Electronic Thesis and Dissertation Repository

6-25-2024 1:00 PM

The imbalanced interaction of verbal ambiguity and pro-drop: The functional hypothesis in homeland and heritage varieties of Calabrian Italian and Ciociaro

Michael Iannozzi, *Western University*

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Italian Linguistics Commons

Recommended Citation

Abstract

In languages like Italian and Spanish, verbal inflection is often able to disambiguate the subject for person and number. These languages also permit a null subject pronoun (pro-drop). In other Romance languages, such as French, verbal paradigms are much more syncretic, and overt subject pronouns are required in most instances.

The *functional hypothesis* proposes a causal relationship between these two aspects of a language: if a verb's inflection disambiguates, an overt subject can be redundant.

In this dissertation, I investigate pro-drop in Ciociaro—a sibling language of Italian, which is spoken in Frosinone, Italy. Ciociaro's verbal inflections are highly syncretic compared to Italian. I compare Ciociaro to the regional Italian spoken in Calabria, which has a more distinctive verbal paradigm.

In comparing the significant factors that constrain pro-drop in these two languages, if the functional hypothesis is supported, Ciociaro should employ overt subject pronouns more than Calabrian Italian to compensate for its more ambiguous verb phrases.

I use three datasets in this dissertation: a heritage corpus of interviews I recorded with the Ciociaro community of Sarnia, Ontario (*CHILS Corpus*); a homeland Ciociaro corpus drawn from a linguistic atlas containing elicitations (*AIS*; Jaberg & Jud, 1928); and a Calabrian Italian corpus of recorded interviews, both the heritage community in Toronto and the homeland variety spoken in Calabria (*HLVC Corpus*; Nagy, 2011). From these corpora I extract 100 tokens from each speaker, which are then coded for linguistic and extralinguistic variables that are significant in comparable pro-drop analyses. Multivariate analyses are then conducted using Rbrul (Johnson, 2009) to support comparisons of the factors' effects in each corpus.

The CHILS corpus comprises 1,736 tokens from 20 speakers, with a 72% null subject rate. The significant variables are: the phrase's subject (SUBJECT), priming from the previous realization, and verb tense.

From the AIS there are 583 tokens from six participants, with a 76% null rate. The significant variables are: priming from the elicitation prompts, SUBJECT, and AMBIGUOUS, meaning the verbal inflection is ambiguous for subject.

The HLVC corpus contains 1,634 tokens from 20 speakers, with an 80% null rate. The significant variables are: priming from the previous referent, SUBJECT, and preverbal elements.

In a combined analysis of all three corpora, SUBJECT and CORPUS are the significant variables, with CHILS disfavouring and HLVC favouring null subjects. This supports the functional hypothesis as Ciociaro has a higher rate of ambiguous verbal inflections. The significance of SUBJECT is consistent with other pro-drop research suggesting a pan-Romance effect.

The combined analysis also reveals that AMBIGUOUS is not a significant variable. Yet, 34% of the CHILS tokens have an ambiguous subject, while it is only 7% for the HLVC corpus. While ambiguous phrases are inversely correlated with null subject rates across the datasets, they are not significantly causally linked. Thus, the functional hypothesis is not supported in this respect.

This dissertation advances our understanding of pro-drop in Romance languages, representing the first variationist analysis of pro-drop in an Italo-Romance *dialetto*.

Keywords

Summary for lay audiences

Subject pronouns in English disambiguate '**I** speak' from '**we** speak' or '**they** speak'. An exception to these identical forms of the verb is 'speak**s**', which has either 'he' or 'she' as its subject. Generally, English requires a subject pronoun be used. In languages like Italian, however, subject pronouns can be omitted (e.g., *parlo* '(I) speak'), which is known as *pro-drop*.

Italian verbs often carry distinct information: *io parlo* ('I speak'), *loro parlano* ('you speak'), and *lui/lei parla* ('he/she speaks'). Each of these is pronounced differently enough in Italian that the subject, in many cases, can be understood without a subject pronoun.

The *functional hypothesis* proposes that the function of subject pronouns is to disambiguate the subject of a phrase. If a verb is clearly inflected for subject, then a subject pronoun is redundant.

I investigate this hypothesis by comparing pro-drop in Calabrian Italian and Ciociaro—a language related to Italian. While Calabrian Italian verbs are often distinct for different subjects, in Ciociaro many verbs have identical forms for multiple subjects, similar to 'speak' in English. According to the functional hypothesis, Ciociaro should require subject pronouns more because its verbs are more ambiguous. I also compare two varieties of each language: the homeland variety in Italy and heritage variety in Canada. By comparing these varieties, I investigate whether subject pronouns are used more in an English-dominant environment.

For the homeland Ciociaro data, I use a linguistic atlas compiled by asking participants to translate phrases from Italian into Ciociaro (Jaberg & Jud, 1928). The heritage Ciociaro data is from interviews I recorded in Sarnia, Ontario (the CHILS). I compare Ciociaro to Nagy's homeland and heritage recordings of Calabrian Italian (the HLVC; Nagy, 2011, 2024), conducted in Italy and Toronto, Ontario. I extract 100 phrases from each participant: six AIS, 20 CHILS, and 20 HLVC (10 heritage and 10 homeland).

I analyse these phrases for 17 factors that may influence subject pronoun usage. Some of the variables include: the subject itself, we could expect fewer 3rd person subject singular pronouns in English because 'speaks' is less ambiguous that 'speak'; the preceding sentence's subject, we'd expect fewer subject pronouns for phrases like 'I listen, *I* speak' than 'I listen, *you* speak'; and ambiguity, more subject pronouns are used if a verb is ambiguous. It is important to note though, that the specific findings differ from one Romance language to the next (e.g., an ambiguous subject in Spanish might be different in Italian).

From my analysis, homeland and heritage Calabrian speakers use subject pronouns nearly identically, suggesting English has had no effect. I find that subject pronouns are used more often by Sarnia's Ciociaro speakers (28%) than Calabrian Italian speakers (20%), but ambiguity is not a significant predictive factor of subject pronoun use. Ambiguity is a significant factor for the AIS data, suggesting CHILS speakers can use context to understand the subject, which is not available to AIS participants who are translating disconnected phrases.

This dissertation contributes to our understanding of pro-drop in Romance languages, including the role of the functional hypothesis, and the importance of studying under-investigated languages.

Acknowledgements

Thank you to my supervisors, David Heap and Naomi Nagy. I am so grateful for all your guidance and support. You have both shown through example. Thank you, David, for giving me the freedom to fail, flail, and then figure things out. Your trust in me made me a better academic. Thank you, Naomi, for giving me so many opportunities to succeed. This dissertation wouldn't exist if you hadn't given me the chance to learn about Faetar, Rbrul, and variation.

Merci au département de Western, j'étais toujours encouragé et soutenu. Merci à Mirela et Chrisanthi pour toute votre aide avec mes questions et requêtes. Merci à mes profs pour les leçons sur la recherche, le comportement professionnel, et le rythme scolaire.

Thank you to Marisa Brook for being so encouraging when I was an unsure undergrad. Thank you to Elaine Gold for letting me learn about linguistics with the Canadian Language Museum. Et merci à Anne-José Villeneuve pour tes conseils avec mes premières études variationnistes.

Thank you to my friends at Western, who heard about my research for 10 years and always were (or at least pretended to be) interested. Julie dP, merci mille fois, je serais moins sans notre amitié. Rae and Sienna, thank you for celebrating me and being my reminder to go for a walk.

Thank you to Sarnia's Ciociari for trusting me with your stories, and especially to Caroline Di Cocco for all you have done for Sarnia's Italian community. This dissertation is proof the Ciociaro language is important and valued.

Finally, thank you to my family. You have always been my confidence. Whenever I was unsure, you were all certain. There is nothing that has been more important than all of you.

Contents

Figures

Tables

Appendices

List of abbreviations and terms

∅	A null subject (in tables, '%∅' indicates the rate [frequency] of null subjects in each level of a variable/variant)
n	Count of tokens
(they)	In glosses – parentheses indicate the expected subject pronoun
they	In glosses – bold text indicates an overt subject pronoun
AIS	*Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* ('The linguistic and ethnographic atlas of Italy and southern Switzerland') is a linguistic atlas compiled from 306 towns throughout Italy and Switzerland from 1919-1925 (Jaberg & Jud, 1928)
CHILS	Ciociaro Heritage Italian Language of Sarnia – Project to record, archive, and study the heritage Ciociaro spoken in Sarnia
CI	Calabrian Italian
Ciociaro	Italo-Romance language spoken in Frosinone (pl. Ciociari is used to refer to community members and/or speakers)
Dialetto	Term for Italo-Romance languages (pl. dialetti)
ELAN	Annotation tool for transcribing and coding audio (ELAN, 2020)
FW	Factor weight indicates how much a constraint favours (>0.50) or disfavours (<0.50) the dependent variable.
HLVC	Heritage Language Variation and Change – Project to document and study the heritage languages spoken in Toronto, Ontario (Nagy, 2011, 2024)
Heritage variety	A variety of a language that is spoken in an area where it is not the dominant language, typically as the result of migration (e.g., heritage Ciociaro)
Homeland variety	A variety of a language that is spoken in what may be seen as its place of origin, typically in contrast to a heritage variety (e.g., homeland Ciociaro)
NP	Noun phrase
NSP	Null subject pronoun
OSP	Overt subject pronoun
Rbrul	Program for running linguistic analyses in R environment (Johnson, 2009)
RI	Regional Italian
SI	Standard Italian
VP	Verb phrase

1. Introduction

In this dissertation, I analyse the variable realization of subject pronouns in Calabrian Italian and Ciociaro—an Italo-Romance language spoken as a heritage language in Sarnia, Ontario. I am particularly interested in investigating if there's a meaningful correlation between Ciociaro's syncretic verbal paradigms and an increase in subject pronoun expression.

Romance languages typically use suffixation of the verbs to mark the tense, number, and person. Variation, however, does exist. Verbs in a language like French, for example, are pronounced such that many of the distinctions between 1st, 2nd, or 3rd person subject; a singular or plural subject; and the verb's tense and aspect can be impossible to distinguish. On the other hand, languages like Spanish are much less syncretic, preserving this information. Also, French often requires an overt subject pronoun (OSP), while Spanish allows a subject pronoun to be omitted, known as a null subject pronoun (NSP), which is shown in example (**1**).

   1)   French: **je** /vɑ̃/   **il** /vɑ̃/ [1]
                  Spanish: (**yo**) /vendo/   (**el**) /vende/
                              '(**I**)sell'   '(**he**) sells'

From these examples, we can see that the verb for a 1SG and 3SG subject are identical in French. French also has few null subjects (15%; Schmitz & Müller, 2008). Spanish, on the other hand, retains a great deal of phonological distinction on the verbs, which reduces the ambiguity of the

---

[1] This thesis uses the Leipzig glossing rules (Comrie et al., 2015), see also
https://www.eva.mpg.de/lingua/resources/glossing-rules.php

subject. Spanish also has a much higher null subject pronoun rate (from 60% [Avila-Jimenez, 1996] to 81% [Padilla, 2020]).

From these examples, we can see that Spanish phonologically marks the subject on many verbs, and it does not require OSPs. Conversely, most of the French verbal paradigm is ambiguous, and it does require OSPs in most contexts. The question is: to what extent are these two phenomena related?

Neither French nor Spanish is entirely reliant on one of these two solutions. Spanish verbs are almost always distinctive in their inflections for subject, yet it still uses them 19-40% of the time. On the other hand, French verb phrases are often unclear for person and number without subject pronouns, yet speakers still use null subjects in up to 15% of their VPs (Leroux & Jarmasz, 2006; Schmitz & Müller, 2008). Trying to understand subject pronoun realization and the variables that condition it in Romance languages has been the subject of many studies, particularly for Spanish.

While there have been fewer studies of pro-drop in Standard Italian, it is similar to Spanish, with distinct verbal inflections for subject, as shown in (**2**). It also has a very high null subject rate of 90% (Schmitz et al., 2016).

2)	Standard Italian:	(**io**) /vendo/	(**lui**) /vende/
	Spanish:	(**yo**) /vendo/	(**el**) /vende/
		'(I) sell'	'(he) sells'

Yet, as I explain in the following chapter, standard Italian is not the native language of many Italians, and never has been. Even today, many Italians speak one of many regional Italians, which are hybrids of the superstrate standard Italian overlaid on the local substrate Italo-

Romance dialetto[2] of each region. Furthermore, until the 1950s, the vast majority of Italians did not speak a regional Italian either, but instead a local Italo-Romance dialetto (Cerruti, 2011:12).

This dissertation is an examination of subject pronoun realization in a regional Italian and a dialetto. The regional Italian I analyse is spoken in the region of Calabria, at the southern tip of Italy. The dialetto is Ciociaro, which was spoken in the Frosinone province of Lazio, between Rome and Naples. In Italy, Ciociaro has been largely supplanted by the regional Italian, but it is still spoken as a heritage language by the Italian diaspora. One of these heritage communities is in Sarnia, Ontario. Of the 2,400 Italian immigrants who came to Sarnia, roughly half of them came from Frosinone, and Ciociaro is the dominant variety for the community (Di Cocco, 1991). This community is described in **2.2.3.2**.

3)   Ciociaro: (**ji**) [venːe]   (**isː ə**) [venːe]
           Standard Italian: (**io**) /vɛnːi/   (**luj**) /vɛnːe/
           French: ʒə /vɛ̃/   il   /vɛ̃/
                         '(**I**) came'   '(**he**) came'
           (Jaberg & Jud, 1928:1700; Serrone [654])[3]

Calabrian Italian has a verbal morphology that is like standard Italian, but Ciociaro has a much more syncretic inflectional paradigm, as shown in (**3**). In this example, the verb *venire* ('to come') is conjugated in the absolute past (*passato remoto*). Comparing the three languages, we

---

[2] In research on the languages of Italy, the convention is to call these varieties *dialetti* (singular *dialetto*). This is not equivalent to the sense of *dialect* used in linguistic studies of English. As Cerruti and Regis explain "The label *dialect* applies not to geographical or social varieties of the national language, but rather to independent language systems" (2014:3). As such, I use the terms dialetto and dialetti.

[3] Jaberg and Jud's linguistic atlas is described in **5.1.2**. Each location is given a number by the researchers, which I include for reference.

see that while 1SG and 3SG have distinct verbal inflections in Italian, they are identical in both French and Ciociaro. Thus, we might expect Calabrian Italian to have a null rate comparable to standard Italian, and Ciociaro to have a null rate closer to French.

In this dissertation, I place Calabrian Italian and Ciociaro along the spectrum of null subject rates for Romance languages and analyse the factors that condition their realization. In the following section, I lay out the research questions that provide the framework for the relevant literature review, data collection, and analyses of this thesis.

1.1. Research questions

In this study, I examine subject pronoun expression in Calabrian Italian and Ciociaro, and the interplay of verbal inflection and subject pronoun realization. This has not been studied in Ciociaro before, but there are many comparable studies that I draw from to aid in the design of this study. Further, in designing the methodological approach of this dissertation, I seek to answer the following specific questions to further our understanding of both Ciociaro and subject pronoun realization in Romance languages more broadly.

1) **Is Ciociaro's null subject rate similar to standard Italian and Spanish or closer to French?**

As introduced in the previous section, one may expect Ciociaro to have a null subject rate closer to French than to Italian because of its more syncretic verbal paradigms. A causal link between an ambiguous verbal inflection and an overt subject pronoun is one instantiation of the functional compensation hypothesis, or the functional hypothesis. According to this hypothesis, when speaking Spanish or Italian an overt subject isn't necessary because the primary 'function' of a subject pronoun, to disambiguate a verb's subject, is partially redundant due to the inflection of

the verb (the subject's gender, for example, is not disambiguated by verbal inflection). However, in languages like French or English, overt subject pronouns are required because the verbal inflection is not functioning to disambiguate the subject. This is discussed in more detail in **3.1.3**.

On the other hand, as described above, the null rate of Spanish or Italian is not 100%. OSPs are not solely used to disambiguate the subject of a VP. For instance, they can be used to signify a change in the subject from the previous VP, even if the verbal inflection is unambiguous. Conversely, the null rate of French is not 0%, and there are times when a null subject can be used, such as certain impersonal phrases (Ø *faut voir*, '[we'll] see'). Therefore, the functional hypothesis, at the least, cannot explain every time an overt subject is used.

The idea that Ciociaro speakers counteract the ambiguity of the verbal paradigm by using more overt subject pronouns is, however, a testable hypothesis. Alternatively, Ciociaro may have a higher rate of null subjects than French, with ambiguity being tolerated or resolved through the larger context of discourse.

Therefore, the functional hypothesis will be tested in this analysis of Calabrian Italian and Ciociaro. A higher rate of ambiguous verb forms will have a corresponding higher rate of overt subject pronouns if the functional hypothesis is supported.

2) **Are the significant variables that correlate to null subjects similar to those found in comparable studies of other Romance languages?**

If the functional hypothesis is true for subject pronouns in Ciociaro, syncretic verbal paradigms should correspond to overt subject pronouns. However, in the context of discourse, numerous factors can help disambiguate a subject. For example, in discussing one's immigration story, the

subject of the story is the 1ˢᵗ person singular. Consequently, in consecutive VPs, the subject

remains clear even if the individual VPs are formally ambiguous.

Subject pronoun realization has not been examined in Ciociaro. Nonetheless, it has been looked

at in many varieties of Spanish, Calabrian Italian, and in some other Romance languages. Studies

have found that certain variables are consistently significant in analyses of pro-drop. For

example, in Spanish, verbal inflections in the imperfect tense are identical for the 1SG and 3SG

subjects. As a result, to disambiguate the subject Spanish speakers produce more OSPs in the

imperfect (Orozco & Hurtado, 2020:19). There are a number of variables that have been found to

be significant in understanding subject pronoun usage in other Romance languages, and these are

discussed in more detail in **3.3**.

Thus, in examining Calabrian Italian and Ciociaro, I draw upon the variables found to be

significant in the realization of subject pronouns in other Romance languages. This approach

offers not only a methodological framework for my analysis, but also an opportunity to compare

the results from my analyses of Calabrian and Ciociaro to those from related languages.

3) **Are null subjects used in the variety of Ciociaro spoken in Sarnia comparably to how they were used in the homeland?**

There is extraordinarily little data available about Ciociaro, and whether it is still spoken by

people in Frosinone today. Anecdotal evidence suggests that only the most rural and elderly of

Frosinone still speak it regularly (Caroline Di Cocco, personal communication, August 15,

2021). Further, studies on the state of Italo-Romance dialetti have found that they are rapidly

disappearing across Italy. Cugno (2008) found that the use of dialetti in the home throughout the nation fell from 32% in 1988 to only 16% in 2006 (2008:158)[4].

Ideally, part of the data collection for this dissertation would have involved going to Italy to conduct fieldwork there. I had intended to travel to Italy in the summer of 2020 to do that, but the worldwide COVID crisis made that impossible. However, there is a language atlas that was compiled in the early 20th century that provides a meaningful comparison: The *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* or AIS[5] ('The linguistic and ethnographic atlas of Italy and Switzerland'; Jaberg & Jud, 1928). This atlas is described in more detail in **5.1.2**.

By using the data contained in this linguistic atlas from Ciociaria,[6] I can compare the dialetto as it exists in Sarnia's heritage community, among speakers who emigrated in the 1950s and 1960s, to how the dialetto was spoken in Italy in the 1920s. While this is not an ideal comparison, it does provide an opportunity to examine Ciociaro at two times and in two places.

When examining any heritage variety of a language, there are questions of how representative it is of the variety spoken in the homeland. However, if it does not exist in the homeland anymore, at least not as the dominant variety, the AIS data may provide the best available comparison to Ciociaro spoken in Sarnia. This is an opportunity to understand Ciociaro's subject pronoun

---

[4] This question did not appear on the census until the 1980s.

[5] I refer to the AIS in this dissertation; however, its original German title is the *Sprach-und Sachatlas Italiens und der Südschweiz* ('Linguistic and ethnographic atlas of Italy and Southern Switzerland').

[6] 'Ciociaro' is the term for the language, 'Ciociaria' is the term for the unofficial region (see **2.2.2**).

realization, ambiguous verbal forms, and the other significant variables as both a heritage variety and as an Italo-Romance dialetto.

To aid in the methodological design and testing of a homeland and a heritage variety, I also compare Calabrian Italian as spoken in the homeland (see **2.2.1**), and as a heritage language spoken in Toronto, Ontario (**2.2.3.1**). There has been research on NSPs of Calabrian Italian as a heritage language (Nagy et al., 2011). In this dissertation I conduct new analyses, but this previous framework provides a meaningful comparison for my research of both Calabrian Italian and Ciociaro as homeland and heritage varieties.

4) **How does the variation of pro-drop in Ciociaro compare to the variation in other Italian varieties?**

As discussed in **3.3**, subject pronoun realization has been the subject of much study in varieties of Spanish. However, this is not the case for standard Italian, the regional Italians, or the Italo-Romance dialetti. There have been very few variationist studies of NSP variation in this region (Heap, 1997, which includes northern Italo-Romance dialetti; Nagy, 2000, for Faetar; Baird et al., 2021; Nagy et al., 2011; Schmitz et al., 2016, for Italian).

However, any results from the analysis of subject pronoun realization in Ciociaro are better understood when compared to other Italian varieties. Fortunately, the *Heritage Language Variation and Change (HLVC)* Corpus (Nagy, 2011) provides a useful means of comparison. This corpus contains a collection of sociolinguistic interviews conducted with speakers of the regional Italian of Calabria. In this dissertation I conduct similar analyses for the Calabrian and Ciociaro recorded interviews. This provides data from another Italian variety to see how Calabrian Italian and Ciociaro are similar and different regarding their respective rates of subject

pronoun realization and significant variables. While Ciociaro is an Italo-Romance dialetto that is spoken increasingly rarely, if at all, in its homeland, Calabrian Italian is a regional Italian and the dominant variety in Calabria today.

This comparison is also a means of testing the functional hypothesis in Italian varieties. Calabrian Italian has a very distinctive verbal paradigm, whereas Ciociaro has a much more syncretic paradigm. Therefore, I expect Calabrian to have a higher null rate, closer to Standard Italian and Spanish varieties, and Ciociaro to have a lower rate. These results may support the functional hypothesis with Ciociaro compensating for the ambiguity of the verbal inflections by using more OSPs. If they have similar rates of subject pronoun realization, but different rates of ambiguity, then the functional hypothesis is not supported.

1.2. Organization

This dissertation has eight chapters. This first chapter provides an overview to the study and lays out the framework of what follows.

Chapter two (**Background**) is an introduction to the languages and the communities relevant to this dissertation. The chapter is organized into two sections: the languages of Italy and the relevant geography for the communities being studied. The first section (**2.1**) is an overview of the complex linguistic situation in Italy: standard Italian and the myth that it has ever been spoken by a sizeable group of native speakers, the regional Italians spoken by the majority today, and the Italo-Romance dialetti they replaced. The second section (**2.2**) introduces the specific regions that are part of this study: Calabria and Ciociaria in Italy, and the establishment of heritage communities in Toronto and Sarnia, Ontario, during the period of mass emigration from Italy following World War II.

Chapter three (**Variables and languages of the dissertation**) provides a description of the relevant linguistic concepts for this dissertation. The first section (**3.1**) is an overview of the key grammatical areas of investigation for this study: subject pronouns, verbal ambiguity, and the functional hypothesis. In the second section (**3.2**) I present the subject pronouns of Calabrian and Ciociaro, the verbal morphology of both, and how differences in the two languages may relate to both VP ambiguity and pro-drop. In the third section (**3.3**), I present significant variables from comparable studies that helped inform the methodological approach of my analysis. This provides a means of informing my hypotheses and understanding my results within the broader context of other Romance languages. The final section (**3.4**), provides an overview of heritage language analysis as it relates to this dissertation, including a discussion of contact effects with English.

Chapter four (**Hypotheses**) is where I lay out the hypotheses for this dissertation, many of which relate to the functional hypothesis, and how my results are used to test them. These are divided across three categories: hypotheses regarding the comparison of homeland and heritage varieties of both Calabrian Italian and Ciociaro (**4.1**), the rate of null subjects (**4.2**), and those that relate to specific independent variables (**4.3**).

Chapter five (**Methodology**) is a presentation of the methodology I developed to conduct the research in this dissertation. The first section (**5.1**) is a description of the three corpora used in this dissertation: the interviews I conducted with Sarnia's Ciociaro community that became the Ciociaro Heritage Italian Language of Sarnia (CHILS) corpus; the *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* (AIS), a linguistic atlas constructed between 1919 and 1925 throughout Italy and southern Switzerland (Jaberg & Jud, 1928); and Nagy's

Heritage Language Variation & Change (HLVC) corpus (2011), from which the Calabrian recordings and transcriptions are used for this dissertation. After a description of the three datasets, the second section (**5.2**) includes the data-coding process for this dissertation, the extraction of verb phrases (VPs) using the ELAN software for the HLVC and CHILS corpora (*ELAN*, 2020), and the transcription process for the maps of the AIS. The third section (**5.3**) is a description of the variables included in the analysis: the dependent variable (an overt or a null subject pronoun for each VP) and the independent variables (linguistic and extralinguistic). The final section (**5.4**) of this chapter is a description of the multivariate analyses in Rbrul (Johnson, 2009), and the techniques used to find the most reliable model for each dataset.

Chapter six (**Results**) is a presentation of the effects of the significant independent variables for subject pronoun realization for each of the corpora included in this dissertation. First, I report the results of pro-drop in Calabrian from the HLVC corpus (**6.1**), including separate analyses for the heritage and the homeland datasets and a combined analysis. In the second section the results from the AIS analysis are presented (**6.2**), providing a homeland comparison for the Ciociaro spoken in Sarnia. In the third section of this chapter (**6.3**), I report the results of the CHILS analysis. In the final section (**6.4**), I compare the results of all three corpora, interpreting the differences as they relate to my hypotheses and the functional hypothesis.

Chapter seven (**Discussion**) is an analysis of the models described in Chapter **5** both as a test of the functional hypothesis and as a comparison to the larger research context of pro-drop in Romance languages. These sections are laid out in the same order as Chapter **4** to determine which hypotheses are supported. The first section (**7.1**) compares homeland and heritage varieties of both Calabrian Italian and Ciociaro. In the second section (**7.2**), I compare the overall

null subject rates across the corpora. The third section (**7.3**) is a description of the significant variables across the analyses, and how these compare to other studies of pro-drop. The final section (**7.4**) of this chapter presents a summary of verbal ambiguity in the varieties analysed in this dissertation, how this compares to varieties of Italian and Spanish that have been previously studied, how it interacts with many of the other variables discussed in this chapter, and how these results may or may not support the functional hypothesis.

Chapter Eight (**Conclusion**) provides a summary of the contributions of this dissertation. First in (**8.1**), I describe how the results of this dissertation contribute to our broader understanding of subject pronoun realization, and the link between pro-drop and ambiguous forms of a verbal paradigm. Second (**8.2**), I present how this research project contributes to future heritage language research: a methodology for engaging with heritage communities, the creation of a linguistic corpus of an under-described dialetto to share with other linguists, and a digital archive for the benefit of the community following the concept of *linguistic gratuity* (Wolfram et al., 2008). In the third section (**8.3**), I lay out future directions of research for Ciociaro: the variable realization of /v/ as [w] and [m], the heavily reduced article paradigm for noun phrases (NPs), the postposition of possessives in Ciociaro, and other aspects of this dialetto.

2. Background

This dissertation is an examination of the variables that condition pro-drop in Calabrian Italian and Ciociaro. To this end, for Calabrian Italian, I analyse recorded interviews from a heritage community in Toronto, Ontario and from a homeland community in Calabria. For Ciociaro, I use recordings from a heritage community in Sarnia, Ontario, and historical atlas data from the Ciociaro-speaking province of Frosinone in Italy. In this chapter, I provide context for both these languages and their respective geographic regions.

In the first section, I present the languages of Italy. There are four broad categories of languages in Italy: standard Italian[7] (*SI*); regional Italians (*RI*), Italo-Romance dialetti, and non-Italian languages. In **2.1.1**, I describe SI, its reach across Italy, and how it differs as a national language from French or Spanish. In **2.1.2**, I explain RIs and how they have become widespread in the last century, including Calabrian Italian. In **2.1.3**, I define dialetti and examine the declining use of these languages, which includes Ciociaro. In the following chapter I describe Calabrian Italian and Ciociaro specifically, but the linguistic complexity of Italy is important to understand the context of these two languages.

In the second section, I describe the geographic regions that are pertinent to the corpora, both homeland and heritage, that I investigate. This helps contextualize the languages and the communities, both in Italy and in the diaspora, within the larger framework of a changing reality for Italy's languages. In the first subsection (**2.2.1**), I present the region of Calabria. In **2.2.2** I

---

[7] In this dissertation I use 'Italian' or 'SI' to generally refer to a general standard Italian. This is in contrast to 'regional Italian' or 'Calabrian Italian'.

describe the region where Ciociaro is spoken, known as Ciociaria. Finally, in **2.2.3** I discuss the emigration from Italy, and the establishment of the Calabrian and Ciociaro heritage communities in Sarnia and Toronto.

2.1. Languages of Italy

We tend to associate a nation with a national language (or national languages): it is commonly believed that the French speak French, in Spain they speak Spanish, in the UK they speak English. These ideas are often associated with the language of the royal courts of what became nations, such as "the Queen's English" or the French spoken in Versailles, etc. However, this ignores the complex history of nations, and the languages that existed within their borders, and that still exist to this day. In France, many spoke Occitan in the south, Francoprovençal in the southeast, and Breton in the northwest, and these languages have survived to the present. In Spain, millions still speak Catalan in the east of the country, Basque is spoken in the north, and Galician in the northwest. In the UK, Irish, Welsh, and Gaelic are still spoken.

However, a standard language has displaced these languages as the maternal language of the majority of each nation's citizens, with that standard reflecting the language spoken in the nation's capital (Paris, Madrid, and London). There are, of course, dialects of French, English, and Spanish, but these dialects are largely mutually intelligible.

This is not the case for Italy. There was no single court that dominated the territory. While kingdoms resembling the present-day nations of France, Spain, and England have existed for centuries, Italy was not unified until 1861 (De Mauro, 1979:51). For most of the history of the Italian peninsula, duchies and republics ruled over city states and small territories, from the

duchy of Savoy in the northwest, to the republic of Venice in the northeast, to the kingdoms of Sicily and Naples in the south.

The terrain of Italy made the borders of these city states easy to delineate, with the Apennine Mountains running north-south through the middle of the country and the Alps across the north. The distinct cultures and relative isolation of each region allowed for the continued use of Italo-Romance dialetti, defined in (**2.1.3**), as the dominant languages of each region until the 1950s (Berruto, 2005:32).

Italian, the language of the state following unification in 1861, was created as a national language. Unlike French, Spanish, or English, SI did not grow from the language of any royal court. Instead, it was created in the 1800s from the literary Italian used by Dante Alighieri (1265-1321), combined with features from the Florentine and Roman languages because of their cultural influence and prestige (Lepschy & Lepschy, 1988:22). This language became the standard written form, and the language of education and politics; however, at the time of its creation it was not the native language of any Italian.

For most Italians at the time of unification, their native language was one of the many Italo-Romance dialetti; sister languages of SI that descended from the same Italic language. These dialetti are partially intelligible to speakers from neighbouring regions, but they are usually not mutually intelligible with SI.

The use of these Italo-Romance dialetti throughout Italy has recently declined. In their place, regional Italians have developed, which are varieties of standard Italian that have incorporated features of a region's Italo-Romance dialetto (Cerruti & Regis, 2014). RIs have supplanted the

Italo-Romance dialetti as the native language of most Italians born since the 1950s (Berruto, 2017).

The following three sections discuss the development and use of these three language groups: SI, RIs, and Italo-Romance dialetti. However, it is important to note that there are, and have been for centuries, communities speaking languages outside the Italo-Romance family. There are communities of speakers of Arbëresh (related to Albanian) and Griko (Italiot Greek) in the south of Italy, and speakers of Ladin (a Rhaeto-Romance language) and high German varieties in the north, with many other non-Italo-Romance communities throughout Italy, such as Faetar (Nagy, 2000). These communities, as well as more recent immigrant communities to Italy, are important to the story of the languages of Italy, but beyond the scope of this dissertation.

2.1.1. Standard Italian

Standard Italian was created as a means of uniting a linguistically (and culturally) diverse people under a national identity following the unification of Italy in 1861. However, SI did not have a base of native speakers to spread it throughout Italy. Therefore, only those in national politics or in the elite of society spoke standard Italian, and even then, it was a language they learned. In fact, De Mauro (1979) finds that in 1870, 97.5% of Italians recorded in the first national census could not speak SI.

Even today, according to some scholars there are no native speakers of standard Italian (Cerruti, 2011:12). However, it has slowly grown to be a second language of most Italians. This change is due to several key factors over the last century: the spread of standardized education, industrialization, two World Wars (alongside mandatory military service), and the growth of mass communication.

The adoption of a standardized education program throughout Italy provided an opportunity to spread a common language. However, before the 1960s instruction was only available to most children up to *la quinta* ('grade 5', for children around 10 years of age), after which the family had to pay for schooling—if it were even available locally. As such, most children did not attend school beyond that age, and, as recently as 1960, almost half did not even complete that (De Mauro, 1979). Further, SI was taught using medieval poetry and narratives that were difficult to grasp and were not relevant to the children's lives, "thus turning the teaching of the national language into a foreign language course little different from that of French or Latin" (Tosi, 2001:63). However, following major reforms in 1963, such as new educational standards and the expansion of free middle school and high school, as well as the widespread acceptance of regional Italian, the proportion of those who could only speak an Italo-Romance dialetto fell to only 5% in 2005 (Gramellini, 2008:190).

For much of the 20th century, the dialetti had continued to be the language of daily use in the rural areas of Italy, where they had little contact with either standard Italian or other Italo-Romance dialetti. However, with the industrialization of the north of Italy, mass migration occurred from the rural agrarian south of Italy to the industrialized north, particularly to the cities of Turin and Milan. At the peak of migration, between 1950 and 1970, two million Italians moved from the south to the urban centres of the north (Bonifazi & Heins, 2000). Different groups of migrants spoke different Italo-Romance dialetti, but now needed a common language in these industrialized cities. This urbanization of the population and mass migration within Italy provided a motivation to use SI that had not existed in the migrants' rural hometowns.

Related to the mass migration of Italians, the military has played a role in expanding the use of standard Italian. Both World Wars and compulsory military service, which existed until 2005, brought men together from across Italy (Wilcox, 2011). The need to communicate with speakers of many different Italo-Romance dialetti further underscored the importance of a language that could be understood by all Italians (Cerruti, 2011).

Finally, the increased access to mass communication has cemented the importance of a national language. By the mid-1960s, 50% of Italians listened to the radio daily, and 34% watched television daily too (Tosi, 2001:13). The desire to engage with national media provided a motivator to learn SI for many Italians.

Therefore, despite being spoken by very few Italians at the time of unification, standard Italian has been established throughout Italy as the language of education, politics, industry, the military, and television and radio. As such, by the end of the 20th century most speakers of Italo-Romance dialetti had learned some form of SI to benefit from the possibilities afforded to those who can speak it. This resulted in the creation of regional Italian varieties throughout Italy.

2.1.2. Regional Italians

As we have just seen, most, if not all, Italians do not speak standard Italian as their native language. Instead it functions in the consciousness of many Italians as "an ideal" (Berruto, 2017:37). While SI has not replaced dialetti, there has been a sweeping change in the native language of most Italians. While at the turn of the 20th century almost all Italians spoke an Italo-Romance dialetto, now regional Italians have become the first language of most Italians.

These RIs have grown out of what Cerruti calls "advergence" between the L2 standard Italian and the native Italo-Romance dialetti (2011). In other words, as Italians transitioned to using SI as the primary language of life outside the home, their Italo-Romance dialetti left their traces on each region's version of standard Italian to create the regional Italians that exist today. For example, the RI of Naples is SI combined with features of the Neapolitan Italo-Romance dialetto. These features are adopted into all aspects of the language: phonology, such as collapsing of [ɛ] with [e] (e.g., *bene* [bene] vs. /bɛne/ in SI, 'well'); morphology, such as the loss of distinction between 1SG and 3SG verbs in the imperfect aspect (e.g., [parlavə] vs. /parlav**o**/ & /parlav**a**/ in SI, 'I was speaking' & 'she/he was speaking'); syntax, such as the omission of the article in possessive NPs (*mia mamma* instead of *la mia mamma* in SI, 'my mom'), and lexical items too (e.g., [penːə] vs. [vɛstiti] in SI, 'clothes') (Lepschy & Lepschy, 1988).

All aspects of standard Italian can undergo changes in a RI. However, what is affected differs from one regional Italian to another. For example, [vɛstiti] might be used for 'clothes' in other RIs, despite a different term existing in the underlying Italo-Romance dialetto.

This makes comparing regional Italians very difficult: they all share the same superstrate language (SI), but they all have different substrate Italo-Romance dialetti, thus different aspects of the grammar are affected in each RI through localized innovation (De Pascale & Marzo, 2016). Eventually, each regional Italian had a "new, established common grammar" (Cerruti, 2011:12).

This change has been rapid and striking. As mentioned above, in 1861, 97.5% of Italians could only speak an Italo-Romance dialetto (De Mauro, 1979). However, by 2005, that had fallen to 5% of Italians, and 73% of Italians only spoke a regional Italian, while 19% spoke both RI and a

dialetto (Gramellini, 2008:190).[8] Most of this change has taken place since the 1950s for the reasons described in the previous section, often within the space of just one generation or two. Parents who spoke an Italo-Romance dialetto as a native language raised their children to speak the local RI exclusively (Cerruti, 2011:12). This was done because knowledge of SI or an RI provided social mobility that a dialetto did not. In the following section I describe these Italo-Romance dialetti.

2.1.3. Italo-Romance languages

As we have seen, standard Italian was created from an amalgam of a literary form of Italian and the local dialetti of the two most influential cities, and it was not the native language of Italians. Instead, until recently almost all Italians spoke Italo-Romance dialetti.

These Italian dialetti are not the same as English dialects. While English dialects are considered varieties of a common language, these Italo-Romance dialetti are not varieties of SI, but rather languages with "autonomous linguistic systems" (Ghimenton, 2013:108). These dialetti descended from the vulgar Latin spoken in Italy, and differ from each other "as much as one Romance language differs from another" (Tosi, 2004:258).

Due to this limited intelligibility and the economic and societal pressures described above, very few Italians today exclusively speak a dialetto, and for those who speak both a dialetto and an RI, the domains of usage for the dialetti have also shrunk. The Italian census has asked questions regarding where dialetti are used since 1988, including among strangers, friends, and family,

---

[8] The remaining 1.5% spoke other languages.

with family being the most common place to exclusively speak dialetti. However, their use among family has fallen from 32% in 1988 to 14% in 2015 (Dota & Panaccione, 2017). So, while these Italo-Romance dialetti were spoken by nearly all Italians at unification, they have been rapidly replaced by RIs, even in the home.

However, this is not to say that Italo-Romance dialetti are not still used. Instead, they are often used alongside RIs. Gramellini (2008) finds that while the exclusive use of Italo-Romance dialetti has fallen when speaking to family, 33% of Italians use some combination of both regional Italians and Italo-Romance dialetti among both family and friends.

The vitality of the Italo-Romance dialetti has continued to influence and shape RIs. However, the reverse is also true, Italo-Romance dialetti are replacing distinctive features with those found in regional Italians, in what Cerruti (2016) calls *l'italianizzazione dei dialetti* ('The Italianization of the dialetti'). This Italianization is happening at all levels of the Italo-Romance dialetti, from the phonological system to the syntax. This is due, at least in part, to the delayed acquisition of the dialetti. Italian children now typically learn the RI first, and then the Italo-Romance dialetto as an L2 as they get older (Ramat, 1995). So, while Italo-Romance dialetti continue to be used daily by many Italians, they are undergoing major changes due to their contact with the RIs.

In the following section, I describe the regions that are relevant to the languages of this dissertation, both in Italy and in their heritage communities in Canada. This provides specific examples of the changes I have described in this section.

2.2. Relevant geographic regions

The linguistic shift from dialetti being the only, or primary, language of most Italians to the rapid adoption of RIs largely occurred following World War II. This is relevant to my dissertation because I am comparing pro-drop in the Ciociaro dialetto to the Calabrian RI. The data I am using comes from homeland communities in Italy and from heritage communities who, at least in part, emigrated during those postwar years. In this section, I provide context for the emigration from the areas of Italy where Calabrian RI (**2.2.1**) and Ciociaro (**2.2.2**) are spoken, the scale of emigration from these regions (**2.2.3**), and the heritage communities where they settled.

2.2.1. Calabria

Figure 1. Region of Calabria, Italy (TUBS, 2011b).

In this section, I provide an overview of the geography and demographics of Calabria to understand the region. Italy today is divided into twenty administrative regions, which in many instances reflect the duchies, city states, and kingdoms that existed prior to unification. As shown in Figure **1**, the Calabrian region is the toe of the Italian 'boot' (TUBS, 2011b).

According to census figures, Calabria has a population of nearly two million (IStat, 2020). However, it has been subject to one of the highest rates of emigration of any of the regions of Italy, both to the industrial north of

Italy and abroad. King et al. estimate that 1.4 million people left Calabria between 1880 and

1970 (1984:113). To explain this emigration from the region, it is important to note that Calabria

has historically been among the poorest of Italy's regions due to its peripheral location from the

large cities and the industrialized north, and its relatively poor farmland. Today, Italy as a whole

has a GDP per capita of €28,900, but in Calabria it is €17,400 (IStat, 2020). The relative poverty

of the region, compared with elsewhere in Italy, provides context for the high rate of emigration

in the pursuit of better economic opportunities.

2.2.2. Ciociaria

Figure 2. Province of Frosinone, Italy (TUBS, 2011a).

Italy is divided into 20 regions, including Calabria. Each region is divided into provinces (the lighter lines in Figure **2**). The area where Ciociaro is spoken roughly corresponds with the province of Frosinone, which is in the region of Lazio (TUBS, 2011a). This province is between Rome and Naples, with the Apennine mountains marking its northern boundary. Frosinone has a population of 486,000, and its largest city is also called Frosinone, which has 45,000 residents (IStat, 2020).

Speakers of Ciociaro often refer to their territory as Ciociaria.[9] The term comes from the sandals, or *ciocie*, traditionally worn by peasants in this area (Devoto & Giacomelli, 1972:93). Ciociaria is not a term that is found on administrative maps of Italy. While Frosinone refers to the geographic province, Ciociaria is an endonym that denotes an identity used by both residents of the province and emigrants across the world to identify themselves, their shared heritage, and their Italo-Romance *dialetto*. Therefore, the population and demographic data from Frosinone is used to discuss the area as an approximation of the less-defined borders of Ciociaria.

Like Calabria, Frosinone has always been a predominantly agricultural area, with many families working as sharecroppers (*contadini*) until recently. However, unlike Calabria, Frosinone is between two major urban centres—Naples and Rome. As such, Frosinone underwent less emigration than Calabria after unification; from 1871-1951 the population of the province grew by nearly 50% (IStat, 2020).

Frosinone was however an area of intense fighting during World War II. During the war, 32 of the 90 towns in the province were completely destroyed (Mariani, 2021:16). As a result, in the twenty years that followed, many left Frosinone. Some went to Rome, while thousands of others left for France, Ireland, and Scotland (Colucci & Sanfilippo, 2006). The majority of those emigrating from Frosinone went to the United States and Canada. In fact, Mariani (2021) estimates that 10% of the population of Frosinone emigrated to the United States alone in the 25

---

[9] *Ciociaria* is used to describe the region. The language is *Ciociaro.*

years following the war. However, many others travelled to Canada. In the following section I describe the mass emigration from both Frosinone and Calabria.

