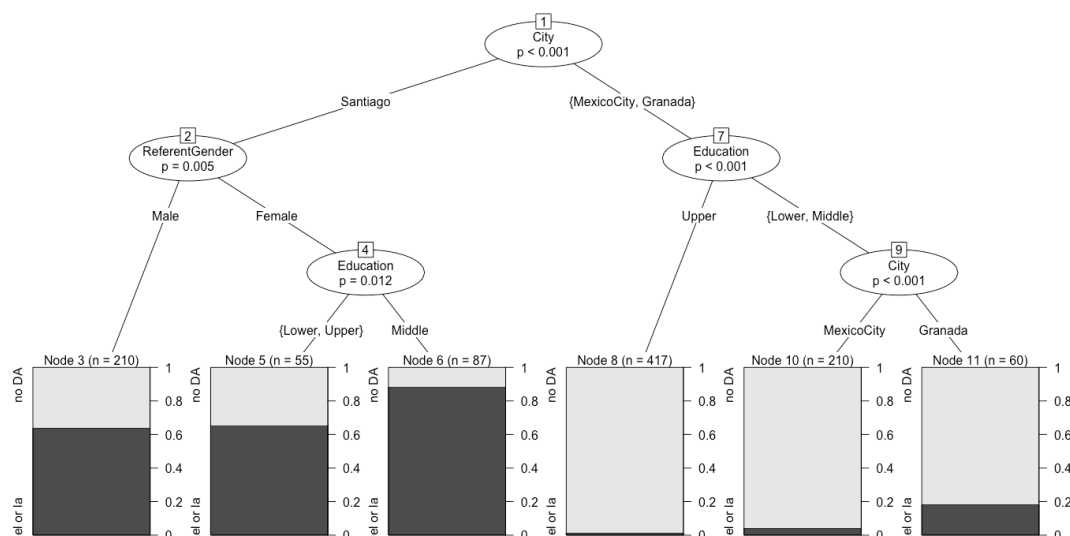


Figure 4. City by referent gender interaction for DA+PN variation.



While one interaction resulted significant, several other interactions were tested in various models and either were not significant or did not allow the model to converge. Thus, a conditional inference tree was conducted to further explore the relationship between the main effects (Figure 5).

Figure 5. Conditional inference tree of the factors that predict DA+PN variation (dark gray = *el* or *la* (DA); light gray = no DA).



The conditional inference tree suggests that the most important independent variable is city, in which the partition separates Santiago from Mexico City and Granada. For Santiago, similar to the regression model, it shows that referent gender is the next most significant factor in which female referents overall favor DA+PN realizations more than male referents. However, referent gender interacts with speaker education in which for female referents, it appears that speakers of middle level educational attainment favor DA+PN realizations more with female referents than those with either lower or upper level educational attainment. On the other half of the tree with Mexico City and Granada, for these communities, the most important factor is education level in which speakers of upper educational attainment from both cities demonstrate nearly no DA+PN realizations. Regarding lower and middle level educational attainment, speakers from Granada demonstrate significantly more DA+PN than speakers from Mexico City. Thus, while education was only a main effect in the regression model, here we can observe that it interacts with referent gender in Santiago and that while speakers of upper level educational attainment from Mexico City and Granada both produce the least amount of DA+PN, these cities do differ in DA+PN realizations among speakers of lower and middle level educational attainment.

5. Discussion

Here we first respond to the research questions, followed by implications, limitations, and future directions of research. *RQ1: What are the overall rates of DA+PN in the respective corpuses of Granada, Spain, Mexico City, Mexico, and Santiago, Chile?* It was hypothesized, based on previous studies (De Mello 1992; Hunt 2021) and ethnographic observations, that Santiago would have the highest rates of DA+PN realizations, followed by Granada, and then Mexico City. The descriptive statistics support this hypothesis. Although the overall rate of DA+PN realizations across the three cities was only 26.3%, in Santiago the rate was 70.2%, in Granada the rate was 10.8%, and in Mexico City the rate of realization was a mere 2.6%. The fact that the

