

Review by Javier Caro Reina

1. Novelty of the contribution

The paper makes a novel contribution embedded in variationist pragmatics examining the role that linguistic and social factors have on the use of the definite article with personal names in three Spanish varieties on the basis of the PRESEEA Corpus materials. Focusing in Santiago de Chile, Mexico City, and Granada, the authors show that the definite article is more frequently employed in Santiago de Chile (especially with female personal names) than in Mexico City and Granada. In this respect, Chilean Spanish is shown to be more innovative than other varieties of Spanish.

2. Empirical evidence

Is the empirical content of the paper sound (i.e. the data are collected and presented properly, the experiments are well designed, the statistics is well done, the examples contain no spelling mistakes, etc.)?

The empirical evidence is based on the PRESEEA Project materials. However, the authors should make clear that they used the online database, which has reduced number of interviews (18 speakers). Note that the original database has a higher number of interviews (e.g. 54 speakers in Granada). Additional comments are given below.

3. Line of argumentation

Is the argument coherent and sound, with no major flaws and/or shortcomings, within the context of the theoretical assumptions made by the author?

First, the theoretical part of the paper has no relevance for the discussion and conclusions. Second, the criteria for selecting Santiago de Chile, Mexico City, and Granada are not straightforward, especially as far as Granada is concerned. Third, the discussion on article usage in Santiago de Chile mainly relies on assumptions of the authors. In this respect, the authors should analyse the role of the social factor 'age' in order to investigate language change in apparent time and the influence of standard language. These issues are addressed in more detail below.

4. References

Are there any relevant scholarly works that have been overlooked by the author? If the answer is YES, you are asked to please provide the full reference(s).

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. (Studia Typologica. Supplements STUF - Language Typology and Universals 29.) Berlin/Boston: 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. (Studia Typologica. Supplements STUF - Language Typology and Universals 29.) Berlin/Boston: De Gruyter.
- Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.
- Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of Language* 7(1). 37-78.
- Hands Schuh, Corinna. 2017. Nominal category marking on personal names: A typological study of case and definiteness. In Tanja Ackermann & Barbara Schlucker (eds.), *The morphosyntax of proper names*. [Special issue]. *Folia Linguistica* 51(2). 483-504.
- Hanssen, Federico. 1913. *Gramática histórica de la lengua castellana*. Halle: Niemeyer.
- 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. (Studia Typologica. Supplements STUF - Language Typology and Universals 29.) Berlin/Boston: De Gruyter.
- Lenz, Rodolfo. 1925. *La oración y sus partes. Estudios de gramática general y castellana*. Madrid: Revista de Archivos.
- Moya Corral, Juan Antonio et al. 2014. Un paso más hacia la convergencia. La /x/ en Granada: Factores sociales. *Revista Española de Lingüística* 44(1). 83-113.
- RAE/ASALE = Real Academia Española & Asociación de Academias de la Lengua (eds.). 2009. *Nueva gramática de la lengua española. Morfología. Sintaxis I. Sintaxis II*. Madrid: Espasa Libros.
- Soler Arechalde, María Ángeles. 2000. Los topónimos. Concordancia de género. *Anuario de Letras: Lingüística y filología* 38. 391-410.
- Stolz, Thomas & Levkovych. 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. (Studia Typologica. Supplements STUF - Language Typology and Universals 29.) Berlin/Boston: De Gruyter.
- Werth, Alexander. 2020. *Morphosyntax und Pragmatik in Konkurrenz. Der Definitartikel bei Personennamen in den regionalen und historischen Varietäten des Deutschen* (Studia Linguistica Germanica 136). Berlin & Boston: Walter de Gruyter.

5. Multiple publication

I have not seen this paper, its content, the proposed analysis, or the conclusions published in other venues.

6. Plagiarism and misconduct

No.

7. [If your recommendation is Accept with (minor) revisions]

8. [If your recommendation is Reject]

If you reject the paper, do you have any suggestions for how to improve it?