2.2.3. Emigration from Italy

The previous sections described the linguistic situation and the pertinent geographic regions of Italy for this dissertation. It is also important to understand the mass emigration from Italy throughout the 20th century. These immigrants were heritage speakers of their mother tongues (i.e., dialetti), and established many heritage communities abroad.[10] Two of these communities are part of my study: the Calabrian Italian corpus contains interviews that were conducted in the Greater Toronto Area (*GTA*), and the Ciociaro corpus was recorded in Sarnia, Ontario.

The population of Italy is today around 59 million (Statistiche IStat, 2022a). However, nearly one in four citizens emigrated between 1876-1976, amounting to 26 million people leaving Italy during this period (Del Boca & Venturini, 2003:2). The destinations of these migrants are broadly divided into two groups: those leaving the north of Italy travelled elsewhere in western Europe (France, Germany, UK, etc.), and those from central and southern Italy emigrated to the Americas or Australia.

The destinations of migrants also changed throughout this period of mass migration. For example, prior to World War I, only 1% of Italian migrants came to Canada, whereas 30% settled in the United States (Daniele, 2021:59). However, during the interwar period, several

---

[10] The Government of Canada considers immigrants as heritage language speakers, not just their descendants (Harrison, 2000). This is also true for most researchers of heritage languages and their speakers. For an overview of current research on heritage languages, see Montrul & Polinsky (2021), and particularly Nagy's chapter on heritage languages in Canada (p. 178-204).

factors changed these immigration patterns. First, emigration from Italy was discouraged due to the fascist government of Mussolini; second, the paths to immigration for many countries became more difficult. Most importantly for the Italian Canadian community was the quota system of immigration passed in the United States in 1924. This reduced the total number of immigrants per year from Italy to about 5,000 (Del Boca & Venturini, 2003:9). Following World War II, many Italians came to Canada instead of the US for this reason. In fact, between 1950-1970 Halifax received 20,000-30,000 Italians immigrants each year (Pier21.ca, n.d.). While these immigrants initially arrived at the port of Halifax, most of them continued to other major cities, such as Toronto and Montreal, or areas with a large need for unskilled labour, such as Sudbury and Sarnia.

In the following sections I describe the two destination cities for these Italian immigrants that are relevant to this dissertation: Toronto and Sarnia. In discussing migration to these cities, it is helpful to use the concepts of *push* and *pull* factors. I have already discussed some of the push factors, the reasons why someone may decide (or feel pushed) to emigrate from a region, such as the relative poverty of Calabria or the destruction of much of Frosinone following World War II. However, once the decision to emigrate has been made, there are also pull factors, which are the reasons why someone will choose to immigrate to (or be pulled toward) one area over another.

2.2.3.1. Toronto

Toronto is the largest city in Canada, with a population of 2.9 million. It is also the most ethnically diverse, with almost half the population being immigrants (Statistics Canada, 2023b). It has drawn immigrants from across the world for decades. Following World War II, tens of thousands of Italian immigrants settled in Toronto, making up 12% of the population by the 1971

census (Tomasi, 1977:496). At the time, the Italian community was the largest immigrant group after those from the UK and Ireland.

The rate of immigration from Italy has fallen since the early 1970s, largely due to a resurgent economy in Italy removing a motivating push factor. Yet, the Italian community is still a significant part of Toronto's population. Today 6% of Torontonians, or nearly 170,000 people, report Italian heritage, and nearly 50,000 still report their mother tongue as Italian (Statistics Canada, 2023b). Unfortunately, however, census information does not distinguish the Italian immigrants by region of emigration, nor is information available to separate "Italian language" into RIs or dialetti. Therefore, how many of those immigrants are from Calabria, or how many in the GTA speak Calabrian Italian, is impossible to say with certainty.

2.2.3.2. Sarnia

Sarnia is a much smaller city than Toronto. It has a population of 72,000, and is located in Southwestern Ontario on the border of Michigan (Statistics Canada, 2023a). It is much less diverse than Toronto, and English is the mother tongue of 88% Sarnians.

In the 1950s and 60s, Sarnia was a city with a lot of economic opportunities. Its economy was booming due to the arrival of petrochemical industries, referred to locally as *the Valley*. This provided a meaningful pull factor for many immigrants from across Europe, with Polish, Ukrainian, Greek, and Italian communities arriving during the postwar period.

Italians in Sarnia

Region	People	%
Abruzzo	201	8%
Aosta	0	0%
Apulia	3	0%
Basilicata	0	0%
Calabria	234	10%
Campania	152	6%
Emilia Romagna	6	0%
Friuli	206	9%
Lazio	1,218	50%
Liguria	5	0%
Lombardy	4	0%
Marche	25	1%
Molise	11	1%
Piedmont	7	0%
Sardinia	66	3%
Sicily	213	9%
Trentino	6	0%
Tuscany	23	1%
Umbria	5	0%
Veneto	39	2%
Total	**2,424**	**100%**

Table 1. Italians who immigrated to Sarnia by region of Italy (Di Cocco, 1991:147).

Many families, including my paternal grandparents, emigrated to Sarnia during this period. They spoke Ciociaro at home and with their friends and family who had also emigrated from Frosinone. Thus, Ciociaro became the maternal language of the children of these families, including my father. Like many of the children of these immigrants, Ciociaro was his first language, and the only one he spoke with his parents, brother, and family friends until he went to school.

Related to the pull of the economic opportunities in Sarnia, the concept of *chain migration* is important. Chain migration refers to the pattern by which immigrants choose where to settle by following family or friends, becoming the 'link' in the chain for the next person who follows (MacDonald & MacDonald, 1964). The chance to have someone you know, who understands Canada, and who had been through a similar experience, helps make the transition easier. This has been true for many immigrants. For example, my grandfather settled in Sarnia by following his brother, who had settled there first. His brother, in turn, had followed his neighbour in Italy, who had emigrated four years prior. Each had arrived, sent back letters and photos of new homes and cars they were able to afford, and provided the pull factors for the next person to make the choice to immigrate to Sarnia.

However, the economic boom provided by the Valley ended in the 1970s, removing the largest pull factor to Sarnia, and the immigration of Italians to Sarnia ended during this period. As a result, in recent decades the community has aged and is shrinking each year.

Laziali in Sarnia

Town	People	%
Alvito	74	6%
Arpino	29	2%
Atina	38	3%
Belmonte	32	3%
Broccostella	163	13%
Casalvieri	280	23%
Cassino	4	0%
Castelliri	13	1%
Ceccano	4	0%
Colfelice	5	0%
Fontechiari	268	22%
Pontecorvo	19	2%
Roccasecca	4	0%
Roma	4	0%
Sora	44	4%
Vicalvi	237	19%
Total	**1,218**	**100%**

Table 2. Italians immigrants to Sarnia from Lazio by hometown (Di Cocco, 1991:148).

Unlike the much larger Italian community in Toronto, we do know a great deal about Sarnia's Italian community because of the work of a community researcher, and childhood immigrant from Ciociaria, Caroline Di Cocco. Her work provides a wealth of detail about the community and provided the demographic information for this dissertation. As shown in Table **1**, between 1950 and 1975 more than 2,400 Italians arrived in Sarnia from across Italy.

We can see from this table that the main regions of emigration are Abruzzo, Calabria, Campania, Sicily, Lazio, and the only region north of Rome, Friuli. While these regions are spread across central and southern Italy, over half of Sarnia's Italian community are Laziali (from the region of Lazio), which contains the province of Frosinone, where Ciociaro is spoken. This is useful in understanding the composition of Sarnia's Italian community, however, Di Cocco's work provides further detail about the hometowns of these immigrants within Lazio, as shown in Table **2**.

It is important to note that all these towns are within Frosinone, apart from Rome. Chain migration explains why we see such homogeneity in the towns of emigration for Sarnia's Italian community. This table shows that 78% of Sarnia's immigrants from Lazio come from only four towns, because they were following someone else in almost every case.

Communities that speak heritage languages are rarely so homogenous. Normally, an immigrant community continues to welcome new immigrants, often from different regions, bringing different dialects. Further, in the case of Italy, immigrants who came after attending standardized education from the 1960s would bring knowledge of an RI (Haller, 1987). However, in Sarnia we have a situation where 50% of Italian immigrants come from one region (Lazio), and 39%[11] come from four towns within a 14 km^2 area (Google, n.d.-b).

Decrease in the population of four villages in Ciociaria, alongside the immigrants to Sarnia from these towns during the same period

Town	Pop. (1951)	Pop. (1971)	Decrease (1951-1971)	to Sarnia	% to Sarnia
Broccostella	2,261	1,729	532	163	31%
Casalvieri	5,658	3,204	2,454	280	11%
Fontechiari	2,143	1,214	929	268	29%
Vicalvi	1,116	663	453	237	53%
Total	**11,178**	**6,810**	**4,368**	**948**	**22%**

Table 3. Decrease in populations for the four towns in Frosinone that make up the majority of Laziali who immigrated to Sarnia (Di Cocco, 1991; Statistiche IStat, 2022a). The final column shows the percentage of the decrease from each town who emigrated to Sarnia.

By combining the demographic information of Sarnia's Italian community with the information from Italy's census, the homogeneity of Sarnia's Italian community is put into even sharper focus. Table **3** shows both the decline in population for these four towns in Frosinone, and that a

---

[11] 39% of the total number of Italians, or 78% of the Laziali.

sizeable fraction of the decrease in population is because they have emigrated to Sarnia. This table provides a useful summary of this chapter. Each of these towns saw significant emigration during the postwar period before, or at the start of, educational reforms and societal changes, which led to RIs beginning to displace dialetti throughout Italy. The push factors included that the towns were impoverished, many of those who emigrated were sharecroppers who did not own land, and the war had brought significant destruction to their towns. This is combined with the pull factors of Sarnia, such as chain migration and the economic opportunities of the Valley. As a result, more than one in five people who left these four neighbouring towns during those 20 years immigrated to Sarnia, where together they made up 39% of Sarnia's Italian community.

This is a rare insight into a heritage community. It is unusual to have such granular detail about the specific villages of emigration for a heritage community, but it provides evidence that most of Sarnia's Italian community emigrated from Ciociaria at the period where RIs still had not replaced the dialetti as the primary language in these rural areas of Italy. As such, there had only been minimal interaction between SI and Ciociaro when the Ciociari of Sarnia's community emigrated from Frosinone.

In this chapter I have provided an overview of the linguistic categories in Italy, and of the regions relevant to this dissertation, both in their homeland areas of Italy and in the heritage communities of Canada. In the following chapter, I present the linguistic concepts that motivated my study and the variables that are relevant in comparable research.

3. Variables and languages

In the previous chapter I explained the linguistic situation of Italy and introduced the immigrant

communities that are relevant to this dissertation. In this chapter, I discuss the linguistic

frameworks and variables that shape my **Hypotheses** and **Methodology**. In **3.1,** I define pro-

drop, subject pronouns, and verbal ambiguity as they are applied in this study. The functional

hypothesis is also defined in this section as it relates to the interaction of pro-drop and verbal

ambiguity. The second section (**3.2**) provides an overview of the pronominal and inflectional

systems of Calabrian Italian and Ciociaro, using standard Italian as a reference for comparing the

two. The third section (**3.3**) contains an overview of comparable studies of pro-drop in other

Romance languages, with a particular focus on significant variables relevant to the

methodological design of this dissertation. In the final section (**3.4**), I discuss Calabrian and

Ciociaro as heritage languages in English-majority Sarnia and Toronto, and what other studies

have found regarding potential contact effects.

3.1. Pro-drop

When a language allows a pronoun to be null, it is said to a *pro-drop language*. However, pro-

drop is a broad term encompassing a few phenomena that vary from one language (or even

variety) to another. Huang (1984) provides an overview of different ways pro-drop is realized in

many languages: object pronouns can be null (e.g., Pashto and Portuguese), as can subject

pronouns (e.g., Spanish and Italian), and some languages allow both object and subject to be null

in the same clause (e.g., Georgian and Swahili). Thus, the term pro-drop does not allow for a

simple dichotomy, with languages neatly organized as pro-drop or not. There has been

significant academic debate surrounding its parametrization (Rizzi, 1986), its place in generative

grammar (Biberauer, 2018), and various typologies for classifying types of pro-drop languages (Holmberg, 2010).

For this study, however, I am investigating pro-drop in Romance languages from a variationist perspective. Through that framework, for both Calabrian Italian and Ciociaro, the dependent variable under investigation is whether a subject pronoun is overt or null in a clause with a tensed verb. As such, for this dissertation, both are pro-drop languages that allow subject pronouns to be overt or null. Categorizing Calabrian and Ciociaro as pro-drop languages does not mean that subject pronouns are never used. Similarly, English is not categorized as a pro-drop language, yet it still has null-subjects (e.g., "Ø looks good"), with Nagy et al. finding English has a null-subject rate of 2% (2011:139).[12] The following sections further specify the scope of the variation.

3.1.1. Subject pronouns

A subject pronoun takes the place of a noun or noun phrase (NP), acting as an anaphor for an understood subject. For example:

    4)   Chris$_i$ says that **she**$_{i/j}$ is a lawyer.

In the sentence presented in (**4**) both *Chris* and *she* can and often do refer to the same person, a 3SG.F subject; the reading where Chris is saying someone else (e.g., Rae$_j$) is a lawyer is possible, but less likely. In a pro-drop language, such as Italian, *lei* ('she') can be omitted in this sort of context. The subject can be recovered from both the proper noun used in the matrix clause, and

---

[12] This includes phrases like 'Ø looks good', but it excludes imperatives like 'look out!'.

from a distinct verbal inflection that indicates that the subject is 3SG and thus is likely the same

subject in the subordinate clause as in the matrix clause, shown in (**5**). Note that the gender

agreement of *avvocata* identifies the null subject as singular and feminine; once again, the most

likely reading co-indexes the null subject with Chris, but another subject is possible in the right

context.

5)   Chris$_i$ dice che $\emptyset_{i/j}$ è un'avvocata.
              Chris$_i$ says that $\emptyset_{i/j}$ is a lawyer (3SG.F).

This study focuses on subject pronouns that represent human subjects. Subject pronouns that

represent non-human animate subjects (e.g., animals), inanimate subjects, meteorological subject

pronouns, and impersonal subject pronouns (or dummy pronouns) are not included in the

envelope of variation of my study. Example sentences of these categories are included below:

6)   Non-human subjects: Sienna$_i$ is barking, **she**$_i$ is hungry.
7)   Inanimate: Buns$_i$ are baking, **they**$_i$ smell delicious.
8)   Meteorological: **It** is raining.
9)   Impersonal: **It** is important to floss.

These types of 3SG and 3PL subjects are regularly excluded from studies of pro-drop because

they tend to have different rates of subject pronoun expression than human subjects. Further, as

described in Heap (1997), these types of 3SG and 3PL subjects also interact with the independent

variables differently than those representing human subjects. As a result, they are not included in

this dissertation's analyses, but are a potential avenue for future research (**8.3**).

3.1.2. Verbal ambiguity

When a subject pronoun is null in Italian, often the subject of the VP is recoverable through some other means. A rich verbal morphology can provide the necessary information of the subject, at least in terms of person and number, though not gender. In other words, those languages that have verbal paradigms that morphologically distinguish the subject for person and number can tolerate a higher rate of pro-drop without necessarily creating ambiguity. On the other hand, languages that do not distinguish one subject from another with distinct verbal inflections require more overt subject pronouns for full disambiguation. Thus, while Italian and French both have subject pronouns, French has a less contrastive verbal paradigm, what Calabrese calls "uninflectedness" (2011). French also has a higher rate of overt subject pronouns than Italian. One of the central questions motivating this thesis is determining if those two aspects of Ciociaro, subject pronoun realization and uninflected (or less inflected) verbal paradigms, are linked.

Table **4** compares the verbal paradigms of English, French, and Italian. As shown in the table, only 1PL and 2PL subjects are disambiguated in the verbal morphology of French. However, for Italian, the inflections disambiguate each of the subjects for person and number. In English, we see almost no inflection, except for 3SG.

Paradigms of the verb 'to speak' in English, French, and Italian

Subj.	English	French	Italian
1SG	I /spik/	Je /paʁl/	(io) /parl+o/
2SG	You /spik/	Tu /paʁl/	(tu) /parl+i/
3SG.M	He /spik+s/	Il /paʁl/	(lui) /parl+a/
3SG.F	She /spik+s/	Elle /paʁl/	(lei) /parl+a/
1PL	We /spik/	Nous /paʁl+ɔ̃/	(noi) /parl+jamo/
2PL	All of you /spik/	Vous /paʁl+e/	(voi) /parl+ate/

| 3PL.M | They /spik/ | Ils /paʁl/ | (loro) /parl+ano/ |
| 3PL.F | They /spik/ | Elles /paʁl/ | (loro) /parl+ano/ |

Table 4. Verbal paradigm of the verb 'to speak' in English, French (parler), and Italian (parlare)

For this dissertation, *verbal ambiguity* refers to cases where the subject cannot be determined from the verb phrase. From the table, while Italian has a much more distinctive verbal paradigm than English and French, there are still cases of verbal ambiguity in the language. The verb *essere* ('to be') has the form *sono* for both the 1SG and 3PL subject ('I am' and 'they are'). Therefore, *sono* is an ambiguously inflected verb because, without an explicit subject, it could refer to one of two subjects.

It is important to note that this syncretism of *sono* does not mean that all phrases where it is used are ambiguous. The subject can be disambiguated by either an adjective (**10**), or by a participle when *essere* is used as an auxiliary (**11**). In these cases, while *sono* is syncretic for 1SG and 3PL subjects, the VP is not ambiguous if the final vowel of either the adjective or participle is pronounced distinctly, and thus showing agreement for gender and number.

10)

    a. *(Io)*   *Sono* *felice*
       (I)   am   happy.SG

    b. *(Loro)* *Sono* *felici*
       (they) are   happy.PL

11)

    a. *(Io)*   *Sono* *venuto*
       (I)   did   come.PPRT.SG.M

    b. *(Io)*   *Sono* *venuta*
       (I)   did   come.PPRT.SG.F

c. *(Loro) Sono   venuti*
   (They) did   come.PPRT.PL.M

d. *(Loro) Sono   venute*
   (They) did   come.PPRT.PL.F

3.1.3. Functional hypothesis

The functional hypothesis is a linguistic concept that can be defined at its most broad as "the function of language is for the speaker (or writer) to communicate meaning to the listener (or reader)" (Labov, 1987:548). In practical terms, this means that if a speaker has a choice between two variants, the one that preserves and transmits more meaning should be favoured. It is important to note that debate around the validity and universality of the functional hypothesis is far from settled. Scholars have examined functionalism through many aspects of language: Givón (2013) provides a good overview of the myriad ways the functional hypothesis has been applied to language, and Labov (1987) provides many examples of its limitations and when it is not supported by variationist research.

When applied to pro-drop the functional hypothesis proposes that there is causal relationship between a less inflected verbal morphology and a lower rate of pro-drop. In other words, the *function* of an overt subject pronoun is to disambiguate a subject in a clause when the verbal inflection does not, and if the verbal inflection is unambiguous, then the subject pronoun is redundant and can be null. A relationship between null subject pronouns and verbal agreement does occur; Huang (1984) provides examples of several languages, such as Pashto and Arabic where subject pronouns can be dropped if there is verbal agreement with the subject. Alternatively, many languages allow null subjects without any verbal agreement at all: Huang (1984) mentions Chinese, Japanese, and Korean; and Yeh (2020) describes Vietnamese

37

similarly. There is, therefore, a spectrum of ways pro-drop and verbal ambiguity interact in languages throughout the world, and the functional hypothesis is not supported in languages like Chinese and Vietnamese, but it may be supported in languages like Pashto and Arabic.

In this study, I test where Calabrian and Ciociaro fall along that spectrum. For example, in the previous section, I showed how the verbal paradigm of the Italian verb *parlare* distinguishes all subjects, except that 3rd person subjects are ambiguous for gender (Table **4**). Spoken French, on the other hand, for the paradigm of *parler*, has ambiguous inflections for all subjects except 1PL and 2PL. The difference in the degree of syncretism in the paradigms of these two languages may play a role in their very different rates of pro-drop, up to 15% for French (Leroux & Jarmasz, 2006; Schmitz & Müller, 2008) and up to 90% for Italian (Lorusso et al., 2005; Schmitz et al., 2016). Thus, in this dissertation the functional hypothesis can be tested through the strength of the inverse relationship between two variables: the higher the rate of ambiguous verbal inflection, the lower the rate of null subject pronouns.

In variationist studies of pro-drop in Spanish, a commonly cited example of the functional hypothesis is from Caribbean varieties that variably delete verb-final /s/. This /s/ distinguishes 2SG from 3SG in the verbal paradigm of many Spanish verbs. As such, when it is lost, according to the functional hypothesis, a lower rate of pro-drop should occur for VPs with a 2SG and 3SG subject to avoid ambiguity. However, the results from many studies of Caribbean Spanish are mixed. Some studies, such as Hochberg's (1986) study of Puerto-Rican Spanish in New York City, find a significant relationship between verbal ambiguity and null-subject rates. Other studies, such as Bayley & Pease-Alvarez (1997), do not find any significant interaction between these variables, instead finding the lower rate of null subjects to be caused by other variables.

Support of the functional hypothesis by pro-drop studies in Spanish varieties is mixed, and it was not supported in a previous study of Calabrian Italian (Nagy et al. 2011).

To the best of my knowledge, support for the functional hypothesis has not been tested in an Italo-Romance dialetto. Therefore, by including verbal ambiguity as a variable in my analyses of both languages, I can test how the functional hypothesis may apply (and apply differently) to both Calabrian Italian and Ciociaro.

3.2. Pro-drop and the languages of this dissertation

In this section I provide a brief overview of how pro-drop applies to the languages that form the basis of this dissertation. This is not meant to provide a complete description of the grammar of standard Italian, Calabrian Italian, or Ciociaro. Instead, I explain how subject pronouns and verbal morphology work in these languages, and how that will inform my hypotheses and methodology.

It is important to examine the verbal paradigms of Calabrian and Ciociaro because differences in syncretism can affect where ambiguity occurs, and thus, potentially, the rate of null subjects. Differences in the pronominal subjects between the languages can also restrict how subject pronouns can be used to disambiguate subjects when speaking. This is relevant because subject pronouns and verbal morphology do not necessarily disambiguate all subjects. As an example, in (**12**) we see that English does not have a means of distinguishing a 3PL.M subject from a 3PL.F, where both the verbal morphology and the subject pronoun are identical, and therefore an overt subject pronoun does not disambiguate a masculine subject from a feminine one. This doesn't mean that an overt subject pronoun won't be used with these VPs, but that ambiguity is not a relevant factor in examining its use.

12)   a. Donald and Cataldo$_i$ haven't arrived, **they**$_i$ are late
                             b. Dorothy and Rita$_i$ have arrived, **they**$_i$ are early

In the following sections I describe the verbal paradigms and the subject pronouns of Calabrian Italian and Ciociaro, and I then describe how these features may affect pro-drop. However, before looking at these aspects of both languages, it is important to first address the lack of available research on both. This context provides an understanding of why many of the examples in this section are from related varieties, and why their inclusion is justified.

Despite the growing use of RIs, both as an L1, and as the language of use outside the home, linguistic studies of regional Italians are scarce, and this is particularly true of Calabrian Italian. As Ledgeway (2010:104) notes, there is very little available for "…the linguist interested in the Italian of Calabria, with existing studies barely scratching the surface and contributing very little towards providing us with a general overview of Italian in the region".

Similarly, Ciociaro, as an Italo-Romance dialetto (**2.1.3**), has not been well-documented. To the best of my knowledge, the only descriptions of the dialetto were produced a century ago. Merlo provides a description of the phonology of Ciociaro, based on interviews in the town of Sora (Merlo, 1919), and there is a brief description of the dialetto in a collection of stories in the vernacular (Merlo & Vignoli, 1920). As a result, information in this section about the grammar of Ciociaro comes from the *AIS* (Jaberg & Jud, 1928), which is supplemented by studies of related languages.

The dialetti of Italy vary a great deal; however neighbouring dialetti tend to share some linguistic features. As a result, most modern linguistic scholarship groups the dialetti according to these clusters of features found in a certain geographic region. However, Ciociaro is found on the

border of two of these groups. Some scholars place Ciociaro firmly within the central group of dialetti, known as *dialetti mediani* (Devoto & Giacomelli, 1972:87; Loporcaro & Paciaroni, 2016:230). Others, however, consider it to be better classified as part of the southern group, or *dialetti meridionali* (Avolio, 2000:15; Vignuzzi, 1997:313). As such, Ciociaro has features that are typical of both the central and southern dialetti (Germani, 2014:140).

This dissertation is not concerned with placing Ciociaro within either of these families. Instead, in this section, I describe the subject pronouns and verbal morphology of Ciociaro, supplemented by relevant examples from both dialetti groups.

In the first section (**3.2.1**), I present the subject pronoun paradigms of standard Italian, Calabrian, and Ciociaro. As there is little research on Calabrian Italian or Ciociaro, I also present the pronominal systems of related varieties to provide context for what may occur in those languages in my dataset. In the following section (**3.2.2**), I describe the verbal morphology of these languages and related varieties to show where ambiguity may occur and how distinctions are made across subjects and tenses. Finally, in the third section (**3.2.3**), I examine how the differences in subject pronouns and verbal morphology may affect pro-drop in both Calabrian Italian and Ciociaro.

3.2.1. Subject pronouns in Italian languages

The paradigm of subject pronouns is an essential part of the grammar for standard Italian, Calabrian, and Ciociaro. The distinctiveness of forms in the paradigms allows for reference to be made to an antecedent and to agree with the verb for person and number. In this section, I present all three varieties.

3.2.1.1. Italian subject pronouns

Subject pronoun paradigm for Italian		
Subject	*Pronoun*	*Written form*
1SG	/io/	Io
2SG	/tu/	Tu
3SG.M	/luj/	Lui
3SG.F	/lɛj/	Lei[13]
1PL	/noj/	Noi
2PL	/voj/	Voi
3PL	/loro/	Loro

Table 5. The subject pronoun paradigm in standard Italian.

Standard Italian has a subject pronoun system with seven forms. This paradigm is shown in Table **5**. There are distinct pronouns for plural and singular subjects in the 1st, 2nd, and 3rd person. There is also a distinct form for masculine and feminine subjects in the 3SG. Like English, Italian distinguishes 3SG.M ('he') from 3SG.F ('she'), but only has one form for 3PL ('they'). From this we can expect that a subject pronoun need not be used to disambiguate a subject across the paradigm, except to disambiguate a 3PL.F antecedent from a 3PL.M. The paradigm presented in this section illustrates the seven-way distinction of subject pronouns in Italian. In the following sections, this pronominal system is compared to Calabrian and Ciociaro.

3.2.1.2. Calabrian subject pronouns

Table **6**, shows the subject pronoun paradigm of Calabrian Italian, as spoken by one participant in the HLVC corpus (Nagy, 2011). The paradigm is the same as standard Italian. Therefore, subject pronouns in Calabrian Italian provide the same seven-way distinction for disambiguating a subject. This is relevant for the functional hypothesis, because if the verbal morphology of a

---

[13] The pronouns *egli* (3SG.M) and *ella* (3SG.F) are almost never used, and only occur in formal writing (Cappellaro, 2017). *Lui* and *lei* have replaced *egli* and *ella* to such a degree that the latter are rarely mentioned in studies of the pronominal system, except to explain how rarely they are used (Berruto 2017:45; Di Domenico, Baroncini, and Capotorti 2020:1).

given VP is ambiguous for subject, the pronominal system can allow for the subject to be

recovered through an overt pronoun.

Subject pronoun paradigm for the regional Italians of Moricone and Calabria, as well as standard Italian

Subject	Taylor (1986:69) Moricone, Latium	Calabria (IXF51A, HLVC)	Italian
1SG	/io/	[io]	/io/
2SG	/tu/	[tu]	/tu/
3SG.M	/isːu/	[luj]	/luj/
3SG.F	/ɛsːa/	[lɛj]	/lɛj/
1PL	/nuj/	[noj]	/noj/
2PL	/vuj/	[voj]	/voj/
3PL.M	/isːi/	[loro]	/loro/
3PL.F	/ɛsːa/		

Table 6. Subject pronoun paradigms in the regional Italian of Latium and Calabria compared to standard Italian.

In Table **6**, I also present the verbal paradigm of another RI, spoken in Moricone, which is spoken in central Italy (Taylor, 1986). I include this paradigm because it contains 3rd person

pronouns for singular and plural subjects that are different from standard Italian. It also has

distinct forms for 3PL.M and 3PL.F, and thus has an eight-way distinction. This is a difference

from standard Italian that Telmon suggests is a part of Calabrian Italian (1993:122). Thus a

distinct 3PL.F subject pronoun may be a part of Calabrian Italian and the dataset I am analyzing;

however, it does not appear in the paradigm of the participant shown in Table **6**, nor in the

recordings available to me, and Ledgeway disputes that it is part of Calabrian Italian at all

(2010:103).

3.2.1.3. Ciociaro subject pronouns

In Table **7**, the subject pronouns of Ciociaro, as used by a participant from the heritage corpus of

Sarnia, are presented alongside the paradigm of two nearby dialetti and standard Italian. This

participant, Antonia, did not use subject pronouns for 2PL. This may be because Ciociaro does

not have distinct subject pronouns for these subjects, but it is also possible the participant did not

produce them in their interview. I have also included the paradigm presented by Iannacito from her description of the dialetto of Villa San Michele (Iannacito, 2000). This town is not in the province of Frosinone, and 40km southeast from the nearest Ciociaro location included in the AIS linguistic atlas (San Donato, 701; see **5.1.2.1**). This paradigm shares many similarities with Antonia's paradigm, except it has a similar 3PL pronoun as standard Italian. I also included the subject pronoun paradigm of the village of Amaseno, which is in Frosinone. This is adapted from Merlo and Vignoli's description of the dialetti of Lazio (Merlo & Vignoli, 1920:68). Although the phonetic alphabet they use is not described, this paradigm from Amaseno appears to be like the other dialetti in the table, with similar 3SG forms to Villa San Michele, and different from standard Italian. Also, the 3PL forms from Amaseno agree with the paradigm used by Antonia over the course of her recording, showing that Ciociaro has distinct feminine forms for both 3SG and 3PL.

Subject pronoun paradigms for the dialetto of Villa San Michele, Ciociaro, and Italian

Subject	Villa San Michele, Molise	Amaseno, Frosinone	Ciociaro (CHILS: Antonia)	Italian
1SG	/i/[14]	⟨i⟩	[i]	/io/
2SG	/tu/	⟨tu⟩	[te], [tə][15]	/tu/
3SG.M	/isːə/	⟨isse⟩	[i], [is]	/luj/
3SG.F	/esːa/	⟨ẹssɑ⟩ ⟨jẹssɑ⟩	[ɛsːa]	/lɛj/
1PL	/nu/	⟨nuɑ⟩ ⟨nuwɑ⟩	[nu]	/noj/
2PL	/vu/	⟨uɑ⟩ ⟨vuɑ⟩ ⟨vuwɑ⟩	--------	/voj/
3PL.M	/lorə/	⟨issi⟩	[isːə]	/loro/
3PL.F		⟨ẹsse⟩ ⟨jẹsse⟩	[isːa]	

Table 7. Subject pronouns in the dialetto of Villa San Michele (Iannacito, 2000:141), the dialetto of Amaseno (Merlo & Vignoli, 1920:68) and Ciociaro, alongside standard Italian.

---

[14] /ji/ after a vowel in the same phrase to avoid hiatus, and /ijə/ is used for emphasis.

[15] Both the 2SG and the 3SG.M forms are used in the CHILS corpus by many participants.

From this table we can see that Ciociaro, as used by Antonia, does not have any of the glides present in the standard Italian paradigm, and that the 3rd person subject pronouns do not distinguish masculine from feminine subjects in the singular or plural. It is also hard to find examples of VPs with a 2PL subject during sociolinguistic interviews, especially with overt pronouns, but the form in Ciociaro is similar to that found in Amaseno and Villa San Michele. I have heard [vu] and [vuwa] used during group interviews, which are not a part of this dissertation.

Table **7** presents the subject pronouns found in one participant's speech in the CHILS corpus. To compare this with the AIS corpus, Table **8** shows the subject pronouns from the four locations within Ciociaria (Jaberg & Jud, 1928). In collecting data from participants throughout the region, the researchers asked participants to provide the subject pronoun paradigm in their dialetto. There are, unfortunately, gaps in the paradigms; however, it appears that the CHILS participant in Table **7** most closely aligns with the paradigm in San Donato.

Subject pronoun paradigm in Ciociaro from four AIS villages

Subject	Serrone (654)	Veroli (664)	San Donato (701)	Ausonia (710)
1SG	[jo], [je], [mi]	[lu]	[ʎə], [ji]	[ʎo]
2SG	[du], [tu], [te]	[tu], [te]	[tu], [te]	[ʧu], [te]
3SG.M	[isːo]	[isːo]	[jisːə]	[isːo]
3SG.F	-------	-------	-------	-------
1PL	[nu]	[nu]	-------	[nuj]
2PL	[vu]	[vu]	[vu]	[vuj]
3PL.M	[isːi]	[isːi]	-------	-------
3PL.F	-------	-------	-------	-------

Table 8. Subject pronoun paradigms from the four Ciociaro AIS villages.

From the subject pronouns presented, none of the participants provide a feminine subject pronoun either 3SG.F or 3PL.F. However, this may not be a complete paradigm, as the paradigms

from Amaseno and Antonia have distinct forms in both cases. Also, there are many different forms in the AIS for 1SG and 2SG subject pronouns.

In this section we have seen how the Calabrian and standard Italian subject pronouns are identical, at least for one HLVC speaker of Calabrian Italian. Further, we have seen how the pronominal system of Ciociaro has fewer distinctions than Calabrian Italian and standard Italian.

3.2.2. Verbal morphology

In this section, I present an overview of the verbal morphology present in the languages under examination in this dissertation. I introduce the way verbs are inflected in these languages, and where morphological ambiguity may arise. This highlights the differences between Calabrian and Ciociaro, and how the concepts of morphological inflectedness and the functional hypothesis may exert different pressures on pro-drop in the two languages.

3.2.2.1. Italian verbal morphology

Standard Italian has a rich verbal morphology. There are clear pronunciation differences between each of the grammatical persons. Thus, within the framework of the functional hypothesis, SI typically does not need to rely on overt subject pronouns because the grammatical information of person and number are made clear through the verbal paradigm.

As can be seen in Table **9** each verb ending for each subject has a clear pronunciation difference, with the exception of the distinction between 1SG and 3PL for *essere* (*sono* meaning either "I am" or "they are"). Therefore, the subject pronoun can be omitted while still conveying who the subject is of a given phrase, except for *sono*.

Verbal paradigms from Italian of the verbs *essere* and *venire* in the present indicative

	Essere		*Venire*	
Subject	*Written*	*IPA*	*Written*	*IPA*
1SG	Sono	/sono/	Vengo	/vɛngo/
2SG	Sei	/sɛj/	Vieni	/vjɛni/
3SG	È	/ɛ/	Viene	/vjɛne/
1PL	Siamo	/siamo/	Veniamo	/vɛnjamo/
2PL	Siete	/sjɛte/	Venite	/vɛnite/
3PL	Sono	/sono/	Vengono	/vɛngono/

Table 9. Verbal paradigms for essere *('to be') and* venire *('to come') in Italian.*

For the tenses, aspects, and moods of verbs, this ambiguity with 1SG and 3PL of *essere* is only present in the present tense. In the imperfect, future, and absolute past, the forms are not the same. In the *passato prossimo*, where *essere* is used as an auxiliary verb, 1SG and 3PL have distinct past participles, which agree in number with the subject. For example, *[io] sono venuto* ('I did come', 1SG.M) is distinct from *[loro] sono venuti* ('They did come', 3PL.M), despite *sono* being identical for both 1SG and 3PL because the agreement on the participle is distinct. Finally, in the present subjunctive the 1SG, 2SG, and 3SG verb endings are identical for regular verbs. Thus, if the functional hypothesis is correct, singular VPs conjugated in the present subjunctive should have a higher rate of overt subject pronouns corresponding to the syncretism in the verbal paradigm for the subjunctive mood.

3.2.2.2. Calabrian verbal morphology

There are a number of features which distinguish Calabrian Italian from standard Italian; however, for the purposes of my analysis, I focus only on the features that apply to the verbal paradigms. The verbal paradigms of *essere* and *venire* are not functionally distinct in the HLVC corpus from those described above for SI. Therefore, the verbal paradigms are not syncretic, and we can expect the same conflict sites for ambiguity as in standard Italian: to clarify feminine

versus masculine 3rd person subjects, 1SG and 3PL *sono*, and singular VPs in the subjunctive mood.

The first and second of these exceptions certainly apply to Calabrian; however, Berruto (2017:42) reports a complete loss of the subjunctive mood, not just in Calabria, but throughout the south of Italy. Therefore, if the subjunctive is not used in Calabrian Italian, this reduces the number of environments in which an ambiguous verb form can occur.

Regarding the phonology and phonetics of Calabrian, there are some aspects worth noting that are distinct from SI. However, they do not create ambiguity in the verbal paradigm. The most distinguishing phonological feature is the lowering of the vowels [e] and [o] to [ɛ] and [ɔ] (Ledgeway, 2010; Telmon, 1993). In standard Italian, a class of regular verbs have 2SG and 2PL verbal inflections that end with -/e/. Therefore, in Calabrian Italian we would expect *tiene* ('he/she holds') to be pronounced /tiɛnɛ/ rather than /tiene/.

However, this lowering of mid-vowels, as well as phonetic differences, such as the loss of [g-] word-initially when followed by [-w-] (e.g., /gwardo/ → [wardo], for *guardo*, 'I watch'), or the voicing of stops following a nasal (e.g., /kɔmpro/ → [kɔmbro], for *compro* 'I buy'), should not create ambiguity in the inflectional morphology of verbs for Calabrian Italian. Overall, while there are certainly differences between Calabrian and standard Italian, the verbal morphology is just as unambiguous for subject.

3.2.2.3. Ciociaro verbal morphology

In the last section, I showed that Calabrian Italian and SI appear to both have little verbal syncretism, and therefore, according to the functional hypothesis, overt subject pronouns are not

expected to disambiguate the subject of a clause in most cases. In this section, I compare those paradigms to the verbal morphology of Ciociaro.

In Table **10**, we see transcriptions from the AIS dataset for the verbs *essere* and *venire* as compared to SI. The paradigms shown here, while showing several differences from SI, still do not have any ambiguity in the verbal morphology, except for 1SG and 3PL *sono*. Based on this distinctiveness of forms, the functional hypothesis should apply in the same way in Ciociaro as in Calabrian.

Verbal paradigms of *essere* and *venire* from the AIS and in Italian

	Essere		*Venire*	
Subject	*AIS (Serrone, 654)*	*Italian*	*AIS (Sonnino, 682)*	*Italian*
1SG	[sɔ]	/sono/	[veŋgo]	/vɛngo/
2SG	[si]	/sɛj/	[ve]	/vjɛni/
3SG	[ɛ]	/ɛ/	[vɛ]	/vjɛne/
1PL	[simo]	/siamo/	[menemo]	/vɛnjamo/
2PL	[sitɛ]	/sjɛte/	[menete]	/vɛnite/
3PL	[sɔ]	/sono/	[vɛnːə]	/vɛngono/

Table 10. Verbal paradigms from the AIS dataset (Jaberg & Jud, 1928:1690) compared to standard Italian.

There are other features of Ciociaro that may be relevant to the analysis of pro-drop, however. From neighbouring dialetti, there are features that may apply to the verbal paradigm of Ciociaro that would lead to more syncretism than in Calabrian or standard Italian. For example, apocope of final segments in the verbal paradigm is a feature of dialetti spoken in neighbouring regions (Iannacito, 2000; Loporcaro, 2009). Similarly, Vignuzzi describes, not a complete apocope, but as the "reduction of unstressed final vowels to [ə]" (1997:314). It is important to note that many researchers describe the verbal paradigm of these dialetti in relation to standard Italian. Thus, apocope or reduction is in reference to the segments that are expected in standard Italian. There is reason to expect this is also true in Ciociaro. Merlo, in his description of Ciociaro as spoken in

the town of Sora in 1919, provides many examples of paradigms with identical forms. For example, *canto* and *canti* ('I sing' and 'you sing') are both transcribed as /kantə/ (Merlo, 1919:141).

This realization of the final vowel in verbal inflections as [ə] in Ciociaro presents the potential for ambiguity between person and number. However, as Ciociaro is understudied, this is possibly an effect of metaphony on the penultimate vowel. This is a widespread phenomenon in Italy described by Calabrese as "a process in which stressed stem vowels are raised, diphthongized, or changed in other ways, before final high vowels" (2016:89). Of particular relevance to Ciociaro, and its reduced word-final vowels, is that many dialetti undergo a second process that lowers the word-final vowel, reduces to /ə/, or is dropped altogether (∅). This process leaves the penultimate vowel inflected for person or number, but the final vowel ambiguous. This occurs in many dialetti, including varieties in Abruzzo and Naples, which border the Ciociaria region to the east and south respectively (Loporcaro, 2016; Sornicola, 2006).

However, while this is a feature of many dialetti to the south of Ciociaria, it may not be part of the verbal paradigms in Ciociaro (Avolio, 2000:26; Germani, 2014:144; Vignuzzi, 1997:313). Merlo states that while it does occur rarely in Ciociaro, is not a productive part of the grammar, unlike "many dialetti in Abruzzo and Puglia" (1919:153).

Significant variables in analysis of /v/ → [w] of Ciociaro

Linguistic factors	FW range	Extralinguistic Factors	FW range
Segment before /v/	58	Speaker	84
Grammatical Category	32	Conversation topic	30
Segment Following /v/	31	Time of occurrence	20
Previous realization	22		

Table 11. Significant factors in the analysis of /v/ → [w] from Iannozzi (2017).

50

Of note from Table **10** is the use of [m-] word-initially in Ciociaro instead of [v-] for 1PL and 2PL

subjects. This appears to be related to what Vignuzzi describes as "widespread 'betacism' "

(Vignuzzi, 1997:314), where /v/ is realized as [b], and can be spirantized, nasalized, or even

deleted intervocalically. I know this occurs in Ciociaro, as the variable realization of /v/ was the

basis of my Master's research (Iannozzi, 2017). In my analysis, I examined the variable

realization of /v/ as [w]. Table **11** shows the significant factors of that analysis.

This variation in the realization of /v/ is specifically relevant for ambiguity of verbal inflection as

[-av-] is added to verbs in the imperfect (e.g., *parla* 'he/she speaks'; *parlava* 'he/she was

speaking'). As the CHILS recordings consist of personal histories, the imperfect is frequently

used by participants. Thus, on its own, if /v/ is realized as [b], [m], or [w], this will not increase

syncretism in the verbal paradigm. However, when combined with the reduction of final vowels

to [ə] or Ø, this becomes much more important to the analysis of pro-drop in Ciociaro, as shown

in Table **12**.

Potential verbal paradigms of *essere* and *venire* in the AIS data compared to Italian

| Subject | Essere | | Venire | |
	AIS (Serrone) /v/ → [w] / Ø V# → [ə] / Ø	Italian	AIS (Sonnino) /v/ → [w] / Ø V# → [ə] / Ø	Italian
1SG	[ɛrə]	/ɛro/	[vɛni(w)ə]	/vɛnivo/
2SG	[ɛrə]	/ɛri/	[vɛni(w)ə]	/vɛnivi/
3SG	[ɛrə]	/ɛra/	[vɛni(w)ə]	/vɛniva/
1PL	[ɛra(w)amə]	/ɛravamo/	[vɛni(w)amə]	/vɛnivamo/
2PL	[ɛra(w)atə]	/ɛravate/	[vɛni(w)atə]	/vɛnivate/
3PL	[ɛranə]	/ɛrano/	[vɛni(w)anə]	/vɛnivano/

Table 12. Potential verbal paradigms of essere and venire in the imperfect for Ciociaro, where /v/ → [w] or Ø and where V# → [ə] or Ø, compared to standard Italian.