DA was realized in a significant majority of cases in the Santiago corpus strongly supports De Mello's (1992) findings, where Santiago was the only speech community of the 10 cities analyzed that demonstrated a high frequency of DA+PN realizations. The present study also supports the findings of Tieperman (2018), which found that overall, in Chilean Spanish, DA+PN is more common than PNs without a DA, at least in informal spoken speech. That is to say, in Santiago, the use of a DA with a PN may be the norm in informal spoken speech. The percentage of DA+PN realizations found here contrasts with the findings from Hunt (2021) in which there were only a total of 1.49% DA+PN realizations. However, despite the low percentage, Hunt (2021) found that Chile's DA+PN realizations were statistically higher than the seven other varieties examined. Thus, as Hunt (2021:39) suggests such differences may reveal differences between written and spoken speech. As previous scholars (Calderón Campos 2015, 2018; Caro Reina 2022; RAE 2005) have suggested that DA+PN occurs more frequently in informal spoken speech, the current study in comparison to Hunt (2021) suggests that prescriptive grammar may influence written speech to overall disfavor DA+PN even in varieties such as Chilean Spanish than generally favor DA+PN. This would loosely support Dixon's (2011) findings that Chilean speakers were more likely to reject the use of DA+PN in more formal contexts. While written versus spoken speech is not the same as comparing informal to formal contexts of spoken speech, one can make the argument that written speech is perceived as more formal than spoken speech. This could be attested by Labov's (1972) *attention paid to speech* model at least with the differences in read speech of word lists and reading passages compared to semi-directed spontaneous speech. Hence, spoken speech may favor more DA+PN realizations than written speech.

RQ2: What are the linguistic and social constraints that govern the use of DA+PN in the respective corpuses of Granada, Mexico City, and Santiago? The mixed effects logistic regression indicates that the three most significant predictors for DA+PN realizations were the three factors of city, education, and referent gender; as well as an interaction between city and referent gender. The main effect for city indicates that in Santiago, DA+PN use is extremely favored, whereas in Granada and Mexico City its use is disfavored, particularly in Mexico City. The main effect of education indicates that DA use with a PN is disfavored by those with higher educational attainment. This supports qualitative prescriptive comments in which "*el habla culta*" generally omits the DA with a PN (Laca 1999: 924; RAE 2005: 249). It is suggested that more years of formal education would present speakers with more prescriptive grammar that may disfavor the use of DA+PN. This aversion could be attributed to a prescriptivist ideology of language use (Del Valle 2007, 2009; Curzan 2014). Thus, with the exception of Chile, DA+PN may be associated with those of less education, or at the very least, those with the most education may avoid the use of DA+PN. This is further supported by the conditional inference tree, particularly for Mexico City and Granada, where those with higher educational attainment demonstrated the least amount of DA+PN realizations. However, as the conditional inference tree indicated, in Chile, DA+PN to be the norm across the three educational levels. In Chile, the middle educational level produced the most DA+PN realizations, although all three levels demonstrate quite a bit of DA+PN realizations. Consequently, while the general trends of education and prescriptive ideologies are found in Mexico City and Granada, the current study provides tentative support to previous

observations that DA+PN can be found across all socioeconomic levels in Chile (RAE 2005: 249).

The main effect of referent gender supports previous findings (De Mello 1992; Aliaga Rovira et al. 2020; Tieperman 2018), which found higher DA realization rates when the referent was female (i.e., *la Mari*) as opposed to male (i.e., *el Pablo*). The city by referent gender demonstrates that this is particularly the case for Santiago, without any gender referent tendencies for Mexico City nor Granada. The conditional inference tree also supported this interaction in which referent gender was only significant for Santiago. Of interest is that Hunt (2021) did not find this main effect for any variety, even for Chile. This could again be a result of differences between written and spoken speech, or as Hunt (2021: 39) indicates, one limitation of a country-wide corpus is that tokens cannot be separated by region or city. Thus, given the current study examined the specific speech community of Santiago, Chile, it could be the case that speakers from Santiago use more DA+PN than other regions. In addition to examining differences across varieties, future studies should examine differences within varieties as we should not assume that DA+PN (or any pragmatic or linguistic feature) is produced the same even throughout one language variety.

The findings of the present study, indicate that DA+PN varies greatly between Chilean, Mexican, and Andalusian Spanish varieties, supporting previous findings (De Mello 1992; Hunt 2021; Tieperman 2018) in which Chilean Spanish most favors DA+PN realizations. It is important to note, that these findings only indicate that *there is a difference* in DA+PN between Chilean, Mexican, and Andalusian Spanish, but does not allow us to explain *why there is a difference* with regards to DA+PN realizations between these three varieties. Here we can merely infer some possible reasons. One interesting finding is that in line with previous studies (De Mello 1992; Tieperman 2018), DA+PN demonstrates a type of gendering of language in which female names favor the DA, while neither Mexico City nor Granada demonstrate the same. As mentioned previously, it appears in Santiago, Chile, the DA+PN is commonly accepted across speakers from different educational levels that this feature does not form part of a “standard language ideology” (Lippi-Green 2012: 67), at least in spoken speech (see Hunt 2021 for written speech). On the contrary, it would appear that speakers from both Granada and Mexico City are subject to a standard language ideology with regards to DA+PN realizations, especially among those with the highest level of education. Given schools teach prescriptive grammar in which proper names “reject, in principle, the presence of an article, as they already constitute a definite nominal phrase in themselves” (Leonetti 1999: 811); those with more years of formal education would avoid this feature even more.