In my view, the paper has, at least, three main problems. First, there is a great difference between the first (Sections 1-2) and second (Section 3-5) part of the paper with respect to style and grammar. This is probably due to the fact that the paper was written in co-authorship. Second, the theoretical issues addressed in the first part of the paper do not feel integrated into the second part of the paper (discussion and conclusions). More specifically, the first sections merely provide a definition of variationist/variational pragmatics, definiteness, etc. However, the link to the use of the definite article with personal names is missing. In this respect, the paper does not succeed in bridging theoretical aspects and the research topic. This problem could be solved if the authors leave aside the theoretical part of the paper and focus on providing a sociolinguistic study working out the role of social factors which have not been treated in detail. This is the case with the social factor 'age' (see below). Third, the paper is devoted to the occurrence of the definite article with personal names. Note, however, that personal names constitute a broad category which includes first names, hypocoristics, family names (or surnames), combination of first and family name, and nicknames. The authors merely talk about "personal proper names" without making a more fine-grained distinction, which has been shown to have an effect on the occurrence of the definite article. Further issues are addressed in more detail below.

In summary, my recommendation is Reject. In what follows, I will list my suggestions for improving the paper according to terminology, content, references, and typos/grammar/style. If the paper is finally accepted, it needs to be proofread by a native speaker.

Terminology

- Please use the term "personal name" instead of "personal proper name". If the authors wish to abbreviate the term, they can employ the abbreviation "PN" (instead of "PPN").
- The term "personal proper name" employed by the authors is not straightforward since it includes first names, hypocoristics, family names, combination of first and family name, and nicknames. The authors should make clear what kind of personal names they are dealing with since the personal name subclass can have an effect on the occurrence of the definite article (see below for examples taken from De Mello 1992).
- The authors sometimes talk about "dialect", sometimes about "variety" when they refer to Chilean, Mexican, and Andalusian Spanish. I suggest using the term variety since the authors are dealing with Santiago de Chile, Mexico City, and Granada Spanish.

Content of the paper

- The title "A corpus analysis of variable definite article" is misleading since it is not the definite article that varies, but rather its usage. I suggest "A corpus analysis of the variation of the definite article with personal names [...]" or "A variationist corpus-based study of the definite article with personal names [...]".
- Section 1 (*Introduction*)
 - Please add the number of the section when describing the structure of the paper (e.g. Section X deals with/is devoted to/etc.).

- Section 2 (*Background*)
 - The title “Background” is too vague. In addition, this section could be transformed into different sections.
 - Subsection 2.1 (*Sociolinguistics and pragmatic variation*) deals with the notions of “variationist pragmatics” and “variational pragmatics”. However, the relevance of these notions for the present study is not straightforward. Instead of giving a definition of the terms, the authors could mention that the paper is embedded in variationist pragmatics (they do at the end of Footnote 1, but this information should be found in the text). Moreover, they should have mentioned previous studies on the use of the definite article with personal names which have been made in such frameworks. If there isn’t any, they should have emphasized the novel contribution of their paper since it would be the first to do so. Interestingly, there has been work on variationist pragmatics, especially on Spanish. However, the authors do not refer to previous work on variationist pragmatics dealing with morphosyntactic phenomena.
 - Subsection 2.2 (*Definite article and definiteness*) is devoted to definiteness. Again, this theoretical subsection is not connected to the use of the definite article with personal names. The authors only refer to Lyons (1999) for a definition of definiteness and further oppose the semantic analyses of proper names made by Geurts (1997) and Abbott (2002). However, the question remains as to why in some languages the definite article is required with personal names and, more importantly, what function the definite article has in such cases.
 - Subsection 2.3 (*Definite article and personal proper names (DA + PPN)*) first presents the use of the definite article with personal names from a cross-linguistic perspective and then focuses on Spanish. The cross-linguistic overview is based on Lyons (1999) and Matushansky (2006). However, there have been a number of recent studies which should be cited (among others: Handschuh 2017, Helmbrecht 2022, Thomas & Levkovych 2022 for cross-linguistic studies; Werth 2020 for German; Caro Reina 2022 for Romance languages). In addition, the occurrence of the definite article with personal names has been shown to be triggered by semantic-pragmatic, lexical, morphosyntactic, phonological, and sociolinguistic factors (see Caro Reina & Helmbrecht 2020: 9-10 for a comprehensive overview). With regard to English, the authors state that “the absence of the definite article is required”. Note that this is the case in unmodified referential contexts. With regard to Spanish, the authors refer to *prescriptive* grammars. However, they cite *descriptive* grammars such as the *Gramática descriptiva del español* (Bosque & Monte 1999). Footnote 2 deals with Portuguese. This information should be placed in the first paragraph. The authors should talk about Brazilian and European Portuguese (instead of Portuguese alone) and cite reference grammars. The *Diccionario panhispánico de dudas* does not need a translation. Please use RAE/ASALE (2009) for the *Nueva gramática de la lengua española*. The quotations from the *Gramática descriptiva del español* and *Nueva gramática de la lengua española* are not really needed. Please try to