51

The table shows that if the final vowel is reduced to [ə], then there is much more ambiguity with the 1SG, 2SG, and 3SG subjects having identical forms. As a result, based on the functional hypothesis, VPs with a singular subject in the imperfect are where we would expect an especially high rate of overt subject pronouns. These two rules, although not specified in Merlo's description of the dialetto of Sora, is reflected in some of the paradigms he provides. When describing the imperfect aspect, he provides the inflection /wə/ for both 1SG and 2SG, where /vo/ and /vi/ respectively would be expected in standard Italian (Merlo, 1919:138-140).

In this section, I have shown how the inflections of verbs in Calabrian do not seem to differ from standard Italian, which means little ambiguity occurs in the paradigms, apart from *sono* and feminine and masculine 3rd person subjects (and potentially the subjunctive present if productive in Calabrian Italian). I have also shown that Ciociaro has much more syncretism, and that the paradigms in the imperfect are particularly uninflected. In the following section, I describe how the pronominal systems and verbal morphology may affect pro-drop in each language.

3.2.3. Pro-drop

To understand how pro-drop is used, I have provided an overview of how subject pronouns and verbal inflections work in Calabrian Italian and Ciociaro. As we have seen, in this respect Calabrian is very similar, if not identical, to standard Italian. As a result, we can expect Calabrian speakers to produce null subjects at a higher rate than Ciociaro as the subject can be recovered from the verb forms, which are not ambiguous for subject. This would support the functional hypothesis as an overt subject pronoun does not provide any grammatical information that is not already included in the verb. However, there are two exceptions to this: *sono* being the identical form for 1SG and 3PL subjects, and disambiguating gender for 3SG subjects (potentially three if

present subjunctive is productive in Calabrian). In these two cases we would expect lower rates of pro-drop.

Now that I have presented the subject pronouns and verbal morphology of Calabrian and Ciociaro, it is important to look at comparable research to understand how pro-drop may work in these languages.

Rates of null subject pronouns in Italian			
Person	Ø	n	%Ø
1st	1,251	1,431	87%
2nd	159	178	89%
3rd	1,209	1,289	94%
total	2,630	2,917	90%

Table 13. Null rates in standard Italian, adapted from Schmitz et al. (2016:113)

There has not been much research of pro-drop in Italian languages from a variationist perspective, but Schmitz et al. (2016) analyse null rates in Italian and found an overall null rate of 90%. They do not look at ambiguity as a factor, nor do the authors provide information about *sono* and its role in subject realization. However, as shown in Table **13**, they do provide the null rates by PERSON, but not by NUMBER. Based on the ambiguity of gender for 3SG subjects as described above, we might expect a lower null rate than for other persons. However, Schmitz et al. report the opposite, with 3rd person subjects having the highest null rates.

Another study was conducted by Lorusso et al. (2005). Their study examined pro-drop in both children and adults, and they found that adults produced null subjects 74% of the time, and children produced null subjects in 75% of their VPs. Unfortunately, they do not report differences according to PERSON or NUMBER at all in their findings.

Finally, Nagy et al. (2011) used the HLVC corpus to analyse pro-drop in heritage Calabrian, as spoken by the community in Toronto, Ontario. They found an overall null rate of 90%, in line

with what Schmitz et al. also report for standard Italian. In their analysis NUMBER was a significant factor, with singular subjects disfavouring null subjects and plural subjects favouring pro-drop. PERSON was not a significant factor. They also found that TENSE was a significant factor, with imperfect disfavouring null subjects, whereas past perfect favoured pro-drop. The results of imperfect and singular subjects disfavouring null subjects may seem to suggest that ambiguity is a factor. However, Nagy et al. did not find any effect for ambiguity in the verbal paradigm, concluding that the functional hypothesis was not supported by their results.

From these studies, it is difficult to establish a clear pattern of how pro-drop may work in Calabrian. We have seen that Calabrian and standard Italian share unambiguous subject pronoun and verbal paradigms, and have overall null rates of 90% in the analyses of Schmitz et al. and Nagy et al., but a lower rate from Lorusso et al. There is also no specific information about ambiguity effects from *sono* or gender for 3SG subjects, but Nagy et al. did find that ambiguity was not a significant factor. Thus, the functional hypothesis may not be supported in Calabrian, or any potential effect is outweighed by more significant factors in the discourse.

For Ciociaro, to the best of my knowledge, there are no variationist studies. However, there are monographs and descriptions of related dialetti. Haller, in describing heritage Italian dialetti in the United States, explains that Neapolitan dialetti frequently have imperfect verbs ending in [-a] for 1SG (1987:400). These Neapolitan dialetti would be on the southern boundary of Ciociaria. This is important to note because this means that in these dialetti both the 1SG and 3SG inflections end in [-a], creating ambiguity in the imperfect verbal paradigm. This same feature is also presented by Iannacito in her grammar of the Molisano dialetto of Villa San Michele (Iannacito, 2000:182), which borders Ciociaria to the southeast.

This potential ambiguity, coupled with the apocope or reduction of -/V/ in verbal paradigms, is relevant to the CHILS corpus because of the focus during the interviews on personal narratives, where the imperfect is often used. Further, because the interviews are mostly conducted one-on-one with participants, the majority of the VPs have either 1SG, 3SG, or 3PL subjects.

There are few variationist studies conducted on pro-drop in Italian languages. Therefore, to better inform the hypotheses proposed in Chapter **4**, in the following section, I organize, compare, and summarize the findings of comparable studies in other languages. While these findings are not from other Italian dialetti, the methodologies of these studies inform the methodology of my dissertation.

3.3. Relevant variables from comparable studies

In the previous section, I described the lack of research on pro-drop in Italian languages. However, there have been many studies that have looked at pro-drop in Spanish varieties, and several that have looked at the Francoprovençal language, Faetar (*cf.* Nagy et al., 2018). There are, of course, many differences between Ciociaro & Calabrian and Spanish varieties, but, comparing pro-drop and the significant factors in Spanish to Italian languages is a reasonable means of informing my methodology, as Filiaci describes:

> Given the morphological similarities and the fact that the two languages are closely typologically related, it seemed plausible to assume that we are dealing with an equivalent phenomenon and much cross-linguistic research has assumed, more or less explicitly, that the pragmatic principles that determine the distribution and interpretation of null and overt pronominal subjects should be roughly equivalent in these two languages. (2010:171)

Therefore, in this section I present variables that have been found to be significant across many studies that are similar to what I propose in this dissertation. This provides background to the

variables that are included in the Methodology (**5**), and informs the hypotheses I propose in the following chapter (**4**).

Table **14** presents a summary of results from studies that are comparable to the one in this dissertation. It is not an exhaustive representation of all studies conducted on pro-drop. Included is the language, variety, overall null subject rate, significant variables from the analysis, and finally those that were not found to be significant. Further, the significant variables are presented in descending order of reported effect size, meaning the range between the highest-ranked variant and the lowest. Presenting the results of many studies in this way makes it easier to see a pattern of variables that are frequently significant.

	Nagy et al., 2018:40		Abreu, 2009:98,125		Carvalho & Child, 2011:23	Orozco & Hurtado, 2020:14	Hernández Constantin, 2021:128	Limerick, 2019:259	Lastra & Butragueño, 2015:10	Padilla, 2020:88
Language	Faetar	Faetar	Spanish	Spanish	Spanish	Spanish	Spanish	Spanish	Spanish	Spanish
Variety	Homeland	Heritage	Monolingual	Biling.	Home.	Home.	Her.	Her.	Home.	L2
Location	Toronto	Toronto	Puerto Rico	Puerto Rico	Uruguay	Columbia	(Columbian) London, Ontario	New Mexico	Mexico	Equatorial Guinea
%Ø	57%	38%	62%	51%	65%	72%	73%	73%	78%	81%
Significant factors	Subject	Subject	Subject	Prev. Real.	Prev. Real.	Subject	Subject	Subject	Subject	Subject
	Preverb. Clitics	Info. Status	Switch ref.	Subject	Subject	Discourse Type	Verb Mood	Switch ref.	Switch ref.	Ambi.
	Age		TAM	Verb class	Switch ref.	Prev. Real.	Switch ref.	Verb class	Mood	Verb class
	Tense		Prev. Real.	Switch ref.	Verb class	TAM	Polarity	Ambi.	Polarity	Switch ref.
	Info. Status		Reflexive	Polarity	Reflexive	Age		Polarity	Discourse Type	Education
				Clause type		Transitivity		Age	Tense	
				Same Tense					Ambi.	
									Age	
Not significant	Polarity	Preverb. Clitics	Verb class	Reflexive	Ambi.	Participant gender	Clause type	Tense	Education	Reflexive
		Age	Clause type	TAM	Clause type	Education	Tense	Mood	Participant gender	Participant gender
		Tense	Same tense	Same mood	TAM		Verb class		Pronoun position	Age
		Polarity	Polarity	Polarity	Participant gender		Participant gender		Freq. verbs	L1

Table 14. Null rates and significant variables from comparable studies ranked by effect size.

3.3.1. Frequently significant variables

From Table **14** above, there are two variables that are significant in every study. First, SUBJECT, meaning the subject of a VP plays a role in whether a subject pronoun is null or expressed. Second is the priming variable, SWITCH REFERENCE, which describes whether the VP has the same subject, or referent, as the previous VP (or INFO. STATUS in Nagy et al., 2018). I describe these next.

3.3.1.1. Subject

There is a consistent finding across the studies summarized above that the factor SUBJECT is always significant. It has the biggest effect in eight of the 10 analyses. This consistency suggests a robust effect that the subject of a VP has upon pro-drop, both in the Spanish varieties and in Faetar.

Within the SUBJECT variable, there is generalization from Spanish varieties that, as Orozco (2015:29) states "Singular pronouns favor overt subjects and plural pronouns disfavor them". Therefore, while the subject of a VP represents both PERSON and NUMBER, in Spanish varieties, NUMBER plays a larger role in subject pronoun realization than PERSON (Carvalho et al., 2015; Hernández Constantin, 2021; Martín-Butragueño, 2020).

However, the only non-Spanish example included in Table **14**, Faetar, a Francoprovençal language spoken in Italy and Toronto, has a different hierarchy than the Spanish varieties. Nagy et al. (2018:40) found that null subjects occur at a higher rate with 3rd person subjects than with 1st person subjects.

Therefore, SUBJECT may be a significant variable in the analyses of pro-drop in Calabrian Italian and Ciociaro, as it is in each of the studies of Table **14**; however, the ranking of the variants within this factor may pattern more closely with either Spanish varieties, with NUMBER having a stronger effect than PERSON, or with Faetar, where the inverse was found.

3.3.1.2. Switch reference

The factor SWITCH REFERENCE indicates whether the previous VP has the same subject as the VP being considered, or whether the VP has a different subject (or referent) than the previous VP, as shown in (**13**). In **a.** the subject of the matrix clause is 3PL.M, while the subject of the embedded clause is 3SG.F, which is a different referent. In **b.** the subject of the matrix and embedded clauses is the same referent. This variable is therefore a *priming* variable. Priming variables seek to measure how much a VP is affected (or primed) by a feature of the preceding VP, in this case the subject of the preceding phrase.

13) a. Donald and John$_i$ are here, **she**$_j$ is happy
     b. Dorothy$_i$ is here, **she**$_i$ is happy

Each of the studies included in Table **14** reports it as a significant factor, although Nagy et al. (2018) used the term INFORMATION STATUS for a similar variable. SWITCH REFERENCE is significant so consistently that Travis (2007:106) refers to the factor as "The most robust and consistent finding across a range of different studies and dialects…" for pro-drop in Spanish.

The general finding is that subject pronouns are more often null if the preceding VP shared the same subject, and a switch to a new subject is more likely to occur with an overt subject pronoun.

59

In other words, if the verbal inflection of a VP is ambiguous, but the subject hasn't changed, the listener can infer the subject, allowing the subject to be recovered from context. The speaker can then mark a change in subject by including an overt subject pronoun, thereby reducing ambiguity for the listener.

3.3.2. Inconsistently significant variables

In the previous section we looked at two variables that are consistently found to be significant across the varieties presented in Table **14**. In this section, we examine two variables, PREVERBAL ELEMENT and TENSE, which are significant in several, but not all, of the studies included.

3.3.2.1. Tense

TENSE, ASPECT, and MOOD are found to be in significant in five of the 10 studies included in Table **14**. This may be due, at least in part, to the inconsistency in how this variable is coded and analysed differently across the studies.

One analysis did not include any version of these factors. TENSE was included as a factor in five of the studies, two of which found it to be significant. Further, of those five studies, three included MOOD as a separate factor, with two analyses also finding it to be significant. However, four analyses combined TENSE, ASPECT and MOOD into a single factor, TAM, and it was found to be significant in two of these. So, there is variation across these studies, both in how these grammatical categories are coded, and the significance of these factors on subject pronoun realization. However, in each of the analyses that included either TENSE or TAM, verbs in the imperfect were included as a variant. None treated ASPECT as a separate variable, instead coding it as a tense of the verb, which is the only way aspect appears in Spanish. Following these methodologies, such as Nagy et al. (2018), I combine all three categories into TENSE as a single

variable. Thus, for this dissertation, the imperfect aspect is a variant of the same variable as the present tense and the subjunctive mood, all are TENSE.

While the findings are inconsistent across the studies presented above, there are some broad generalizations that can be drawn from them that can be tested in my study. For example, the imperfect tends to favour overt pronouns because of its more ambiguous verbal inflections across persons (Orozco & Hurtado, 2020:19). This suggests that TENSE may overlap with verbal ambiguity in terms of its role in subject pronoun expression, which may also be significant in Calabrian Italian and Ciociaro.

3.3.2.2. Preverbal element

It is important to note that within the envelope of PREVERBAL ELEMENTS there are several distinct grammatical categories: negation, direct and indirect objects, and reflexive clitics.

Within the framework of the functional hypothesis, one might hypothesize that reflexive preverbal elements would lead to a higher null rate because the reflexive clitic disambiguates the person of subjects (e.g., a 1SG from a 3SG subject). A reflexive clitic can also disambiguate, in some cases, number—but never grammatical gender. This is true of some of the studies cited in Table **14**, but other studies do not find reflexive pronouns to be a significant predictor of subject pronoun realization (Abreu, 2009; Erker & Shin, 2015; Padilla, 2020).

Non-reflexive object pronouns, direct or indirect, are not considered in any of the studies of Spanish varieties discussed in this chapter. However, in Faetar, "…subject pronouns are less likely to surface when other material fills the preverbal space in linear surface order, even with different syntactic roles" (Nagy et al., 2018:41).

Finally, preverbal negation markers have been examined in some of the studies in Table **14**, with inconsistent significance. The general trend across those studies is that negative VPs have a higher null rate. This may be due to the negation occupying the preverbal space as Nagy et al. (2018) noted.

In this section, I have presented an overview of methodologically similar studies, focusing on the variables they examined, and which were found to be significant. The findings of these studies help inform the hypotheses and methodology of this dissertation. Of particular note is that across the 10 studies, almost none of the demographic variables are found to be significant, except AGE, which is significant in three out of the five analyses that reported it, and EDUCATION, which is significant in Padilla (2020). While nothing that was found to be significant—or insignificant— in other research should be assumed to hold true for either Ciociaro or Calabrian Italian, it helps provide a research context within which they can be analysed.

3.4. Calabrian and Ciociaro in contact with other languages

Both the HLVC and CHILS corpora contain recordings of participants who live in Canada, speaking Calabrian and Ciociaro as heritage languages in English-dominant communities. As such, an obvious question is to what extent has the language been shaped by contact effects with other Italian languages and English. This is an especially pertinent question when looking at verbal ambiguity as English has very syncretic verb forms. In this section, I first present the linguistic complexities of the heritage languages in contact with other Italian languages, and how that may apply to the Calabrian and Ciociaro communities I am examining. I then give some examples of research on heritage languages in contact with English to inform my hypotheses in the following chapter.

In Italy, as described in **2.1.2**, regional Italians have become the mother tongues of most Italians since the 1950s. This increased contact with RIs has led to the *Italianization* of some dialetti over the previous 60 years as they adopt features of standard Italian, especially at the lexical level (Cerruti, 2016; Dal Negro & Vietti, 2011). However, for the Italians who immigrated to Canada in the 1950s and 1960s standard Italian was not a part of their lives. They emigrated as standardized education and RIs were just beginning to spread across Italy.

There are, though, some differences between the Ciociaro community in Sarnia and the Calabrian community in the GTA. As discussed in **2.2.3.2**, most of the Ciociaro of Sarnia have had little to no contact with standard Italian, either in Canada or in Italy. Therefore, contact effects from SI are unlikely to be found in the Ciociaro community. Regarding contact with other dialetti, the Italian community of Sarnia is also majority Ciociaro. So, the dialetto is the language of use among the community, and English is used with non-Ciociaro speakers, either other Italian immigrants or anglophones of any origin. This narrows the use of Ciociaro, but also means it has been in only limited contact with other dialetti, as English is used among speakers of different dialetti rather than using an Italian koine established by the community. This initially seems to run counter to what Haller (1987) finds looking at dialetti speakers in New York City, where there is evidence of a "dialectal lingua franca". However, in describing the domains of use of participants to the study, Haller finds "all informants concur in claiming the use of dialect with parents and grandparents, dialect or English with other family members, standard Italian with Italian friends and acquaintances, and English at the workplace" (1987:397). This is similar for the Italian community of Sarnia; except they do not have access to standard Italian to use with their Italian friends and acquaintances, and so use English instead. Therefore, I expect there to be little contact between Ciociaro and other dialetti, RIs, or standard Italian.

63

The Calabrian community of Toronto may have experienced more exposure to dialetti, other regional Italians, and standard Italian. However, in this chapter I have shown that for the relevant features, namely subject pronouns and verbal paradigms, Calabrian and standard Italian are very similar. Thus, while contact may have affected the heritage Calabrian of the GTA in other ways (e.g., lexically), those changes fall outside the scope of this dissertation.

Contact effects with English have been the subject of much scholarship for minority languages, including its effect on pro-drop in those languages. Because English has a null subject rate of just 2% (Nagy et al., 2011:139), one expected effect on a pro-drop language in contact with English would be a reduced rate of null subjects.

While this is an incredibly difficult question to answer definitively, Nagy et al. (2011) compared homeland and heritage varieties of Cantonese, Russian, and Calabrian from the HLVC corpus for contact effects with English, and found "no correlation between either rate or constraint hierarchies and either generation or any ethnic orientation measure" (2011:143). This is similar to our analysis of homeland and heritage Faetar, including speakers from three generations, which also found no significant difference in NSP rates across generations for heritage speakers (Nagy et al., 2018:43).

The CHILS analysis only includes first-generation speakers of Ciociaro. So, while they are a heritage community in Sarnia, they did not learn Ciociaro in Canada. However, Schmitz et al. (2016) compared monolingual Italian speakers from Campania (southern Italy) to heritage speakers raised in Germany, and found an identical null subject rate of 90% for both heritage and monolingual Italian speakers. This is a useful comparison as both German and English are non-pro-drop languages.

This is not to say that contact with English has not been found to influence pro-drop. A study on bilingual English and Spanish speakers in the United States found a higher rate of overt subject pronouns among those who were more proficient in English (69%) than monolingual Spanish speakers (43%) (Montrul, 2004). However, the participants in her study were born in the United States, started speaking English by the time they started primary school, and the "intermediate" speakers "felt more proficient in English than in Spanish" (Montrul, 2004:131). In the case of the Ciociaro speakers included in this study, they were born in Italy and emigrated at an average age of 18.7 years (with a range of six to 33) and only one participant believed they spoke better English than Ciociaro.

Therefore, it might be tempting to assume there must be a difference between the heritage variety and the homeland variety of Calabrian and Ciociaro regarding either the overall rate of pro-drop or the significant factors. However, research comparing homeland and heritage language communities for this effect have been mixed.

In this chapter I have provided the research context for this dissertation. I have described the central grammatical features of the languages I am investigating. Specifically, I have described pro-drop, which relates to the subject pronominal system and verbal morphology of Calabrian and Ciociaro. Further, I presented the functional hypothesis, and how it may relate to pro-drop based on the ambiguity of a VP and the ability of the interlocutor to recover the subject from elsewhere in the discourse. I have also provided a brief overview of comparable research to better inform my hypotheses and methodology for this project.

4. Hypotheses

In the preceding chapters, I have provided the relevant background to present my own research. I have situated both Calabrian Italian and Ciociaro within the larger framework of the languages of Italy and have described comparable studies which inform this dissertation. In this chapter, I lay out the hypotheses I am testing.

In my study, I test the functional hypothesis and how it relates to subject pronoun usage in Calabrian and Ciociaro. To examine its potential effects, each of the following hypotheses is tested by interrogating the data from a different viewpoint. By analyzing the corpora and their results with these hypotheses in mind, I can better understand how pro-drop works in these languages, how the findings relate to the functional hypothesis, and how the results compare to other research.

I have grouped my hypotheses into three subsections: first are hypotheses related to differences that may occur between homeland and heritage varieties of both Calabrian and Ciociaro (**4.1**). Second are hypotheses regarding the null rates of these languages and comparing those rates to those of other languages presented in the previous chapter (**4.2**). In the third section I propose hypotheses about the significant variables that constrain subject pronoun realization in Calabrian and Ciociaro, and how they compare to other studies (**4.3**). In this chapter, I number the hypotheses as they are presented, and that numbering is repeated when the hypotheses are tested in Chapter **6**.

4.1. Homeland and heritage varieties

In this thesis, I examine Calabrian and Ciociaro, using corpora from Italy (*homeland*) and from a heritage community in Canada for each. As such, I examine to what extent the homeland and heritage varieties differ in their usage of null subjects.

If they differ a great deal, several possibilities can explain the changes both languages may have undergone. It could be due to contact the heritage communities have had with their respective dominant English communities (in Toronto and Sarnia). It may also be divergent evolution of the homeland and heritage varieties. Another possibility is koineization and accommodation with related Italian migrant communities, due to internal migration in Italy or mixing of Italian heritage communities in Canada. Of course, differences between the homeland and heritage varieties can also be a mix of these possibilities.

(1a) I hypothesize that there is a similar overall null subject rate for the homeland and heritage varieties of both Calabrian Italian and Ciociaro

I believe the homeland and heritage varieties of each language will have similar null subject rates. I do not think that the effects of contact, divergent evolution, or accommodation will have altered either variety to a significant degree because, as described in **3.4**, the domains of use for both Calabrian and Ciociaro are distinct. In Canada, both languages are used among members of the heritage community, and English is used elsewhere. As such, I expect that the grammars of these languages will be unaffected from contact with English or another Italian variety.

This hypothesis is tested by comparing the overall null subject rates of the heritage Calabrian corpus to that of the homeland variety (*HLVC corpus)*, and then by comparing the rates of the heritage Ciociaro corpus (*CHILS*) to that of the atlas data (*AIS*).

This hypothesis is supported by previous variationist sociolinguistic analyses (in contrast to numerous experimental studies, see discussion in Nagy, 2015), which found little difference in rates of subject pronoun usage between homeland and heritage varieties of Faetar (Nagy et al., 2018) and Spanish (Avila-Jimenez, 1996; Carvalho & Child, 2011; Montrul, 2004; *inter alia*). There are exceptions to this, however, with Abreu finding overt subject pronouns used at a higher rate for bilingual Spanish-English speakers than monolingual Spanish speakers (Abreu, 2012).

Once this hypothesis has been tested, and the differences (if any) between the null rates of the homeland and heritage varieties have been established, I then compare the significant variables of both varieties, for both languages.

My first hypothesis is related to the idea that the homeland and heritage varieties of both Calabrian and Ciociaro have not diverged very much in terms of null rates. My next hypothesis relates to their constraint hierarchies.

(1b) I hypothesize that, for Calabrian Italian and Ciociaro, each language's homeland and heritage varieties share significant variables.

The significant variables may differ between Calabrian and Ciociaro, but I hypothesize that between the homeland and heritage varieties of each the significant variables will be the same. This is a further means of examining how the homeland and heritage varieties of both languages have, or have not, diverged from each other. The reasoning of this hypothesis is the same as (1a): I do not believe the two varieties have become distinct in how subject pronouns are used, despite the distance of space and time since emigration.

If the homeland and heritage varieties have similar null subject rates and similar significant variables, this would suggest that they have not diverged from each other, and that the participants in each corpus are using pro-drop in a similar way. This would justify combining the data from each language into a single analysis, and then comparing both languages to other comparable studies. On the other hand, if the rates and variables differ significantly, then the homeland and heritage varieties will be analysed and compared separately to examine potential contact effects with English (heritage varieties) or with other varieties of Italian (homeland varieties). To test this hypothesis, the significant variables of the four corpora will be compared: AIS to CHILS, and the homeland to the heritage Calabrian Italian variety.

4.2. Null subject rate

The next set of hypotheses are tested by comparing Calabrian and Ciociaro to each other. The first step in this comparison is to examine how their overall null subject rates compare.

(2a) I hypothesize that Ciociaro has a lower null-subject rate than Calabrian Italian.
This is because Ciociaro, as described in the previous chapter (**3.2.2**), has a more syncretic inflectional system than Calabrian. This hypothesis is tested by comparing the overall null rate between the two languages.

If Ciociaro has a lower null-subject rate, this alone does not support the functional hypothesis. To test if the lower rate of null subjects is related to a more ambiguous verbal morphology, the following hypothesis is proposed.

(2b) I hypothesize that a lower null-subject rate correlates with a higher rate of ambiguous verbal forms.

This inverse correspondence between the two rates would support the functional hypothesis as an overt subject pronoun reduces the need for distinctive verbal morphology to identify a VP's subject. This correlation between lower null-subject rates and higher rates of verbal syncretism does not specify the direction of causality, if any. In other words, even if the hypothesis is confirmed, it would not be clear whether Ciociaro has more ambiguous verbal morphology, which resulted in a lower null-subject rate, or that Ciociaro has a higher rate of overt subject pronouns, which allowed for less distinctive verbal inflections.

To test this hypothesis, the null subject rate of each corpus will be compared to its rate of VPs with ambiguous verbal forms. If both hypotheses (2a) and (2b) are supported, Ciociaro will have a lower null-subject rate and a higher rate of ambiguous verbal forms than Calabrian. If this hypothesis is confirmed, the results would support the functional hypothesis by showing that subject pronouns are more likely to be used when a language has more verbal ambiguity.

I believe this will be the case between Calabrian and Ciociaro; however, the AIS corpus is likely to be an outlier. The AIS corpus was collected using disconnected translation prompts, rather than the spontaneous narratives of the HLVC and CHILS corpora. As a result, it does not allow for (or require) the subject to be recovered from a previous VP.

(2c) I hypothesize that AIS has a lower null-subject rate than CHILS, yet a similar ambiguous verbal morphology rate.

This hypothesis will be tested by examining the rates of ambiguous verbal forms for the AIS data and for the CHILS data in (2b). If confirmed, (2c) further supports the functional hypothesis because each phrase is independent from the preceding and following VP, meaning the subject is either disambiguated by an overt subject pronoun, the verbal morphology, or is left ambiguous.

A potential complication in testing this hypothesis is how faithfully the participants translate the prompts they are given in Italian, which is addressed in hypothesis (3b).

Once the rates of both Calabrian and Ciociaro have been examined, both languages will be compared to the studies presented in the preceding chapter. Among the languages I described, French has the lowest null subject rate (15%; Schmitz & Müller, 2008), Spanish has a wide range of reported null rates (51-81%; see **3.3**), and finally Italian had the highest null rate of the (90%; Nagy et al., 2011; Schmitz et al., 2016).

(2d) I hypothesize that both Calabrian Italian and Ciociaro have higher null rates than both French and Faetar, similar rates to some varieties of Spanish, and lower rates than Italian.

As shown in the previous chapter, both Calabrian Italian and Ciociaro have much less syncretic verbal paradigms than French, and thus the functional hypothesis would predict that they will have higher null-subject rates. Ciociaro is also more syncretic than Italian, so it should have a comparatively lower null-subject rate. Further, given the wide range of rates presented for the Spanish varieties examined in **3.3**, I hypothesize that some varieties of Spanish have higher null rates than Ciociaro. However, given the more distinct verbal inflections of Calabrian, which appear to be very similar to standard Italian, I suspect it will have a higher null-subject rate than all the Spanish varieties discussed in **3.3**.

This hypothesis is tested by comparing the null rates of Calabrian and Ciociaro (and each of their homeland and heritage varieties, if it is appropriate to maintain separate analyses), to the reported null rates of these languages from methodologically comparable studies. If the null rates are ranked in this way across languages, this would support the functional hypothesis because those

languages with more ambiguous verbal morphology have a correspondingly lower null-subject rate to clarify the subject of a VP.

4.3. Relevant Variables

Once the overall null rates of Calabrian and Ciociaro are examined, the next step is to compare the significant variables in each of the analyses. These constraint hierarchies are the factors that relate to subject pronoun realization, as discussed in **3.3**. I make several hypotheses about these variables, how they are tested in this dissertation, and how particular outcomes will support or counter the functional hypothesis.

(3a) I hypothesize that SUBJECT is a significant variable for each of the corpora in this dissertation, with subjects containing ambiguous verbal inflections disfavouring null subjects, and subjects with unambiguous verbal inflections favouring null subjects.

To test this hypothesis, the multivariate analyses will show the effect a VP's subject has on whether a subject pronoun is null or overt for both Calabrian Italian and Ciociaro. The SUBJECT of a VP has been found to be significant in every study introduced in **3.3.1.1**. Certain subjects, such as 1PL and 2PL, have a much more distinctive verbal inflection in each of the Romance languages presented in the preceding chapter. Other subjects tend to have very similar inflections, such as 1SG and 3SG.

The consistency of the effect of SUBJECT supports the functional hypothesis insofar as the distinctions in the inflectional system are related to a VP's subject. Therefore, I hypothesize that SUBJECT will be a significant variable for each of the corpora. As shown in **3.2.2**, the inflection of verbs with, for example, 1PL subjects is much less ambiguous than those with 1SG or 3SG subjects. If the functional hypothesis is supported, that difference should correspond to a lower rate of null subjects for 1SG or 3SG subjects than for those with 1PL subjects.

This hypothesis is supported if SUBJECT is a significant variable in each of the analyses, and with a ranking of the variants that shows subjects that have ambiguous inflections disfavour null subjects and those with unambiguous inflections favour null subjects.

(3b) I hypothesize that SWITCH REFERENCE is a significant variable for both the HLVC and CHILS corpora, and PRONOUN IN PROMPT is a significant variable for the AIS data.

My second hypothesis for the significant variables is related to priming variables. For the HLVC and CHILS corpora, I hypothesize that SWITCH REFERENCE is a significant variable. I predict a higher null rate when the subject of a VP is unchanged from the previous VP, and a lower null rate when the subject has changed. I expect the same effect for both the HLVC and CHILS corpora.

However, AIS contains translations of phrases from elicitations in Italian provided by the fieldworker, and so does not have SWITCH REFERENCE as a variable. As a result, this hypothesis cannot apply to that corpus. Instead, I consider the effect of a different priming variable: the presence or absence of a pronoun in the elicitation prompt, and how participants translate these prompts into Ciociaro. In other words, do the participants translate the semantic information of the prompt or translate word-by-word, including pronouns? For example, they may add an overt pronoun when the prompt had a null subject. While this is not the same as the SWITCH REFERENCE variable, it is a priming variable. I hypothesize that participants will significantly disfavour null subjects when the prompt contains an overt subject pronoun, favouring a direct translation.

This hypothesis will be tested through the results of multivariate analyses to measure the strength of the effect of these priming variables (SWITCH REFERENCE and PRONOUN IN PROMPT) on subject pronoun realization.

If SWITCH REFERENCE is significant, the ranking of variants could support the functional hypothesis. If a subject pronoun is more often overt when the subject of a narrative changes, this overt subject pronoun disambiguates a change in subject, when the verb forms may be similar or identical. This will be tested by examining the null rate for a VP with a 3SG.M subject, when the preceding VP had a 3SG.F (or a different 3SG.M subject) subject versus an unchanged subject or another subject (e.g., 1SG). In this case, a change from 3SG.M to 3SG.M would have the same verbal inflections, yet the subject has changed, so the functional hypothesis posits that an overt subject pronoun should be used. However, if a VP has a 1SG subject, and the previous VP had a 3SG.M, the verbal morphology, if distinctive, already clarifies this change in subject, and an overt subject pronoun is not required.

Regarding the remaining variables that were described in **3.3.2**, I expect that neither TENSE nor PREVERBAL ELEMENT will be significant variables for either the CHILS or HLVC analyses. However, the AIS results may differ as the prompts cover a wide variety of tenses, aspects, and moods.

(3c) I hypothesize that the TENSE variable is significant in the AIS analysis, and the ranking of the variants reflects their relative ambiguous verbal morphology.

Because the AIS data is not based on narratives, which tend to favour past tenses and 1st and 3rd person subjects and past tenses, there are more VPs in the present and future tenses, providing

more variants to consider. Also, different tenses have varying amounts of syncretism. I hypothesize this will be reflected in their ranking in the analysis.

This hypothesis is supported if TENSE is a significant variable in the AIS analysis. It is further supported if, within the ranking of the variants, tenses with fewer syncretic forms, such as those with auxiliary verbs, favour null subjects, while those tenses with more ambiguous verbal paradigms, such as the imperfect or present tense, disfavour null subjects.

For the variable PREVERBAL ELEMENT, divided into object pronouns, reflexive pronouns, and negation particles, may be correlated with a higher null-subject rate in comparable studies, I expect any effect will be much smaller than the other variables considered in this section for all the corpora. Within the ranking of variants for this variable, however, there is a means of testing the functional hypothesis.

(3d) I hypothesize a lower null rate for VPs with a reflexive pronoun and a 3rd person subject than for those with a 1st or 2nd person subject.

If VPs with reflexive pronouns have a much higher rate of null subjects than the other categories of preverbal elements, this would support the functional hypothesis, as a subject pronoun is not necessary when a reflexive pronoun clarifies the subject. In standard Italian, 1SG, 2SG, 1PL, and 2PL subjects each have distinct reflexive pronouns in standard Italian. However, there is no distinction made between 3SG.M, 3SG.F, 3PL.M, and 3PL.F.

Therefore, if 3rd person reflexive pronouns similarly do not disambiguate the subject for gender or number in Calabrian and Ciociaro, this hypothesis can be tested by examining the difference between the rates of null subjects for 1st and 2nd person versus 3rd person in VPs with a reflexive pronoun. The functional hypothesis suggests that VPs with reflexive pronouns and 1st or 2nd

person subjects should favour null subjects, as the overt subject pronoun is redundant. But VPs with reflexive pronouns and 3rd person subjects are still ambiguous for gender and number, and thus should disfavour null subjects.

In this chapter, I have presented the hypotheses I test to interpret the results of this dissertation. I present which hypotheses are supported, and which are not, in Chapter **1**. By testing these hypotheses through my analysis of pro-drop, I can better understand any variation between homeland and heritage varieties; how Calabrian Italian and Ciociaro compare to each other and to other Romance languages; and how significant variables, and the ranking of variants within them, can inform us about the realization of subject pronouns in these languages. Further, the hypotheses I have laid out in this chapter, and whether the results support them, are a test of the functional hypothesis and its potential role in the use of subject pronouns in Calabrian Italian and Ciociaro.

In the following chapter, I describe the methodology of this study: the participants of each corpus, the variables that are coded, and the modeling to analyse the data.

5. Methodology

Now that the background and hypotheses have been established, in this chapter I describe the methodology used to analyse pro-drop in Calabrian Italian and Ciociaro. The first section (**5.1**) is a description of the three corpora used in this dissertation: Nagy's *Heritage Language Variation & Change* (HLVC) corpus (2011), from which the Calabrian subcorpus is used; the *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* (AIS), a linguistic atlas collected in the early 20th century throughout Italy and Switzerland (Jaberg & Jud, 1928), from which the homeland Ciociaro data is extracted; and the Ciociaro Heritage Italian Language of Sarnia (CHILS) corpus, which contains interviews I conducted with Sarnia's heritage Ciociaro community.

After a description of the corpora and participants, I present the coding process (**5.2**). Specifically, I describe how I track and code the examples of subject pronoun use in the recordings using the ELAN software (*ELAN*, 2020) for the HLVC and CHILS corpora, and how I extract the data from the AIS maps. In **5.3**, I lay out the variables that are coded and analysed. In the final section of this chapter (**5.4**), I describe the multivariate analyses conducted using the Rbrul program (Johnson, 2009), and the methods I use to find the most reliable model for each corpus.

5.1. Corpora

This thesis investigates the relationship between verbal morphology and subject pronoun usage in Calabrian Italian and Ciociaro. This section describes the three sources of data. I use audio recordings from a Calabrian Italian subcorpus of Nagy's HLVC project (2011). This is a corpus of interviews conducted with homeland and heritage speakers. I also use two sources of recorded

Ciociaro: the homeland data is from the AIS linguistic atlas, which was collected in Ciociaria in 1923-1924 (Jaberg & Jud, 1928); and the heritage data is from recordings of semi-directed sociolinguistic interviews I have conducted since 2015.

In this section I describe each of these corpora and the participants that are included in my dissertation.

5.1.1. HLVC

The CHILS and AIS corpora provide a means of comparing the homeland and heritage varieties of Ciociaro. Ciociaro is a dialetto, however, and it is unclear how distinct it is from Italian. To provide a comparison of Ciociaro to a regional Italian, I use Calabrian data from Nagy's 2011 corpus *Heritage Languages Variation & Change* (HLVC), which contains recorded interviews from many heritage languages in Toronto (Nagy, 2011), as well as recorded interviews from speakers who live in the homeland of each language. For the Calabrian Italian subcorpus, this means participants are either heritage speakers living in the greater Toronto area (GTA), or they are living in Calabria at the time of the recording (the homeland variety).

These interviews were conducted between 2009 and 2013. The interviewers for both the homeland and heritage interviews are heritage speakers of Calabrian Italian, and the interviews are conducted in Calabrian Italian. Interviews consist of three parts: participants respond to an ethnic orientation questionnaire, describe the images and scenes of a children's book in the target language, and an interview (Nagy, 2011). During the interview, participants are asked questions about their neighbourhood, community, childhood, work, and relationships.

Figure 3. Example of transcribed interview from the HLVC corpus (IXF35A).
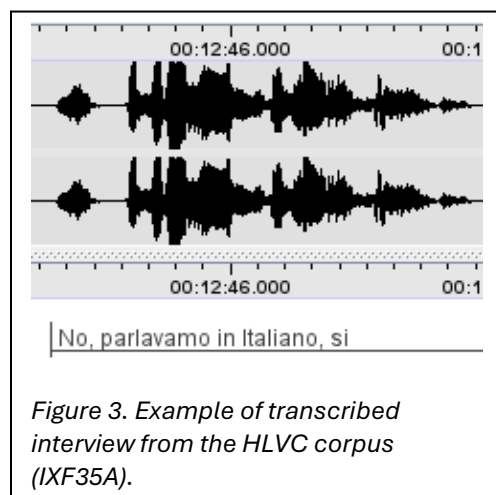
The interviews are semi-directed by the interviewer and are similar in structure to those in the CHILS. Further, the interviews, as part of the HLVC project, are already transcribed orthographically in ELAN by Nagy's research assistants (see

https://ngn.artsci.utoronto.ca/HLVC/3_2_active_ra.php,

https://ngn.artsci.utoronto.ca/HLVC/3_3_former_ra.php). These transcribed interviews, as shown in Figure **3**, were generously shared with me, which made the process of identifying VPs for coding much faster than for the AIS or CHILS corpora.

5.1.1.1. HLVC participants

Homeland participants					Heritage participants				
Speaker	*Gender*	*Age*	*Duration*	*Tokens*	*Speaker*	*Gender*	*Age*	*Duration*	*Tokens*
IXF35A	F	35	24:34	87	I2F57A	F	57	30:06	95
IXF38A	F	38	33:41	66	I1F61A	F	61	1:18:31	98
IXF51A	F	51	44:38	101	I1F65A	F	65	52:51	86
IXF61A	F	61	33:41	79	I1F71A	F	71	54:53	96
IXF94A	F	94	1:15:04	102	I1F73A	F	73	49:29	100
IXM35A	M	35	50:47	93	I1M60A	M	60	1:16:09	112
IXM47A	M	47	56:32	94	I1M61A	M	61	39:42	109
IXM52A	M	52	56:27	100	I1M61B	M	61	57:29	106
IXM61A	M	61	32:29	95	I1M62A	M	62	57:22	102
IXM64A	M	64	1:10:21	103	I1M75A	M	75	45:45	99
Average		**54**	**47:49**	**92**	**Average**		**65**	**47:56**	**100**

Table 15. Table showing the speakers from the HLVC corpus used in this study.

From the Calabrian subcorpus, 10 homeland and 10 heritage interviews were chosen, balanced for gender (Table **15**). These participants were chosen to be as similar in age to the CHILS

participants as possible. Thus, the eldest speakers from the homeland and heritage corpora were chosen, with an average age of 59 when they were interviewed. As near to 100 tokens as possible from each participant is extracted from the recordings, which range from 20 to 80 minutes in length.

These interviews were conducted by HLVC research assistants who were members of the Toronto Italian community with family from Calabria. In many cases, the participants knew the interviewer already (a relative or family friend). Both the homeland and heritage interviews took place in quiet cafes, restaurants, or in the participants' homes. There were several occasions where third parties would enter the conversation for a while, and then leave. While these were interesting chances to hear the participant interact with others, thereby reducing the performative aspect of being interviewed ("Observer's Paradox"; Labov, 1972: 209), these portions were excluded from the analysis to be consistent with the coding of the CHILS interviews.

5.1.2. AIS

The CHILS corpus provides data for Sarnia's heritage variety of Ciociaro, equivalent to the heritage Calabrian recordings from the HLVC corpus. It would be ideal to have data from the 1950s in Italy, the approximate period of immigration for Sarnia's Ciociaro community, to compare how the variety has changed over time. Or, to have recordings of Ciociaro speakers collected in the past couple decades, which would be a more direct comparison to the HLVC corpus. To the best of my knowledge, neither corpus exists, making the data from the AIS project the most comparable data available (Jaberg & Jud, 1928). To compile this linguistic atlas, fieldworkers travelled to 306 towns throughout Italy and the Italian-speaking cantons of Switzerland and found informants in each location to take part in their study. In southern Italy,

Gerhard Rohlfs conducted 81 interviews between 1919 and 1925 (Jaberg & Jud, 1987:4[16]).

Rohlfs asked each participant to translate the Italian prompts to the "local patois" (Jaberg & Jud,

1987:2). The researchers conducted interviews in Ciociaria in 1923 and 1924, and their data is

the most reliable available source of how Ciociaro was spoken at the time.

Figure 4. Example of the images used in the AIS questionnaire, showing grapevines (Jaberg & Jud, 1928:1304).
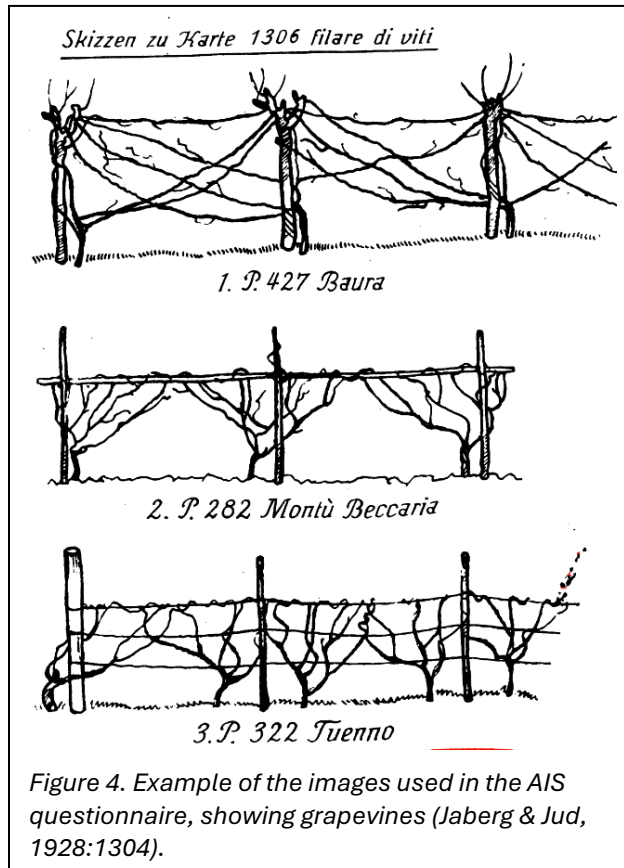
The standard version of the interview consists of 2,000 questions. Some of the tasks are to provide verbal paradigms or identify pictures (cf. Figure **4**). Most of the questions, however, are direct translation requests; the interviewer provides a term or phrase in standard Italian and asks the informant to repeat it in their local variety. For example, prompt[17] #698 is *ho la voce rauca* ('I have a hoarse voice'). The informant then repeats the prompt in their local variety. Almost all these prompts are disconnected from the adjacent

prompts but tend to be of a related topic. Prompt #697 is *ho la febbre* ('I have a fever'), for

example, and #699 is *sto per svenire* ('I am going to faint').