One limitation in the current study is not having more information about the participants to be able to reliably code for pragmatic factors (closeness of the participant’s relationship with the referent, quality of the participant’s relationship with the referent, quality of the conversational context). This would have allowed a more nuanced examination of factors that govern the use of DA+PN and truly explore the speech act of DA+PN. These pragmatic variables are important to know as qualitative observations indicate that DA+PN has a “familiar” connotation (Fernández-Leborans 1999: 112) and may imply “intimacy” or “closeness” (Calderón Campos 2015: 90). Future work should make sure to reliably incorporate these pragmatic variables, whether with sociolinguistic interviews or using an experimental socio-pragmatics task. Even though DA+PN may be relatively infrequent in varieties

such as Andalusian Spanish, having insights into these pragmatic variables would shed light into the social contexts when DA+PN realizations are most favored. Future studies should also examine the social perceptions of DA+PN following the third-wave sociolinguistics (Eckert 2008) to better understand the social and pragmatic values associated with DA+PN across varieties. This would provide insights into how pragmatics-politeness varies across varieties (Brown & Levinson 1987). Based on the production data from this study, it is hypothesized that Chilean listeners would perceive the use of DA+PN much more favorably than Mexican and Andalusian listeners, but such evaluations would likely be subject to contextual factors as well.

6. Conclusion

Thus, the current study revealed that DA+PN realizations are most favored by speakers from Santiago, Chile. Speakers of Santiago demonstrate that DA+PN realizations are the norm across all educational levels, but that there is a gendering of language in which female reference names favor DAs over male names. It is suggested that speakers from Mexico City and Granada may demonstrate a stronger standard language ideology with regards to DA+PN realizations in which the overall realizations are quite few, and nearly absent altogether of those with more years of educational attainment. Thus, not only do these varieties differ in overall frequency in which DA+PN appears to be the norm in Santiago and much less common in Granada and Mexico City, they also differ in terms of which social factors favor DA+PN realizations in each variety.

Acknowledgments

This article is an elaborated follow-up to Tieperman (2020). The authors would like to thank the PRESEEA research teams in Santiago (Chile), Mexico City (Mexico), and Granada (Spain) for providing open-access to the sociolinguistic interview transcripts. We are also indebted to three reviewers for their insightful comments. All errors remain are our own.

References

- Abbott, Barbara. 2002. Definiteness and proper Names: Some bad news for the description theory. *Journal of Semantics* 19(2): 191–201. <https://doi.org/10.1093/jos/19.2.191>
- Aliaga Rovira, Leonardo, Avilés Retamal, Patricia, Román Montes de Oca, Domingo, Serrano Insunza, María José, Colicheo Cabrera, Camila, & Soto Cáceres (2020). ‘Nicanor y la Violeta’: Un análisis crítico del uso del artículo definido ante hombres propios masculinos y femeninos. *Logos: Revista de Lingüística, Filosofía y Literatura* 30(2): 224–235. [https://doi.org/10.1016/0378-2166\(87\)90153-6](https://doi.org/10.1016/0378-2166(87)90153-6)

Allerton, D. J. 1987. The linguistic and sociolinguistic status of proper names: What are they, and who do they belong to? *Journal of Pragmatics* 11(1): 61–92. [https://doi.org/10.1016/0378-2166\(87\)90153-6](https://doi.org/10.1016/0378-2166(87)90153-6)

Barron, Anne. 2017. Variational pragmatics. In Anne Barron, Yueguo Gu, & Gerard Steen (eds.), *The Routledge Handbook of Pragmatics*, 91–104. London/New York: Routledge.