paraphrase when possible. The end of the section feels, again, unconnected to the rest of the paper. In this respect, the authors could mention that they provide empirical evidence for the descriptions found in the literature (especially for Chilean Spanish).

- Subsection 2.4 (*Previous studies of DA + PPN in Spanish*) gives an overview of previous studies on the use of the definite article with personal names in varieties of Spanish. Lima is missing in the description of De Mello's (1992) study. The authors sometimes talk about Ciudad de México, sometimes about Mexico City. The name of the city should be homogeneous throughout the paper. The authors do not accurately summarize the results from De Mello (1992) since they concentrate on the occurrence (or lack) of the definite article. However, De Mello (1992) shows the relevance of the following factors (albeit not in combination with each other): personal name subclass, gender (masculine vs. feminine names) and dimension of knowledge (famous vs. ordinary names). For example, in Caracas the definite article is more frequently attested with famous names (13/17) than with ordinary names (4/17) while in Buenos Aires it is more frequently attested with ordinary names (13/15) than with famous names (2/15). In addition, in Santiago de Chile and Buenos Aires the definite article occurs more frequently with feminine names (10/15 and 90/98, respectively) while in Caracas it occurs more frequently with masculine names (14/17). These results should motivate the hypotheses given in Section 2.5 (see below). The study by Christodulelis (2016) can be left out since her findings are only based on two speakers and therefore are not representative. Footnote 4 is irrelevant and should be deleted since the section is devoted to Spanish, and not to Portuguese. Finally, an overview of the use of the definite article with personal names in varieties of European and Latin American Spanish is given in Caro Reina (2022: 75-76).
- Section 2.5 (*Research questions*). The authors state that the results from De Mello (1992) and Hunt (2021) motivate the research questions. If this is the case, the question arises as to why Andalusian and Mexican Spanish were selected at all. As shown by De Mello (1992), the definite article is not employed in Sevilla (Andalusian Spanish) and Mexico City. This problem can be solved if we consider that De Mello's (1992) findings are based on the *Habla Culta* project while the PRESEEA project includes up to three levels of education, which can contribute to completing the picture given by De Mello (1992). The authors claim that "Based on the authors' participant observations throughout Andalucía, it was hypothesized that Granada would have the next highest overall rates of DA+PPN". In my view, this claim is unsubstantiated. Finally, the hypotheses should not involve frequency, but rather the impact of the factors detected in the literature (see above). In addition, the research questions should be kept apart from the hypotheses.
- Section 3 (*Methods*)
 - The name of the section should be "methodology" rather than "methods". One important issue revolves around the question of

whether the PRESEEA Corpus is well suited for the study of the morphosyntactic behaviour of personal names with respect to the use of the definite article. As previously observed in the literature (among others Dahl 2000), personal names are not always frequent in corpora. With regard to the PRESEEA Corpus, the subcorpora considerably differ with respect to the frequency of personal names: 102 in Granada, 352 in Santiago de Chile, and 585 in Mexico City. Another issue concerns the search for personal names in the corpus: Was the search made manually? Finally, the PRESEEA Corpus could be described in more detail (data collection, semi-structured interviews, etc.). More importantly, has the PRESEEA Corpus been used for studies on variationist pragmatics?