---

[16] The original Volume 1 was published in 1928; however, here I cite the information about the project's methodology provided in the 1987 Italian translation.

[17] I use "prompt", rather than "map" when the interviewer provided an oral prompt to elicit a translation from the participant. These correspond to the map number.

The responses to the prompts were transcribed and published as maps of Italy and southern Switzerland. The prompt is at the top of each page, and a map of Italy is included with the transcribed response of each participant aligned to their geographic location. I used an online version of the AIS maps for this research (Tisato, 2009). Of the 2,000 maps, I identified 258 prompts containing an inflected verb. From these, I excluded the phrases which had a non-human or an inanimate subject, an NP subject, or were in the imperative, in keeping with the methodology for the HLVC and CHILS corpora. This left 184 prompts that could have an overt subject pronoun when translated by the participant into Ciociaro (See Appendix **G** for a list of the prompts used in this study).

The AIS presents the responses of the participants as transcribed by the researchers in a phonetic alphabet that is distinct from the International Phonetic Alphabet (IPA). However, the website that manages the AIS data provides documentation on this phonetic alphabet (the link provided by the Istituto dell'Atlante Linguistico Italiano reproduces their description of their transcription conventions). I was able to use this documentation to create correspondences between the phonetic transcriptions of the AIS and IPA to allow the tokens to be transcribed, coded, and compared to the CHILS (see Appendix **F** for these correspondences).

5.1.2.1. AIS participants

AIS participants

AIS Point	Location	Gender	Parents' P.O.B.	Work	Date	Age	Tokens
654	Serrone, Frosinone	M	Serrone	Sharecropper	18-25 Sept., 1924	51	99
664	Veroli, Frosinone	M	Father: Santa Francesca Mother: Isola Liri	Sharecropper	27-30 Sept., 1924	56	105
701	San Donato, Frosinone	M	San Donato	Municipal usher	12-15 Sept., 1924	60	72

710	Ausonia, Frosinone	M	Ausonia	Sharecropper	9-12 Oct., 1924	51	101
656	Scanno, Aquila	M	Scanno	Sharecropper	13-16 Sept., 1923	44	101
682	Sonnino, Latina	M	Sonnino	Homemaker	2-6 Oct., 1924	56	105
Average						**53**	**97**

Table 16. Summary of demographic information from the AIS points included in this study (Jaberg & Jud, 1987:143-148).

Each village or town included in the AIS is recorded as a point on the map and given a number. There are four towns within the province of Frosinone in the AIS data: Serrone (#654), Veroli (#664), San Donato (#701), and Ausonia (#710). Further, two other locations fall outside the boundaries of Frosinone, but are close to the hometowns of participants of the CHILS corpus and were thus included: Scanno, in the province of Aquila, is northeast (#656), and Sonnino, in Latina, is southwest (#682). See Figure **5**.

Population of AIS locations in 1921 and 2021

Location	1921	2021
654 - Serrone, Frosinone	2,519	3,016
664 - Veroli, Frosinone	15,527	19,849
701 - San Donato, Frosinone	4,760	1,897
710 – Ausonia, Frosinone	3,334	2,431
656 – Scanno, Aquila	4,070	1,723
682 – Sonnino, Latina	5,231	7,440

Table 17. Population of the locations from the AIS corpus included in this study, using census data from the period of the interviews and the present (Statistiche IStat, 2022b).

Table **17** provides population data of the AIS locations from the census conducted the nearest to the date of the interviews, as well as from the most recent census to provide a sense of the size of these towns. Each of these towns saw significant emigration during the post-war period, and most of them now have a smaller population than they did a century ago (see **2.2.3**).

The participants for each location, referred to as *informants* by Jaberg and Jud (1987), are briefly described alongside a short description of the village they represent in the atlas. The information that is provided is basic, but useful. Each informant is described based on their occupation, the

hometown of each of their parents, their age at the time the interview was conducted, and a brief description of the strengths and weaknesses of their linguistic knowledge. As an example, the informant from Serrone is described by Rolhfs, "Never left the countryside, good ethnographic knowledge…absolutely faithful dialectologically speaking. Less confident toward morphology…" (Jaberg & Jud, 1987:143). The other informants are similarly described, a brief mention of their strengths and weaknesses in speaking their dialetto, and their relationship to the area. For example, the informant from Veroli is described as a passionate hunter, and the informant from Ausonia is described as influenced by the written language, meaning standard Italian. These descriptions are summarized in Table **16**. Note that all the participants from these points were born in, and have continuously lived in, the village they represent as informants. Therefore, the location and their place of birth are the same. It is important to note that although 184 prompts are included in my dataset, AIS does not record a response for every prompt for every participant. This is why the total number of tokens for each location is less than 184. It also means that while there are 101 tokens for both Ausonia and Scanno, these 101 responses are not for the same 101 prompts.
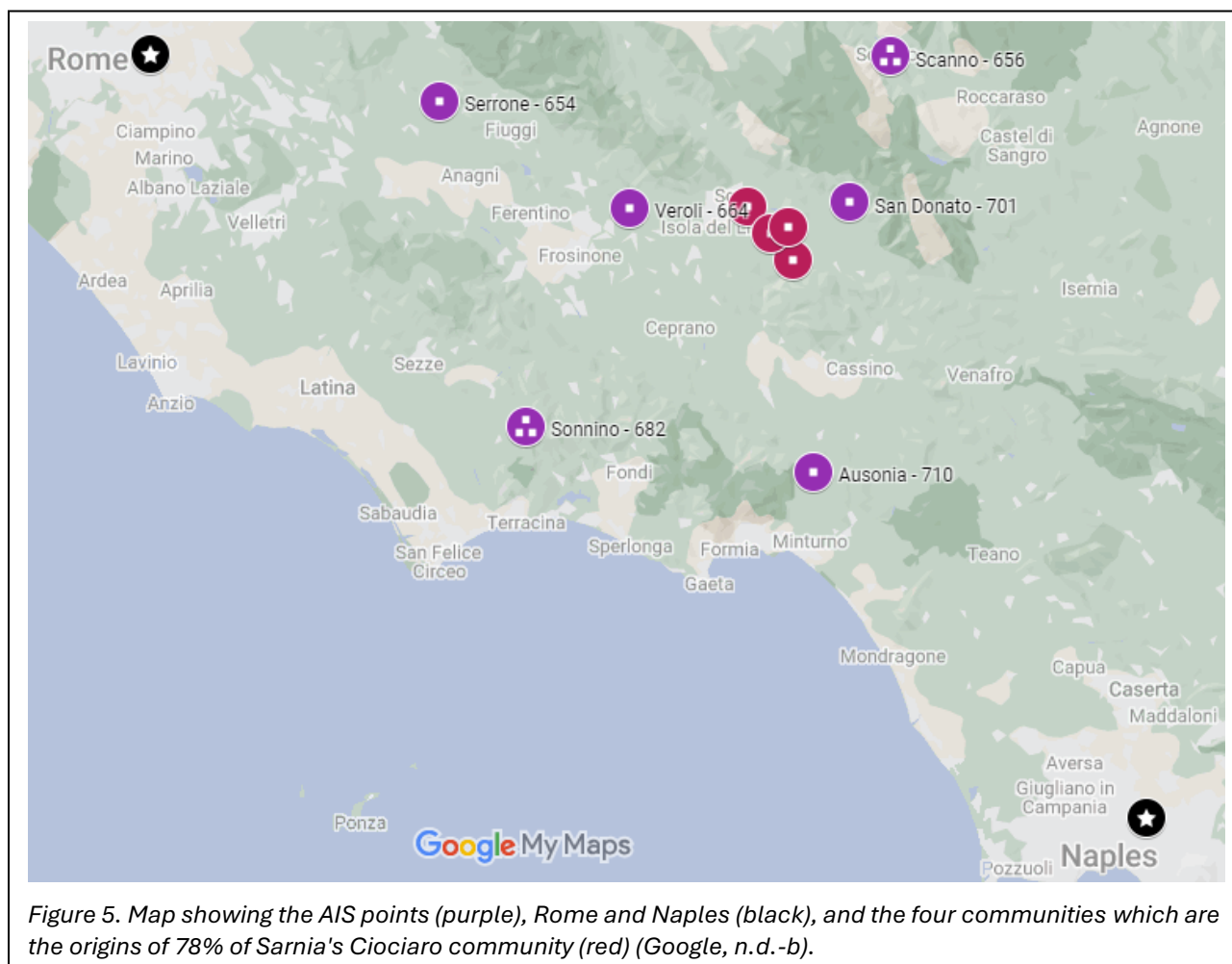
Figure 5. Map showing the AIS points (purple), Rome and Naples (black), and the four communities which are the origins of 78% of Sarnia's Ciociaro community (red) (Google, n.d.-b).

In Figure **5**, the AIS points included in this study are shown on a map of Italy (Google, n.d.-b). The AIS points are indicated with single squares for those in Frosinone and three squares for those outside the region but adjacent to the province. Also, the cities of Rome and Naples are shown with black circles to give a sense of the area's location within Italy. Finally, I also included the communities where the majority of Sarnia's Ciociaro community come from (in red; see **2.2.3.2**). Of the AIS points, Veroli (#664) and San Donato (#701) are the closest to this cluster of villages from which 78% of Sarnia's Ciociaro community emigrated. These two AIS villages are approximately 20km away from the villages of the majority of Sarnia's Ciociari. Therefore, while the AIS speakers are from a generation or two before the CHILS participants,

85

they do come from the same area, and are an appropriate homeland community for comparison with the CHILS.

5.1.3. CHILS

I began my research with the Sarnia Ciociaro community in 2015. I applied for and received ethics approval for this project in 2016 (see Appendix **A**). I recruited participants from the community using family contacts (my father and grandmother are Ciociari) and working with community historian Caroline Di Cocco.

I have interviewed 37 Ciociari in Sarnia, and each interview is approximately one hour long. Each interview has two parts: a wordlist and a semi-directed conversation. For the wordlist, I ask participants to translate English prompts into Ciociaro to study the phonetic variable that was the subject of my MA research: /v/ is often realized as [w] (Iannozzi, 2017). For the semi-directed conversation, I ask questions in English, and the participant responds in Ciociaro (see Appendix **B** for the questionnaire). The pro-drop tokens are extracted from these conversations.

Recordings are conducted in the participants' homes, often at the kitchen table. I use two Zoom recorders during the interviews, an H5 near the participant and a backup H2 off to the side. [18] The recordings are written onto SD cards, which are then transferred and stored on two external harddrives in separate locations. After explaining my research, receiving the participant's consent to participate, and turning on the recorders, I encourage the participants to speak in Ciociaro. Participants often express uncertainty about why they should be interviewed, that they

---

[18] In cases where the participant is rather expressive with the hands and knocks one of the recorders, or keeps banging on the table, I try to keep the backup recorder on a different table.

speak Ciociaro and not "good" or "proper" Italian. However, after some talking, and assuring

them that I did want to hear and record Ciociaro, the interviews begin.

Participants discuss topics such as their childhood in Italy, their emigration experience, what

their first winter was like, etc. As the interviews are generally about their life, the verbs used by

the speakers are typically in the past tense (often imperfect), and there are very few uses of the

future tense. Also, as participants are typically talking about their own experiences, the 1st person

is especially frequent, especially the 1st person singular (e.g., 'I had never seen snow'), which is

also the case for the HLVC recordings. Stories of the participants' children, parents, brothers,

and sisters also provide many examples of verbs in the 3rd person singular and plural (e.g., 'he

didn't realize how big Canada was'; 'they didn't think I'd stay'). The interviews were historical

and autobiographical, with minimal input from me, so there are few examples of verbs with a 2nd

person subject, plural or singular.[19]

From these interviews, 20 participants were chosen, balanced for gender.

---

[19] Except for several examples of 2SG imperatives urging me to drink more espresso or eat more cookies.

5.1.3.1. CHILS participants

CHILS participants

Name[20]	Gender	Place of birth (POB)	Partner's POB	Age	Age of arrival	Duration	n
Annita[21]	F	Fontechiari	Fontechiari	58	20	41:32	91
Antonia	F	Fontechiari	Canada	69	9	49:17	83
Antonietta	F	Posta Fibreno	Posta Fibreno	89	26	31:21	88
Arcangela	F	Casalvieri	Casalvieri	82	18	24:34	86
Assunta	F	Alvito	Alvito	86	24	1:26:26	84
Caroline	F	Fontechiari	Fontechiari	66	6	35:01	85
Clara	F	Castelliri	Castelliri	82	17	28:13	80
Joanna[c]	F	Fontechiari	Fontechiari	57	8	30:36	85
Nunciata	F	Fontechiari	Fontechiari	85	27	33:30	77
Palma	F	Fontechiari	Fontechiari	84	21	52:10	90
Angelo F.	M	Vicalvi	Castelliri	80	17	1:20:52	90
Ennio	M	Veroli	Veroli	84	20	25:42	85
Ernie	M	Fontechiari	Fontechiari	58	10	29:33	84
Frank D'A.	M	Casalvieri	Casalvieri	78	19	43:50	91
Frank L.	M	Fontechiari	Fontechiari	62	18	28:04	89
Isy	M	Casalvieri	Canada	65	10	40:48	87
Nello	M	Casalvieri	Canada	74	10	43:00	93
Raffaele	M	Alvito	Alvito	95	32	56:58	94
Renzo	M	Alvito	Alvito	81	19	1:04:16	100
Sante	M	Posta Fibreno	Posta Fibreno	89	28	48:07	74
Average				**76**	**18**	**43:42**	**87**

Table 18. Participants included in this study from the CHILS corpus.

---

[20] First names have not been changed. Participants consented to their recordings being shared on a publicly available archive, which is open online (www.italiansinlambton.ca). Those with an initial of their surname share a first name with another participant who may included in this study (e.g., Frank L. and Frank D'A.), or not (e.g., Angelo F.)

[21] The following participants were married and interviewed together: Annita and Frank L, Antonietta and Sante, and Joanna and Ernie. However, all the tokens included in this dissertation for Antonietta are from her interview with her sister-in-law, Nunciata.

For the research presented in this thesis, I use 17 recordings, in which there are 20 participants and 14.6 hours of audio, as shown in Table **18**. The 20 participants are 10 men and 10 women ranging in age from 57-95 at the time of the interview, and with an average age of 76. Three of these interviews were conducted in pairs, one with sisters-in-law (Antonietta and Nunciata), and the other two with married couples (Annita & Frank L, Joanna & Ernie). Sante was interviewed with his wife Antonietta, but her interview with Nunciata is used in this dissertation. In each of these cases, I interviewed each participant separately, and then had a conversation with them both. To ensure consistency, only the individual portion is included in this study. However, as the other participant was nearby, there are also examples of 1PL (e.g., 'we didn't know what to expect'). As can be seen in Table **18**, the participants are a homogenous group, with many participants, as well as their spouses, sharing the same place of birth (P.O.B.).

Each of the participants lived a similar life in Italy before emigrating. They all attended school in Italy until they finished 'la quinta', roughly corresponding to grade five, when they were age 10. Joanna, Ernie, Isy, and Nello emigrated with their parents after finishing la quinta. The exceptions are speakers Caroline and Antonia, who are sisters, and emigrated with their parents before they had finished la quinta. The other participants, the next youngest at the time of emigration being Angelo F. and Clara (17), worked the farm with their families before emigrating.

These participants were all from rural farming communities, and often lived a several hours walk from the town centre. Some families owned small farms and homes. Other families did not own the farms they worked on, nor homes they lived in. Most families worked as sharecroppers: they

did not own the land but farmed it on behalf of someone else. In most cases, they reported not knowing anyone from outside their small communities of several hundred people.

With so few years of standardized education, and little to no communication with people outside their small communities, the participants did not acquire Italian. They lived their early lives in Ciociaro in Italy, and continue to live their lives in Sarnia, with only passive exposure to standard Italian on television.

In previous presentations of this fieldwork, I have been asked about the influence of ongoing contact with Italy and (a more standardized form of) Italian. Most participants have family in Italy; however, until the early-to-mid 1990s, calls to Italy were expensive and rare. Even once those in Sarnia had the means to call internationally, their families in Italy didn't have telephones until much later. Many participants reported sharing news of births, marriages, etc. by calling someone in the nearest town and asking them to share the news with their families, or by scheduling a time to call back when their family could be in town to answer the phone.

Further, of the 20 people included in this study, everyone returned to Italy at least once, but only Caroline has done so regularly. The other participants returned to Italy for a marriage, funeral, or other significant event, but described the area as too changed to be recognizable. The scale of emigration, as described in **2.2.3**, helps explain the lack of interactions with people in Italy.

Italian television was not available to the participants until relatively recently. A channel called *Telelatino* (Now *TLN*) was added to basic cable around 1995. This channel broadcast programs in Italian and Spanish like soccer games, cooking shows, and half-hour news programs. However, when asked, only a couple of the participants said they had watched the watched soap

operas on the channel. By the late-2000s TLN launched separate Italian-language and Spanish-language speciality channels, and none of the participants reported paying for these.

Sarnia's Ciociaro community has been in much more prolonged contact with English. Research on contact effects of English on pro-drop in other heritage languages has been mixed (see **3.4**). However, extralinguistic variables are included in my analysis to measure any possible effect. It is also possible that participants in the CHILS corpus who arrived in Canada at a younger age would have a larger contact effect with English. Those who arrived in Sarnia when they were younger than 18 attended some schooling in English, and so more transfer from English might be possible. To account for this possibility, the age of arrival (AOA) of speakers is included as an extralinguistic variable. These extralinguistic variables measure, at least in part, the Ciociaro domains of usage for each participant. One may expect a stronger transfer effect from English for participants who used English at their workplace (most participants are now retired). If participants use English at home with their spouse, a stronger transfer effect from English may occur. A possible consequence of this contact with English would be a lower rate of null subjects. Any effect from these different experiences can be included in the analysis of Ciociaro's pro-drop as extralinguistic variables.

5.2. Coding

In the previous section, I described the different corpora and the participants from each one that are included in my analysis. To test the hypotheses presented in Chapter **4**, verb phrases from the recordings must be identified. These are then coded for the presence or absence of an overt subject pronoun, as well as the independent variables. To annotate the audio files I use ELAN, a free software produced by the Max Planck Institute for Linguistic Analysis (2020).

Once the VPs have been identified and coded, distributions are examined, and multivariate analyses are conducted. In **5.4**, I describe how I use Rbrul within the coding environment of R (R Team, 2020) to perform these analyses. Rbrul (Johnson, 2009) is a free software within R specifically designed for linguistic studies that require logistic or linear regression models.

5.2.1. Coding of HLVC and CHILS

Figure 6. Example of tiers in ELAN used to code variables (HLVC, IXF35A, 12:46).
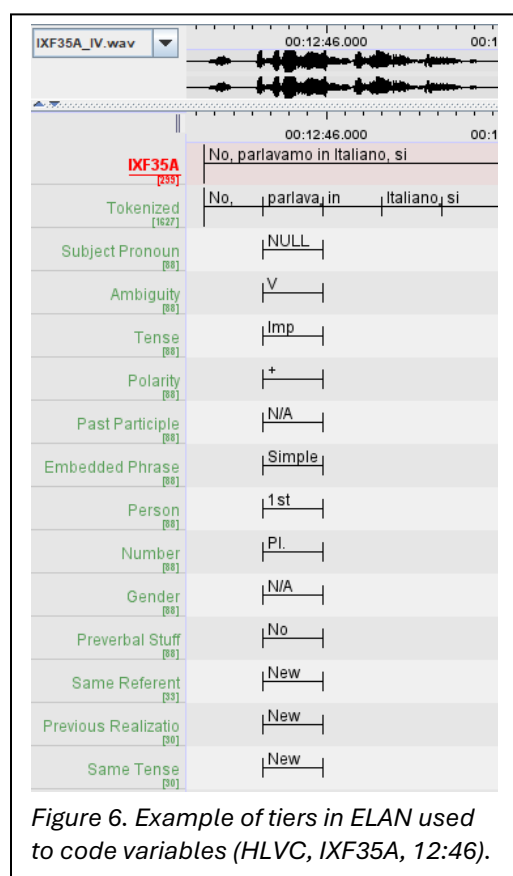
This study uses ELAN (2020) to transcribe and code the audio recordings of the HLVC and CHILS datasets. Once an audio file is opened in ELAN, time-aligned annotations can be made. These annotations can be placed on separate tiers, which can be independent of one another or aligned. A separate tier is made for each of the linguistic variables. Each of the tiers uses a controlled vocabulary to ensure variants are coded consistently across the interviews. Controlled vocabularies restrict coding by requiring a variable to be coded according to predetermined variants. For example, when coding a token for the variable TENSE, a drop-down menu appears, from which one of the variants is selected (e.g., imperfect, absolute past, future, etc.). This consistency removes the possibility of typos, using abbreviations (e.g., IMPERFECT vs. IMP.), and other coding issues. Figure **6** shows an example of the coding (see Appendix **C** for the complete controlled vocabulary).

The first tier allows for a transcription of the verb phrase and is labelled according to the participant's name. The second tier contains each individual word within the verb phrase ('Tokenized'). The inflected verb and the tiers beneath it are aligned to ensure all variables are time-aligned to an individual token. The details of each of the variables, and the variants chosen for each, are explained in **5.3**.

For each of the transcribed recordings, as near to 100 verb phrases (VPs) as possible are coded. I begin coding after the first 10 minutes of the interview to reduce the impact of initial participant nervousness, and preliminary questions also elicit short answers (e.g., 'When were you born?').

Only verb phrases with a human subject are included for each of the HLVC, AIS, and CHILS corpora. This means verbs with non-human subjects are excluded, such as *la scuola sia importante* ('school is important' HLVC, IX35A, 14:26). Further, verbs with an impersonal subject are also excluded. These are phrases where the subject does not have a referent. These include phrases about the weather or phrases like *è vero* ('[it] is true' IX35A; 11:25). Also phrases that have a dummy (expletive) subject pronoun, like *c'è*, *ci sono*, *c'era*, *c'erano* ('there is', 'there are', 'there was', 'there were'), are also excluded because they always have an overt subject pronoun, and so also have no variability in pronoun realization. Finally, constructions with verbs like *mi pare* and *mi piacere* ('it seems to me', 'I like it' [literally 'it pleases me']) were also excluded as these constructions typically do not allow an overt preverbal subject pronoun. As shown in (**14**) and (**15**) these phrases contain postverbal noun phrases and preverbal object pronouns:

14)   *Mi   piacciono   le   varianti*
        OBJ.1SG   like.PRES.3PL   the   variants
        'I like variants'

93

15)   *Mi*   *pare*   *figo*   *questo dialetto*
         OBJ.1SG   seems.PRES.3SG   cool   this   dialect
         'This dialect seems cool to me'

Once the coding is done for 100 VPs (or as close to that number as possible), the file is exported

as a tab-delimited text file (.txt). Each of these text files is an individual participant's tokens.

These files are then imported into an Excel file (Microsoft Corporation, 2021) that combines all

the participants. A separate Excel file is created for each corpus. These spreadsheets have the

tokens arranged in rows, with the tiers' contents arranged in columns. This is the token file that

is used for the Rbrul analysis, as described in **5.4**. The coding process to extract tokens from the

AIS is different than for the coding of the HLVC and CHILS and is described next.

5.2.2. Coding of AIS

As the AIS data is already transcribed, ELAN is not used for that dataset. Instead, all

transcriptions are converted to IPA and coded within Microsoft Excel tables (Microsoft

Corporation, 2021).

Figure 7. Example of an AIS map showing the participants' transcribed response for prompt #649 (Jaberg & Jud, 1928:649).
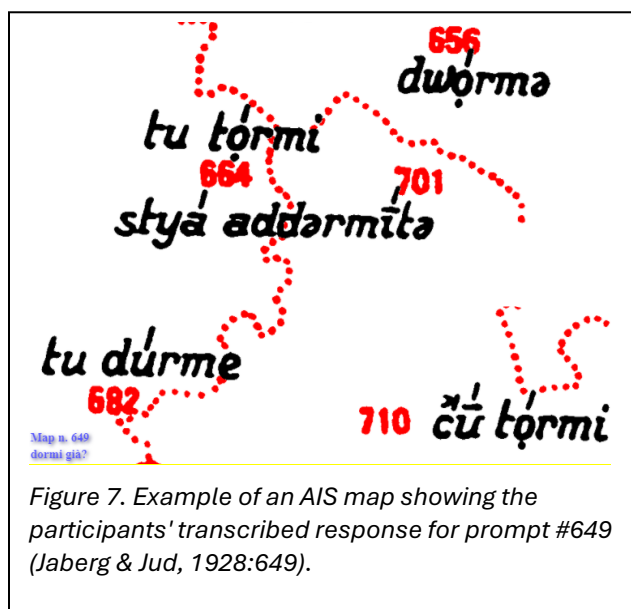
The AIS tokens are extracted from the publicly-available maps online (Tisato, 2009). To do this, I identify the prompts that have a tensed verb. I then refine the list of prompts by removing imperatives, which typically do not take a subject pronoun in Italian varieties (Iannacito, 2000:142). These do not occur often in the HLVC or CHILS recordings because of the narrative structure, but they are

94

a part of the AIS prompts (e.g., prompt #6 is *guarda!* ['look!']). I further refine this list by

removing prompts that had non-human or nominal subjects (*la pecora è in caldo*; 'the sheep is in

heat'; prompt #1070), as well as nonreferential subjects (*è piovuto*; 'it rained'; prompt #367)

because they cannot have an overt subject pronoun.

I then take a screenshot of the relevant region of the map for each of the prompts from the list of

184 prompts with a human subject that met the other criteria. Figure **7** is an example of these

screenshots (Jaberg & Jud, 1928:649).[22] A spreadsheet was created to track the responses

provided by the participants. Each prompt is recorded on a row with the map number, and each

participant's response in a separate column. Using the screenshots, the participants' responses

are transcribed into IPA from the phonetic alphabet using correspondences I created (see

Appendix **F)**. For example, the responses from Figure **7** are transcribed Table **19**. From these

transcriptions we see three participants used an overt subject pronoun (in bold), and two repeated

an equivalent of *già*.

Transcription of AIS prompt #649

Map	Prompt	654	664	701	710	656	682
649	Dormi già?	[ga dormi]	[**tu** tormi]	[stja adːərmitə]	[**tɐu** tormi]	[dwormə]	[**tu** durme]

Table 19. Transcription of responses to prompt dormi già? *('[are] [you] sleeping already?') for the six locations of the AIS dataset.*

The AIS provides narrow transcriptions with more detail than is required for my analysis. For

example, [ó] is an acute accent marking primary stress, and the macron, [ō], is above a vowel

to mark that it is long. Vowel length does not disambiguate in the AIS data and is not consistent

[22] The AIS maps divide Italy in half. This separates Serrone (654) from the other locations, so it is not included in the screenshot.

based on subject. As an example, prompts #359 and #1655 both contain *vieni* ('come.2SG'), and the response is transcribed [ve] or [veː] for four of the six participants—two use the same form in both instances, two switch between long and short. As a result, prosodic stress and vowel length are not pertinent to this study, and they are not included in the transcriptions.

A particular feature of the AIS dataset is that because the participants are asked to repeat disconnected prompts, there is no context through which to resolve ambiguity. This presents a problem when coding a VP with a null subject pronoun and trying to determine if it has a 3SG.M or 3SG.F subject. Many of the prompts do not indicate grammatical gender for 3SG prompts. For example, the prompt #1606 *non ha mai fretta* ('[3SG.M/F] is never in a hurry') contains no information about the gender of the subject of the phrase (Jaberg & Jud, 1928:1606). To this prompt, two of the six participants responded with an added overt 3SG.M pronoun ([isːo], Serrone; [ɣisːə], Scanno). The other four responses, aligning with the prompt, contained no gender agreement. As such they were coded as 3SG.AMBI to indicate that the gender of the subject ambiguous. Because these tokens have a null subject and form their own variant (3SG.AMBI) for SUBJECT, they create an invariably null group that cannot be included in the overall analysis. Within the dataset there are 149 responses across 29 prompts that were coded in this way, and they are not included in the dataset or token counts in this chapter.

After I transcribe the responses from the participants, I code each response for the variables based on the coding schema described in **5.3**.

5.3. Variables

The following sections explain the variables that are coded in this study, as well as which variants are included for each. The following sections describe why each variable is included in

this study. There has never been a variationist study on Ciociaro, so the variables that are included have been significant in comparable research for other languages (see **3.3**). However, some of those factors may not prove significant, or the direction of effect may be different from studies. By casting a wide net in looking at independent variables, the goal is to not miss any important factors in the study of Ciociaro's subject pronoun usage.

Thus, I code the linguistic and extralinguistic variables. The large number of variables being coded means it is possible, and in fact likely, that many will not be selected as significant in the final analysis.

First, I describe the dependent variable, which codes the presence or absence of a subject pronoun (**5.3.1**). Second, the independent variables are coded according to which variant applies to each instance of the dependent variable. These fall into two broad categories: the linguistic variables that are related to the participants' speech during the interview, and the extralinguistic variables related to characteristics of the participants themselves.

The linguistic variables are categorized according to their focus: the variables related to the verb's inflection are described in **5.3.2**, those related to the subject of the VP in **5.3.3**, those related to the phrase's structure in **5.3.4**, and the priming variables describing the preceding VP are described last in **5.3.5**. Unless noted otherwise, these variables apply to the HLVC, AIS, and CHILS corpora. The extralinguistic variables are presented in **5.3.6**. For example, the participants' hometown, age of emigration, and other factors may account for the differing rates of subject pronoun use.

5.3.1. Dependent Variable

Subject realization for the HLVC and CHILS datasets

Code	Description	HLVC - Home	HLVC - Her	CHILS
OVERT	Subject pronoun is realized	151	174	491
Ø	Subject pronoun is null	597	712	1,245
NOUN	Subject pronoun is a common or proper noun	164	109	127
Total	***All tokens that were coded***	***912***	***995***	***1,863***
Without NP	**Tokens without an NP subject included in the analysis**	**748**	**886**	**1,736**

Table 20. The dependent variable and its variants for the HLVC and CHILS datasets.

The dependent variable in this analysis is the presence or absence of a subject pronoun for VPs with a human referent. The controlled vocabulary used to code this variable is shown in Table **20**. The dependent variable of the AIS dataset has a separate set of variants and is described in **5.3.1.1**. For the HLVC and CHILS datasets, each verb phrase is coded as one of these three variants: OVERT for an overt subject pronoun, Ø for when a subject pronoun is not used, and NOUN for when the subject is a noun phrase, including proper nouns and common nouns.

The tokens coded as NOUN are used in the coding of the priming variables (see **5.3.5**). However, as stated above in **5.1.2**, they are not a part of the final count of tokens for each dataset. Only VPs that *can* have an overt subject pronoun are included, and there are no examples, in any of the recordings, where a VP has both an NP and an overt subject pronoun. Therefore, the tokens coded as NOUN in Table **20** are not included in the token count for each corpus and are not included in the description of the variables below.

An example of a token coded as NOUN phrase is shown in (**16**).[23] The phrase that directly follows it, shown in (**17**), has an overt subject pronoun and is coded as OVERT. Finally, (**18**) shows a phrase with a null subject, coded as Ø.

16) [i   majistɹi komɛndʒaɪʌnə   paɹla]
       the teacher   begin.PRET.3PL   speak.INF
       i maestri cominciavano a parlare
       'the teachers began to speak'
       (CHILS, Assunta, 51:41)

17) [**ɛlə**   dɪskʁevevəno]
       they   describe.IMP.3PL
       loro descrivevano
       '**they** described'
       (CHILS, Assunta, 51:44)

18) [dovevam   andaɹ]
       must.IMP.1PL   go.INF
       dovevamo andare
       '(we) had to go'
       (CHILS, Assunta, 52:13)

It was expected that there would be many examples of postverbal subject pronouns, which would be coded as a separate variant. Both postverbal noun phrases, such as shown in (**19**), and postverbal subject pronouns, such as in (**20**) are found in all three corpora. However, they do not occur often, and the few examples of postverbal subject pronouns are included with the preverbal ones.

---

[23] Examples are transcribed on the first line and a gloss on the second. The Italian translation is provided in italics for reference on the third, and an English translation on the fourth. Subject pronouns are bolded.

19) [e   mɔrt   mi marito]
     is.AUX.3SG die.PPRT my husband
     È morto mio marito
     'my husband is dead'
     (HLVC, IXF35A, 4:38)

20) [no   sjam   andat:ə   **noj**]
     NEG is.AUX.1PL   go.PPRT   we
     Non siamo andati noi
     '**we** did not go'
     (HLVC, I1F65A, 8:48)

5.3.1.1. AIS - Dependent Variable

Subject pronoun realization for the AIS dataset

Code	Description	n
ADD	Subject pronoun is null in prompt, but is overt in response	48
PROMPT	Subject pronoun is overt in both the prompt and the response	91
REMOVE	Subject pronoun is in prompt, but is not present in response	26
NULL	Subject pronoun is neither in prompt nor response	418
Total		**583**

Table 21. The dependent variable and its variants for the AIS dataset.

Participants in the AIS corpus were given prompts in standard Italian to translate into Ciociaro.

Their responses are the VPs which are coded and analyzed; however, some of the prompts have

an explicit subject pronoun, while others do not. Unlike the HLVC and CHILS corpora, where

participants spoke freely, the prompts may play a priming effect that must be accounted for in

my analysis. Therefore, each VP is coded as one of four variants: ADD, an overt subject pronoun

that was not in the prompt; PROMPT, the response and prompt both have an explicit subject

pronoun; REMOVE, a null subject that was overt in the prompt; and Ø, both the response and the

prompt had a null subject.

The token count (n) of each variant shows that the participants did not directly translate the prompts. Instead, sometimes the participant's response contains an overt subject pronoun that was null in the prompt. Tokens such as this are coded as ADD, as shown in (**21**). In this case, the participant also did not include *già* ('already') in their response.

21)　　　Response:　　[**tɕu**　tormi]
　　　　　　　　　　　2SG　　sleep.PR.2SG
　　　　　Prompt:　　*dormi già?*
　　　　　　　　　　　'(are) (**you**) sleeping already?'
　　　　　　　　　　　(AIS, Ausonia, prompt #649)

In other instances, as shown in (**22**), both the prompt and the response have an overt subject pronoun, the token is coded as PROMPT.

22)　　　Response:　　[**is**ːo　diʃe]
　　　　　　　　　　　he　　say.PR.3SG
　　　　　Prompt:　　*egli dice*
　　　　　　　　　　　'**he** says'
　　　　　　　　　　　(AIS, Serrone, prompt #384)

Tokens are coded as REMOVE if the response has a null subject pronoun, but the prompt contains an overt pronoun. An example is included in (**23**) where the participant from Ausonia gives a response with a null subject, yet the prompt contains an overt 3SG.M subject pronoun.

23)　　　Response:　　[　　lɛge]
　　　　　　　　　　　Ø　　read.PR.2SG
　　　　　Prompt:　　*egli lecce*
　　　　　　　　　　　'he reads'
　　　　　　　　　　　(AIS, Ausonia, prompt #768)

Finally, most tokens in the dataset are coded as ∅. This means that both the token and the prompt

contain a null subject, as is the case in (**24**). It is interesting to note that a standard pronunciation

in Italian would be /vɔʎːono/ with -/no/ being the standard inflectional suffix for regular verbs

with a 3PL suffix.

24)   Response: [vwotə]
                        ∅   want.PR.3PL
          Prompt: *vogliono*
                        '(they) want'
                        (AIS, San Donato, prompt #835)

It may seem that because the participants in all six locations are being given the same prompts,

their responses are similar. This is not the case. For example, the prompt *vogliono* from (**24**) had

three null subjects and three overt pronouns. The participants from Serrone ([**is**ːi vodo]), Veroli

([**jis**ːi vote]), and Scanno ([**kil**ə vwonə]) all had overt subject pronouns. The participants from

Ausonia ([vɔnːo]) and Scanno ([volo]), as well as from San Donato [vwotə], produced null

subjects.

Subject pronoun realization

Code	Description	HLVC - Home	HLVC - Her	CHILS	AIS
OVERT	Subject pronoun is realized	151	174	491	139
∅	Subject pronoun is null	597	712	1,245	444
Total		**748**	**886**	**1,736**	**583**

Table 22. Variants for the dependant variable.

As shown in this section, the dependent variable for the AIS dataset has additional variants that

do not apply to the HLVC and CHILS datasets. However, to analyse the dependent variable, and

to allow comparability, REMOVE and ∅ were collapsed into ∅, and ADD and PROMPT were

collapsed to OVERT.

Table **22** shows the tokens included for each of the datasets based on a unified dependent variable, a binary factor indicating whether a subject pronoun is overt or null. These are the tokens that are coded for the independent variables. In the following sections I describe the linguistic independent variables, organized by category. Each group of independent variables begins with an example coded according to the relevant variables, followed by a description of each variable and its variants.

5.3.2. Linguistic Variables – Verb

Verb form variables coded for:
[lo pɔrtat a mia sɔrɛl] ('(I) brought it to my sister')

Variable	Code	Description
VERB TENSE	CP	The verb is in the compound past
PAST AUXILIARY	AVERE	The auxiliary is *avere*
FINAL SEGMENT OF VERB	V	[o] is pronounced for *ho*
AMBIGUOUS VERB FORM	N/A	The verb is unambiguously inflected

Table 23. The verbal form subgroup of linguistic variables coded using an example from HLVC, I1M60A, 40:09.

The first group of linguistic variables describes the form the verb takes. There are four factors in this group: the tense of the verb (**5.3.2.1**), auxiliary choice for VPs in the *passato prossimo* (past perfect) (**5.3.2.2**), the final segment of the verb (**5.3.2.3**), and whether the verb form is ambiguous (**5.3.2.4**).

Table **23** provides an example of coding for a phrase from the Calabrian heritage dataset. For TENSE, the verb phrase, [o pɔrtat] (*ho portato*, literally 'have brought'), has both an auxiliary and a participle. For PAST AUXILIARY, [o] (*ho*, literally 'have') is the 1SG conjugation of the auxiliary *avere*. The FINAL SEGMENT of the participle is a consonant, lacking distinction for gender or number, but the auxiliary is [o], which is unambiguous in the paradigm of *avere* for a 1SG

subject. This example, shown in (**25**), has a null subject pronoun. These variables, as well as

their variants, are described in more detail below.

25)　　　　[l　　　o　　　　　　　pɔrtat　　　　a　　　mia　　　sɔrɛl]
　　　　　　DOBJ　have.AUX.1SG　bring.PPRT　to　　　my　　　sister
　　　　　　L'ho portato a mia sorella
　　　　　　'(I) brought it to my sister'
　　　　　　(HLVC, I1M60A, 40:09)

5.3.2.1. Tense

Tense

Code	Description		HLVC - Home	HLVC - Her	AIS	CHILS
AP	Absolute Past		0	(7)[24]	5	353
CP	Compound Past (Verb has an auxiliary and participle)		152	207	77	260
FUT	Future		11	6	42	(2)
IMP	Imperfect		199	296	39	804
PR	Present		355	365	347	319
OTH	Other verb tenses (e.g., subjunctive, etc.)					
	COND	Conditional	17	9	45	
	SUBJ	Subjunctive	7	3	6	(10)
	IMP.SUBJ	Imperfect subjunctive	6		12	
	COND.SUBJ	Conditional subjunctive	1			
	SUBJ.PST	Subjunctive past			6	
	COND.PST	Past conditional			4	
Total			**748**	**886**	**583**	**1,736**

Table 24. Variants and token count for TENSE.

As presented in **3.3.2.1**, I collapse TENSE, ASPECT and MOOD into the variable TENSE. The

participants mostly talk about their past, as can be seen in Table **24**,so the majority of the VPs are

---

[24] All seven of these tokens from the heritage HLVC corpus have a null subject, and so were excluded.

in a past tense, especially the imperfect and compound past (*passato prossimo*). It is also interesting to note that in the HLVC dataset there are almost no tokens in the absolute past (*passato remoto*). This is supported by Ledgeway, who describes the regional Calabrian of the north of the region as having the compound past as "the sole perfective past paradigm" (2010:103). The CHILS dataset is relatively balanced between the absolute and compound past tenses. For both the HLVC and the CHILS datasets, most of their tokens in the present and past tenses. The exception to this pattern is the AIS corpus and its question-and-response format, which produces many more examples of verbs in the subjunctive, conditional, and future than occur during the interviews of either the CHILS or HLVC recordings. Example (**26**) shows the different forms the verb *fare* ('to do/make') takes in each of these tenses for standard Italian. While both Calabrian and especially Ciociaro would see pronunciation differences from these standard forms, the structure of the tenses is the same.

26)
Example	**Tense**
(io) faccio il pane	Present
(io) facevo il pane	Imperfect
(io) ho fatto il pane	Compound past
(io) feci il pane	Absolute past
che (io) faccia il pane	Subjunctive
(io) farei il pane	Conditional
(io) farò il pane	Future
'(I) make bread'	

Some research has found that the tense of a VP is a significant variable in subject pronoun realization, but other studies have not. Research on pro-drop in Faetar found verbs in the past-tense are less likely to be null, but the effect was small (Nagy et al., 2018). In Spanish varieties the imperfect, in particular, has been shown in some studies to disfavour null subject pronouns

(Otheguy et al., 2007; Travis, 2007). An explanation for this, related to the functional hypothesis, is that those tenses have more ambiguous verb forms, and thus necessitate subject pronouns for disambiguation. For example, *trabajaba* ('I worked' or 's/he worked') is the form for both first and 3rd person singular imperfect indicative in Spanish (Cameron, 1993).

Previous research of Calabrian Italian using the HLVC corpus also report this effect, whereby the compound past was the most likely to have a null subject and the imperfect past was the least likely (Nagy et al., 2011). This may be explained by the compound past having both an inflected auxiliary verb and a past participle, which is inflected for gender and number. This difference between imperfect disfavouring and compound past favouring null subjects is especially relevant for the HLVC and CHILS corpora, as they make up the majority of the tokens for both datasets.

5.3.2.2. Past Auxiliary

Auxiliary for compound past VPs

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
AVERE	*Avere* is the past participle	81	120	31	43
ESSERE	*Essere* is the past participle	71	87	50	176
STARE	*Stare* is the past participle	0	0	0	41
N/A	The VP does not have a past participle	596	679	502	1,476
Total		**748**	**886**	**583**	**1,736**

Table 25. Variants and token count for AUXILIARY.

As discussed in the section above on verb tenses, COMPOUND PAST tokens may behave differently from other tenses in how they constrain subject pronoun realization. The forms of the auxiliaries are often distinct enough that the ambiguity that may arise for the IMPERFECT does not arise. Therefore, the TENSE variable accounts for any interaction between auxiliary verbs and subject pronoun realization. However, it is important to examine which auxiliary is being used for tokens

coded as COMPOUND PAST. Of the two primary auxiliary verbs that account for most of the tokens in Table **25**, *avere* has distinct forms across the paradigm, but *essere* does not. The 1SG and 3PL inflections of *essere* are both *sono* ('I am' and 'they are'). Therefore, within the tokens that are in the compound past, we should expect those with a form of *avere* to favour null subjects more than those with *essere*.