Barron, Anne, & Klaus P Schneider. 2009. Variational pragmatics: Studying the impact of social factors on language use in interaction. *Intercultural Pragmatics* 6(4): 425–442. <https://doi.org/10.1515/IPRG.2009.023>

Bates, D., M. Maechler, B. Bolker, & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48. <https://doi.org/10.48550/arXiv.1406.5823>

Brown, Penelope, & Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Burge, Tyler. 1973. Reference and proper names. *Journal of Philosophy* 70(14): 425–439. <https://doi.org/10.2307/2025107>

Calderón Campos, Miguel. 2015. El antropónimo precedido de artículo en la historia del español. *Hispania* 98(1): 79–93.

Calderón Campos, Miguel. 2018. Intersubjectification and textual emphasis in the use of definite article + proper name in Spanish. In M. Bouzouita, I. Sitaridou, & E. Pato (eds.), *Studies in historical Iberoromance morphosyntax*, 75–98. Amsterdam: John Benjamins.

Cameron, Richard, & Scott Schwenter. 2013. Pragmatics and Variationist Sociolinguistics. In R. Bayley, R. Cameron, & C. Lucas (eds.), *The Oxford Handbook of Sociolinguistics*, 464–483. Oxford: Oxford University Press.

Caro Reina, Javier. 2022. The definite article with personal names in Romance languages. In Javier Caro Reina & Johannes Helmbrecht (eds.), *Proper names versus common nouns: Morphosyntactic contrasts in the languages of the world*, 51–92. Berlin: De Gruyter.

Caro Reina, Javier & Johannes Helmbrecht. 2022. Proper names versus common nouns: an introduction. In Javier Caro Reina & Johannes Helmbrecht (eds.), *Proper names versus common nouns: Morphosyntactic contrasts in the languages of the world*, 1–20. Berlin: De Gruyter.

Carranza Brito, Rocío. 2008. Artículo ante nombre propio: Matices de significado. *LL Journal*, 3(2).

Christodulelis, Eleni. 2014. *Pragmatic Constraints on Definite Article Use with Anthroponyms in Brazilian Portuguese*. M.A. thesis, Ohio State University.

Christodulelis, Eleni. 2017. Definite Articles with Anthroponyms Across Varieties of Spanish: Quantitative Support of Conventional and Conversational Implicature. Hispanic Linguistics Symposium. Paper presented at Hispanic Linguistics Symposium, Texas Tech University, Lubbock, TX.

Clancy, Brian. 2011. Complementary perspectives on hedging behaviour in family discourse: The analytic synergy of variational pragmatics and corpus linguistics. *International Journal of Corpus Linguistics* 16(3): 371-390. <https://doi.org/10.1075/ijcl.16.3.05cla>

Corpus diacrónico del español (CORDE). Real Academia Española. <https://corpus.rae.es/cordenet.html>.

Cunha, Celso, & Lindley Cintra. 2017. *Nova gramática do português contemporâneo*. 7th edn. Rio de Janeiro: Lexikon.

Curzan, Anne. 2014. *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139107327>

Davies, Mark. 2016-. *Corpus del español: Web/dialects*. <https://www.corpusdelespanol.org/web-dial/>

De Mello, George. 1992. El artículo definido con nombre propio de persona en el español hablado culto contemporáneo. *Studia Neophilologica* 64(2): 221-234. <https://doi.org/10.1080/00393279208588100>

Del Valle, José. 2007. Embracing diversity for the sake of unity: Linguistic hegemony and the pursuit of total Spanish. In A. Duchêne, & M. Heller (eds.), *Discourses of endangerment: Ideology and interest in the defense of languages*, 242-267. Continuum International.

Del Valle, José. 2009. Total Spanish: The politics of a Pan-Hispanic grammar. *PMLA* 124(3): 880-886. <https://doi.org/10.1632/pmla.2009.124.3.880>

Dixon, Ian James. 2011. *A study of language attitudes concerning the de-affrication of /tʃ/, the pronunciation of the /tr/ consonant cluster, and the use of the definite article with proper names in Santiago, Chile*. M.A. thesis, Brigham Young University.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12: 453-76. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>

Eckert, Penelope, & Sally McConnell-Ginet. 2003. *Language and Gender*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511791147>

Escalona Torres, J. 2019. The effects of social distance in service encounters in Puerto Rican panaderías. In J.C. Félix-Brasdefer & M.E. Placencia (eds.), *Pragmatic variation in service encounter interactions across the Spanish-speaking world*, 230-247. Oxford/New York: Routledge.

Félix-Brasdefer, J. César. 2015. *The language of service encounters: A pragmatic-discourse approach*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139565431>

Félix-Brasdefer, J. César, & A. Yates. 2019. Regional pragmatic variation in small shops in Mexico City, Buenos Aires, & Seville, Spain. In J.C. Félix-Brasdefer & M.E. Placencia (eds.), *Pragmatic variation in service encounter interactions across the Spanish-speaking world*, 15-34. Oxford/New York: Routledge.