- Section 3.1 (*Participants*). The first paragraph begins with “Data was collected from the three cities of Granada, Spain; Mexico City, Mexico; and Santiago, Chile”. This suggests that the authors collected the data themselves. Instead, the paragraph should begin with “This study is embedded in the PRESEEA project”. The authors explain that “the PRESEEA corpus is a searchable online database comprised of 18 sociolinguistic interviews per city”. This may hold true for the online database. Originally, the number of sociolinguistic interviews varies from city to city. For example, in the Granada subcorpus there are 54 speakers – that is, 3 speakers for each combination of age, gender, and educational level (see Moya Corral et al. 2014: 92). This implies that the PRESEEA Corpus materials were only partially analysed. The authors should mention that they did not have access to the whole corpus although in the acknowledgments they explicitly thank the PRESEEA teams in Santiago de Chile, Mexico City, and Granada for providing access to the sociolinguistic interview transcripts.
- Section 3.2 (*Dependent variable, data collection, & data coding*). The authors should specify that they concentrated on personal names in referential use (i.e. as arguments) in unmodified contexts – that is, without modification by means of an adjective phrase, a prepositional phrase, or a restrictive relative clause. The examples should, when possible, be shortened (e.g. “...y yo me acuerdo que merendado, cuando iba a buscar a Ø Eulalia...” > “cuando iba a buscar a Ø Eulalia...”). The paper contains only two examples of personal names with definite article from the corpus. It would be valuable to have examples from all three cities. The authors mention that personal names were transcribed differently in the PRESEEA Corpus: full form for Granada and abbreviated form for Santiago de Chile and Mexico City. However, example (1), which is from Santiago de Chile, exhibits a full form. It is not clear whether example (1) is from Santiago de Chile, as indicated in the transcription, or from Granada, as indicated by the authors in the text. The fact that personal names were abbreviated poses some problems. First, the authors could have checked the personal name in the recordings (if available). Second, the abbreviated form of a personal name can stand for different personal name subclasses. The implications derived from this issue should be addressed in the text, and not in a footnote. Note that nicknames should

be excluded from the analysis since they are always, at least in Romance languages, accompanied by the definite article. There is also a tendency for hypocoristics to take the definite article (*la Mari*). The authors point out that “Last names (i.e. *los Sánchez*) were also excluded as all varieties produce the DA with a plural last name”. However, this only applies for family names in the plural. This raises the question of whether family names in the singular were considered at all. With regard to the deferential titles *don* and *señor*, both meaning ‘mister’, there is diatopic variation. For example, in Spain the title *don* does not require the definite article (*don Marco*) while *señor* does (*el señor García*) (see RAE/ASALE 2009 for details).