Auxiliary choice in Italian is dictated by the grammar of the language, with verbs like *venire* ('to come') taking the auxiliary *essere* and verbs like *vedere* ('to see') taking *avere*, together creating a phrase like *sono venuto e ho visto* ('[I] came and [I] saw'). However, in Ciociaro in many cases where *avere* is the expected auxiliary in Italian, *essere* is used. For example, from the AIS data, prompt #1537 is **hai** *cucito bene* ('[you] sewed/knit well') with *avere* as the auxiliary in the prompt, but the Veroli participant responds with [**se** kuʃito bɛne] with *essere* taking the place of *avere* (Jaberg & Jud, 1928:1537). This is also found throughout the CHILS dataset, which uses *essere* much more often than *avere*. In example (**27**) we see the verb *vedere* in the compound past with *sono* used as the auxiliary form for a 1SG subject. The dialetto of Villa San Michele, which borders Ciociaria, also uses *essere* in almost all cases where *avere* is expected in standard Italian (Iannacito, 2000:178).

> 27) [lo   son   wistə]
>   DOBJ AUX.1SG see.PPRT.1SG
>   *L'ho visto*
>   '(I) did see it'
>   (CHILS, Nello, 40:57)

It is also possible that auxiliary selection can be related to subject. Legendre (2010) finds that for dialetti, 1st and 2nd person subjects favour *essere* and 3rd person subjects favour *avere*. This is

also described by Casalicchio and Cordin for a northeastern dialetto (2020:268). By coding this variable, the nuances of auxiliary selection can be better understood as they may relate to subject pronoun realization.

Finally, although it only occurs in the CHILS dataset, the auxiliary *stare* can be used in standard Italian with a present participle for the past progressive. In Italian, this verb, like *avere*, does not have any ambiguity in its paradigm. An example is provided in (**28**).

28)   [**nu**   stavam   pagand   la   ɹɛnt]
        we   AUX.1PL   pay.PPR   the rent
        Noi stavamo pagando l'affitto
        '**We** were paying the rent'
        (CHILS, Nello, 16:25)

5.3.2.3. Final segment of verb

Final segment of verb

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
[ə]	Final vowel of verb is realized as [ə]	40	55	139	430
ø	The conjugated verb ends in a consonant	33	42	90	568
V	The conjugated verb ends in the same final vowel as the equivalent in standard Italian	675	789	354	738
Total		**748**	**886**	**583**	**1,736**

Table 26. Variants and token count for FINAL SEGMENT.

One of the central questions of this dissertation is whether the syncretic verbal paradigms of Ciociaro are related to a lower null-subject rate than in Calabrian. To test this question, I include two different variables related to ambiguity. The first is coded according to the final segment of the verb. In standard Italian, the final segment is a vowel that typically disambiguates the subject. This is also the case in Calabrian Italian. Table **26** shows that almost all the HLVC tokens end in a vowel that disambiguates the subject.

However, in Ciociaro the final vowel is often a /ə/ or is absent, ending the verb in a consonant. Example (**29**) shows a verb that is produced with the expected final vowel, which is coded as 'V'. The phrase that follows it, shown in (**30**), is in the imperfect and instead of a final /o/ that would be expected in Italian for a 1SG subject, a [ə] is produced. A final /ə/ can occur for any of the subjects, as shown in (**31**) where a 3SG subject ends in [ə] instead of the expected [a] in Italian.

29)   [o   fat:o   ɪ   manuale]
        AUX.1SG   do.PPRT.1SG   the manual(labour)
        ho fatto il (lavoro) manuale
        '(I) did manual labour'
        (CHILS, Ennio, 12:37)

30)   [faʧevə   i   matoni]
        do.IMPF.1SG   the bricks
        facevo mattoni
        '(I) made bricks'
        (CHILS, Ennio, 12:39)

31)   [abitavə   viʃin a   nu   kaza]
        live.IMPF.3SG   near   to our house
        Abitava vicino a casa nostra
        '(she) lived near our house'
        (CHILS, Ennio, 15:06)

In examples (**30**) and (**31**) the subject is ambiguous, as the inflection is identical for both Ennio describing his work (1SG), and where his wife lived (3SG). However, this variable does not necessarily equate to ambiguity. If a final vowel is realized as [ə] or is not realized at all, there can still be segments which disambiguate a subject. Specifically, the plural subjects typically contain a penultimate consonant that distinguishes person of the subject (see **3.2.2**). For regular

verbs, the plural subjects are inflected with +/Vm(V)/ for 1PL, +/Vt(V)/ for 2PL, and +/Vn(V)/ for 3PL. This applies to all tenses. For example, (**32**) shows a verb in the imperfect that is coded as ø because it has no final vowel. However, the 1PL is not ambiguous because of the final [m].

32)   [doveva**m**   andaɪ]
         must.IMPF.1PL go.INF
         Dovevamo andare
         '(we) had to go'
         (CHILS, Ennio, 11:39)

5.3.2.4. Ambiguity

Ambiguous

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
1SG/2SG/3SG	Verb has ambiguous form shared with at least the 1SG, 2SG, and 3SG subjects	26	22	91	324
1SG/3SG	Verb has ambiguous form shared for verb with 1SG and 3SG subjects	3	0	22	119
SONO	Verb is a form of "sono", either in PRESENT or an auxiliary where the participle lacks disambiguating agreement	28	33	5	143
N/A	Subject is clear from the verb form	691	831	465	1,150
Total		**748**	**886**	**583**	**1,736**

Table 27. Variants and token count for AMBIGUOUS.

Ciociaro's "uninflectedness" (Calabrese, 2011) increases the ambiguity of the subject inflected on the verb. According to the functional hypothesis, this should cause speakers to increase their rate of usage of subject pronouns to disambiguate the subject of VP when compared to Calabrian. Table **27** shows how rarely the subject is ambiguous in Calabrian, with most of the tokens unambiguous and coded as N/A. Most of the ambiguity in Calabrian comes from variants coded as SONO. These tokens are coded as ambiguous only when neither any past participle, nor

any adjective modifying the subject is inflected for gender and number that agree with the subject. For example, in (**33**) while *sono* could refer to either a 1SG or 3PL subject, the final [o] on the participle makes clear the subject is 1SG, and thus this token is not ambiguous and is coded as N/A. In (**34**), however, the participle does not clarify the subject, and is coded as SONO.

33)   [sɔn   andat**o**]
        be.AUX   go.PPRT.1SG
        sono andato
        '(I) went'
        (HLVC, IXM35A, 20:42)

34)   [sɔn   arivat]
         be.AUX   go.PPRT
        sono arrivato
        '(I) arrived'
        (HLVC, IXM52A, 21:50)

Another possibility for ambiguity is a levelling of 1SG and 3SG subjects, which is coded as 1SG/3SG. In Ciociaro many verbs, particularly in the imperfect, have identical forms for 1SG and 3SG subjects. For example, in Italian the paradigm for the verb *essere* in the imperfect has three distinct forms for singular subjects: /ɛro/ for 1SG, /ɛri/ for 2SG, and /ɛra/ for 3SG. In the AIS though, the participant from Serrone produces [ɛra], [ɛri], and [ɛra] for these same subjects (Jaberg & Jud, 1928:1690). Thus, while 2SG is unambiguously inflected, the other two forms are identical. This is also seen in examples (**30**) and (**31**), where instead of /a/, [ə] is the final segment for both 1SG and 3SG subjects.

These two variants SONO and 1SG/3SG are for ambiguity between two persons. The final variant, 1SG/2SG/3SG is for VPs that are ambiguous for more than two persons. This is especially frequent in the CHILS dataset. This is because in many cases, even VPs with a plural subject are

111

ambiguous for subject if the final disambiguating consonant is not produced. For example, in the previous section (**32**) is coded as N/A for this variable because the 1PL subject is clear due to the final [m]. However, in (**35**) we see a verb with a 1PL subject that has neither a final vowel nor [m]. In this case, not just the final vowel is reduced, but the entire syllable /mV/ is absent:

    35)   [non   kɔnoʃevə   ɪl   late]
                   NEG   know.IMPF.1PL the   milk
                   non conoscevamo il latte
                   '(we) didn't know milk'[25]
                   (Clara, 12:54)

In (**35**), Clara is discussing the cultural shock of arriving in Canada and how different the food was. The lack of inflection on the verb means that the subject cannot be understood from the inflection alone. However, the phrase that immediately precedes (**35**) is "at first, it was just me, my dad, and my brother", which makes it clear that Clara is saying that the three of them did not know about milk as a drink. According to the functional hypothesis, regardless of context, phrases like this should disfavour null subjects because the verb is ambiguously inflected. In research that included AMBIGUOUS as a variable, the results have been mixed. In Nagy et al.'s analysis of Calabrian Italian it was not significant (2011). However, it has been significant in analyses of pro-drop in some Spanish varieties (Lastra & Butragueño, 2015; Limerick, 2019; Padilla, 2020).

---

[25] Clara is discussing how her family had to adjust their eating habits when they first arrived in Canada. They did not drink milk in Italy because it was not pasteurized. Her family did not "know milk" as a drink. While it is possible Clara is just talking about herself, the preceding and following phrases point to a 1PL subject.

5.3.3. Linguistic Variables – Subject

Subject variables coded for:
[isə e mɛnudə sulə] ('He came alone')

Variable	Code	Description
PERSON	3RD	The VP has a 3rd person subject
NUMBER	SING	The VP has a singular subject
GRAMMATICAL GENDER OF 3RD PERSON SUBJECT	MASC	The 3rd person subject is masculine
SUBJECT	3SG.M	The subject is 3rd person singular masculine.

Table 28. The subject subgroup of linguistic variables coded using an example (CHILS, Nello, 15:27).

The next group of variables are those that relate to the subject of the VP. There are four variables in this group: PERSON (**5.3.3.1**), NUMBER (**5.3.3.2**), GENDER for 3rd person subjects (**5.3.3.3**), and SUBJECT (**5.3.3.4**), which combines the three. For example, the phrase in (**36**) is coded as 3rd, SINGULAR, and MASCULINE, meaning it has a 3SG.M subject referent. The phrase also has the overt subject pronoun [isə].

36)   [**isə**   e   mɛnudə   sulə]
        he   is.AUX.3SG   come.PPRT.SG   alone
        lui e venuto solo
        '**He** came alone'
        (CHILS, Nello, 15:27)

The variable SUBJECT is significant in each of the studies of pro-drop that are discussed in (**3.3.1.1**). I hypothesize that it will also be a significant variable for my analyses. But, to better understand how SUBJECT interacts with subject pronoun realization, I code for each of the separate aspects of it. By analysing them separately and together, I can better analyse and interpret both significance and direction of effect for the various aspects of SUBJECT. This is described in detail in the following sections.

5.3.3.1. Person

Person

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
1ST	The verb's subject is in the 1st person	406	548	293	711
2ND	The verb's subject is in the 2nd person	65	46	103	51
3RD	The verb's subject is in 3rd person	277	292	187	974
Total		**748**	**886**	**583**	**1,736**

Table 29. Variants and token count for PERSON.

Grammatical person is coded separately from number to examine each aspect of the subject separately. From Table **29**, few tokens take a 2nd person subject, especially the HLVC and CHILS datasets. This is expected due to the narrative structures of these interviews, where the participant is talking about themselves, as well as their family and friends. For PERSON, the subject of the VP is collapsed into three variants: 1SG & 1PL (**37**), 2SG & 2PL (**38**), and 3SG & 3PL (**39**):

37) [faʧamə   lə   makaɹun]
     do.IMP.1PL   the macaroni
     Facevamo i maccheroni
     '(we) made macaroni'
     (CHILS, Nunciata, 27:32)

38) [lə   makaɹun   non   avɛvat   ijɛk]
     the   macaroni   NEG   have.IMP.2PL here
     I maccheroni, non avevate qui
     'the macaroni, (you [2PL]) didn't have it here'
     (CHILS, Nunciata, 27:48)

39) [tənevan   lə   makaɹun]
     do.imp.3PL   the macaroni
     Tenevano i maccheroni
     '(they) had macaroni'
     (CHILS, Nunciata, 27:55)

114

Studies on Spanish (Carvalho & Child, 2011) and on Calabrian (Nagy et al., 2011) find that

PERSON is not a significant variable. However, it is significant in homeland and heritage Faetar.

Nagy et al. (2018) report 3rd person being the most likely to be Ø and 2nd person the least likely.

I expect 3rd person subjects will have a higher null rate than 1st or 2nd person subjects in both

Calabrian and Ciociaro. This is because any VP with a 1st or 2nd person subject has either an

overt or null subject pronoun, but for 3rd person subjects, an NP is also possible. That being said,

I expect person will not be significant, and instead subject will be, which combines person and

number (see **5.3.3.4**)   Tokens with NP subjects are not included in my analysis, leaving a higher

proportion of null subjects.

5.3.3.2. Number

Number

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
SING	Subject of the verb is singular	512	564	419	1,148
PL	Subject of the verb is plural	236	322	164	588
Total		**748**	**886**	**583**	**1,736**

Table 30. Variants and token count for NUMBER.

NUMBER has two variants: SING, combining 1SG, 2SG, and 3SG; and PL which combines 1PL, 2PL,

and 3PL. In each of the datasets, singular subjects are more than twice as common as plural.

NUMBER is a significant variable in Calabrian, with plural subjects being more likely to have null

subjects in both (Nagy et al., 2011). In Spanish varieties, NUMBER is not typically included as a

variable. However, SUBJECT is significant in all the studies included in **3.3.1**, and plural subjects

favour null subjects while singular subjects disfavour them (Abreu, 2009; Cameron, 1993;

Hernández Constantin, 2021).

I expect my analyses of Calabrian and Ciociaro will find the same. As described in **5.3.2.4**, verbs with a singular subject are often identical except for the final vowel, while plural subjects almost always have more distinct inflections. Because of these clear differences in the plural subjects, I believe overt pronouns will be more likely to occur in singular subject than plural ones.

5.3.3.3. Grammatical gender of 3rd person subject

Grammatical gender of 3rd person subject

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
MASC	The VP has a masculine 3rd person subject	217	237	149	838
FEM	The VP has a feminine 3rd person subject	63	55	38	136
N/A	The VP has a 1st or 2nd person subject	468	594	396	762
Total		**748**	**886**	**583**	**1,736**

Table 31. Variants and token count for GENDER.

Verb phrases that have 3rd person subjects are coded as either MASC or FEM. Most of the tokens included in each dataset have a masculine subject, with around 20% being feminine.

Unfortunately, none of the Spanish studies I included in **3.3** report 3rd person subjects divided by grammatical gender, so it is not possible to see if there is a difference in these Spanish varieties. In Faetar, 3rd person favours null subjects; however, in heritage Faetar MASC favour null subjects more than FEM (Nagy et al., 2018).

Heap used the AIS and ALF[26] to examine subject pronouns in northern Italo-Romance dialetti and Occitan varieties (Heap, 1997). In his analysis of grammatical gender, he found that masculine subjects also favour null subjects whereas feminine subjects disfavour. He

---

[26] The *Atlas linguistique de la France* is a similar project to the AIS, conducted by Gilliéron and Edmont across France with the atlases published from 1902-1910.

hypothesizes that this is due to MASC being "unmarked", and it is therefore assumed when a 3rd

person subject is ambiguous for gender. As a result, an overt subject pronoun disambiguates FEM

subjects from the default MASC. In my datasets, there are many more MASC tokens than FEM. As a

result, I expect this will also be true for both Calabrian and Ciociaro, with FEM having a lower

rate of null subjects than MASC.

5.3.3.4. Subject of VP

Subject of VP

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
1SG	The subject is 1st person and singular	302	356	240	518
2SG	The subject is 2nd person and singular	65	46	59	61
3SG.M	The subject is 3rd person singular and masculine	108	118	82	464
3SG.F	The subject is 3rd person singular and feminine	47	44	38	116
1PL	The subject is 1st person and plural	104	192	53	195
2PL	The subject is 2nd person and plural	(3)	(3)	44	(10)
3PL.M	The subject is 3rd person plural and masculine	106	119	67	374
3PL.F	The subject is 3rd person plural and feminine	16	11	0	20
Total		**748**	**886**	**583**	**1,736**

Table 32. Variants and token count for SUBJECT.

The previous three variables have shown the aspects of a VP's subject, but SUBJECT combines

them. Table **32** shows the distribution of the tokens for each of the datasets. Most of the tokens

fall into three variants: 1SG, 3SG.M, and 3PL.M. Also, 2PL and 3PL.F tokens are particularly

infrequent. However, the distribution of subjects is much more balanced for the AIS dataset

because participants were given prompts instead of the sociolinguistic interviews of the HLVC

and CHILS corpora.

This is the factor with the biggest effect in all the comparable studies, except for two where it has

the second biggest effect. As I described in the previous sections, plural subjects favour null

subjects and singular subjects disfavour them.

117

Pulling together the variables in this section, I expect that 3rd person subjects will favour null subjects more than the other PERSON variants, that plural subjects will favour null subjects over singular subjects, and that 3rd person masculine subjects will favour null subjects over 3rd person feminine subjects.

5.3.4. Linguistic Variables - Phrase

Phrase variables coded for:
[ma i no lə maŋgə] ('but he did not miss it')

Variable	Code	Description
POLARITY	-	The phrase has a negative particle
PREVERBAL ELEMENT	OBJ	The phrase has a preverbal direct object
CLAUSE TYPE	SIMPLE	The phrase is not a matrix or embedded clause
VERB	MANCARE	'to miss'
VERB CLASS	COGNITION	Within the psychological/cognition class

Table 33. The subgroup related to phrase structure, coded using an example (CHILS, Antonia, 14:28).

This group of variables are coded according to the phrase which contains the inflected verb. There are five variables in this group: the polarity of the phrase (**5.3.4.1**), any PREVERBAL ELEMENT (**5.3.4.2**), PHRASE TYPE (**5.3.4.3**), VERB (**5.3.4.4**), and VERB CLASS (**5.3.4.5**). For example, the phrase in (**40**) is negative, has a preverbal object, and is a simple clause. The verb of the phrase is *mancare* ('to miss'), which is part of the COGNITION class.

40)   [ma   **i**   no   lə   maŋgə]
        but   he   NEG   DOBJ   miss.PRET.3SG
        ma lui non lo mancò
        'But **he** did not miss it'
        (CHILS, Antonia, 14:28)

The variables in this group are described in more detail below.

118

5.3.4.1. Polarity

Polarity of VP

Code	Description	HLVC - Home	HLVC - Her	AIS	CHILS
+	The VP is in the affirmative	629	738	530	1,469
-	The VP is in the negative	119	148	53	267
Total		**748**	**886**	**583**	**1,736**

Table 34. Variants and token count for POLARITY.

Polarity is coded to account for any effect a preverbal negative marker has on the realization of a subject pronoun. In Calabrian and Ciociaro the negative particle occurs preverbally and typically takes the form /no(n)/. Approximately 15% of the tokens in my datasets are negative (10% for AIS).

In Spanish varieties, negative polarity is a significant variable in some studies, although the effect is usually not very large (Abreu, 2009; Hernández Constantin, 2021; Limerick, 2019). The direction of effect is that phrases with a negative particle favour null subjects. Further, previous research on heritage Calabrian did not find a significant effect (Nagy et al., 2011).

Heap's analysis of Gallo-Romance varieties offers insight into why a preverbal negative particle favours null subjects. In Occitan varieties and Italo-Romance dialetti negation is marked preverbally, postverbally, or with a particle before and after. In his analysis, negative phrases favour null subjects if there is a preverbal element, but there is no effect if only a postverbal element is present (Heap, 1997:209). He proposes that this is because a preverbal element, in this case a negation marker, fills the same position that the subject pronoun would normally occupy, 'blocking' the realization of an overt subject pronoun.

An example of the preverbal subject position being 'occupied' by a negative marker is shown in
(**41**), which is also shown in **5.3.1**. In this case, from a heritage Calabrian speaker, the subject
pronoun is overt, but it is postverbal.

41)   [no   sjam   andat:ə   **noj**]
        NEG   is.AUX.1PL   go.PPRT   we
        Non siamo andati noi
        '**we** did not go'
        (HLVC, I1F65A, 8:48)

5.3.4.2. Preverbal Elements

Preverbal element

Codes	Description	HLVC - Home	HLVC - Her	AIS	CHILS
OBJ	Preverbal object	100	115	111	141
REFL	Preverbal reflexive pronoun (affirmative or negative)	68	88	46	76
NEG	negative preverbal element	83	92	47	112
NEG + OBJ	both a negative particle and an object pronoun	31	48	6	82
N/A	No clitics or objects	466	543	373	1,325
Total		**748**	**886**	**583**	**1,736**

Table 35. Variants and token count for PREVERBAL.

Preverbal elements are coded according to their category. The variable POLARITY codes only
whether a VP has a negative particle or not. The PREVERBAL ELEMENTS variable codes for
different types of preverbal elements, separating negative phrases into other variants if they have
another preverbal element. This will show if certain types of preverbal elements favour null
subjects over others. It is important to note that because this variable and POLARITY both code
phrases that have negative preverbal elements, they cannot be both included in the same
multivariate analysis due to collinearity (the same tokens will be grouped together in the same
variables). Table **35** shows the distribution of these preverbal categories in each dataset. Roughly

35% of the HLVC and AIS tokens have at least one preverbal element, while they are a bit less frequent in the CHILS dataset (24%). The variants include: PREVERBAL OBJECTS, which do not agree with the subject; REFLEXIVE PRONOUNS that do agree with the subject, which may have a negative particle as well or not; NEGATIVE PARTICLES; and phrases that have both a negative particle and an object pronoun. In both Calabrian and Ciociaro these elements occur in a regularized order, with a negative particle occurring before the object or reflexive pronoun, followed by the verb. This is shown in (**42**), which has a direct object, and in (**43**), which contains both a negative particle and an object:

42)   [lə   kɔmpɹavə   a la lablawz]
         DOBJ buy.IMP.3SG   at the Loblaws
         Li comprava a la Loblaws
         '(He) bought them at the Loblaws'
         (CHILS, Arcangela, 11:57)

43)   [no   l   o   saʧ]
         NEG   DOBJ   AUX.1SG know.PPRT
         Non   l'ho saputo
         '(I) did not know that'
         (CHILS, Arcangela, 12:20)

There are examples of reflexive pronouns being coded, with two of the four studies in **3.3.2.2** that included it as a variable finding it significant (Abreu, 2009; Carvalho & Child, 2011). The presence of reflexive pronouns favours null subjects. Under the functional hypothesis, I suggest this is because the reflexive pronoun already agrees with the subject, rendering an overt subject pronoun redundant.

The effect of preverbal objects does not appear to be a frequently investigated variable in pro-drop research for Spanish varieties. However, for Calabrian, Nagy et al. found a significant effect

with the presence of a preverbal object favouring a null subject (2011). In Faetar, preverbal

objects favour null subjects in the homeland variety, but not in the heritage variety (Nagy et al.,

2018). Heap finds that, contrary to the Spanish studies mentioned above, reflexive pronouns do

not favour null subjects in Gallo-Romance varieties (Heap, 1997:210). However, he does find

that both objects and negative particles favour null subjects. Interestingly, he finds that VPs with

more than one preverbal element have a stronger effect than those that just have a single object

or negative particle (p.183). This supports the hypothesis that these preverbal elements occupy

the space, blocking subject pronoun realization.

In support of this idea, as shown in phrase (**44**) for the previous variable, preverbal objects also

co-occur with postverbal subject pronouns:[27]

44)   [o i dʒokatəli li   faʧevamə   noj]
        Oh the toys   DOBJ make.IMP.1PL   we
        Ah i giocattoli, li facevamo noi
        'oh the toys, **we** made them'
        (HLVC, I1F61A, 31:31)

5.3.4.3. Clause Type

Clause type

Code	Description	HLVC - Home	HLVC - Her	CHILS
EMBED	The VP is in the subordinate clause	67	49	71
MATRIX	The VP is in the matrix clause	56	34	64
SIMPLE	Simple phrase	625	803	1,601
Total		**748**	**886**	**1,736**

Table 36. Variants and token count for CLAUSE.

---

[27] for a more thorough discussion of postverbal subjects in Italian languages, see Cardinaletti, 2018.

This variable codes VPs according to their clause type. This variable cannot be coded for the AIS dataset because the participant is translating prompts with simple clause structure, rather than producing narrative speech, so there are no complex clauses. Table **36** shows that most of the tokens in the HLVC and CHILS datasets are simple clauses. (**45**) is an example of the other two variants, a matrix clause ('he had an idea'), and an embedded clause ('of what he could do').

45)   [**isə**   tənevə   ʌn   ideja kə   poteva   fa]
                   he   take.IMPF.3SG DET   idea CP   can.IMPF.3SG do.INF
                   lui teneva un'idea de che poteva fare
                   '**He** had an idea of what (he) could do'
                   (CHILS, Antonia, 16:22)

Otheguy et al. found a significant effect for this variable in New York Spanish, with embedded clauses favouring null subjects (2007). However, it has also been found to not be significant in other Spanish varieties (Abreu, 2009; Carvalho & Child, 2011; Hernández Constantin, 2021). It was not reported as a variable in studies of Faetar or heritage Calabrian.

While coding for the variable may still yield interesting results, there is a possibility there may be interactions between this variable and others. For instance, embedded phrases may occur more often with the same subject as the matrix phrase, as in (**45**), leading to an interaction between this variable and SWITCH REFERENCE.

5.3.4.4. Frequent Verbs

This variable and the following are only coded for the CHILS analysis. To code this factor, I group VPs according to the infinitive form. Table **37** shows the 15 most common verbs in the CHILS analysis, which account for 56% of the tokens in the dataset. For VPs in the compound past tense, the participle is coded, not the auxiliary (e.g., *sono arrivato* is coded for *arrivare* and not *essere*). It is important to note that for *avere* and *essere* these tokens do not include those instances where these verbs are used as auxiliaries because they do not carry semantic meaning.

Most common verbs of CHILS dataset	
Verb	*n*
fare	137
dire	92
parlare	87
venire	86
andare	71
sapere	70
essere	67
tenere	63
arrivare	62
volere	59
lavorare	56
vedere	51
abitare	45
avere	43
pensare	36
total	**974**

Table 37. *The 15 most common verbs in the CHILS dataset.*

Most studies of pro-drop organize verbs into semantic classes. This is addressed in the following section. However, it is also possible that especially frequent verbs condition subject pronoun expression differently than less-frequent verbs. Frequency effects of verbs on pro-drop has been examined in two Spanish varieties by Erker & Guy (2012). While they did not find that it was a significant factor, they did find that factors that are significant interact with frequency effects. For example, *sapere* ('to know') likely has most of its occurrences with a 1SG subject, with the participant describing what they do (or did) know. Therefore, we can see if these tokens affect subject pronoun realization differently than other 1SG subjects. I do not expect this variable to be significant in my analysis of the CHILS dataset, but nuances in the significant factors may be revealed by examining frequent verbs.

5.3.4.5. Verb class

In order to account for the verbal semantics of verbs, many studies include a variable for verb classes (Abreu, 2009; Posio, 2011; Travis, 2007). The number of verb classes, and which verbs are in which class, varies from study to study, but I use the five groups proposed by Travis (2007). In Table **38** I organize the 15 most frequent verbs in the CHILS dataset by verb class.

Verb classes for the most frequent verbs

COGNITION	SPEECH	COPULA	MOTION	OTHER
sapere	dire	essere	venire	fare
volere	parlare	avere	andare	lavorare
pensare			arrivare	abitare
vedere				tenere

Table 38. The 15 most common verbs in the CHILS dataset, organized by verb class.

The first group, COGNITION, contains verbs that express thought, feelings, or emotion. This class tends to most favour overt subject pronoun (Carvalho & Child, 2011; Torres Cacoullos & Travis, 2019). Verbs grouped in SPEECH occur frequently in most datasets, and sometimes favour and sometimes disfavour null subjects (Limerick, 2019; Travis, 2007). A potential reason for why these classes of verbs favour overt subject pronouns is because the participant is centring themselves in the narrative. For example, in (**46**), Arcangela uses an overt subject pronoun with the verb *parlare* when describing how, when she worked at a restaurant, most of her coworkers chatted in English, but her and her friend, '*we* talked in Italian'.

46) [**nu**   paɹləvam   ən   ɪtalijən]
      we   speak.IMP.1PL   in   Italian
      noi parlavamo italiano
      '**we** talked in Italian'
      (CHILS, Arcangela, 11:35)

This is a complication in including VERB CLASS and other variables that are collinear with it in a single analysis. COGNITION-type verbs are much more common with 1SG (or 1PL) than 3SG or 3PL subjects because it is easier to know (and more common to talk about) our own thoughts and wants than those of others. There are also what Torres Cacoullos & Travis (2019) call "lexically particular constructions". An example from both Calabrian and Ciociaro is *no lo so* ('I don't know'). As a common construction, *sapere* co-occurs with both 1SG subjects and PREVERBAL ELEMENTS. To avoid interactions, VERB CLASS must be run in a separate analysis from these variables.

The third group are COPULA verbs, which occur at a high rate and have unique properties as auxiliary verbs compared to the other groups. These verbs tend to slightly favour overt pronouns. The fourth group, MOTION, expresses actions or movement. Verbs like 'go', 'come', and 'leave' tend to least favour overt pronouns (Carvalho & Child, 2011; Limerick, 2019). Finally, a general 'other' group is used for all the verbs that don't fit neatly into any of the previous groups.

As with the variable FREQUENT VERBS, I do not expect verb class to be a significant variable in the CHILS analysis. However, by examining VERB CLASS alongside the significant variables, a deeper understanding of pro-drop in the CHILS data can be gained.

5.3.5. Linguistic variables - Priming Variables

Priming variables coded for:
[kwana papa ɛntɹav la kaza isːə salutava sɛmbɹə]
('When dad came into the house, he always greeted [us]')

Variable	Code	Description
PREVIOUS REALIZATION	NOUN	Overt subject pronoun in preceding VP
SWITCH REFERENCE	SAME	The subject of the verb is the same as the preceding VP

SWITCH TENSE	SAME	The verb is in the same tense as the preceding VP

Table 39. The priming subgroup of linguistic variables coded using an example (HLVC, IXM61A, 18:36).

The final group of linguistic variables is the priming variables. Priming effects test the effect the preceding phrase may have on the VP under consideration. Priming effects are well-established for their effect on subject pronoun realization, and at least one priming variable is significant in each of the studies described in **3.3**.

In this study I include three priming variables for the HLVC and CHILS corpora and one for the AIS. For the HLVC and CHILS these are: previous realization (**5.3.5.1**), meaning whether the subject pronoun in the preceding VP was overt or null; switch reference (**5.3.5.2**), whether the previous VP was coreferential with the VP under consider consideration; and switch tense (**5.3.5.3**), whether the previous VP was coded as having the same tense (see **5.3.2.1**).

The phrase in (**47**) shows two inflected verbs.[28] The second VP, [is:ə salutava], may be primed by the first, [papa ɛntɹav]. Thus, for the second VP, where the previous realization is the NP subject [papa], the referent is the same, and the tense is imperfect in both VPs.

47)   [kwana   papa ɛntɹav   la kaza **is**:ə salutava   sɛmbɹə]
                When   dad   enter.IMP.3SG the house he   greet.IMP.3SG always
                Quando papa entrava la casa, lui salutava sempre
                When dad came into the house, **he** greeted [us] always
                (CHILS, Antonietta, 11:51)

---

[28] The first VP is not included in the dataset as a dependent variable token because it has a NP subject (see **5.3.1**)

The AIS data cannot be coded for these variables because the participant is only producing one phrase at a time. However, as described in (**5.3.1.1**), the prompts the participants are given to translate may contain an overt subject pronoun or not, and this may prime them to reproduce that. As such, the presence or absence of a subject pronoun in the prompt is coded as a priming variable for the AIS dataset (see **5.3.5.4**).

5.3.5.1. Previous Realization

Previous Realization

Code	Description	HLVC - Home	HLVC - Her	CHILS
OVERT	The previous VP has an overt subject pronoun	85	124	773
NULL	The previous VP has a null subject	385	484	272
NOUN	The previous VP has an NP subject	94	66	101
NEW	New Topic/Turn	184	212	590
Total		**748**	**886**	**1,736**

Table 40. Variants and token count for PREVIOUS REALIZATION.

The previous realization variable is a means of testing for a priming effect caused by the subject pronoun realization of the preceding phrase (or phrases). The expected effect is that we are primed to repeat the structure that was used in the preceding phrase. For example, if the previous VP had an overt subject pronoun, this primes an overt subject pronoun in this VP.

I include four variants for this factor: OVERT, the preceding VP had an overt subject pronoun; NULL, the previous VP had a null subject; NOUN, the priming VP has an NP subject; and NEW, which means the participant has paused, changed topics, or is starting a new turn. Table **40** shows how the tokens are distributed across these variants.

There are various methods of coding this variable, with some studies including the 10 preceding phrases in their analysis of priming effects (Travis, 2007). I discuss some of these methods as

avenues of future research in **8.3**. However, for this study I only include the VP that immediately

precedes the one being coded. Also, many studies only consider a previous realization that is

coreferential. This is the case with the studies of Spanish varieties included in **3.3** (Abreu, 2009;

Carvalho & Child, 2011). Thus, they do not include phrases like the one shown in (**48**), where

the referent in the first VP is 3SG.F, and in the second phrase is 3SG.M.

48)　　　　[la moʎə ɛɹa　　　də soɹa　e　**is** ɛɹa　　　də pɛskʊɹə]
　　　　　　The wife　is.IMP.3SG　from Sora　and　he is.IMP.3SG　from Pescara
　　　　　　La moglie era di Sora, e lui era di Pescara
　　　　　　'The wife was from Sora, and **he** was from Pescara'
　　　　　　(CHILS, Filomena, 17:43)

However, I include tokens that are non-coreferential with the previous VP alongside those that

share the same referent, as shown in (**49**):

49)　　　　[kwana **nu** sə sɛtəvam　a la tawolin sə paɹlavam]
　　　　　　When　we REFL sit.IMP.1PL at the table　　REFL talk.IMP.1PL
　　　　　　Quando noi ci sedevamo a tavola, ci parlavamo
　　　　　　'When **we** were at the table, we would talk to each other'
　　　　　　(CHILS, Caroline, 14:05)

Studies looking at Spanish varieties have found that the priming effect of the previous realization

is significant, even without coreferentiality (Orozco & Hurtado, 2020). Often these studies

combine previous realization with switch reference, as mentioned above. However, both methods

show the same direction of effect with an explicit pronoun in the previous VP disfavouring a null

subject. A new turn, or change in topic, also disfavours null subjects. If the preceding VP has a

null subject, a null subject is favoured. I expect this will also be true in my analyses, independent

of coreferentiality.

129

5.3.5.2. Switch reference

Switch reference

Code	Description	HLVC - Home	HLVC - Her	CHILS
SAME.REF	Same referent as previous VP	260	331	592
SWITCH.REF	Different referent from previous VP	304	343	554
NEW	New Topic/Turn	184	212	590
Total		**748**	**886**	**1,736**

Table 41. Variants and token count for SWITCH REFERENCE.

An often-overlapping variable with previous realization is switch reference. Each VP has a referent, the subject of the phrase, and when the referent is different from the preceding VP, a 'switch' has occurred. I code three variants for this factor: SAME.REF, when the previous VP had the same subject; SWITCH.REF, when there is a switch between the preceding VP and the current one; and NEW, which means the VP does not have a priming VP due to a pause, change in topic, or new turn. In (**50**), for example, Filomena has just said that when her family made wine, they wouldn't try it until the first full moon following the feast of Saint Martin. This is the start of a new turn: after she describes how the wine is made, I ask her when her family would first drink it. I said that in my family we first drank the new wine at the feast of Saint Martin. She responds:

50)   [**is** l   a   pris   a la   luna]
        he DOBJ AUX.3SG take.PPRT at the moon
         lui l'ha preso alla luna
        '**he** took it at the moon'
        (CHILS, Filomena, 26:01)

51)   [**i**   m   ɪmbaɹeva   lu t͡ʃɛkəɹz   bikʊz   ɹ ɛɹə   sɛmbɹə kjuɹioza]
        he OBJ learn.IMP.3SG the checkers because I be.IMP.1SG always curious
        lui m'imparava (a giocare) a dama perché io ero sempre curiosa
        '**He** taught me to play checkers because **I** was always curious'
        (CHILS, Caroline, 22:50)

130

This is a significant variable in many studies of pro-drop across Romance languages. In fact, switch reference is a significant factor in each of the studies discussed in **3.3**. The direction of effect is typically that co-referentiality favours null subjects, and a 'switch' in referent disfavours null subjects, as shown in (**51**). This is the result found in Spanish varieties (Carvalho & Child, 2011; Otheguy et al., 2007; Torres Cacoullos & Travis, 2019), and heritage Calabrian (Nagy et al., 2011). Tokens which do not have priming VP, such as a new turn, typically disfavour null subjects, but to a lesser degree than when there is a switch in reference.

This consistent finding in Spanish and Calabrian is often said to be a means of reducing ambiguity—especially between two 3SG or 3PL referents, where the verbal inflection is the same (Hernández Constantin, 2021:58). Interestingly, Faetar is an exception, with the finding that a switch actually favours a null subject (Nagy et al., 2018). However, I expect that both Calabrian and Ciociaro will have the more common direction of effect, with a switch in referent disfavouring null subjects and coreferential VPs favouring null subjects.

5.3.5.3. Same Tense

Switch tense

Code	Description	HLVC - Home	HLVC - Her	CHILS
SAME.TENS	Same tense as previous VP	343	422	626
DIFF.TENS	Different tense from previous VP	221	252	520
NEW	New Topic/Turn	184	212	590
Total		**748**	**886**	**1,736**

Table 42. Variants and token count for SWITCH TENSE.

The previous two variables have looked at priming effects caused by the subject of the preceding VP: the realization of the subject pronoun, and the referent. This variable looks at the potential priming effect caused by the tense of the previous VP. I code three variants for this factor:

SAME.TENS, when the previous VP is in the same tense; DIFF.TENS, when the previous VP's tense is different; and NEW, when, like the previous two variables, the VP is at the start of a new topic or turn. The distribution for this factor is included in Table **42**, showing that the slight majority of tokens from each dataset has the same tense.

This variable has been much less studied than PREVIOUS REALIZATION and SWITCH REFERENCE. The only study from **3.3** that includes it is Abreu (2009). In their analysis of pro-drop in Puerto Rican bilingual and monolingual Spanish speakers, it was a significant factor for the bilingual group, but not the monolingual.

Travis (2007) also found a significant effect, although they only included 1SG subjects in their study on the strength of different priming effects on subject pronoun realization, so it is not comparable to my analysis. In both studies the direction of effect is that the same tense as the previous VP favours a null subject, whereas a change in tense disfavours null subjects. Both studies found that this factor has a strengthening effect on SWITCH REFERENCE. In other words, if both the referent and tense are the same, null subjects are significantly more favoured than when either the tense or the referent has changed.

In this section, I have described the three priming variables that are included in the HLVC and CHILS datasets. I include them because comparable studies have found them to be significant. I expect the direction of effect to be the same in my analyses: null subjects prime null subjects, continued reference to a subject primes null subjects, and VPs in the same tense prime null subjects.

5.3.5.4. AIS linguistic variable - Pronoun in prompt

Pronoun in prompt

Codes	Description	Tokens
PRESENT	An overt subject pronoun was in the prompt	118
ABSENT	No subject pronouns were in the prompt	465
Total		**583**

Table 43. Variants and token count for PRONOUN IN PROMPT.

As described in **5.3.1.1**, because the AIS data is based on translations of single phrases, given to the participants as prompts, the priming variables based on the previous VP produced by the participant cannot be included. However, the prompts themselves can serve as a priming variable of sorts. Specifically, if the prompt contains an overt subject pronoun, coded as PRESENT, the participant may be primed to produce an overt subject pronoun in their response. On the other hand, if the prompt contains a null subject, coded as ABSENT, the participant may be primed to give a response that has a null subject. Table **43** shows that most of the prompts contain a null subject, with 20% of them having an overt subject pronoun.

This variable does not have a direct equivalent in the studies mentioned elsewhere in this section; most of which extracted their tokens from recorded interviews. However, Abreu examined priming effects at both the intra-speaker and inter-speaker level, finding that participants produced more overt subject pronouns when their interlocutor had just used one (Abreu, 2009:110).

I expect this variable will have the same direction of effect in the AIS analysis, with participants favouring null subjects when the prompt contains a null subject and disfavouring them when the prompt has an overt subject pronoun.

5.3.6. Extralinguistic Variables

Example of extralinguistic variables

Variable	*Example*
SPEAKER	ARCANGELA
AGE AT RECORDING	82
PARTICIPANT GENDER	FEMALE
AGE OF ARRIVAL	18
HOMETOWN	CASALVIERI

Table 44. Extralinguistic variables included in the study with examples from Arcangela (CHILS participant).

Extralinguistic variables are those that are not directly related to subject pronoun realization. In other words, they do not vary by token. For example, the age, gender, and other social factors that may affect a participant's subject pronoun usage but are not speech related. I include five extralinguistic factors in my analyses. SPEAKER (**5.3.6.1**) and AGE (**5.3.6.2**) are extralinguistic factors for all three datasets. Participant gender (**5.3.6.3**) is a variable for both the HLVC and CHILS datasets. Because all the AIS participants are male, P-GENDER is not a variable for that analysis. Because I conducted the CHILS recordings, more information is available to me, and I include the final extralinguistic variables: AGE OF ARRIVAL (**5.3.6.4**) and HOMETOWN (**5.3.6.5**) for only this corpus.

As an example of these variables, the CHILS participant Arcangela was 82 when I interviewed her, she was born in Casalvieri and emigrated from there when she was 18. These extralinguistic variables are described in more detail below.

5.3.6.1. Speaker

Every person uses language in a distinct way. During sociolinguistic interviews, participants may react differently to the experience, have a certain way of speaking, prefer certain phrase structures, or have other idiosyncrasies to their speech. To capture this inter-speaker variation

within the multivariate analysis, each participant is coded as a SPEAKER for the HLVC and CHILS. For the AIS data, each speaker represents a location, which is equivalent to SPEAKER for HLVC or CHILS. This variable is included as a random intercept in the analyses of all three datasets. This is described in more detail in **5.4.2**.

5.3.6.2. Age

Of the extralinguistic factors included in studies of pro-drop, age is often the factor that is significant. Languages change over time, and a significant difference can be found across age groups. Four of the studies discussed in **3.3**, three of Spanish varieties and one of Faetar, find that subject pronoun usage differs significantly across different age groups. The three studies of Spanish varieties find that subject pronoun usage is favoured for older participants and disfavoured for younger (Lastra & Butragueño, 2015; Limerick, 2019; Orozco & Hurtado, 2020). The opposite is true for homeland Faetar (Nagy et al., 2018).

However, age is often less significant for studies of heritage languages, as is the case in heritage Faetar (Nagy et al., 2018), despite it being significant in the homeland variety. The reasons for this are largely explainable by the domains of use for the heritage language. In her description of the Columbian Spanish heritage community in London, Ontario, Hernández Constantin finds that the first and second-generation speakers do not have significant differences with regard to pro-drop (Hernández Constantin, 2021). In her description of the community, she explains that the language is mainly used by the second-generation to communicate with the first. As such, older speakers mostly use the language to communicate with other older speakers, which may impede innovation and change.

Almost all of Sarnia's Ciociaro community emigrated from Italy within a 20-year window (1950-1970), most were young, and very few passed the language to their children (see **2.2.3.2**). As a result, AGE, rather than GENERATION, is used as the extralinguistic variable for my analyses. Table **45** shows the range and median age for each corpus included in this study.