Fernández-Leborans, María Jesus. 1999. Capítulo 2: El nombre propio. In I. Bosque & V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española: Sintaxis básica de las clases de palabras*, 77-128. Madrid: Espasa.

Fuentes-Rodríguez, Catalina, María Elena Placencia, & María Palma-Fahey. 2016. Regional pragmatic variation in the use of the discourse marker pues in informal talk among university students in Quito (Ecuador), Santiago (Chile), and Seville (Spain). *Journal of Pragmatics* 97, 74-92. <https://doi.org/10.1016/j.pragma.2016.03.006>

García, Carmen. 2008. Different realizations of solidarity politeness: Comparing Venezuelan and Argentinean invitations. In Klaus P. Schneider & Anne Barron (eds.), *Variational Pragmatics: A focus on regional varieties in pluricentric languages*, 269-305. Amsterdam: John Benjamins.

Geurts, Bart. 1997. Good news about the description theory of names. *Journal of Semantics* 14(4): 319-348. <https://doi.org/10.1093/jos/14.4.319>

Geurts, Bart. 2002. Bad news for anyone? A reply to Abbott. *Journal of Semantics*: 19(2): 203-207. <https://doi.org/10.1093/jos/19.2.203>

Grice, H.P. 1969. Utterer's meaning and intention. *The Philosophical Review* 78: 147-177. <https://doi.org/10.2307/2184179>

Hawkins, John A. 1978. *Definiteness and Indefiniteness*. London: Croom Helm.

Helmbrecht, Johannes. 2022. Proper names with and without definite articles: preliminary results. In Javier Caro Reina & Johannes Helmbrecht (eds.), *Proper names versus common nouns: Morphosyntactic contrasts in the languages of the world*, 51-92. Berlin: De Gruyter.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl, & Achim Zeileis. (2020). *Party: A Laboratory for Recursive Partytitioning*, R package ver. 1.3-4, <https://cran.r-project.org/web/packages/party/index.html>.

-
- Hunt, Keenan. 2021. *Definite articles preceding anthroponyms in Spanish: A multi-country corpus study*. M.A. thesis, University of Georgia.
- Kuznetsova, Alexandra, Per B. Brockhoff, & Rune H. B. Christensen. 2017. *lmerTest Package: Tests in Linear Mixed Effects Models*. *Journal of Statistical Software* 82(13): 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Labov, William. 1966. *The social stratification of English in New York City*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511618208>
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1978. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44. Austin, TX: Southwest Educational Development Laboratory.
- Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Malden, MA/ Oxford, U.K.: Blackwell.
- Laca, Brenda. 1999. Capítulo 13: Presencia y ausencia de determinante. In Ignacio Bosque & Violeta Demonte (Eds.), *Gramática Descriptiva de la Lengua Española: Sintaxis básica de las clases de palabras*, 891-928. Madrid: Espasa.
- Lavandera, Beartz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7(2): 171–182. <https://doi.org/10.1017/S0047404500005510>
- Lenth, R., H. Singmann, J. Love, P. Buerkner, & M. Herve. 2018. *Emmeans: Estimated marginal means, aka least-squares means*. <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>.
- Leonetti, Manuel. 1999. Capítulo 12: El artículo definido. In I. Bosque & V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española: Sintaxis básica de las clases de palabras*, 787-890. Madrid: Espasa.
- Lippi-Green, Rosina. 2012. *English with an accent: Language, ideology, and discrimination in the United States*. 2nd edn. New York: Routledge. <https://doi.org/10.4324/9780203348802>
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326. <https://doi.org/10.1093/jos/4.4.279>
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511605789>

Matushansky, Ora. 2006. Why Rose is the Rose: On the use of definite articles in proper names. In O. Bonami & P. Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics* 6, 285–307.

Moreno-Fernández, Francisco. 2005. Project for the sociolinguistic study of Spanish from Spain and America (PRESEEA): A corpus with grammar and discourse bias. In T. Takagaki, S. Zaima, Y. Tsuruga, & F. Moreno-Fernández (eds.), *Corpus-based approaches to sentence structures*, 265-288. Amsterdam/Philadelphia: John Benjamins.