- Section 3.3 (*Independent variables*) lists the extra-linguistic and linguistic independent variables. Note, however, that the terms social and linguistic factors are more common in sociolinguistic studies. Gender was classified as an extra-linguistic factor. In my view, it should be classified as a linguistic factor. The linguistic factors are comprised of position in utterance and preceding segment. It is not clear why the position in utterance should be relevant. Instead, the syntactic function (subject, direct object, etc.) might be more important. In addition, the authors could have coded personal names depending on whether they are famous or ordinary since this factor has been shown to have an effect in the literature (see above for De Mello 1992 and Caro Reina 2022 for Romance languages). The last paragraph can be deleted since the additional linguistic factors laid out by the authors were finally excluded from the study.
- Section 4 (*Results*)
 - Section 4.1 (*Inferential statistics*). It is not clear why ‘participant’ (rather than ‘city’) is the most important predictor. I suppose this is due to the fact that the number of participants from Mexico City and Granada, who do not employ the definite article with personal names, overwhelm the number of participants from Santiago de Chile, who employ it (36 speakers from Mexico City and Granada vs. 18 speakers from Santiago de Chile). If this is the case, this predictor could be excluded. The authors do not address inter-speaker variation. In other words, do all speakers from Santiago de Chile (or Mexico City and Granada) behave homogeneously with respect to the use of the definite article with personal names?
 - Section 4.2 (*Results*). The percentages in the first paragraph can be rounded up (decimal numbers are not really needed). The authors illustrate the results for the sociolinguistic factors by means of figures: Figure 3a for educational level and Figures 3b and 3c for gender. However, only Figure 3c shows the results according to city. Similar to Figure 3c, I suggest presenting the results for educational level according to city in a new figure. Figure 3b can be deleted since Figure 3c is more informative. The new figures should contain the token frequency of personal names with and without definite article within the bars. For example, the authors do not quantify the occurrence of the definite article with feminine personal names in Santiago de Chile (nor do they

mention it in the discussion). Crucially, the authors do not discuss the role of age. In this respect, I suggest presenting the results for age group according to city in a new figure and discussing the implications derived from this social factor, especially with respect to language change (see below for comments). Similarly, the remaining factors (utterance position and preceding segment), which were previously shown in Figure 2, are no longer treated.

- Section 5 (*Discussion & Conclusions*)

- Discussion and conclusions should be dealt with in separate sections. The discussion makes reference to the hypotheses presented in Section 2.5. Again, the hypothesis that the use of the definite article with personal names is more frequent in Granada than in Mexico City is rather speculative since it is based on “ethnographic observations” of the authors.
- When the authors refer to De Mello’s (1992) findings, they say that “Santiago was the only capital city in the Spanish-speaking world that demonstrated a high frequency of DA realization”. This claim (“the only capital city in the Spanish-speaking world”) should be reformulated since De Mello (1992) focused on ten cities.
- The authors point out that “one can make the argument that written speech is perceived as more formal than spoken speech”. Certainly, the formality of the speech situation increases from casual style (casual speech) to more formal style (formal speech, reading-passage, word-list) (see Chambers & Trudgill 1998: 59-61).
- The discussion contains conjectures regarding the development of the occurrence of the definite article with personal names (“It is suggested that more years of education would present speakers with more prescriptive grammar that would disfavor the use of a DA+PPN”). In this respect, it is necessary to know the effect of the social factor ‘age’ in Santiago de Chile in order to investigate language change in apparent time. If article usage is restricted to older age groups, then we could observe an effect of standardization (or language prescription in schools as assumed by the authors). Conversely, if article usage is restricted to younger age groups, we could observe a change in progress. For this reason, the authors should examine the social factor ‘age’ in more detail in Section 4.2.
- Some parts of the section contain a repetition of the same ideas.
- The continuous mention of future studies can be avoided.
- Finally, there are two issues that could be addressed in the discussion. First, the lack of the definite article with personal names is standard while its presence is non-standard. However, the data from Santiago de Chile suggest that the occurrence of the definite article with personal names is becoming standard in this variety. However, as already mentioned, the authors do not present the results with respect to age, which would enable us to investigate language change in apparent time. Second, Chilean Spanish is more innovative with respect to the use of the definite article with personal names. However, as for the morphosyntax

of proper names, this is not the only feature that distinguishes Chilean Spanish from other varieties of Spanish. An additional feature involves the gender assignment rules of city names, which in Chilean Spanish are masculine (Hanssen 1913: 174; Lenz 1925: 111-112), while in standard European Spanish they are masculine or feminine depending on the ending due to phonological gender (Soler Arechalde 2000).