Age of participants

Description	HLVC – Home	HLVC – Her	AIS	CHILS
Youngest participant	35	57	44	57
Oldest participant	94	75	60	95
Median	**51.5**	**61.5**	**53.5**	**80.5**

Table 45. Range of ages for participants from each corpus.

From the table, we see how homogenous these groups are. For the Rbrul analyses, AGE is coded a continuous factor. Due to the homogeneity of the corpora, AGE is not expected to be a significant variable.

5.3.6.3. Participant gender

While the age of the participants is very homogenous, both the HLVC and CHILS samples are balanced for participants' gender: five women and five men for each of homeland and heritage Calabrian, and 10 women and 10 men for the CHILS. All the AIS participants are male, and so this variable does not apply.

Of the comparable studies of Spanish and Faetar discussed in **3.3**, two find P-GENDER[29] to be a significant variable: Carvalho & Child (2011) report females disfavouring null subjects, and

---

[29] P-GENDER is used in the coding schema for my analyses to distinguish the extralinguistic variable from the linguistic variable GENDER, which describes grammatical gender agreement, for example of 3rd person subjects (see **5.3.3.3**).

males favouring; however, they are quick to point out that they are only using 12 participants (eight males, four females). The same direction of effect is found by Abreu (2009) with six females and four males. However, every other comparable study referenced above did not find participant gender to be a significant factor. I do not expect it to be a significant factor in my analyses either.

5.3.6.4. Age of arrival

The CHILS participants, as heritage speakers of Ciociaro, arrived in Canada at different ages. It is possible that those who arrived at a younger age have had more contact with English and may have a higher rate of overt subject pronoun use. The age of arrival of the CHILS participants is shown in Figure **8**. The participants arrived in Canada between the ages of six (Caroline) and 32 (Raffaele), with a median age of 18.5.

Figure 8. Chart showing the age of arrival (AoA) of the CHILS participants.
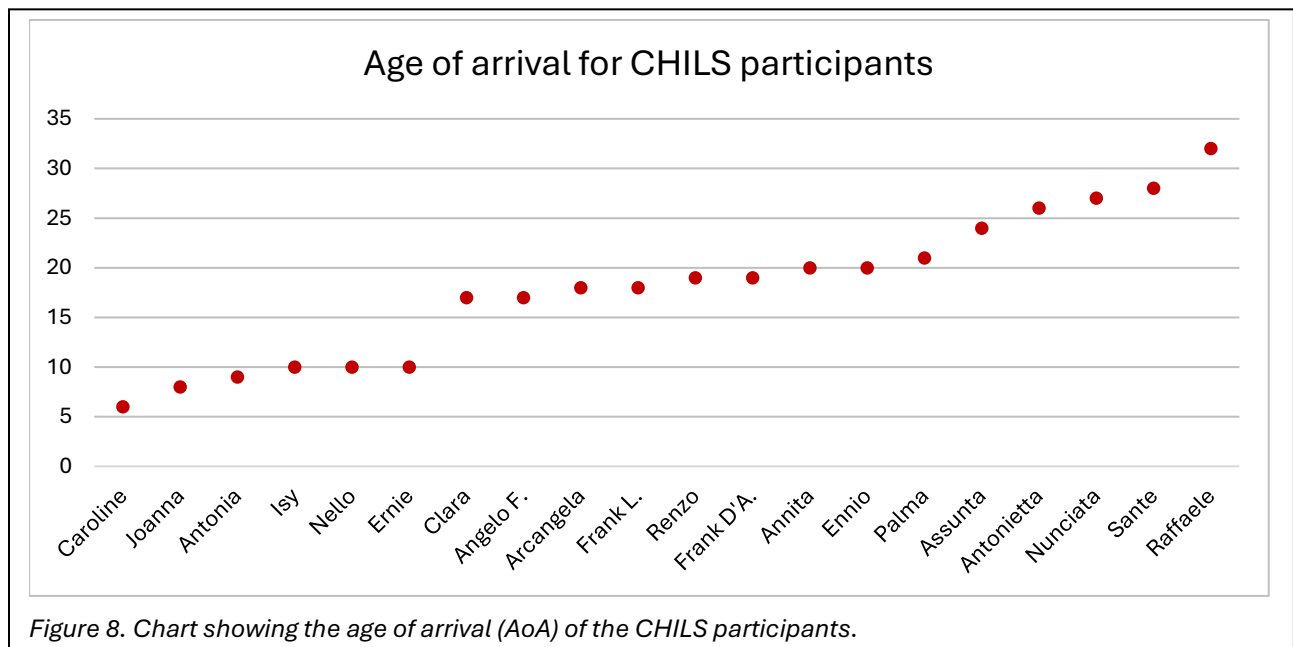
The age of arrival of heritage speakers has been found to be a significant variable in some studies of Spanish varieties. In their study of New York Spanish, Otheguy et al. (2007) found that the older the speaker was when they arrived in New York, the fewer overt pronouns they used, suggesting a contact effect with English. However, this was not the finding of Limerick (2019). In their study of Spanish speakers in Georgia, age of arrival was not a significant factor, and the participants with the oldest ages of arrival had the highest rates of subject pronoun use.

For Sarnia's heritage Ciociaro community, I do not expect this to be a significant factor.

5.3.6.5. Hometown

A factor I was able to include for the CHILS corpus is the hometown of each participant. There could be differences in subject pronoun usage based on the speaker's hometown. Also, the six participants of the AIS corpus are each from different towns, so I wanted to include the hometown of the CHILS participants to examine if any inter-speaker variation could be associated with hometowns, which are summarized in Table **46**.

Hometown of CHILS participants

Town	Participants
Alvito	3
Casalvieri	4
Castelliri	1
Fontechiari	8
Post Fibreno	2
Veroli	1
Vicalvi	1

Table 46. Hometowns of CHILS participants compared to nearest AIS location of ages for participants from each corpus.

However, due to chain migration, most members of Sarnia's Ciociaro community were born in only a few towns (see **2.2.3.2**). This is reflected in the participants included in this study. In fact, Figure **9** shows just how small the area is that all the CHILS participants come from (Google, n.d.-b). As a result, I do not expect the hometown of the CHILS participants to be a significant variable.

Figure 9. Map showing the hometown of the CHILS participants in red, and the AIS locations in purple (Google, n.d.-b).
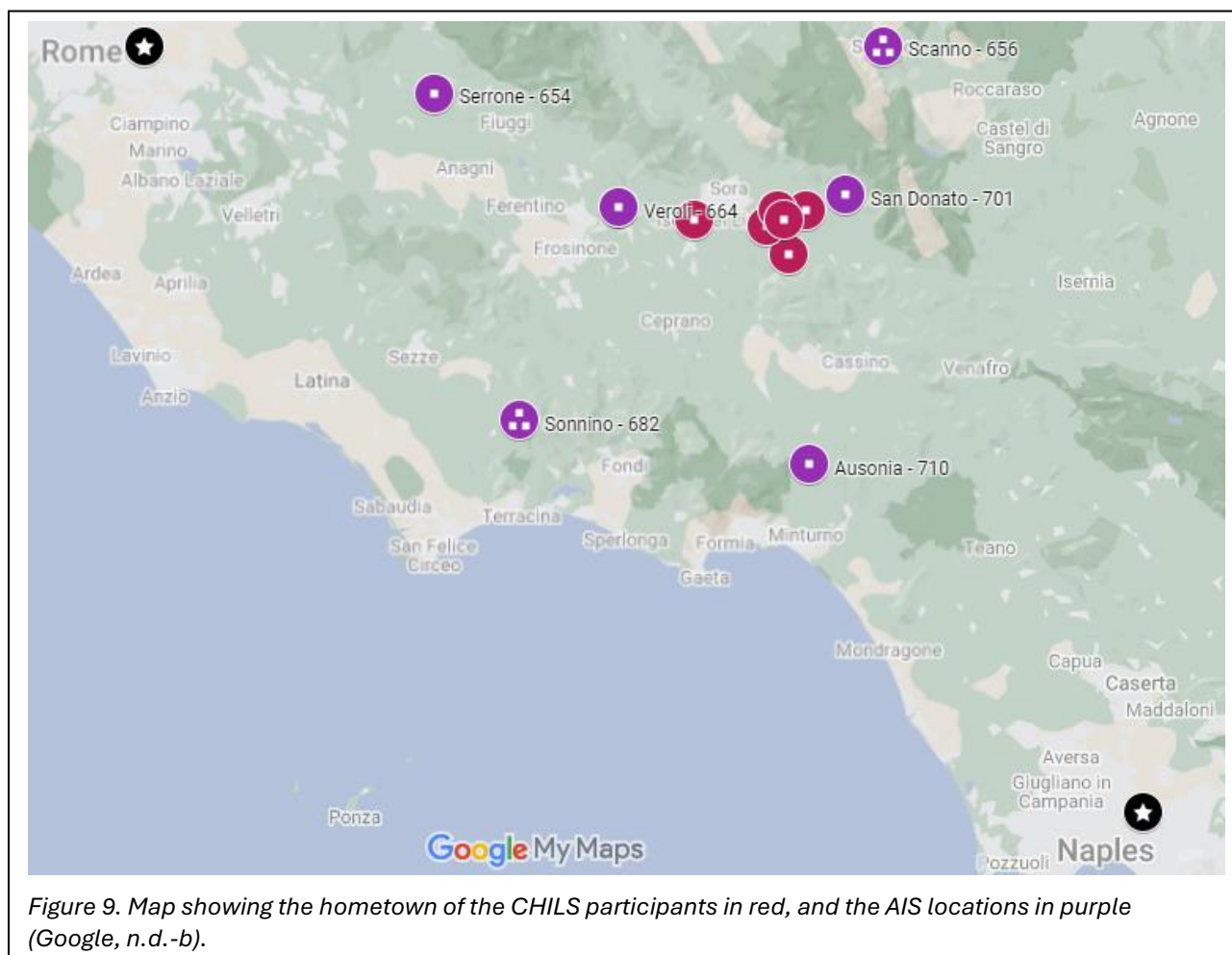
5.4. Analysis of data

Once the VPs are transcribed and the variables are coded for each of the tokens included in (**5.3**), I analyse the data using Rbrul (Johnson 2009), which runs within the statistical software R (R Core Team, 2021). Rbrul performs multivariate logistic regression analyses to predict which independent variables (linguistic and extralinguistic) have a significant impact upon the dependent variable—the presence or absence of a null subject. I explain this process by first explaining how the tokens are organized, the analyses refined, and the model of best fit is selected (**5.4.1**). I then briefly discuss why SPEAKER is treated as a random intercept, and the

other independent variables are not (**5.4.2**). Finally, I explain how I interpret the results of my analyses, and how they are presented in the following chapters (**5.4.3**).

5.4.1. Rbrul analyses

In order to conduct my analyses, I export the tokens from ELAN that have been coded for the variables into my token file as described in (**5.2**). There are four token files: homeland variety of Calabrian, heritage variety of Calabrian, AIS, and CHILS. Each one is first analysed in Rbrul separately. To ensure consistency across the analyses, the same approach must be followed for each.

First, I select the dependent variable (**5.3.1**), and its application value for the analyses. The application value for all analyses presented throughout this dissertation is ∅, and not OVERT. This means that all significant variables presented from Chapter **6** show whether independent variables favour or disfavour a null subject. This is presented within each variable as a factor weight (*FW*) for each variant. FWs range from 0 to 1. As I selected "centred FWs" for all my analyses, a FW above 0.5 means the variant favours the application value, a null subject in this dissertation. A FW below 0.5 means the variant disfavours a null subject. The further a FW is away from 0.5, the stronger the effect. As an example, for POLARITY the two variants would be given FWs: AFFIRMATIVE (0.25) and NEGATIVE (0.75). In this example, affirmative phrases disfavour null subjects, and negative phrases favour them.

Second, the independent variables are selected. I run an analysis of each independent variable with the dependent variable before running multiple independent variables together. This is for two reasons. It ensures that the controlled vocabulary is consistent across all the participants that make up each token file. For example, Rbrul treats lower-case and upper-case letters as different

variants. So, if (hypothetically) there are a couple of tokens where Excel's auto-correct

unhelpfully capitalized IMP, where the rest of the imperfect tokens are coded as IMP, this could be

quickly noticed and corrected in the token file. The other reason to run an analysis of the

dependent variable with each independent variable before combining them is to notice and

address knock-outs and variants with especially low token counts. For example, as shown in

(**5.3.3.4**), there are very few instances of 2PL subjects in the HLVC and CHILS datasets. And

every VP with a 2PL subject has a null subject. Because of this, this variant is perfectly predictive

of subject pronoun realization—every time a VP has a 2PL subject, it also has a null subject. This

is known as a knock-out. These knock-outs can be addressed in two ways, they can be collapsed

into another variant that does vary for subject pronoun realization, or they can be removed from

the analysis. If they are removed from the analysis, the token count is still reported, but the

values are surrounded by parentheses in the tables, whereas the token count for the other variants

is not (see Table **32** in **5.3.3.4**). Variants with especially low token counts should also be

addressed at this stage to determine if they should be collapsed with another variant within the

variable. As an example from the AIS dataset, the variable TENSE only has five tokens coded as

AP. These five tokens should be collapsed into one of these other variants, whichever is patterned

similarly (for rate of null subject) and is justified by grammatical similarity. In this case they

were collapsed into the IMPERFECT variant because they are both past tenses that do not have an

auxiliary.

Third, once the distribution of each independent variable has been examined and refined, a

combined analysis is done to determine which variables are significant and which are not. There

are several types of combined analysis, of which the first is a *step-up*. For this type of analysis,

Rbrul creates a model with the dependent variable and one independent variable. This model

calculates the significance of the independent variable on the dependent variable. It then repeats this with each of the independent variables. The most reliably predictive independent variable is selected. Rbrul then creates another model, the next *step up*, by adding one more independent variable, and calculates the significance of these two variables' relationship to the dependent. It repeats this again with each independent variable and the most reliable is chosen to be the second independent variable included in the model. This model, now with two independent variables, is compared to the first with only one. Rbrul compares those models for which best corresponded to the actual data in the token file. If the two variables are more predictive than one independent variable, Rbrul then adds a third. The analysis does this many times, adding variables that strengthen the reliability of the model, and stopping if the odds of predictive error increase. The result is a model that shows which variables, when added, most faithfully reflect the dataset.

A *step-down* analysis is conducted in the opposite way. A model is first created based on all the independent variables, and each 'step' removes an independent variable that contributes the least in accounting for the variability. A variable is removed from the model, and the remaining variables are modelled. The predictive abilities of the new model are compared with the previous model. If the reliability of that model is not worse, the variable is not affecting subject pronoun realization. Or, if it is, it complicates the model more than it helps predict the dependent variable.

The final type of analysis is a *step-up/step-down*, which conducts both types and selects a best model for each, known as a *best run*. The two models should converge on the same significant variables. If there is not a match, it is likely that some variables have unpredicted interactions, and require adjustments.

The reliability of an analysis can be improved by selecting models with a lower Akaike

information criterion (*AIC*). The AIC estimates the likelihood that the model faithfully fits the

data. As the number of variables included in a model increase, the probability that some are

significant increases. However, this may not reflect the reality of the data. It may just be an effect

of the quantity of the independent variables and their variants, and not the quality of their

relationship with the dependent variable. Therefore, when comparing models, a lower AIC is

preferred. A lower AIC reflects an increase in probability that the significant variables are

meaningful. Therefore, any independent variables that are shown to have very little impact on the

dependent variable in the modelling should be excluded from the final 'step-up/step-down'

analyses. Even if they were not selected in the models, they may still have an effect.

Fourth, any interactions between independent variables must be identified and addressed. An

interaction occurs when groupings of certain variants pattern too closely with the groupings of

variants of another factor, making it impossible to tease apart which factor actually has an effect.

These can be identified in the analysis by examining the variants and their rankings for a given

factor. Rbrul ranks variants based on their FWs: from the variant with the FW closest to 1 to the

variant with the FW closest to 0. An indication that an interaction is occurring is if the ranking of

the FWs is markedly different than the ranking of the percentage rates. For example, if for

polarity negative phrases have a FW of 0.75 but are null 20% of the time; and affirmative

phrases have a FW of 0.25 but are null 60% of the time, an interaction is influencing those FWs.

These interactions can be addressed by creating interaction factors. These are factors which

combine the variants of the two variables. For example, creating an interaction factor of SWITCH

REFERENCE and CLAUSE TYPE creates variants that test the significance of SAME.REF+EMBED,

SAME.REF+MATRIX, SAME.REF+SIMPLE, SWITCH.REF+EMBED, etc. In this case, I found that the

143

combined variant SAME.REF+EMBED was causing an interaction between the two variables—
many embedded tokens have the same referent as the preceding token (the subject of the matrix
clause). As a result, I ran two more step-up/step-down analyses, removing one the variables in
each, and found that the interaction was resolved—CLAUSE TYPE was not significant anymore,
and was removed from future analyses.

Fifth and finally, once a step-up/step-down analysis has been conducted for each HLVC, AIS,
and CHILS dataset, and any issues with the variables have been addressed, the significant factors
from the best runs are tested in a *one-level* analysis. As explained above, a step-up/step-down
analysis is comparing one model to another, as such the significance the of a variable is relative
to the strength of the previous variables in the previous model. In a one-level analysis each
independent variable is tested for the significance of its relationship with the dependent variable.
In the results presented in the following chapters, the significant variables have a p-value below
the standard threshold of $p<0.05$. However, in a one-level analysis each variable is measured for
significance independently. As such the ranking of variables presented is not by significance, but
by *range* or *effect size*. The range (or magnitude) of effect is the range between the highest-
ranked variant and the lowest-ranked of a given significant factor. Returning to the example
above for POLARITY: if NEGATIVE has a FW of 0.75, and AFFIRMATIVE has a FW of 0.25, the
range is 50.

The significant variables of the one-level analysis, alongside the FWs and ranges of the variants
for each, are presented in the following chapter.

5.4.2. Speaker as a Random intercept

The process of selecting the significant variables was described in the previous section. SPEAKER is included in each model presented in the following chapter. This is to show the level of inter-speaker variation within each corpus. However, interactions would likely arise when SPEAKER is included in a model with many linguistic variables and all extralinguistic variables. For example, if feminine subjects especially favour null subjects, then the participants who produce more VPs with feminine subjects will appear to favour null subjects more than participants who have comparatively few VPs with feminine subjects. Also, a participant who has a wife and daughters, might be expected to produce more phrases with 3SG.F and 3PL.F subjects than another participant who has a husband and sons. To account for these interactions, in this example between SPEAKER and SUBJECT, SPEAKER is included in the Rbrul analyses as a *random intercept*.

Treating SPEAKER as a random intercept means that, as Padilla describes, Rbrul measures the inter-speaker variation that is "over and above (or 'under and below')" what the significant variables would predict (Padilla, 2020:83). In other words, Rbrul gives each speaker a FW that shows how much they favour or disfavour null subjects beyond what the model would predict based on the included factors. This means that in the example above, a participant who has an especially high number of feminine subjects is predicted to have a high rate of null subjects. If that participant still has a FW that favours null subjects, this means that their higher rate of null subject is above and beyond what their use of feminine subjects would predict.

Importantly, the model uses the complete dataset to give a participant a predicted FW, but it also works in reverse. The inter-speaker variation is accounted for when giving FWs to the other variables. Meaning that if the participant does use an especially high rate of null subjects, beyond

what is predicted by their distribution of 3SG.F and 3PL.F subjects, the model neutralizes this

effect in assigning a FW to feminine subjects.

5.4.3. Interpreting results

Once the model with the best fit to the raw data has been determined, the null rate, significant

factors, and inter-speaker variation can be examined. In Chapter **6**, I interpret the results through

the lens of how they respond to my hypotheses presented in Chapter **4**.

First, I examine and report the results of each individual analysis. The overall null rate is not

especially meaningful in isolation. The significant factors are examined next, from the largest

range of effect to the smallest. The variant ranking and direction of effect for each significant

factor is interpreted next. Specifically, I look for evidence of interactions, and perform additional

Rbrul analyses to find and understand their cause. This can be done, as described above, by

creating interaction factors of the significant variables, or by examining the cross tabs of the two

variables to show how the tokens are distributed across the variants of the two factors. This can

reveal details that were missed in the larger analysis. For example, if SUBJECT is interacting with

TENSE, by examining the crosstabs, we may discover that most tokens in the PRESENT have a 1SG

subject, whereas very few of those in the IMPERFECT do. This can help with interpreting the

variant ranking of a variable.

I then examine the SPEAKERS, to see how much inter-speaker variation there is within a given

corpus. Variation can reveal patterns that are not included in the significant factors of the model,

and outliers can be examined more closely. SPEAKER as a random intercept accounts for variation

due to the distribution of the factors included in the model. But it is possible that AGE, while not

selected as a significant variable, may have a slight effect that is revealed in the inter-speaker variation.

Second, I compare the results of the heritage and homeland varieties. If the null rate is different in the two samples, this may be evidence of contact-induced change (either from standard Italian in the homeland varieties or English from the heritage varieties). This can be further examined by looking at the significant factors. If the two varieties do not share significant variables, or the size of effect is markedly different this may further point to a divergence in the two varieties. Even if the significant variables are similar between the homeland and heritage varieties, the variant hierarchy or direction of effect may be different. If it is, this suggests that the significant variable is functioning differently in relation to pro-drop for the two varieties. The two varieties are then combined in a one-level analysis with their shared significant variables, and CORPUS is tested as an independent variable, with the variants HOMELAND and HERITAGE. The FW of these two variants can reveal the degree of difference between the varieties.

Third, I compare the Calabrian and Ciociaro results. I hypothesized that the overall null rate of Ciociaro would be lower, and this can be examined. And, if it is, further investigation of the role verbal ambiguity may play, and potential support for the functional hypothesis. The significant variables, both their ranges of effect and variant rankings in what ways the two languages are similar and distinct. Specifically, if they share significant variables, with similar directions of effect, it may point to broader patterns common throughout Italian or Romance languages.

Finally, I compare the results from my analyses to the results of other pro-drop research. The null rates of both Calabrian and Ciociaro can be compared to the null rates of Faetar, Italian, or many

147

Spanish varieties. The significant results can also be compared, both in terms of how big of an

effect they have on pro-drop, but also on the direction of that effect.

The variables described in this chapter were included because they have been significant in other

studies. Now the significant variables of the HLVC, AIS, and CHILS datasets can be compared

to the studies that motivated their inclusion in my methodology. By interpreting the results of

each corpus's analysis through how they respond to my hypotheses and how they align with

comparable research, I can contextualize them within the larger field of inquiry into pro-drop in

Romance languages.

6. Results

The previous chapter discussed the coding of the variables included in my analyses, as well as the multivariate analysis methods used to shed light on subject pronoun variation in Calabrian Italian and Ciociaro. In this chapter I present the results of these analyses in four sections. In **6.1**, I present the results of pro-drop in Calabrian Italian from the HLVC corpus. The homeland and heritage varieties are reported first, followed by a combined analysis of both groups. In **6.2**, I report the results from the AIS corpus, providing insight into homeland Ciociaro. In **6.3**, I share the results of the CHILS corpus. In the final section **6.4**, I present combined analyses, first of AIS and CHILS, and finally of the three corpora.

For each of the sections, the significant variables of the one-level analysis, alongside the FWs and ranges of the variants for each corpus are included. The significant variables are described one-at-a-time, but each one should not be thought of as independent from the other significant variables. The results for each analysis show which variables, when included together in a mixed-effect model, are most reliably predictive of the null subject rates found in my datasets. For all the results presented, as described in **5.4**, the factor weight (*FW*) represents whether a variant is more or less likely to occur with a null subject pronoun, given the constellation of independent variables included in that model. A FW of 0.5 is neutral, meaning a variant that neither favours nor disfavours null subject pronouns; a FW below 0.5 means a variant is less likely to co-occur with an NSP; and above 0.5 favours an NSP. The range is also calculated for each significant factor by calculating the difference between the highest factor weight and the lowest. This range provides a measure of the strength of each factor's effect: the larger the range between highest FW and lowest FW, the stronger the effect.

6.1. HLVC Results

For this analysis, as presented in **5.1.1.1**, the 10 eldest homeland speakers and 10 eldest heritage speakers were chosen, balanced for gender. The five male and five female homeland participants were interviewed in Calabria, and the heritage speakers were interviewed in the Greater Toronto Area of Ontario.

In the following sections, I first present the results of the homeland dataset, followed by the heritage group. Following these separate analyses, I present the results of a combined analysis, and provide more detail about the factors with the largest effect.

6.1.1. Homeland speakers

This analysis of the homeland participants from the HLVC corpus contains 748 tokens (*n*), with an overall null subject rate (*%Ø*) of 80%. The significant variables are SUBJECT, SWITCH REFERENCE, and PREVERBAL ELEMENT. This means that for homeland Calabrian speakers, null subjects are most likely to occur in phrases that have the top-level variant of each significant independent variable included in the model. In other words, the results presented in Table **47** show that a null subject is most likely in phrases that have a 3PL subject (FW = 0.68), have the same referent as the preceding phrase (FW = 0.69), and have a preverbal element (FW = 0.55).

For the SUBJECT variable, there is an interaction effect whereby 3SG.F has a lower null rate but a higher FW than 3SG.M, though the difference is very slight. Apart from this exception, 3rd person subjects favour null subject pronouns, while 1st and 2nd person subjects disfavour null subjects.

One-level analysis of significant variables for Homeland participants[30]

Tokens (n):	748	%Ø:	80%
Subject			
	n	%Ø	FW
(2PL)[31]	(3)	(100%)	N/A
3PL	122	90%	0.68
3SG.F	47	87%	0.60
3SG.M	108	88%	0.59
1PL	104	81%	0.47
1SG	302	75%	0.38
2SG	65	63%	0.28
Range			**40**
Switch Reference			
	n	%Ø	FW
SAME.REF.	260	90%	0.69
SWITCH.REF.	304	79%	0.48
NEW	184	67%	0.32
Range			**37**
Preverbal Element			
	n	%Ø	FW
YES	277	83%	0.55
NO	471	78%	0.45
Range			**10**
Speaker (Random)			
Std. Dev.			0.44

Table 47. One-level analysis of significant variables for homeland participants from the HLVC corpus.

For SWITCH REFERENCE, if the VP has the same subject as the preceding VP (SAME.REF), a null subject is favoured, whereas a different referent slightly disfavours a null subject (SWITCH.REF). The start of a new turn or topic disfavours a null subject (FW = 0.32).

Finally, the variants for the PREVERBAL ELEMENTS factor, which combines negation particles, reflexive pronouns, and object clitics, show that the presence of a preverbal element favours a null subject pronoun. However, it is important to note that the range is small.

This is especially relevant when we compare these results to those of the heritage speakers.

6.1.2. Heritage speakers

This analysis contains 886 tokens, and an overall null rate of 80%, the same rate as the homeland analysis. As Table **48** shows, the two significant variables chosen by this analysis of heritage

---

[30] The standard p-value threshold of 5% (p <0.05) was used to test for significance in these analyses.

[31] Variants presented in parentheses were not included in an analysis because of their invariability.

speakers are shared by the homeland analysis, SUBJECT and SWITCH REFERENCE. The difference is

that PREVERBAL ELEMENTS is not a significant variable for heritage speakers. However, the

homeland analysis shown in the previous section also has PREVERBAL ELEMENTS close to the p-

value threshold of 0.05 (p=0.04), and therefore the two results are similar despite having a

different number of significant variables.

One-level analysis of significant variables for heritage participants

	n:	886	%Ø:	80%
Subject				
		n	%Ø	FW
(2PL)		(3)	(100%)	N/A
3SG.M		118	93%	0.72
3PL		130	90%	0.62
1PL		192	85%	0.52
2SG		46	80%	0.45
3SG.F		44	77%	0.39
1SG		356	71%	0.30
	Range			**42**
Switch Reference				
		n	%Ø	FW
SAME.REF.		331	89%	0.68
SWITCH.REF.		343	78%	0.44
NEW		212	71%	0.37
	Range			**31**
Speaker (Random)				
Std. Dev.				0.46

Table 48. One-level analysis of significant variables for heritage participants from the HLVC corpus.

There are, nonetheless, some differences in the two analyses. While both analyses present SUBJECT as having a bigger effect than SWITCH REFERENCE, there are differences in the order of the variants. There is an outlier in the hierarchy of the variants for subject: 3SG.F disfavours null subjects, while the other 3rd person subjects (3SG.M and 3PL) favour null subjects. Possible reasons for this are discussed in the combined analysis that follows. We otherwise find very similar results to the homeland analysis. This strengthens the

representativeness of the combined analysis of both groups presented in the following section.

6.1.3. Combined analysis of both HLVC datasets

Table **49** shows the results of the multivariate analysis of all Calabrian speakers—homeland and

heritage—together. These two subcorpora of the HLVC have similar null rates and significant

variables. Further, when CORPUS is included as an independent variable, it is not significant with identical factor weights of 0.50 for both homeland and heritage speakers (see Appendix **H. 5** for the Rbrul results).

One-level analysis of pro-drop in the HLVC corpus

	n:	1,634	%∅:	80%
Switch reference				
		n	%∅	FW
SAME.REF		591	89%	0.69
SWITCH.REF		647	79%	0.46
NEW		396	69%	0.35
		Range		**34**
Subject				
		n	%∅	FW
(2PL)		(6)	(100%)	N/A
3SG.M		226	91%	0.66
3PL		252	90%	0.65
1PL		296	83%	0.52
3SG.F		91	82%	0.49
1SG		658	73%	0.34
2SG		111	70%	0.34
		Range		**32**
Preverbal Element				
		n	%∅	FW
YES		617	83%	0.54
NO		1017	79%	0.46
		Range		**8**
Speaker (Random)				
Std. Dev.				0.43

Table 49. Significant variables of pro-drop realization in the HLVC corpus.

Across 20 speakers, the number of tokens is 1,634. The overall rate of subject pronoun omission is 80%. Although we might expect there to be a difference in the omission of subject pronouns between those participants who live in Italy (HOMELAND), and those who are heritage speakers in the GTA (HERITAGE), there is not. Of the 14 independent variables that were included in the dataset, three are significant: SWITCH REFERENCE, SUBJECT, and PREVERBAL ELEMENT.

The speakers were included in the analysis as random intercepts to account for any effects of an individual speaker's idiolect (as described in **5.4.2**). The variation across speakers is described below, but first I summarize the significant variables.

6.1.3.1. Switch reference

Of the factors selected by the analysis, SWITCH REFERENCE has the largest effect (range = 34). In summary, a subject pronoun is most likely to be realized at the start of a turn (FW = 069). A

subject pronoun is also favoured when the subject is different from the previous phrase. A VP is most likely to have a null subject pronoun when the subject remains the same as the previous phrase.

Nagy et al. (2018) found similar results: continued reference to the same subject increases the likelihood of a null subject. A potential explanation is that the subject hasn't changed, so any ambiguity from similar verb forms is reduced, while an overt pronoun may signal to the listener that the referent of this phrase may be different from the previous VP.

A new turn is usually marked by a preceding pause and often a change in topic. Therefore, there is no previous sentence or context from which the listener can infer the subject. An explicit pronoun therefore helps clarify the subject of the VP.

Comparison of subject element used in NEW TURN tokens by subject

	NP	Pronoun	Ø	Total	%Ø
1PL	4	19	53	**76**	70%
1SG	0	75	118	**193**	61%
3PL	20	9	46	**75**	61%
2SG	0	11	11	**22**	50%
3SG.M	28	6	31	**65**	48%
3SG.F	17	2	15	**34**	44%
Total	**69**	**122**	**274**	**465**	**59%**

Table 50. Subject of VPs used in NEW TURN tokens (HLVC).

In addition to an explicit pronoun, the speaker can use a noun phrase (NP) to disambiguate the subject for 3rd person subjects. A 3rd person subject can be referenced with someone's name, demonstratives (e.g., 'those guys'), and other noun phrases to specify the subject of a VP. Although they were not included in the dependent variable, NPs are often used at the start of new turns to disambiguate a 3rd person subject, as Table **50** shows.

This is not possible for 1ˢᵗ and 2ⁿᵈ person singular VPs, which explains their higher rate of overt pronouns. A minor exception is 1PL, which can be used with a NP to specify who is the subject of a VP, as shown in example (**52**):

52)   [kɔ   la   muʎa stavam   a wɪndzʌɹ]
         with the wife   stay.IMP.1PL at Windsor
         Con mia moglie stavamo a Windsor
         'with my wife, (we) stayed in Windsor'
         (CHILS, Angelo, 1:14:09)

However, 1PL subjects have a higher null rate than any other subject. This can be explained by its unambiguous verbal suffix, which makes the subject clear for the listener, even at the start of a new turn.

It is also important to note that for the separate analyses, SWITCH REFERENCE was second to SUBJECT in terms of effect size. However, in the combined analysis it is the factor with the largest effect.

6.1.3.2. Subject

The SUBJECT variable is the factor with the second-largest effect. The ranking of the variants' FWs suggests that subjects that represent people who are present for the interview, 1ˢᵗ and 2ⁿᵈ person singular (speaker and interviewer), are less likely to be represented by null subject pronouns (i.e. have a FW near or below 0.5); whereas 3ʳᵈ person subjects, who are likely not present at the time of the recording, are more likely to occur in null subject VPs. However, there are two variants do not follow this pattern of 3ʳᵈ person subjects favouring null subjects, and 1ˢᵗ and 2ⁿᵈ person subjects disfavouring null subjects: 3SG.F and 2PL subjects. For VPs with a 3SG.F subject, this is only true for the heritage speakers, while the homeland speakers had 3SG.F in the

variant hierarchy next to 3SG.M, with both favouring null subjects. Possible reasons for these outliers are proposed below.

Throughout the recordings, the 1SG subject pronouns are used to foreground the participant in each narrative, or to introduce an opinion or thought outside their personal stories. Overt subject pronouns for 2SG subjects signal a shift from recounting stories, which have 3rd or 1st person subjects, to addressing the interviewer. This often occurs when the participant wants to confirm the interviewer understands, remembers, or knows what the participant is speaking about.

There were six tokens of 2nd person plural subjects; however, all six occurrences had null subjects, so those VPs were excluded from the final analysis. Each of these tokens is quoted speech, as shown in (**53**). 1PL slightly favours null subjects (FW = 0.52), and is ranked above 1SG and 2SG, both of which disfavour null subjects.

53)   [na siɲora a   djɛtːo   vɔl   aseʤare i   pranz ke   avɛte fatːo]
        a   woman AUX.3SG say.PPRT Ø want.PR.1SG try.INF   the lunch that Ø aux.2PL make.PPRT
        Una volta na signora ha detto, "voglio assaggiare il pranzo che avete fatto"
        one time a woman said, "(I) want to try the lunch that (you [2PL]) had made"
        (HLVC, I1F73A, 18:47)

There are two potential reasons for this. The first is that the verbal inflection for 1PL, like that of 2PL, is unambiguous, meaning the VP's subject is less likely to be ambiguous to the listener (see **3.2.2**). With the disambiguating inflection of verbs with 1PL subjects, a subject pronoun is not needed to clarify the subject.

The second is that the 1PL is often used in a more generalized sense than 1SG. For example, in (**54**), the speaker is describing the way her community used to live—with their doors open to everyone. In terms of the division of those present and absent during the interview, the 1PL may be partially absent, but the 1SG is always present. This more generalized sense, and its distinctive verbal morphology, may explain why 1PL subjects slightly favour null subjects (FW = 0.52), and have such a different FW from 1SG (0.34).

54)   [vivevamo   kɔn lə   pɔrt apɛrte]
              live.IMP.1PL   with the door open.PPRT.PL.F
              vivevamo con le porte aperte
              '(we) lived with the doors open'
              (HLVC, IXF51A 18:56)

Third person subjects are more complicated. 3SG.M has the highest FW for null subjects (0.66), with 3PL being very similar (0.65). 3PL.M and 3PL.F were combined in the analysis because there were very few feminine plural subjects (n = 22), the two subjects had similar rates of null subjects (89% for 3PL.M, and 92% for 3PL.F), and the subject pronoun is the same (*loro*). This shows that the subject pronoun paradigm is like that of standard Italian, without a distinction for gender for 3PL subjects (see **3.2.1**). This high rate of null subjects for 3rd person subjects is at least partially accounted for by the options, other than an overt subject pronoun, to refer to these parties. As described in the previous section, speakers can use proper nouns or NPs, instead of a subject pronoun, to disambiguate the subject.

However, the 3SG.F does not fit this pattern. In the analysis, 3SG.F slightly disfavours null subjects (FW = 0.49). This requires further investigation because these subjects might be expected to behave similarly to 3SG.M as they have identical verbal morphology. This position of

3SG.F subjects in the variant hierarchy is also found in the heritage analysis, but not in the homeland analysis, where it is almost identical to 3SG.M (FW = 0.60 and 0.59, respectively). A possible explanation for this difference in FW between 3SG.M and 3SG.F is that the masculine variant is the presumed 3rd person singular subject (Heap, 1997:208). According to this hypothesis, VPs with a null subject and a 3SG subject agreement are assumed to be masculine, and an overt subject pronoun is used to disambiguate 3SG.F from 3SG.M. This possibility is supported by the data presented in Table **50**, which shows that new turns with a 3SG.F subject are also more likely to have an NP subject (50%) than those with a 3SG.M subject (43%). This is only a possibility, and further research will need to be done to determine why 3SG.F patterns so differently for heritage speakers than for homeland speakers.

6.1.3.3. Preverbal Element

The final significant factor from the analyses is PREVERBAL ELEMENT: the presence of an object clitic, negative marker, or reflexive pronoun. Reflexive pronouns already make person and number of the subject explicit, so an overt subject pronoun does not further clarify the subject, except for 3rd person subjects, which share a common reflexive pronoun for 3SG.F, 3SG.M, 3PL.F, and 3PL.M subjects. However, the presence of any one of these elements favours a null subject.

Object clitics and negative particles also occur preverbally and, as discussed in **3.3.2.2**, have been found to co-occur with a higher null rate (Heap, 1997:182-183).

Preverbal elements in the HLVC analysis

	n	%Ø	FW
OBJECT PRONOUN	293	84%	0.53
REFLEXIVE PRONOUN	157	82%	0.51
NEGATIVE PARTICLE	175	82%	0.51
N/A	1,009	78%	0.45

Table 51. Categories of preverbal elements, and the ranking of the variants in the HLVC analysis (n=1,634).

Before combining the preverbal elements into just two variants (presence or absence), null subject rates for each type of preverbal elements were compared. Table **51** shows the range across these preverbal types. The delta between the categories of preverbal elements is small (0.02). In fact, reflexive pronouns have the same FW as negative particles, despite reflexive pronouns clarifying PERSON in all cases, and both PERSON and NUMBER for 1st and 2nd person subjects.

After multiple Rbrul analyses, and different methods of collapsing variants in this factor group, the most reliable analysis came from combining all the types of preverbal elements into a single variant, and the absence of any preverbal element as the other variant, as shown in Table **49**.

6.1.3.4. Speaker

Speaker was run as a random intercept in the analyses (see **5.4.2**). As a random intercept, Rbrul accounts for the variation that exists across speakers for the dependent variable, the presence or absence of a null subject pronoun, in building a model of the significant independent variables that does not vary across speakers.

Table **52** shows the ranking of the speakers based on FWs. Homeland speakers are marked with an 'X' following the initial 'I' (e.g., I**X**M61A), and heritage speakers are indicated by a number representing their generation in this position (e.g., I**1**F91A is a first-generation heritage speaker). First-generation participants emigrated from Italy, second-generation participants were born in Canada to parents who are first-generation, and the children of second-generation parents would be third generation. However, as I selected the oldest participants from each group, there is only

one second-generation participant included in the analysis (see **5.1.1.1**). Because of this homogeneity in the heritage participants chosen for this study, generation was not a significant factor.

Variation in pronoun usage across HLVC participants

	n	%Ø	FW
IXM47A	69	90%	0.66
IXF38A	61	93%	0.59
I1F65A	80	84%	0.59
I1F61A	93	88%	0.59
I1M61A	95	91%	0.59
IXF35A	70	87%	0.58
I1M75A	86	85%	0.55
I2F57A	89	80%	0.51
IXM52A	83	78%	0.51
IXM64A	89	82%	0.51
I1M62A	80	83%	0.51
I1F71A	89	82%	0.51
I1M60A	97	77%	0.47
IXF61A	63	78%	0.46
IXF94A	78	77%	0.46
IXM35A	85	72%	0.44
IXM61A	68	71%	0.41
IXF51A	82	74%	0.40
I1M61B	90	73%	0.39
I1F73A	87	61%	0.30

Table 52. Inter-speaker variation in the HLVC analysis.

The variation between speakers (from 61% to 93% Ø) may be due to each participant's specific speech style, or it may be due to the topics that were discussed during the interview. If, for example, one participant spent much of the interview discussing their brother, they would have more 3SG.M subjects and many instances of the referent remaining the same (SAME REFERENT). These two factors, as shown above, would favour null subjects, leading the participant to have a higher rate of null subjects. However, by including the speakers as a random effect, this is accounted for in the ranking of participants by FW. In other words, the inter-speaker variation in Table **51** is distinct from what can be ascribed by the significant variables. This explains why there are discrepancies in the hierarchy between overall null rate and FW. For example, participant IXF38A has a much higher null rate (93%) than I1F65A (84%), yet they have the same FW (0.59), which is less than the FW of IXM47A (0.66). This shows that, removing the effects of SWITCH REFERENCE, SUBJECT, and PREVERBAL ELEMENT, IXM47A favours null subjects the most, even though he has the third-highest modeled rate.

The homeland and heritage groups are evenly distributed across the range of the Table **52**.

Similarly, males and females (denoted by *M* or *F* in the speaker code) are not clustered at either

end of the range. AGE also is not significantly predictive for subject, with a 47-year-old having

the highest FW (IXM47A) and a 51-year-old having the third-lowest FW (IXF51A), for

example. These findings suggest that subject pronoun usage does not vary significantly based on

the participants' gender, age, or community (homeland vs. heritage) in this corpus.

To further confirm that there were no significant differences between the homeland and heritage

analyses, I conducted one more analysis that included the significant variables listed above and

included a variable for the varieties. This variable had two variants: HOMELAND and HERITAGE.

In this analysis, the FW for both was 0.5, showing there is no significant difference between the

two corpora (see Appendix **H. 5**).

6.2. AIS Results

Table **53** is a one-level analysis of the significant variables of the AIS data. The overall null rate

is 76%, and the dataset contains 583 tokens taken from the six locations. As described in Table

53, 149 tokens had to be removed from the dataset because the gender of the 3rd person subject

could not be determined from the response provided by the participant (3SG.AMBI) with null

subjects used categorically.

Three variables are significant in the Rbrul analysis: whether the prompt provided to the

informant for translation had an overt subject pronoun (PRONOUN IN PROMPT); the person,

number, and gender of the subject of the phrase (SUBJECT); and whether the verbal morphology is

ambiguous for person (AMBIGUOUS). Also, LOCATION was included as a random effect to account

for inter-speaker variation.

One-level analysis of significant variables for AIS participants

	n:	583	%∅:	76%

Pronoun in prompt

	n	%∅	FW
ABSENT	465	89%	0.78
PRESENT	118	27%	0.22
		Range	**56**

Subject

	n	%∅	FW
3SG.AMBI	(149)	(100%)	N/A
3SG.F	38	97%	0.75
1PL	53	89%	0.60
2SG	59	71%	0.58
3PL	67	93%	0.53
2PL	44	66%	0.45
1SG	240	77%	0.37
3SG.M	82	52%	0.23
		Range	**52**

Ambiguous verbal form

	n	%∅	FW
DISTINCTIVE	465	87%	0.63
AMBIGUOUS	118	34%	0.37
		Range	**26**

Speaker (Random)

Std. dev.	0.20

Table 53. The significant variables from a one-level analysis of the AIS.