Murphy, Bróna. 2012. Exploring response tokens in Irish English- a multidisciplinary approach: Integrating variational pragmatics, sociolinguistics and corpus linguistics. *International Journal of Corpus Linguistics* 17(3): 325-348.
<https://doi.org/10.1075/ijcl.17.3.02mur>

Nakagawa, Shinichi, & Holger Schielzeth. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution/British Ecological Society* 4: 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>

Placencia, María Elena. 2008. Requests in corner shop transactions in Ecuadorian Andean and Coastal Spanish. In Klaus P. Schneider & Anne Barron (eds.), *Variational Pragmatics: A focus on regional varieties in pluricentric languages*, 333-354. Amsterdam/Philadelphia: John Benjamins.

Placencia, María Elena. 2012. Regional pragmatic variation. In G. Andersen & K. Aijmer (eds.), *Pragmatics of society*, 79–113. Berlin: DeGruyter.

Placencia, María Elena, & Carmen García. 2019. ‘No gracias amigo’: Refusals of bargaining offers in e-service encounters in Mercado Libre Ecuador and Mercado Libre Venezuela. In J.C. Félix-Brasdefer & M.E. Placencia (eds.), *Pragmatic variation in service encounter interactions across the Spanish-speaking world*, 55-76. Oxford/New York: Routledge.

PRESEEA (2014-). *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. [<http://preseea.linguas.net>].

Poplack, Shana, & Sali A. Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org

RAE (Real Academia Española). 2005. *Diccionario panhispánico de dudas* <https://www.rae.es/dpd/>

RAE/ASALE (Real Academia Española & Asociación de Academias de la Lengua) (eds.). 2009. *Nueva gramática de la lengua española. Morfología. Sintaxis I*. Madrid: Espasa Libros.

Roberts, Craig. 2003. Uniqueness in definite noun phrase. *Linguistics and Philosophy* 26: 287-350.

Roos Film, Sony Pictures Television International (Production Companies). 2006. *Casado con hijos* [Television series].

Schneider, Klaus P. 2021. Variational Pragmatics. In Michael Haugh, Dániel Z. Kádár, & Marina Terkourafi (eds.), *The Cambridge Handbook of Sociopragmatics*, 663-686. Cambridge: Cambridge University Press.

Schneider, Klaus P., & Anne Barron. (eds.). 2008. *Variational pragmatics: A focus on regional varieties in pluricentric languages*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.178>

Schneider, Klaus P. 2010. Variational pragmatics. In M. Fried, J. Ostman, & J. Verschueren (eds.), *Variation and change: Pragmatic perspectives*, 239– 267. Amsterdam/Philadelphia: John Benjamins.

Stolz, Thomas & Nataliya Levkovich. 2022. On *Special Onymic Grammar (SOG)*: Definiteness markers in Fijian and selected Austronesian languages. In Javier Caro Reina & Johannes Helmbrecht (eds.), *Proper names versus common nouns: Morphosyntactic contrasts in the languages of the world*, 237-264. Berlin/Boston: De Gruyter.

Tagliamonte, Sali A., & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practices. *Language Variation and Change* 24: 135–178. <https://doi.org/10.1017/S0954394512000129>

Terkourafi, Marina. 2011. The pragmatic variable: Toward a procedural interpretation. *Language in Society* 40(3): 343–372. <https://doi.org/10.1017/S0047404511000212>

Terkourafi, Marina. 2012. Between pragmatics and sociolinguistics: Where does pragmatic variation fit in? In J. César Félix-Brasdefer & D. Koike (eds.), *Pragmatic Variation in First and Second Language Contexts: Methodological issues*, 295-318. Amsterdam/Philadelphia: John Benjamins.

Tieperman, Robin. 2018 (October). Investigating constraints on uses of definite articles with proper names in Chilean Spanish. Paper presented at the Hispanic Linguistics Symposium. Austin, TX.

Tieperman, Robin. 2020. *¿Cómo está el Pablo?* Examining definite article use with personal proper names in three regional varieties of Spanish. M.A. thesis, Texas Tech University.

Van Langendonck, Willy. 2007. *Theory and typology of proper names*. Berlin: De Gruyter.

Walker, James A. 2010. *Variation in Linguistic Systems*. London: Routledge.

Werth, Alexander. 2020. *Morphosyntax und Pragmatik in Konkurrenz. Der Definitartikel bei Personennamen in den regionalen und historischen Varietäten des Deutschen*. Berlin: De Gruyter.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Wotjak, Gerd. 1985. Zur Semantik der Eigennamen. *Namenkundliche Informationen* 48: 1–17.