References

- The place of publication is missing in some references: Brown & Levinson (1987), Calderón Campos (2018), Cameron (2013), etc.
- The number of pages is missing in some references: Carranza Brito (2008), Labov (1966), etc.

Typos/Grammar/Style

- "but across" should read "but also across" (p. 1).
- "with regards to" should read "with regard to" (p. 2).
- "which provides insights into the frequency across dialects" can be deleted since it is a repetition of the sentence before.
- "prior to" should read "before" (passim).
- "as well to provide an" should read "as well as an" (p. 2).
- Please delete comma in "variation, utilizes" (p. 2).
- The term *variationist pragmatics* is given in quotation marks. The same should apply for the term *variational pragmatics* (p. 2).
- "context the" should read "context of the" (p. 3).
- The quotation mark at the end of the first paragraph should be deleted (p. 2).
- "arenas" should read "areas" (p. 3).
- Style: "and inform what is occurring" (p. 3).
- "either of" should read "either" (p. 3).
- "variables is" should read "variables are" (p. 3).
- Style: "working definition works" (p. 3).
- Style: "bring some resolution to this dilemma" (p. 3).
- Begin new paragraph for "Thus, rather" (p. 4).
- Verb missing in ", as Terkourafi (2012)," (p. 4).
- Style: "In spite of this fact" (p. 5).
- "required, such as" should read "required, as in" (p. 5).
- Style: "Then are other languages" (p. 5).
- "the presences" should read "the presence" (p. 5).
- "in themselves" should read "by themselves" (p. 5).
- "normally are not" should read "are not usually" (p. 5).
- "by an article" should read "by the article" (p. 5).
- "connotations of *"familiar o coloquial"*" should read "familiar or colloquial connotations" (not in italics) (p. 5).
- "Laca (1999)" should read "(Laca 1999)" (p. 6).
- "The preposition of the article" should read "preponed articles" (p. 6).
- Style: "DA+PPN realizations" (passim).
- Style: descriptive corpus analysis (p. 6).
- Style: "did not display one realization" (p. 6).

- “they agreed of” should read “they agreed on” (p. 6).
- Check syntax before “and whether” (p. 6).
- Style: “Using a diachronic analysis, Calderón Campos [...] analyzed [...] utilizing” (p. 6).
- The information contained in Footnote 3 should appear in the paragraph in parentheses.
- Style: “It was found that [...] neither respondent found”.
- Grammar: “in which with female referents [...] favored DA realizations” (p. 7).
- “they suggest such” should read “they suggest that such” (p. 7).
- Style: “realizations based on city” (p. 8).
- Style: “binary realization” should read “realization” (p. 9).
- Style: “when the DA was realized with the PPN” should read “when the PPN was realized with the DA” (p. 9).
- Style: “in the PRESEEA corpus” should read either “for the PRESEEA corpus” or “in the PRESEEA project” (p. 9).
- Grammar: “corpuses” should read “corpora” (passim).
- Please translate the example “Vi a S” (p. 9).
- Style: “referent names” should read “personal names in referential use” (p. 9).
- Style: “other determiners such as *mi* occurred directly in front of the PPN” (p. 10).
- Grammar: “because it already occupied” should read “because they already occupy”
- Style: “occupied the space that other determiners such as *el* or *la* might otherwise occupy” (p. 10).
- Style: “referent gender refers to the gender of the individual being referred to as a third person within the context of the conversation” (p. 10).
- Style: “a plural last name” should read “a last name in the plural” (p. 10).
- Style: “to determine the variable importance” (p. 11).
- Grammar: “in order of most to least importance” (p. 11).
- “those with the most education” should read “speakers with higher educational level” (p. 16).
- “would also worthwhile” should read “would also be worthwhile” (p. 18).
- “for proving access” should read “for providing access” (p. 18).

COMPETING INTEREST

I have no competing interests.