The AIS corpus is not made up of semi-directed interviews like the HLVC and CHILS corpora. Instead, data was collected by an interviewer who provided prompts in standard Italian and asked the participant to translate into Ciociaro (see **5.1.2**). Thus, the priming variables of the HLVC and CHILS analyses (e.g., SWITCH REFERENCE) are not included in the analysis. However, the variable PRONOUN IN PROMPT, while not equivalent to a factor in the CHILS or HLVC analyses, is a priming variable of sorts, and is described next.

6.2.1. Pronoun in prompt

The variable with the biggest effect in this analysis is whether there was an explicit pronoun in the prompt. While in almost all cases a subject pronoun is optional in both Ciociaro and SI, when asked to translate a prompt that contains an overt pronoun, the participant is much more likely to reproduce that pronoun in their response (27% null) than if there is no pronoun in the questionnaire prompt (89% null).

However, there are exceptions. For instance, there are cases where an explicit pronoun in the prompt is dropped by the participant, or where a prompt has a null subject pronoun that is

162

realized in Ciociaro. In example (**55**), informants were given the prompt #649 *dormi già?* ('[are] [you] sleeping already?'), which has a null subject:

55)	**Form**	**Location**
	[ʤa dormi]	Serrone (654)
	[**tu** tormi]	Veroli (664)
	N/A	San Donato (701)
	[**tɕu** tormi]	Ausonia (710)
	[dwormə]	Scanno (656)
	[**tu** durme]	Sonnino (682)
	Dormi già?	
	(are) (**you**) sleeping already?	
	(AIS, prompt #649)	

In this example, when given a prompt with a null subject: three informants produced an overt subject pronoun, two did not add a pronoun, and one response is missing from the dataset. It is not clear why those three informants added an overt subject pronoun when translating a prompt with a null subject. It is not to disambiguate a subject, because the three participants who produced overt pronouns also produced unambiguous verbal inflections (while the participant from Scanno did not). Regardless, it is an exception to the general effect of a prompt with a null subject, which is that the participants almost always also produce a null subject.

On the other hand, when a prompt contains an overt subject pronoun, 27% of the responses from participants are null. This is one of the complications of the AIS dataset when compared to the HLVC and CHILS. Each prompt, and any lexical or semantic effects contained within that phrase, is amplified within the results, because each prompt has up to six tokens representing the different participants' responses.

There are 32 tokens where an overt subject pronoun that was included in the prompt is removed in a participant's response: 20 for a 2SG or 2PL subject, nine for 1SG, and three for 3SG.M. However, these 32 tokens come from only 11 prompts. Two prompts in particular, #1624 *voialtri chiudete la porta* ('**2PL** close the door'), and #1346 *bada! **Tu** versi il vino* ('Look! **2SG** are spilling the wine'), have overt pronouns in the prompt. However, all six informants produced a null subject for #1624, and for #1346 four informants produced a null subject, one response is missing, and one kept the overt subject pronoun.

These examples provide potential reasons as to why the subject pronouns were removed in the responses. Response #1624 has a 2PL subject, and the +/ete/ suffix is sufficiently distinctive to avoid any ambiguity of the subject, meaning an overt subject pronoun does not further disambiguate. Also, while #1346 is not an imperative phrase, it does follow an imperative (*bada!* ['look!']), and it is an exclamation of surprise or a reprimand that someone is spilling wine. This may result in fewer overt subject pronouns because the subject is understood from the context.

Aside from these exceptions though, the vast majority of the responses from the AIS participants match whether the prompt contained an overt subject pronoun or not.

6.2.2. Subject

The second significant variable in the analysis is the subject of the phrase. The ranking of the variants for SUBJECT in the overall analysis above appears to be without a clear pattern: 3SG.F has the highest FW (0.75), while 3SG.M has the lowest (0.23). First and 2nd person subjects also appear to have no clear order. This also is counter to the results of the HLVC analysis, where 3SG.M had the highest FW (0.66), with 3rd person subjects favouring null subjects and 1st and 2nd person subjects disfavouring.

This ranking of the variants can be explained, in large part, by interactions: for the 3rd person singular, the prompts themselves often have other grammatical information about the subject, and the variables interact strongly within the model. As explained in **5.2.2**, many of the 3SG prompts do not indicate whether the subject is masculine or feminine. These were coded as 3SG.AMBI. For example, the prompt *non ha mai fretta* ('[3SG.M/F] is never in a hurry') contains no information about the subject of the phrase (Jaberg & Jud, 1928:1606). To this prompt, two of the six participants responded with an added overt 3SG.M pronoun (Serrone used [isːo] and Scanno added [ɣisːə]). Therefore, while these two responses were not coded as 3SG.AMBI, the other four responses with null subjects were. Because these tokens have a null subject and form their own group (3SG.AMBI) for SUBJECT, they create an invariably null group that cannot be included in the overall analysis. Within the dataset there are 149 responses across 29 prompts that were coded in this way and excluded from the analysis. Thus, the 3SG.M tokens have such a low rate of null subjects because, at least in part, many of the null 3rd person subjects are not included in the analysis.

Conversely, 3SG.F subjects have a high rate of null subjects (n = 37/38) because all seven of the prompts with a feminine singular subject have agreement after the verb that makes the subject clear. Prompt #74, for example, *è una buona serva* ('[she] is a good maid'), does not contain an explicit pronoun in any of the responses, but the gender is not ambiguous because of the agreement on the NP that follows the verb, contrasted with *è un buon servo* for a 3SG.M subject. The only response with an explicit 3SG.F pronoun is from prompt #707, *è guarita* ('[she] is healed/recovered'), where *guarita* contains agreement for 3SG.F (versus *guarito* for a male

subject). The participant from Scanno gives the response: [ɣɛsːɑ ʧe rətsɛtɑ].[32] This is the only example of a 3SG.F pronoun in the AIS dataset. It appears that [ɣɛsːɑ] is a variant of the 3SG.F pronoun [ɛsːa] that is attested in neighbouring dialetti (see **3.2.1**).

Therefore, the very different FWs for 3SG.M and 3SG.F subjects can be attributed to the ambiguous 3rd person singular subjects that are not included in the analysis. A 3SG response with a null subject could only be disambiguated by agreement elsewhere in the phrase, or if the prompt itself contains an overt pronoun. There are seven 3SG.F prompts, none of which have an overt pronoun. There are 14 3SG.M prompts, six of which have an overt pronoun in the prompt. Taken in consideration with the factor PRONOUN IN PROMPT, described in the preceding section, this has an important effect on the null rate of both subjects.

One-level with SUBJECT as the sole variable

	n	%∅	FW
3SG.AMBI	(149)	(100%)	N/A
3SG.F	38	97%	0.89
3PL	67	93%	0.72
1PL	53	89%	0.62
1SG	240	77%	0.40
2SG	59	71%	0.34
2PL	44	66%	0.29
3SG.M	77	52%	0.19

Table 54. Results of one-level analysis from the AIS corpus with SUBJECT as the sole variable.

Another issue to be addressed in the ranking of the variants of SUBJECT is the interaction of this factor with the other significant factors. Table **54** shows the ranking of the variants within the factor when it is run in Rbrul as the sole variable. This table shows a ranking of variants that is similar to the one presented in the HLVC analysis: 3rd person is the most likely to be null, followed by 1st person and finally 2nd person,

except the outlier of 3SG.M that has already been addressed. Further, in both the HLVC analysis and the table above, 1PL is the only non 3rd person subject to favour null subjects (FW = 0.62). In

---

[32] The adjective [rətsɛtɑ] is a dialectal term in Ciociaro of unclear etymology meaning 'better/healed'.

the analysis of significant variables that is presented at the start of this section, the ranking of the variants for SUBJECT is very different from the ranking in Table **54**. For example, 3PL has one of the highest FWs when SUBJECT is the only variable (0.72); however, in the overall analysis, its FW falls to (0.53). This is explained by the interaction between SUBJECT and AMBIGUOUS, and it is addressed in the following section.

6.2.3. Ambiguous verb forms

The final significant factor from the analysis is whether the participant uses a verb with an ambiguous inflection. In **5.3.2.4**, I describe this variable and its variants—the different kinds of ambiguity that occur in the dataset. However, I collapsed the variants into a binary distinction between AMBIGUOUS and DISTINCTIVE when conducting the multivariate analysis. This allowed for a more reliable model.

Ambiguous verb forms disfavour null subjects (FW = 0.37), whereas unambiguous verb forms favour null subjects (FW = 0.63). In other words, when a verbal inflection is not distinct for subject, an overt pronoun is more likely. This supports the functional hypothesis because when the verb can be relied upon to determine the subject, a null subject is favoured, but when the verb is ambiguous for subject, a subject pronoun is more useful and more likely to be used.

A token was coded as ambiguous if its inflection or verb form was also used for a different subject in the same location (i.e., by the same participant). As an example, in (**56**) and (**57**) the participant from Veroli (664) gives the following responses:

56)   [adːo vaj]
Dove vai ?
Where (are) (you) going ?
(AIS, prompt #821, Veroli)

57)   [vaj a kombra]
Vado a comprare
(I) (am) going to buy
(AIS, prompt #822, Veroli)

These two tokens are coded as ambiguous because the verbs take identical forms despite having different subjects.

This factor also interacts with SUBJECT because certain grammatical subjects have a higher rate of ambiguous verbal morphology than others. For example, none of the 53 1PL tokens are ambiguous because the verb always ends in [m], [mə], or [mo]. Thus, a verb ending with an [m(V)] indicates 1PL. The responses for prompt #1646 are shown in (**58**):

58)

Response	**Location**
[ɛ poko dembo ke sim arivadi]	Serrone
[sem arːivati poku tɛmpu]	Veroli
N/A	San Donato
[sim arːivaʧi na tanʤo]	Ausonia
[semə nuntə mamaw]	Scanno
[semo arːevato mo]	Sonnino

Siamo arrivato poco fa
'(we) arrived a little bit ago'
(AIS, prompt #1664)

The 3PL tokens, however, are not uniformly unambiguous like the 1PL tokens. For example, *sono*, even in SI, can be both 'I am' and 'they are'. Therefore, while 3PL is ranked higher when subject is analysed as an individual variable, the 1PL tokens are given a higher ranking in the

168

overall analysis because, when combined with the AMBIGUOUS factor, they have a stronger

interaction with another significant variable than 3PL.

6.2.4. Location

One-level analysis of LOCATION as the sole variable

Location	n	%∅	FW
Ausonia	101	81%	0.57
Veroli	105	79%	0.54
Sonnino	105	78%	0.52
San Donato	72	78%	0.52
Serrone	99	74%	0.46
Scanno	101	67%	0.39

Table 55. Results of one-level analysis with LOCATION as the sole variable for the AIS analysis (n=583).

The extralinguistic variable included in the AIS analysis is the location, which represents the participant from each of the villages included in the analysis. This is included as a random intercept, as SPEAKER was for the HLVC and CHILS analyses.

Table **55** shows the overall rate of null subjects across the six locations included in the analysis, as well as

their FW. Each location is represented by a single speaker, so extralinguistic factors such as age or gender cannot be included in this analysis.
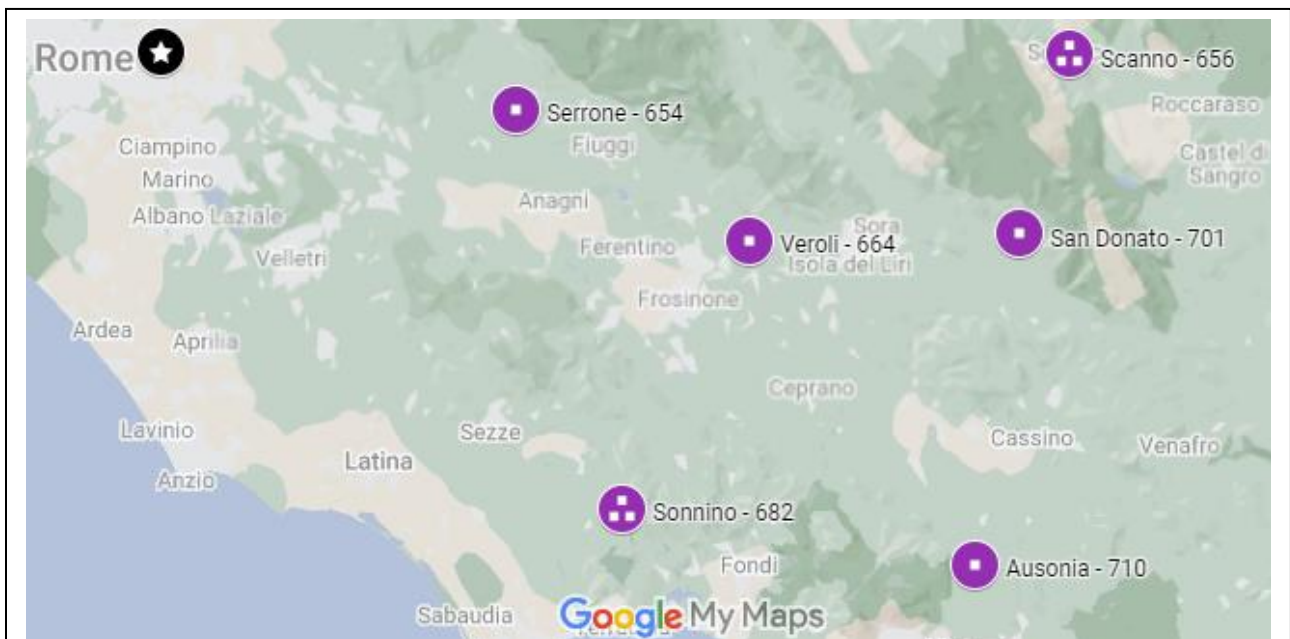
Figure 10. Map showing the location of each of the six locations included in the AIS analysis, two outside Frosinone marked with three squares (Google, n.d.-a).

169

Four of the six participants fall within a 3% range of null subjects (81% - 78%) and favour null subjects, with Serrone and Scanno disfavouring null subjects. However, when compared to Figure **10**, there is no obvious pattern between the locations and their null rates (Google, n.d.-a). Serrone (location #654) and Scanno (location #656) are not geographically closer to each other than they are to Veroli or San Donato.

When comparing each location's rate of ambiguous verb forms, there does not appear to be an effect either, as shown in Table **56**. The Spearman's rho test, included in the table, suggests a slight negative correlation where an increase in ambiguity leads to fewer null rates, but the p-value does not reach the level of significance. The lowest null rate (67%) is from Scanno, which also has a high number of ambiguous verbs forms (31%). However, San Donato has the highest rate of ambiguity at 32%, but it has an 11% higher null rate (78%). Further, the remaining four locations all have nearly the same number of ambiguous verbal forms (15-16%).

Comparison of the null rate and the ambiguous verb forms of each location

Location	n	%Ø	%Ambi
Ausonia	101	81%	16%
Veroli	105	79%	15%
Scanno	105	78%	16%
San Donato	72	78%	32%
Serrone	99	74%	15%
Scanno	101	67%	31%

Spearman's r_s = -0.22; p = 0.67

Table 56. Results of the comparison of the null rate and ambiguous verb forms based on location from the AIS analysis (n=583).

The analysis of the AIS data presented here serves as a means of understanding the varieties of Ciociaro spoken in Italy prior to the emigration of the CHILS speakers. The factor PRONOUN IN PROMPT does not have an equivalent in the recordings used for CHILS; however, the ambiguous verb forms and the significance of SUBJECT across both the HLVC and AIS are compared to CHILS in the following section.

6.3. CHILS Results

As described in **5.1.3.1**, 10 male participants and 10 female participants from the CHILS corpus are analysed, ranging in age from 57 to 95.

One-level analysis of pro-drop in the CHILS corpus			
n:	*1,736*	*%Ø*	72%
Subject			
	n	*%Ø*	*FW*
(2PL)	(10)	(100%)	N/A
3PL.M	374	93%	0.82
3SG.F	116	77%	0.57
3PL.F	20	75%	0.55
1SG	518	68%	0.45
3SG.M	464	64%	0.40
2SG	51	59%	0.34
1PL	193	57%	0.32
	Range		**50**
Previous Realization			
	n	*%Ø*	*FW*
NULL	773	77%	0.59
NEW	590	69%	0.51
OVERT	272	67%	0.51
NOUN	101	58%	0.40
	Range		**19**
Tense			
	n	*%Ø*	*FW*
(SUBJUNCTIVE)	(10)	(100%)	N/A
(FUTURE)	(2)	(100%)	N/A
IMPERFECT	804	76%	0.56
PRESENT	319	69%	0.53
ABSOLUTEPAST	353	72%	0.50
COMPOUNDPAST	260	62%	0.41
	Range		**15**
Speaker (Random)			
Std. Dev.			0.31

Table 57. Significant variables of pro-drop realization in the CHILS corpus.

The same independent variables that are included in the analysis of the HLVC dataset are considered in the CHILS analysis. A further two variables, FREQUENT VERBS and VERB CLASS, were included in separate runs to avoid interactions, but neither was significant. They are briefly described later in this chapter (**6.3.5**).

Table **57** shows the results of the multivariate analysis. Across 20 speakers, the number of tokens (*n*) was 1,736. The overall rate of subject pronoun omission is 72%, 8% less than the HLVC dataset, and 4% less than the AIS analysis. The differences and similarities between the datasets will be discussed further in the following section (**6.4**). From the table we can see that the three significant variables are: SUBJECT, PREVIOUS REALIZATION, and TENSE.

The participants were included as random intercepts to account for each person's idiolect, just as SPEAKER was included in the HLVC analysis, and LOCATION was included in the AIS analysis.

6.3.1. Subject

The SUBJECT variable has the largest effect in the analysis. The ranking of the variants shows that 3rd person subjects are ranked as the most likely to be null, with FWs that favour null subjects. 1st and 2nd person subjects are the least likely, with FWs that disfavour null subjects, which is broadly the same pattern found for the AIS and HLVC analyses. However, there are two exceptions to this. First, the 10 tokens with a 2PL subject, which are categorically null, were not included in the analysis. There is also the exception of 3SG.M, which is the only 3rd person subject that disfavours null subjects. 1SG has a lower FW than 3PL and 3SG.F, but higher than 3SG.M.

Why there is this difference between 3SG.M and 3SG.F requires more analysis. As shown for the AIS results (**6.2.2**), 3SG.F subjects often disambiguate the subject through other means, such as agreement on the past participle or an adjective. Therefore, an overt subject pronoun is not as useful for VPs with a 3SG.F subject as there are other means of recovering the subject from the discourse. However, an analysis looking for interactions with AMBIGUITY was not significant, and both 3SG subjects are ambiguous 33% of the time.

Finally, there were only 10 tokens of 2PL subjects—all of which had null subjects, so those VPs were excluded from the final analysis. Each 2PL token comes from narrative speech using the present tense. For the verbal paradigm, 2PL has a distinct verbal ending, with a final syllable +/te/. This /t/ sound makes these verb forms distinct, even with a complete elision of the final

172

vowel, as described in **3.2.2**. For example, in (**59**) the verb has no final vowel, but the inflection is unambiguous:

59) [**no** tut:i pɛnsiamə ke siɛt amɛɹikano]
we all think.PR.1PL that is.PR.2PL American
Noi tutti pensiamo che siete americani
'**We** all think that (you [2PL]) are American'
(CHILS, Assunta, 64:34)

6.3.2. Previous Realization

The PREVIOUS REALIZATION variable has the second-largest effect in this analysis. This variable shows the priming effects of the previous VP on subject pronoun realization for the current VP. In other words, if a subject pronoun is overt or null for a given phrase, it has an impact on the realization of the next VP—even if the referent of the VP changes.

In the results of the CHILS analysis, if the preceding VP had a null subject, the current VP favours a null subject (FW = 0.59). If the preceding VP had an overt subject pronoun, there is only a slight effect (FW = 0.51). The same is true at the start of a turn (FW = 0.51). Finally, if the previous phrase had a noun phrase subject, then a null subject is disfavoured (FW = 0.40).

Although NPs were not included as a variant in the dependent variable, clauses that had NP subjects were coded for their priming effect on the following VP token. Only NPs with human subjects were coded, such as *mia mamma* ('my mum') or *Carmela*, as described in **5.3.5.1**.

60) [kaɹmɛla l ɪnvito dɛntɹə kaza e **lui** dise]
Carmela OBJ invite.AP.3SG into house and he say.AP.3SG
Carmela l'invitò dentro la casa e lui disse
'Carmella invited them into the house, and **he** said'
(CHILS, Assunta, 58:12)

Example (**60**) shows why overt subject pronouns occur more frequently after a NP. Both VPs have a 3SG subject, so an overt pronoun clarifies a switch in referent to a different 3SG subject because the verb forms are the same.

At the other end of the ranking of this variable, if the previous VP had a null subject, then a null VP is favoured. This priming effect has been found in a number of previous studies see Torres Cacoullos & Travis (2019), whereby a null subject primes the next subject to be null. This effect is often strengthened when there is no change in referent in subsequent subjects. However, unlike the HLVC data, the CHILS results do not suggest a similar effect. The variable SWITCH REFERENCE was tested in this analysis but was not significant.

Interaction between SWITCH REFERENCE and PREVIOUS REALIZATION in the CHILS dataset

Prev. Real.	*SWITCH.REF*		*SAME.REF*	
	n	%Ø	n	%Ø
NULL	422	76%	353	79%
OVERT	87	61%	185	69%
NP	47	57%	54	59%
Total	**556**	**72%**	**592**	**74%**

Table 58. Interaction between SWITCH REFERENCE and PREVIOUS REALIZATION for the CHILS corpus.

When examining the two variables together, as shown in Table **58**, if the previous subject was null, a null subject is only 3% more likely if the referent is the same compared to a changed referent (79% and 76%).

Similarly, the other variants of PREVIOUS REALIZATION also have a slight increase in the null rate when the previous subject refers to the same referent, but the direction of effect is as expected, and is not significant.

Finally, both OVERT PRONOUN and NEW TURN have very similar null rates and have identical FWs (0.51).

6.3.3. Tense

The final significant variable is TENSE. From the ranking of the variants, VPs in the imperfect and present tense favour a null subject, VPs in the absolute past (*passato remoto*) neither favour nor disfavour null subjects, and VPs with a compound past tense, meaning with an auxiliary verb (*passato prossimo* or *trapassato prossimo*), disfavour a null subject. Finally, there are two tokens in the future tense and 10 tokens in the subjunctive, but all 12 are null and were not included in this analysis.

The ranking of these tenses reflects their relative ambiguity. With the reduction of word-final vowels in Ciociaro, the necessary distinctions between two inflected verbs in the compound past can be difficult to parse. In SI, *ha fatto* and *hanno fatto* ('s/he made' and 'they made'), are pronounced differently; however, in Ciociaro [a fatə] is often used for both inflections. Another example, which occurred many times in the interviews, is *sono venuto/a* ('I came') and *sono venuti/venute* ('they came' [3PL.M/3PL.F]). However, in Ciociaro this present perfect verb is normally pronounced /so wɛnudə/[33] regardless of the subject, corresponding to an increased rate of overt pronouns to disambiguate a 1SG and 3PL subject. This ambiguity explains why the compound past has the lowest FW of the tenses and disfavours null subjects.

As described for the results of the variable SUBJECT (**6.3.1**), the imperfect tense and 3rd person subjects interact due to the relative ambiguity of plural and singular subjects. Table **59** shows the interaction between number and tense. Specifically, the null rates have a clear pattern across the

---

[33] Four of the speakers pronounced both as [so mɛnutə], (see **3.2.2** and (Iannozzi, 2017) for a description of the variable realization of /v/ as [m] in Ciociaro)

Interaction of TENSE and SUBJECT in CHILS analysis

Tense	Subject	n	%∅
PRESENT	3SG.M	27	100%
	3PL.M	50	90%
COMPOUND PAST	3SG.M	93	55%
	3PL.M	29	62%
IMPERFECT	3SG.M	217	67%
	3PL.M	232	96%
ABSOLUTE PAST	3SG.M	128	60%
	3PL.M	63	100%

Table 59. Interaction of the variables of TENSE and SUBJECT for 3rd person subjects (CHILS, n=839).

tenses: present and compound past VPs have similar null rates for 3SG.M and 3PL.M subjects, and imperfect and absolute past tokens show much lower null rates for the singular than the plural. This difference in null rates between 3SG and 3PL for the imperfect and absolute past is because the verbal morphology is ambiguous for these two subjects in those tenses. However, for present and compound past VPs, the 3SG and 3PL forms tend to be distinct, which explains their similar null rates. Instead, ambiguity in the paradigms of these tenses occurs between 3PL and 1SG subjects, with *sono* appearing in most of those ambiguous tokens.

The TENSE variable does not interact only with 3rd person subjects. There is a mismatch in the hierarchy between PRESENT and ABSOLUTE PAST for FW and null rate. This is only a slight effect, where tokens coded as PRESENT have a FW of 0.53 with a 69% null rate, whereas those coded as ABSOLUTE PAST have a FW of 0.50 and a 72% null rate. This can be partially explained by the interaction of VPs in the present tense and with 1SG and 2SG subjects. As discussed in the previous section, 3rd person subjects broadly favour null subjects and 1st and 2nd person subjects disfavour null subjects.

Proportion of subjects within each tense

Person	Present	AP	CP	Imp
1+2	72%	31%	49%	37%
3	28%	69%	51%	63%

Table 60. Distribution of subjects across tenses for the CHILS corpus.

As Table **60** shows, the present tense is the only variant where the majority of tokens have either a 1st or 2nd person subject (during an interview, these subjects refer to either

the participant or the interviewer). For all the other tenses, VPs with a 3rd person subject are much more common. This interaction between the present tense and subjects that disfavour null subjects likely accounts for the slight mismatch of the variants for TENSE in the analysis. The remaining fixed independent variables were not found to be significant.

6.3.4. Speaker

Variation in pronoun usage across CHILS participants

Speaker	AOA[34]	n	%Ø	FW
Angelo F	17	90	87%	0.61
Rafaelle	32	94	82%	0.59
Frank D'A	19	91	79%	0.56
Isy	10	87	78%	0.54
Sante	28	74	76%	0.54
Caroline	6	85	78%	0.53
Nello	10	93	72%	0.52
Antonietta	26	88	77%	0.52
Palma	21	90	73%	0.51
Nunciata	27	77	75%	0.51
Annita	20	91	75%	0.51
Joanna	8	85	71%	0.50
Frank L	18	89	70%	0.49
Assunta	24	84	71%	0.48
Clara	17	80	69%	0.47
Ennio	20	85	67%	0.46
Renzo	19	100	62%	0.43
Ernie	10	84	60%	0.42
Arcangela	18	86	59%	0.41
Antonia	9	83	53%	0.37

Table 61. The speakers in the CHILS analysis with their Age of arrival (AOA) (n=1,736).

The participants were included in the analysis of the CHILS data as a random effect. This allows the modelling to account for variation across speakers, and any interactions an individual speaker might have with a variable. For instance, VPs with a 1PL subject have the lowest FW, and 18 of Ernie's 84 VPs have a 1PL subject, while only one of Frank L's 89 VPs have a 1PL subject. By treating the SPEAKER variable as a random effect, this variation is taken into account by the model.

There is a wide variation in the null rate among the speakers (53-87% Ø). While the model accounts for the variation across speakers and

the significant factors described above, it is tempting to ascribe the differences to extralinguistic

[34] AOA = Age of arrival in Canada as an immigrant

factors. However, interpreting what this range means will require further investigation and analysis. An explanation could be an increased contact effect with English for some of the speakers, such as from exogamous marriages. For example, Antonia is the speaker with the lowest null rate (53%), and she is married to an anglophone. She only speaks Ciociaro with her sister, Caroline, who is married to a Ciociaro speaker and has a null rate of 78%. While it might be tempting to suggest that Antonia's use of overt subject pronouns is influenced by contact with English, it is not that simple. Contrary to this hypothesis, Nello and Isy are also married to anglophones and have higher rates of null subjects (72% and 78%, respectively).

There are also three married couples within the dataset: Ernie (60%) and Joanna (71%) have rather different overall null rates, Antonietta (77%) and Sante (76%) have almost identical null rates, and Annita (75%) and Frank L (70%) have a delta that falls in between the two. Therefore, contact effects from exogamous marriages cannot explain the variation between the speakers.

It is also possible that participants in the CHILS corpus who arrived in Canada at a younger age would have a larger contact effect with English. Those who arrived prior to age 18 and received some schooling in English could experience more transfer effects on subject pronoun realization in Ciociaro. However, the age of arrival (AOA) of speakers who were included in the analysis is provided alongside their respective rates of null subjects. As can be seen, AOA does not correlate to a higher (or lower) null rate. This is similar to the findings of Limerick (2019:263), who found that length of residency in the United States, age of arrival, and English proficiency were not related to a significant change in subject pronoun realization for heritage Spanish speakers. Further, as discussed in **2.2.3.2**, most of the Ciociaro of Sarnia have had very little to

Variation in pronoun usage across CHILS participants compared to ambiguous verbal forms used

Speaker	n	%Ø	%Ambi
Angelo F	90	87%	43%
Rafaelle	94	82%	52%
Frank D'A	91	79%	38%
Caroline	85	78%	42%
Isy	87	78%	33%
Antonietta	88	77%	27%
Sante	74	76%	42%
Annita	91	75%	31%
Nunciata	77	75%	34%
Palma	90	73%	33%
Nello	93	72%	29%
Assunta	84	71%	13%
Joanna	85	71%	35%
Frank L	89	70%	42%
Clara	80	69%	30%
Ennio	85	67%	34%
Renzo	100	62%	32%
Ernie	84	60%	21%
Arcangela	86	59%	28%
Antonia	83	53%	33%
Average	**87**	**72%**	**34%**

Spearman's r_s = 0.52, p= 0.02

Table 62. Variation in null pronoun use compared to ambiguous verbal forms across CHILS participants (n=1,736).

no contact with standard Italian, either in Canada or in Italy. Therefore, any potential contact effects from SI are not relevant to this dissertation.

Finally, a link between more ambiguous verbal forms produced by the participant and a lower null rate would account for the variation. However, as can be seen in Table **62**, this is not the case. The Spearman's coefficient shows the opposite effect. Speakers who use more ambiguous forms actually have a significantly higher null rate, which is the opposite of the expected result. Therefore, the direction of effect runs counter to the functional hypothesis.

So, further research is needed to determine what may account for the variation of overt subject pronoun usage among these participants.

6.3.5. Verbal semantics

As described in the methodology chapter, VERB and VERB CLASS were only coded for the CHILS dataset. Coding for these variables in the HLVC and AIS datasets can be explored in future research. Although this variable was not found to be significant in the overall analysis, there are meaningful conclusions we can draw from looking at it in more detail. As shown in the table of comparable studies (Table **14**), six of 13 studies examined VERB CLASS, with four of these studies

179

finding it significant. One study looked at FREQUENT VERBS, but frequency of verbs was not a significant factor.

Most common verbs of CHILS dataset		
Verb	*n*	*%Ø*
parlare	87	78%
fare	137	78%
arrivare	62	76%
sapere	51	76%
vedere	70	74%
lavorare	56	73%
volere	59	69%
andare	71	68%
venire	86	67%
abitare	45	67%
dire	92	63%
avere	43	60%
essere	67	60%
tenere	63	57%
pensare	36	47%
total	**1,025**	**68%**

Table 63. Variation of null rates across the most common verbs (CHILS).

From those studies, the generalized finding of VERB CLASS is that 'cognition' verbs (e.g., *pensare* 'to think') favour overt subject pronouns and 'dynamic' verbs (e.g., *fare* 'to do') favour null subjects, with 'stative' verbs (e.g., *abitare* 'to live in/inhabit') falling between the two other groups (Torres Cacoullos & Travis, 2019:676). Further, in terms of FREQUENT VERBS, Erker & Guy (2012) examined high-frequency verbs and their interaction with other variables, such as SWITCH REFERENCE, TAM, and SUBJECT. They found a more nuanced effect: "all other constraints on pronoun use are weak or nonexistent among low-frequency items, but strong and significant among high-frequency items" (2012:548). In other words, the verbs that occur the most frequently amplify the significance of other variables.

In coding this variable in the CHILS data, I first identified 35 verbs that occurred more than 10 times, representing 79% of the tokens (n=1,368). Here, I look at the 15 most frequent verbs. Table **63** shows these verbs which account for 56% of the total tokens of the dataset. There appears to be a significant relationship between the specific verb and pro-drop. The dark line between the verbs *lavorare* ('to work') and *volere* ('to want') represents the overall null rate of the CHILS analysis (72%), thus those verbs above this line have a higher-than-average rate of

180

null subjects, and those below have a lower rate. From these frequent verbs, it appears that verb class is not a useful variable for understanding pro-drop in the CHILS data.

For example, in previous research outlined in **5.3.4.5**, verbs like *pensare* ('to think') and *sapere* ('to know') are commonly assigned to the 'cognition' class of verbs. Based on the findings of Carvalho & Child (2011) and Torres Cacoullos & Travis (2019), this class could be expected to disfavour null subjects, yet the opposite appears to be true. Though, it is important to note that these are simply null rates, and not weighted in a multivariate analysis.

As shown in this section, verb frequency and verb class may correlate to pro-drop in Ciociaro. A more fine-tuned analysis, such as laid out by Erker & Guy (2012), is an area for future research. Specifically, their finding that VERB CLASS and SUBJECT are not significant variables for infrequent verbs, but only for high-frequency verbs, is an exciting area for future research. This is of particular interest because it appears to be related to ambiguous verbal forms, specifically that "rarer forms provide insufficient or unreliable evidence about lexically specific aspects of variable processes—in this case, the cooccurrence of a verb with an overt subject pronoun" (Erker & Guy, 2012:553). Therefore, we would expect that infrequent verbs would disfavour null subjects. This, and other areas of future research are presented in **8.3**.

6.4. Combined results

In the previous sections, we've seen the analyses of the separate corpora. In **6.1.3**, I presented the analysis of the combined datasets for homeland and heritage Calabrian Italian. In this section, I show the results of the combined Ciociaro corpora, AIS and CHILS, to see how those results compare to their separate analyses, how the participants of the two datasets compare, as well as the appropriateness of combining the datasets. There are important differences in the

methodologies of these two corpora, but useful information can nonetheless be learned from the combined analysis.

Following this, the final section is an analysis of the combined datasets for Calabrian Italian, homeland Ciociaro, and heritage Ciociaro. This analysis provides more insight into the shared significant factor, SUBJECT, and how the participants compare across datasets.

6.4.1. Combined Ciociaro results

One-level analysis of pro-drop in the CHILS & AIS corpora			
Tokens (n):	2,287	%Ø	73%
Subject			
	n	%Ø	FW
2PL	(54)	(72%)	N/A
3PL	461	92%	0.81
3SG.F	154	82%	0.60
1SG	758	71%	0.43
2SG	120	68%	0.39
1PL	248	64%	0.38
3SG.M	546	63%	0.36
	Range		**45**
Corpus			
	n	%Ø	FW
AIS	539	77%	0.55
CHILS	1,748	72%	0.45
	Range		**10**
Speaker (random)			
Std. dev.			0.33

Table 64. Analysis of pro-drop in both CHILS and AIS corpora with SUBJECT and CORPUS as variables.

The significant variables of the AIS analysis are: PRONOUN IN PROMPT, SUBJECT, and AMBIGUOUS. For the CHILS analysis, they are: SUBJECT, PREVIOUS REALIZATION, and TENSE. To conduct an analysis that combines these two corpora, only SUBJECT, their shared significant variable, is included as a linguistic variable. The other variable included in this analysis is CORPUS, with the two variants AIS and CHILS. This variable tests whether the differences between the corpora are significant. Finally, SPEAKER is included as a random intercept. The results presented in Table **64** show an overall null rate of 73% from 2,287 tokens (539 from the AIS dataset and 1,748 from CHILS).

The analysis selected both independent variables as significant, SUBJECT and CORPUS.

The SUBJECT factor had to be modified to accommodate the differences in the two datasets. First, 2PL tokens had to be excluded from the analysis as they are invariably null in the CHILS dataset

182

and would have created an interaction between SUBJECT and CORPUS if included. Second, the 3PL.M and 3PL.F tokens had to be collapsed into 3PL in the CHILS corpus to align with the variants included in the AIS analysis. Despite these modifications, the ranking of variants is consistent with the separate AIS and CHILS results: 3rd person subjects favour null subjects, except the outlier of 3SG.M, as in the CHILS corpus, and 1st and 2nd person subjects disfavour null subjects.

CORPUS was also selected as a significant factor in the analysis. This does not necessarily mean that the varieties are too distinct or that homeland and heritage have drifted apart due to contact with English. Rather, the differences in data collection between the semi-directed and prompt-and-response interviews may account for this. However, these results show that combining the AIS and CHILS into a single Ciociaro corpus is not justified.

Finally, SPEAKER was included as a random intercept, with the results presented in Figure **11**.[35] The distribution of participants above and below 0.5 shows that the CHILS and AIS speakers vary in favouring and disfavouring null subjects. From the CHILS participants, 12 of the 20 favour null subjects, with eight disfavouring. This is nearly the same proportion as the AIS participants, with four out of the six favouring null subjects. There is no pattern suggesting an effect of AGE, which is confirmed by the analysis. These results point to CORPUS being selected as a significant variable due to interaction with SUBJECT, and not because the participants individually behave in significantly different ways. For example, 1PL subjects has a FW of 0.60 for the AIS analysis, and 0.31 for CHILS. These differences in the two datasets and the

---

[35] Complete Rbrul analysis is provided in Appendix **H. 3**.

significant factor of CORPUS in this combined-corpus model mean that it is not appropriate to combine the AIS and CHILS to compare a Ciociaro dataset to the combined Calabrian Italian.

Figure 11. Scatterplot showing FWs of AIS and CHILS participants from the combined analysis, arranged by age.
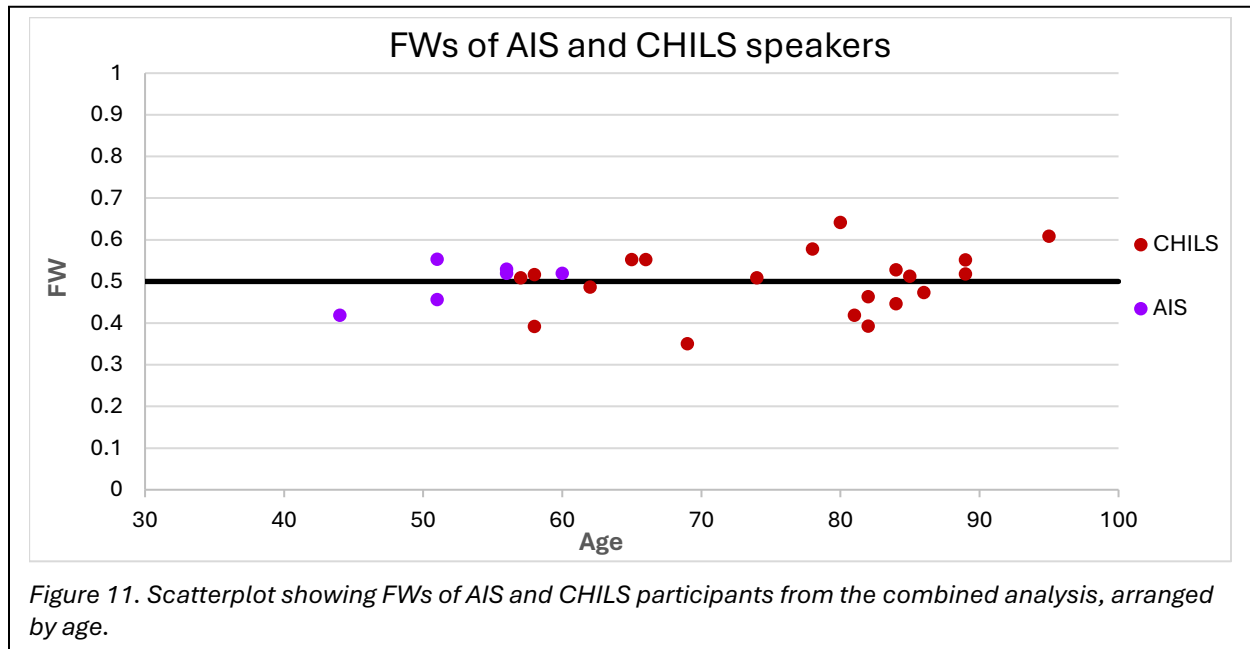
6.4.2. Combined HLVC, AIS, and CHILS results

Thus far, I have shown the differences between the corpora included in this study through separate analyses. I have also tested whether combining the corpora to provide larger analyses was justified.[36] In the case of homeland and heritage Calabrian Italian, it is justified. However, in the case of homeland and heritage Ciociaro, represented by the AIS and CHILS corpora, it is not. Therefore, in this section I present a combined analysis of HLVC, AIS, and CHILS datasets. The results of this analysis are included in Table **65**. The results show an overall null rate of 76% and a total of 3,921 tokens. Also, as in the previous section, the only independent variable that was

---

[36] See Appendix **H** for analyses of the remaining combinations that were not described in this chapter: AIS & HLVC, and CHILS & HLVC.

significant in each of the separate analyses was SUBJECT, which is included in the combined

analysis along with CORPUS and SPEAKER included as a random intercept.

One-level analysis of pro-drop across all corpora			
Tokens (n):	*3,921*	*%Ø*	*76%*
Subject			
	n	*%Ø*	*FW*
2PL	(60)	(75%)	N/A
3PL	713	91%	0.77
3SG.F	245	82%	0.58
1PL	544	75%	0.44
3SG.M	772	71%	0.42
1SG	1,416	72%	0.40
2SG	231	69%	0.37
	Range		**40**
Corpus			
	n	*%Ø*	*FW*
HLVC	1,634	80%	0.56
AIS	539	77%[37]	0.52
CHILS	1,748	72%	0.42
	Range		**14**
Speaker (random)			
Std. dev.			0.36

Table 65. One-level analysis using tokens from the HLVC, AIS, and CHILS datasets with the variables SUBJECT and CORPUS.

For the variable SUBJECT, 2PL tokens had to be excluded from the analysis because they are invariably null in both the HLVC and CHILS corpora. If not, an interaction between the AIS variant of corpus and the 2PL variant of subject would have occurred. Further, from the CHILS dataset, the 3PL tokens were collapsed to remove a distinction between 3PL.M and 3PL.F because the other two datasets did not include this distinction, as explained earlier in the chapter. Despite these changes, there is a robust range between the variants in this analysis, and their ranking follows a similar

pattern found in their separate analyses. Broadly speaking, 3rd person subjects favour null

subjects, while 1st and 2nd person subjects disfavour them. The outliers are explainable by

interactions. For instance, 3SG.M has a higher FW, yet a lower null rate than 1SG. This can be

explained by 1SG VPs consistently disfavouring null subjects (HLVC, 0.34; AIS, 0.37; CHILS,

---

[37] The difference between the number of tokens and the rate of null subjects presented here for AIS (n=539 and %Ø=77%), and the tokens and rate presented in **6.2** (n=583 and %Ø=76%) is due to 2PL being excluded from this analysis.

0.45), and 3SG.M VPs being ranked differently across corpora (HLVC, 0.66; AIS, 0.23; CHILS, 0.4). Further, 1PL favours null subjects in the HLVC and AIS (0.52 and 0.60, respectively), while it disfavours them in the CHILS dataset (0.31). Despite these outliers, the overall ranking of the variants of 3rd person subjects favouring null subjects, and 1st and 2nd person subjects disfavouring them, is generally consistent with the individual analyses presented in this chapter.

CORPUS is the other independent variable that was included in the Rbrul analysis, and it is also significant, though with a smaller range than SUBJECT. The ranking of the variants is consistent with the differences in the rates of pro-drop, with HLVC and AIS favouring null subjects and CHILS disfavouring. The last variable included in the analysis was speaker as a random intercept to see the distribution of participants.

Figure 12. Scatterplot showing FWs of HLVC, AIS, and CHILS participants from the combined analysis, arranged by age.
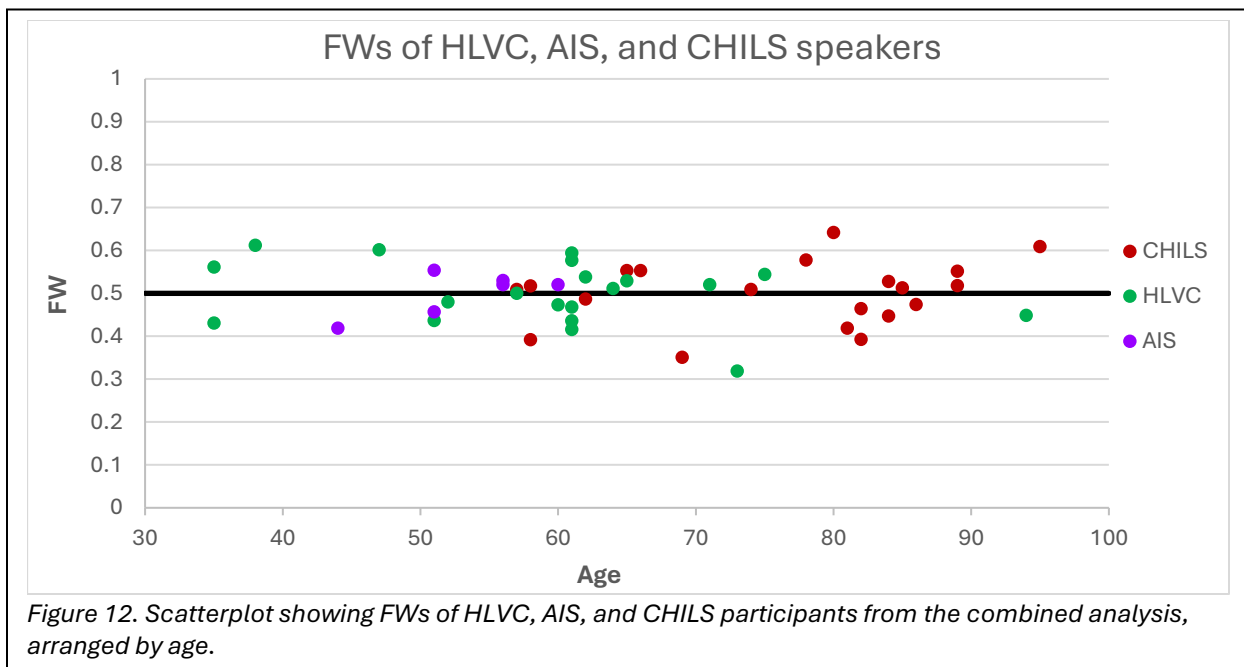
Figure **12** places these participants on a scatter plot, showing their individual FWs, organized by age.[38] As confirmed by the analysis, AGE is not a significant factor in subject pronoun expression. The figure shows that for each of the corpora, most of the participants have a FW above 0.5, with 11 of the 20 speakers from the HLVC, four of the six AIS participants, and 12 of the 20 from CHILS. There is one participant from each of the HLVC and CHILS corpora with an especially low FW. These two speakers are also reflected in the individual analyses presented above, speaker I1F73A from HLVC and ANTONIA from CHILS. These two participants have the lowest null rates in each of their respective datasets: 61% for I1F73A, and 52% for ANTONIA. However, aside from these two speakers with especially low rates of pro-drop, the speakers are distributed in similar ways for each of the corpora.

	Calabrian Italian (HLVC)	Homeland Ciociaro (AIS)	Heritage Ciociaro (CHILS)
	SWITCH REFERENCE	*PRONOUN IN PROMPT*	*SUBJECT*
	range = 34	range = 56	range = 51
Significant variables	*SUBJECT*	*SUBJECT*	*PREVIOUS REALIZATION*
	range = 32	range = 52	range = 19
	PREVERBAL ELEMENT	*AMBIGUOUS*	*TENSE*
	range = 8	range = 26	range = 15

Table 66. Comparison of significant variables and their ranges across three corpora.

In this chapter, I have shown the results of Rbrul analyses for each of the corpora. I have also examined each of the significant variables. These analyses were conducted for two languages, Calabrian Italian and Ciociaro, and homeland and heritage varieties of each. Despite the differences in the languages and methodologies, there are results that are consistent across all three results. Table **66** provides a summary of these results. In particular, SUBJECT and a priming

---

[38] The table of the complete Rbrul analysis is included in Appendix **H. 4**.

variable are significant variables in each analysis. In the following chapter I compare these

results to my hypotheses and place them in the broader context of pro-drop studies.

7. Discussion

In this dissertation I analysed pro-drop in Calabrian Italian and Ciociaro, used that evidence to test the functional hypothesis as it applies to subject pronoun usage in these languages, and contextualised my results by comparing them with similar studies. My hypotheses related to these goals are laid out in Chapter **4**. In this chapter, the results of the preceding chapter are considered through the framework of those hypotheses. The sections of this chapter mirror those of Chapter **4**, and each hypothesis is restated as it is tested. Table **67** provides a summary of the hypotheses, as well as if they were supported (✔), refuted (✘), or had mixed results (✱).

Hypothesis	HLVC	CHILS	AIS
(1) Heritage and homeland			
(a) Similar null subject rate	✔	✘	✘
(b) Similar significant variables	✔	✘	✘
(2) Null subject rates			
(a) Ciociaro has a lower null rate than Calabrian		✔	
(b) Lower null rate correlated with higher rate of ambiguity	✱	✱	✔
(c) AIS has lower null rate than CHILS, similar rate of ambiguity		✘	✘
(d) i. Higher null rates than French and Faetar	✔	✔	✔
(d) ii. Lower null rates than Italian	✔	✔	✔
(d) iii. Similar null rates to Spanish	✱	✔	✔
(3) Significant variables			
(a) i. Subject is a significant variable	✔	✔	✔
(a) ii. Ambiguous inflections disfavour null subjects	✱	✱	✔
(b) Switch reference / pronoun in prompt is significant	✔	✘	✔
(c) Tense is significant for AIS			✘
(d) VPs with reflexive pronouns favour null subjects	✘	✘	✘

Table 67. Table summarizing the hypotheses that are supported (green), refuted (red), and mixed (blue).

In the first section (**7.1**), I interpret the results of the homeland and heritage analyses. I present the HLVC analyses, the justification for combining those analyses as described in **6.1.3**, and how that supports my first hypothesis. Following that, I re-examine the CHILS and AIS results, why those results were not combined, and how that refutes my first hypothesis.

In the second section (**7.2**), I compare the null subject rates of Calabrian Italian and Ciociaro, how they correlate with ambiguous verbal inflections, and how that partially supports my second set of hypotheses. Following this examination, I compare the rates of pro-drop in Calabrian Italian and Ciociaro to other Romance languages.

In the third section (**7.3**), my hypotheses about significant variables are tested against the results of the HLVC, AIS, and CHILS analyses. I examine how my results for these variables relate to the functional hypothesis, how they compare to studies with similar methodologies, and how they support some aspects of my hypotheses, and challenge others. I then examine the ranking of variants within the significant variables to see how they differ, and what that reveals about how these variables may work differently in Italo-Romance languages compared to more-studied Spanish varieties.

The final section of this chapter is a discussion of the overarching feature of Ciociaro that motivated this study, verbal ambiguity (**7.4**). In this section, I compare the rates of verbal ambiguity in the four datasets of this dissertation, how that ambiguity interacts with other variables, and how the results both support and refute the functional hypothesis. Since ambiguous verbal forms were hypothesized to play a large role in subject pronoun realization at the start of this dissertation, I end this section with a discussion of why this factor was not selected as a significant variable in the CHILS analysis.

7.1. Homeland and heritage varieties

In this section, I evaluate my hypotheses regarding differences between homeland and heritage varieties in Calabrian and Ciociaro (see **4.1**). I compare the homeland to the heritage subcorpora of the HLVC, and I compare the AIS to the CHILS. I first compare the overall null rates of the

varieties for each language, testing hypothesis (1a). I then examine the significant variables of

each analysis to test hypothesis (1b).

Hypothesis (1a): I hypothesize that there is a similar overall null subject rate for the homeland and heritage varieties of both Calabrian Italian and Ciociaro.

The results for the homeland and heritage varieties of Calabrian Italian are very similar. Both

varieties have an overall null subject rate of 80%. Thus, my hypothesis that the homeland and

heritage varieties will be similar is confirmed for Calabrian. This identical rate suggests that, at

least regarding pro-drop, the two varieties have not diverged since the heritage community

emigrated during the latter half of the 20th century. This is confirmed by an Rbrul analysis that

includes SUBCORPUS as a variable—homeland and heritage having FWs of 0.50 (see Appendix

H. 5).

The comparison between the results of the AIS and CHILS corpora is more complicated. CHILS

is a corpus of participants providing oral histories, whereas AIS is a corpus composed of

translation elicitations into Ciociaro from prompts in Italian. There are differences in the null

rates of the two varieties, but these are complicated by the differences between the corpora. The

null rate of the AIS data, which represents the homeland variety, is 76%. The null subject rate for

CHILS is 72%.

However, to test how similar the CHILS and AIS results are, I ran a combined analysis in Rbrul,

as described in **6.4.1**. When CORPUS is included as a variable, with CHILS and AIS as the two

variants, it is a significant variable, even if the effect isn't large (range = 10). The CHILS corpus

disfavours (FW = 0.45) and AIS favours null subjects (FW = 0.55). This refutes my hypothesis

(1a) for Ciociaro because the corpora are significantly different when the effects of conditioning

191

factors are accounted for while comparing rates. These differences are expanded upon in testing the next hypothesis.

(1b): I hypothesize that, for Calabrian Italian and Ciociaro, each language's homeland and heritage varieties share significant variables.

I first consider the homeland and heritage varieties of Calabrian Italian, which have identical null subject rates. Both the homeland and heritage varieties have SUBJECT as the variable with the largest range of effect (homeland range = 40; heritage range = 42), followed by SWITCH REFERENCE (homeland range = 37; heritage range = 31). There is one more significant variable for the homeland variety, PREVERBAL ELEMENT, which has a much smaller effect (range = 10). This variable was not significant for the heritage analysis. When the two corpora are combined, as in **6.1.3**, PREVERBAL ELEMENTS is the significant variable with the smallest effect (range = 8).

Therefore, hypothesis (1b) is partially supported by the homeland and heritage varieties of CI. They share the same significant variables, with similar ranges of effect, except that PREVERBAL ELEMENT is significant for the homeland variety, and it is not significant in the heritage analysis. Despite this minor difference, the two varieties are not very different.

Returning to Ciociaro, there is a difference in the null subject rates, and there are also differences in the significant variables of the CHILS and AIS analyses. Both the homeland and heritage analyses have three significant variables, but they only have one in common. For both analyses, SUBJECT is a significant variable with a large range of effect (AIS = 52; CHILS = 51). For the CHILS analysis, this is the largest range, and it is the second-largest range for the AIS. The other two significant variables are distinct for the two varieties. The variable with the largest range for AIS is PRONOUN IN PROMPT—whether the prompt the participant was asked to translate into

192

Ciociaro contained an overt subject pronoun (range = 56), which cannot be a factor in the heritage analysis. For CHILS, the variable that has the second-largest effect (range = 19) is PREVIOUS REALIZATION, which cannot be tested in the AIS data.

The narrative structure of CHILS means the preceding phrase can be analysed with the PREVIOUS REALIZATION factor, while the elicitation of phrases in the AIS allows for the prompts to be analysed. However, these two factors are both kinds of priming variables. In the CHILS corpus, the realization of an overt or null subject pronoun in the preceding VP has the expected priming effect. The AIS results show that the presence of a subject pronoun in the Italian prompts primes the participant to reproduce them in Ciociaro. Thus, while these are different variables, each reflecting their specific research methodology, both Ciociaro corpora do exhibit a significant priming effect.

Finally, both analyses produced a third significant variable. For AIS this was whether the verb form itself was ambiguous (range = 26), and for CHILS this was the tense of the VP (range = 15). Ambiguity of verb forms was not a significant variable in the CHILS analysis, and tense was not a significant variable in the AIS analysis. A potential reason for this is that the narrative structure of the CHILS recordings reduces the effect of the ambiguous verbal inflections. When a VP occurs in isolation, as is the case in the AIS prompts, the subject cannot be retrieved from elsewhere as there is no dialogue, so a subject pronoun is required to disambiguate the subject. For the CHILS dataset, even if the verbal inflection is ambiguous, the subject can sometimes be recovered from the larger context of the narrative. Also, as I explained in both the description of AMBIGUITY for the AIS analysis (**6.2.3**) and in the CHILS results of TENSE (**6.3.3**), the two variables are closely linked.

Despite the similarities between the CHILS and AIS results, continuing from the previous chapter, I consider these analyses separately because the combined analysis of both varieties (**6.4.1**) found CORPUS to be a significant variable.

Returning to the hypotheses at the start of this section, I have shown how the homeland and heritage varieties of the Calabrian corpus are very similar. This confirms my hypotheses, both in terms of the overall null subject rate (Hypothesis 1a) and the significant variables (Hypothesis 1b). This can be attributed to the methodological structure shared by the corpora, and to the homeland and heritage varieties of Calabrian Italian not being significantly different in their use of pro-drop. Despite heritage Calabrian being in an English environment, there does not appear to be contact effects that are leading to a significant divergence.

The CHILS and AIS analyses, on the other hand, cannot be combined and refute hypotheses (1a) and (1b). The two varieties have different null rates, different significant variables, and a combined analysis shows the difference between the corpora is significant. This may be due to divergences in the two varieties; however, these differences in subject pronoun usage may also be attributed to the materially different research contexts in which the fieldwork was conducted.

7.2. Null subject rate

In this section, I first compare the overall null rates of Calabrian Italian and Ciociaro (hypothesis 2a), and I test the hypothesis that, across language varieties, a lower null rate will correspond with a higher rate of ambiguous verbal morphology (2b). I then examine whether that correspondence differs for the AIS data due to its distinct data collection methodology (2c). Finally, I place the overall null rates of Ciociaro and CI within the context of comparable studies of other Romance languages (2d).

(2a): I hypothesize that Ciociaro has a lower null-subject rate than Calabrian Italian.

One-level analysis of pro-drop across HLVC, AIS, and CHILS			
Tokens (n):	*3,921*	*%Ø*	*76%*
Subject			
	n	*%Ø*	*FW*
2PL	(60)	(75%)	N/A
3PL	713	91%	0.77
3SG.F	245	82%	0.58
1PL	544	75%	0.44
3SG.M	772	71%	0.42
1SG	1,416	72%	0.40
2SG	231	69%	0.37
	Range		**40**
Corpus			
	n	*%Ø*	*FW*
HLVC	1,634	80%	0.56
AIS	539	77%	0.52
CHILS	1,748	72%	0.42
	Range		**14**
Speaker (random)			
Std. dev.			0.36

Table 68. One-level analysis using tokens from the HLVC, AIS, and CHILS datasets with the variables SUBJECT, CORPUS, and SPEAKER (random).

Ciociaro does indeed have a lower null-subject rate than CI. The homeland and heritage varieties of CI have the highest null rate (80%). For Ciociaro, the AIS data has a higher null rate (76%) than CHILS (72%). This supports hypothesis (2a). To confirm that this hypothesis is supported, I ran an Rbrul analysis with CORPUS and SUBJECT as independent variables, and with SPEAKER as a random effect. SUBJECT was included as an independent variable because it is the only one that is significant for all three datasets.

CORPUS is a significant variable, as shown in Table **68**, with Calabrian Italian favouring, the AIS slightly favouring, and CHILS disfavouring null subjects. To test if the difference of pro-drop in Calabrian and Ciociaro is related to the more syncretic inflectional system of Ciociaro, next I test if there's an inverse correlation between the two.

(2b): I hypothesize that a lower null-subject rate correlates with a higher rate of ambiguous verbal forms.

Hypothesis (2b) is supported when considering rates of ambiguity and null subjects. Calabrian Italian has a higher null-subject rate (80%) and a lower rate of ambiguous verb forms (7%) than CHILS, which has a lower null-subject rate (72%), and a much higher rate of ambiguous verb forms (34%). This correspondence between a lower null rate and a higher rate of ambiguous verb forms supports the functional hypothesis, as it suggests that subject pronouns are used more

195

often in languages where verbal inflections cannot be relied upon to disambiguate a VP's subject.

One-level analysis of pro-drop with AMBIGUOUS **and** CORPUS **as an interaction factor for HLVC and CHILS**

Tokens (n):	3,382	%Ø	76%
Subject			
	n	%Ø	FW
2PL	(16)	(100%)	N/A
3PL	646	91%	0.78
3SG.F	207	79%	0.54
3SG.M	690	73%	0.46
1PL	491	73%	0.43
1SG	1,176	71%	0.39
2SG	172	69%	0.37
	Range		**41**
Ambiguous*Corpus			
UNAMBI:HLVC	1,522	81%	0.56
AMBI:CHILS	591	69%	0.56
UNAMBI:CHILS	1,157	73%	0.45
AMBI:HLVC	112	70%	0.45
	Range		**11**
Speaker (random)			
Std. dev.			0.37

Table 69. One-level analysis of HLVC and CHILS with factors SUBJECT, AMBIGUOUS *CORPUS, and SPEAKER (random).

To test if this correlation is significant, I ran a logistic regression analysis to examine if verbal ambiguity is correlated with the different corpora. AIS could not be included because the strength of the effect of AMBIGUOUS caused model collapse. The other two corpora, HLVC and CHILS, are included in an interaction factor of AMBIGUOUS and CORPUS. I also included SUBJECT in the analysis as it is a significant variable in both the HLVC and CHILS results. Table **69** shows the results of this analysis. Despite the correspondence between *rates* for null subjects and ambiguous verbal inflections, the FWs show that, while AMBIGUOUS has the expected direction of effect for HLVC, CHILS actually favours null subjects when the VP is ambiguous for subject. Therefore, whatever effect ambiguous verb forms have on subject pronoun realization in the CHILS corpus, the ranking of variants does not support the functional hypothesis. Therefore, hypothesis (2b) is only supported inasmuch as the *rate* of ambiguous forms is higher for CHILS while the *rate* of null subjects is lower than for the HLVC. The functional hypothesis is not supported though, as the FW for CHILS shows that ambiguous phrases favour null subjects.

(2c): I hypothesize that AIS has a lower null-subject rate than CHILS, yet a similar ambiguous verbal morphology rate.

Due to the disconnected nature of the prompts in the AIS data, I hypothesized that more subject pronouns would be used to disambiguate subjects in the AIS data than in the CHILS data, despite having similar rates of ambiguous verbal inflections. The rate of ambiguity for AIS (20%) falls between the rates for the CHILS (34%) and HLVC (7%), as does the null subject rate (76% for AIS, 80% for HLVC, and 72% for CHILS). As a result, the AIS results do not support hypothesis (2c).

However, as discussed above, the rates cannot show a meaningful correlation between these two variables. The AIS analysis is also the only corpus where AMBIGUITY is a significant variable. In fact, the results for the AIS corpus, as shown in Table **69**, do support the functional hypothesis as verbs coded as ambiguous disfavour null subjects (FW = 0.22), and those that are unambiguous favour null subjects (FW = 0.78). This issue is investigated further in **7.4**. Now that the null rates and corresponding ambiguous verb form rates have been compared between the three corpora, in the next section I consider how the null rates in these corpora compare to other studies.

(2d): I hypothesize that both Calabrian Italian and Ciociaro have higher null rates than both French and Faetar, similar rates to some varieties of Spanish, and lower rates than Italian.

To provide context for the analyses presented in this dissertation within the larger field of pro-drop in Romance languages, Figure **13** shows the null rate and language variety of comparable studies, listed in decreasing order of null rate. These studies were described in **3.3** because they provide a range of methodological approaches to analyzing pro-drop, resulting in a diversity of significant variables in their analyses of homeland and heritage varieties of Italian (in red), Spanish (in green), as well as Faetar and French (in blue).

Figure 13. Stacked column graph showing null subject rates reported from this and comparable studies, grouped by language family.
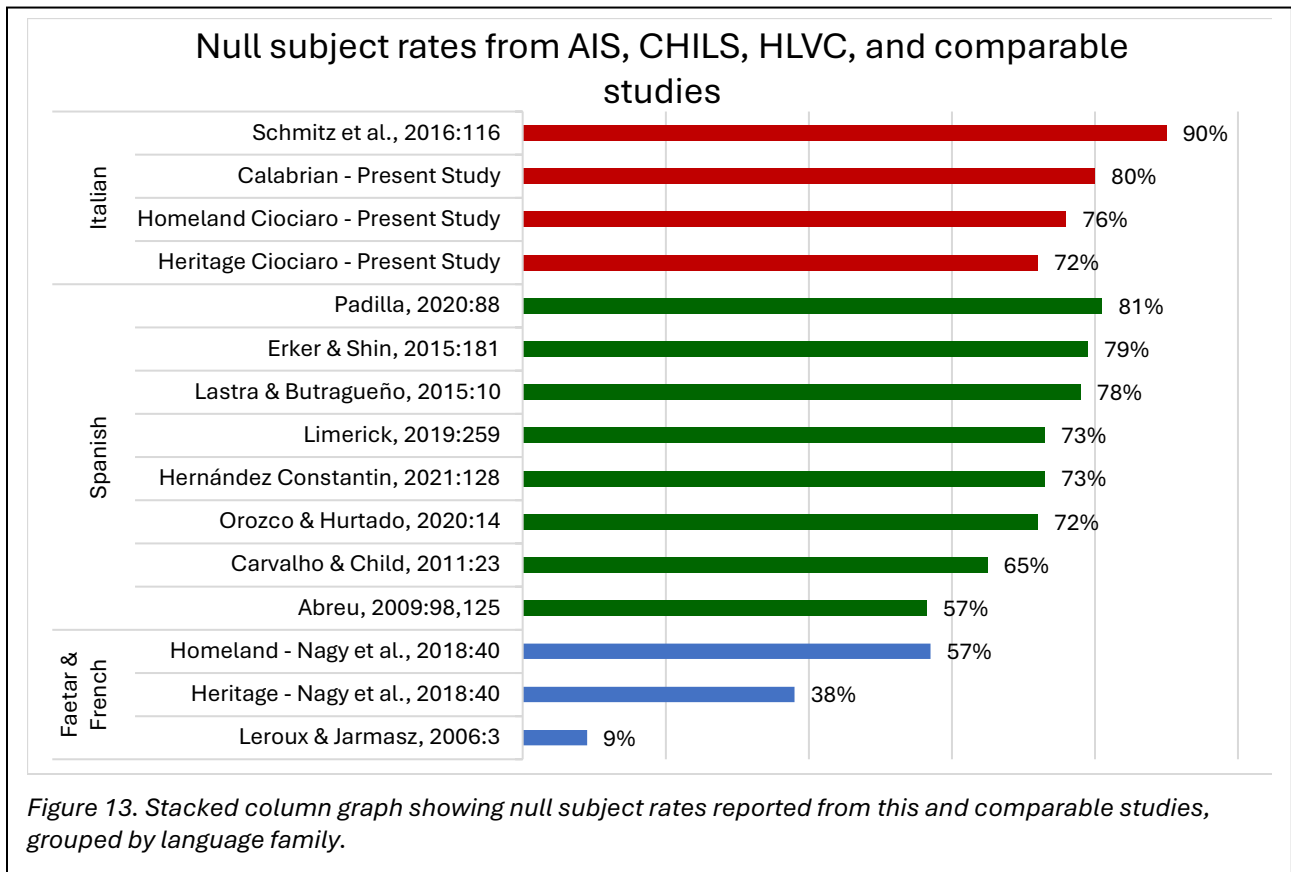
Returning to (2d), I hypothesized that both Ciociaro and Calabrian would have higher null-subject rates than French and Faetar and lower rates than Italian, which is supported by the data presented in Figure **13**. I also hypothesized that Calabrian Italian and Ciociaro would have null subject rates that fell within the broad range of the varieties of Spanish, which is also borne out by the data. I further predicted that none of the Spanish varieties would have a higher null-subject rate than Calabrian. This prediction is nearly supported by the data. Almost all the Spanish varieties have a lower null rate, except Equatorial-Guinean Spanish which has a 1% higher null rate than Calabrian Italian. Thus, the hypothesis was mostly supported by the data. Both Ciociaro and Calabrian have higher null-subject rates than French or Faetar, and lower null rates than Italian, but they fall within the large range presented in **3.3** for Spanish varieties (51% to 81%).

Taken as a whole, the cross-linguistic pattern supports the functional hypothesis. Those languages, such as standard Italian and Calabrian Italian, which have distinct verbal morphology marking the subject, person, and number, have higher rates of null subjects. Those languages, such as Ciociaro and Faetar, that have more ambiguous verbal inflections, produce more overt subject pronouns. The range found in the Spanish varieties also supports the functional hypothesis (see **3.1.2**). For example, Puerto Rican Spanish has identical inflections for many verbs with 2SG and 3SG subjects (Abreu, 2009; Hochberg, 1986), and that ambiguity corresponds with a lower null-subject rate than other varieties. However, as we have seen in this section, a lower null rate and a higher rate of ambiguous verb forms can both be true without one having a significant effect on the other, as is the case in my CHILS analysis.

In this section, four hypotheses have been tested. Hypothesis (2a) was supported by the data, as Ciociaro (both AIS and CHILS) has a lower null-subject rate than CI. Further, that lower null-subject rate corresponds to a higher rate of ambiguous verbal inflections, which supports hypothesis (2b). However, AIS does not have a lower null-subject rate than CHILS, nor a similar rate of ambiguous verbal morphology, which contradicts (2c). Finally, both Ciociaro and Calabrian have higher null-subject rates than French and Faetar and lower rates than Italian, which supports hypothesis (2d). Ciociaro also has a null rate within the range of Spanish varieties, aligning with my hypothesis, and Calabrian Italian has a higher null rate than all but one of the Spanish varieties (Equatorial-Guinean Spanish; Padilla, 2020). As a result, hypothesis (2d) is mostly supported by my study.

7.3. Relevant variables

In this section I discuss the significant variables of the HLVC, AIS, and CHILS analyses, how they confirm or counter my hypotheses 3a – 3d, and how they compare to the significant variables introduced in **3.3**.

In the first section, I discuss the SUBJECT variable. This was a significant variable in each of my analyses. It was also a significant factor for each of the comparable studies discussed in **3.3.1.1**. I examine this variable, the ranking of variants across the results, and show how the results support the functional hypothesis.

In the second section, I discuss the priming variables, presented in **5.3.5**. The variables SWITCH REFERENCE and PREVIOUS REALIZATION were significant in the HLVC and CHILS results. I discuss these variables and how they relate to ambiguous verb forms and the functional hypothesis, as well as how these results compare to similar studies.

In the final two sections, I discuss the variables TENSE and PREVERBAL ELEMENT, which were significant in the AIS and HLVC results, respectively. However, they were not significant in the CHILS analysis. I propose reasons for these differences, how those results compare to studies in other varieties, and how it relates to the functional hypothesis and my hypotheses 3c and 3d.

7.3.1. Subject

(3a): I hypothesize that SUBJECT is a significant variable for each of the corpora in this dissertation, with subjects containing ambiguous inflections disfavouring null subjects, and subjects with unambiguous inflections favouring null subjects.

There is a consistent finding in studies of pro-drop in Romance languages that the factor SUBJECT, which distinguishes subjects according to PERSON and NUMBER, is significant, and it was the factor with the biggest range of effect in 11 of the 13 analyses included in Table **14**.

This is also true of the CHILS, AIS and HLVC analyses. SUBJECT was the variable with the biggest range of effect for CHILS (range = 51), and the second biggest for HLVC (range = 32) and AIS (range = 52). This supports my hypothesis that SUBJECT would be a significant factor for all three analyses.

Within this variable though, there are differences across languages in the ranking of the variants. For Spanish varieties, there is a pattern that singular subjects are less likely to be null than plural subjects. This ranking of the variants, where NUMBER has a larger and more consistent effect than PERSON, is found across varieties of Spanish (Carvalho et al., 2015; Hernández Constantin, 2021; Martín-Butragueño, 2020).

However, this is not the case for Faetar. Nagy et al. (2018:40) found that null subjects occur at a higher rate with both singular and plural 3rd person subjects, (35-68% null subjects for the heritage variety and 21-51% for the homeland variety) than with 1st person subjects (heritage: 6-20% null subjects and homeland: 3-7%).

Table **70** shows the hierarchy of subjects across the three analyses of this dissertation. Unfortunately, there are not enough 2PL subjects to be included in the analysis, but of the 10 examples in the CHILS data and six in the HLVC, all 16 had null subjects. This table shows that all three analyses share similarities with both the Faetar pattern (3rd person subjects favour null subjects, 1st and 2nd person subjects disfavour null subjects), and the pattern found in Spanish

varieties (plural subjects favour null subjects). They also each have exceptions within their individual patterns, which are discussed in Chapter **6**.

Ranking of variants across corpora for SUBJECT

		Favours null subjects			*Disfavours null subjects*			
CHILS	(2PL)	3PL.M	3SG.F	3PL.F	1SG	3SG.M	2SG	1PL
AIS		3SG.F	1PL	2SG	3PL	2PL	1SG	3SG.M
HLVC	(2PL)	3SG.M	3PL	1PL	3SG.F	1SG	2SG	

Table 70. Hierarchy of variants for SUBJECT ordered by factor weight.
Note: those subjects left of | favour null subjects, whereas those right of | disfavour null subjects.

An important consideration for the ranking of variants for SUBJECT is its potential interaction with the ambiguity of the verbal paradigm. Certain subjects have more ambiguous inflections, and if that correlates to an increase in overt subject pronouns, this will support the functional hypothesis. For Ciociaro's paradigms, the 1SG, 2SG, and 3SG, often have identical inflections. This aligns with the table; those subjects disfavour null subjects. There are exceptions, however, with 3SG.F favouring null subjects for both AIS and CHILS, and 2SG favouring null subjects in the AIS analysis.

For the 3SG.F outlier in the AIS data in particular, this is a result of agreement found elsewhere in the VP (**6.2.2**), where six of the seven prompts for VPs with a 3SG.F subject contain gender agreement on an adjective or an NP. This may explain the large difference in the factor weights for 3SG subjects by gender for CHILS (3SG.F = 0.57 and 3SG.M = 0.40) and AIS (3SG.F = 0.77 and 3SG.M = 0.22). For both, agreement plays a disambiguating role, but to a larger extent for AIS. This explanation does not account for the HLVC results, where the opposite ranking occurs (3SG.F = 0.49 and 3SG.M = 0.66), when we would expect it to be comparable to CHILS. Further, this finding cannot be compared to Spanish varieties as most studies of pro-drop in Spanish do not separate 3rd person subjects by gender. Further, ambiguity does not relate to these differences

202

of ranking for 3SG.F and 3SG.M, as the verbal inflection is identical, regardless of gender. This suggests that further refinement in a future study is needed to fully understand the hierarchy of variants for each of the analyses.

Despite these exceptions, however, the hierarchy of each of the analyses broadly supports the functional hypothesis as those subjects with the most ambiguous verbal inflections (1SG, 2SG, 3SG) disfavour null subjects. However, 1PL has an unambiguous verbal inflection, yet it is the subject with the lowest factor weight (0.31) in the CHILS analysis, disfavouring null subjects. This is in opposition to the expected finding, which is found in the AIS and HLVC hierarchies, where 1PL favours null subjects. This is an outlier that refutes the functional hypothesis and cannot be explained by an interaction with ambiguity, priming variables, disproportionate use by a certain speaker, or any other variable. Further research may reveal why it is ranked differently to the other analyses and contrary to what is expected.

As this section has shown, hypothesis (3a) is supported by the analyses, and SUBJECT is a significant factor for the HLVC, AIS, and CHILS corpora. However, despite the consistency of SUBJECT as a significant variable in pro-drop analyses of Romance languages, there is heterogeneity within the ranking of the variants within this factor. Specifically, the predominance of Spanish varieties in similar pro-drop analyses do not necessarily describe pan-Romance phenomena. This emphasizes the importance of including lesser-studied languages in analyses of linguistic phenomena. While studies of pro-drop in Spanish varieties find that "…since the conditioning effect of SPP person and number affects virtually all varieties of Spanish, we are dealing with a general tendency that has a universal linguistic explanation" (Orozco, 2015:29),

the internal constraints of the variable may vary across Romance languages, as in Faetar, Ciociaro, and CI.

7.3.2. Priming variables

In this section, I consider the priming variables SWITCH REFERENCE and PREVIOUS REALIZATION. As these priming variables rely on narrative (the VP has the same referent as the previous VP, or the previous VP had an overt subject pronoun, for example), the AIS corpus is not included in this discussion. However, I will return to the AIS corpus near the end of this section to discuss its priming variable, whether the provided prompt had an overt subject pronoun or not.

(3b): I hypothesize that SWITCH REFERENCE is a significant variable for both the HLVC and CHILS corpora, and PRONOUN IN PROMPT is a significant variable for the AIS data.
This hypothesis was only partially supported by the results. For the HLVC results, SWITCH REFERENCE is a significant variable, and has the largest range of effect (range = 34) of the significant variables, which supports my hypothesis.

However, in the CHILS results, PREVIOUS REALIZATION is a significant variable (range = 19), while SWITCH REFERENCE is not, which runs counter to my hypothesis. The significance of PREVIOUS REALIZATION in the CHILS data suggests that the speaker continues with either null or overt subject pronouns based on what was used in the previous clause, regardless of a changed referent.

To test whether there was an interaction between these two priming variables for the two datasets, I ran another analysis that combined the variants of both variables. This analysis would show differences between, for example, a switched referent when the preceding VP had an overt subject pronoun compared to when the preceding VP had a null subject. However, the interaction

of these variables was not significant in either analysis. For the CHILS analysis, if the previous VP has a null subject, the FW is identical whether the preceding VP has the same referent or a switched referent (FW = 0.5). For the HLVC results, however, a pronoun is slightly more likely to be overt if the previous VP's null subject is also a change in referent (0.54) than if it is not (0.46).

This may be related to the syncretism of the verbal forms in Ciociaro. For example, if two adjacent VPs have identical verbal inflections and the first VP has a null subject, an overt subject pronoun in the second VP is just as useful to show the subject has not changed, as when it has switched. On the other hand, for the HLVC corpus, even with two null subjects, the rate of ambiguous forms is low (7%), so, if the subject has not changed, the subject can be inferred by the listener from the verbal inflection. This interaction of switch reference and ambiguity is also argued by Cameron "…the effects of morphological ambiguity will be seen most effectively not as independent constraints on null subject variation, but as effects that interact with switch reference…" (1993:307). It appears that the higher rate of ambiguous forms in the CHILS data reduces the significance of SWITCH REFERENCE as a variable. For both HLVC and CHILS, the direction of effect is the same, a change in referent has a lower null rate than the same one, but the size of the effect is reduced for CHILS, potentially due to its higher rate of ambiguous verbal inflections.

Finally, the AIS dataset does not include either of these priming variables because of the prompt and translation structure of the elicitations. However, whether each prompt does or does not contain an overt subject pronoun is the significant variable with the largest effect (range = 56). Although this variable is different from the two others discussed in this section, it is a priming

variable. The presence of an overt subject pronoun in the prompt primes the participant to include one in their translation as well. This is made clear in Table **71**, which shows just how often the participant replicates the presence or absence of a subject pronoun from the prompt in their translation. The participants only included overt subject pronouns when the prompt didn't contain one in 8% of the tokens, and only used null subject pronouns to translate prompts that contained overt subject pronouns 4% of the time.

Subject pronoun realization in relation to prompts in the AIS dataset

	n	% of corpus
NULL	418	72%
REMOVED	26	4%
ADDED	48	8%
OVERT	91	16%
Total	**583**	**100%**

Table 71. Rates of agreement between presence and absence of subject pronouns in prompts and responses for the AIS corpus.

Returning to the hypothesis, priming variables are significant across all three analyses, and have the largest (HLVC and AIS) or second-largest (CHILS) effect. However, I hypothesized that SWITCH REFERENCE would be significant in both the HLVC and CHILS results. While this was the case for the HLVC, it was not true for CHILS. The high rate of ambiguous forms in Ciociaro seems to reduce the effect of SWITCH REFERENCE in favouring or disfavouring null subjects.

Also, while AIS is not coded for the same priming variables, its priming variable is significant, supporting my hypothesis. Thus, my hypothesis is supported by the HLVC and AIS corpora, but not by the CHILS corpus.

All three of these analyses show the importance of including variables outside the specific sentence that is being analysed when examining pro-drop. These results also support the functional hypothesis. For SWITCH REFERENCE in the HLVC results, marking a change in subject from one VP to the next with an overt pronoun ensures that change is understood by the listener. However, in the case of the CHILS results, if the subject may already be unclear due to a much

higher rate of verbal syncretism than in Calabrian, then clarifying with an overt pronoun ensures the subject of a VP is clear, regardless of whether the subject has changed or not.

7.3.3. Tense

(3c): I hypothesize that the TENSE variable is significant in the AIS analysis, and the ranking of the variants reflects their relative ambiguous verbal morphology.

Of the three corpora, I hypothesized that the TENSE variable would only be significant for AIS. This hypothesis was not correct. TENSE was a significant factor in the CHILS results, though it had the smallest range of effect of the three significant variables (range = 15). For both the HLVC and AIS analyses, TENSE was not significant. I also hypothesized that tenses with auxiliary verbs would favour null subjects, as they have less syncretic paradigms, and imperfect and present-tense verbs would disfavour null subjects.

The results, however, were the opposite. For the CHILS analysis, imperfect tokens favour null subjects (FW = 0.56), and so do those in the present tense (FW = 0.53), whereas VPs in a compound past tense, those using an auxiliary verb and a past participle, disfavour null subjects (FW = 0.41). As described in **6.3.3**, these differences between my hypotheses and the results can be explained by interactions with SUBJECT and ambiguous verbal morphology. For example, in the compound past, the auxiliary verb *essere* ('to be') has the same form, *sono*, for both 1SG and 3PL subjects. Also, Ciociaro uses *essere* as the auxiliary verb much more than Calabrian (see **5.3.2.2**), which means the ambiguity of *sono* is more frequent in Ciociaro. Further, the participle in Ciociaro often lacks explicit agreement. For all these reasons, 39% of tokens in the compound past are ambiguous as either a 1SG or a 3PL subject. Thus, an overt subject pronoun helps disambiguate tokens in the compound past, at least in these cases.

Interaction of TENSE and AMBIGUITY for CHILS

Interaction	n	%Ø
AP:UNAMBI	234	77%
IMP:UNAMBI	547	77%
CP:AMBI	118	75%
PR:UNAMBI	227	75%
IMP:AMBI	257	74%
AP:AMBI	119	62%
PR:AMBI	92	54%
CP:UNAMBI	142	50%

Table 72. Analysis of the interaction of TENSE and AMBIGUITY (CHILS, n=1,736).

If verbs in the compound past that have ambiguous inflections had a lower rate of null subjects than phrases with unambiguously inflected verbs, this would support the functional hypothesis. However, as Table **72** shows, this is not the case. In fact, the opposite is true. Verbs in the compound past (CP) with ambiguous inflections have a higher rate of null subjects (75%) than those that are unambiguous (50%). Verbs that are in the present (PR) and absolute past (AP) do have the expected results, with ambiguous verbs having a lower rate of null subjects. However, verbs in the imperfect and compound past (which together make up 62% of the dataset) do not behave as expected. In fact, in the analysis presented in **6.3**, the compound past is the tense that most disfavours null subjects (FW = 0.41), yet ambiguous VPs in the compound have a higher-than-average rate of null subjects. Therefore, in this case the functional hypothesis is not supported by the results of the CHILS analysis.

7.3.4. Preverbal element

(3d): I hypothesize a lower null rate for VPs with a reflexive pronoun and a 3rd person subject than for those with a 1st or 2nd person subject.

I hypothesized preverbal elements would not be a significant factor for any of the analyses, but I also thought that within the variable there was a means of testing the functional hypothesis. For 1st and 2nd person subjects, a reflexive pronoun disambiguates the subject's number but not gender. However, for 3rd person subjects, it does not disambiguate the gender or number of the subject. Thus, the functional hypothesis would predict a lower null-subject rate for 3rd person subjects with reflexive pronouns than those in the 1st or 2nd person.

For the CHILS and AIS corpora, PREVERBAL ELEMENT was not a significant variable on subject pronoun realization, but it was for the HLVC corpus. For the combined analysis of Calabrian Italian, it was the significant variable with the smallest range (8), and a direction of effect as expected: VPs with preverbal elements favour null subjects (FW = 0.54), and those without disfavour null subjects (FW = 0.46).

Preverbal elements separated by role

	CHILS		AIS		HLVC	
	n	%Ø	n	%Ø	n	%Ø
REFL	78	81%	46	82%	170	84%
OBJ	217	77%	117	70%	340	86%
NEG	116	62%	53	71%	191	82%
NONE	1,325	71%	367	84%	933	78%
Total	**1,736**	**72%**	**583**	**76%**	**1,634**	**80%**

Table 73. Preverbal elements and the corresponding null rates across the three analyses.

Hypothesis (3d) was based on the presumption that VPs with reflexive pronouns would behave differently from VPs with other preverbal elements. However, as described in **6.1.3.3**, the similar FWs of object and reflexive pronouns for the HLVC datasets led to these categories being collapsed in their analyses. Table **73** shows the null rates of these different preverbal elements across the three corpora of this dissertation. The table shows the heterogeneity of the rates across the types of preverbal elements. However, it also shows that, except for the AIS results, VPs with a reflexive pronoun and those with an object pronoun have very similar null rates.

This pattern, where a preverbal object or reflexive pronoun corresponds with a higher null-subject rate is also found in Faetar, where "…subject pronouns are less likely to surface when other material fills the preverbal space in linear surface order, even with different syntactic roles" (Nagy et al., 2018:41).

Returning to hypothesis (3d), to measure the significance of reflexive pronouns on subject pronoun realization and for the hypothesis to be supported, regardless of their overall null rates,

VPs with a 3rd person subject and a reflexive pronoun should disfavour null subjects because they do not disambiguate the subject between masculine and feminine subjects, whereas those with 1st or 2nd person subjects should favour null subject pronouns because the referent is unambiguous.

One-level analysis of pro-drop with PREVERBAL*PERSON as an interaction factor for HLVC, AIS, and CHILS

Tokens (n):	3,975	%Ø	76%
Preverbal*Person			
	n	%Ø	FW
NEG:3rd	91	93%	0.63
N/A:1+2	1,579	73%	0.59
Oth:1+2	329	71%	0.53
Refl:3rd	138	89%	0.50
Refl:1+2	141	72%	0.50
Oth:3rd	307	84%	0.47
N/A:3rd	1,194	78%	0.41
NEG:1+2	196	66%	0.38
Range			**25**
Speaker (random)			
Std. dev.			0.43

Table 74. One-level analysis of HLVC, AIS, and CHILS with the interaction factor PREVERBAL*PERSON and SPEAKER (random).

However, it is important to account for the variation that exists in the SUBJECT variable. For example, VPs with a 1SG subject disfavour null subjects in each of the analyses, regardless of the presence of a preverbal pronoun. So, the significance of reflexive pronoun presence must be tested, separate from the significance of SUBJECT. Thus, I ran an analysis with an interaction variable of PERSON and PREVERBAL ELEMENT, while keeping SPEAKER as a random intercept. This variable measures whether reflexive pronouns affect null subject rates differently from other preverbal elements (or the absence of any).

Table **74** shows this is the case, though not with the expected direction of effect. In this table, PERSON is used instead of SUBJECT because VPs with 3rd person subjects use the same reflexive pronoun for singular or plural subjects, so NUMBER is not relevant. Further, as it is comparing distinct and non distinct reflexive pronouns, 1st and 2nd person are combined as distinct in contrast to 3rd person, which is non distinct. The results show that reflexive pronouns have identical FWs for PERSON (FW = 0.50). Of note, we also see that the other types of preverbal

elements *do* behave differently for PERSON. This requires further investigation as to why, for example, a negative particle favours null subjects for 3rd person subjects (FW = 0.63), but disfavours for 1st or 2nd person subjects (FW = 0.38). The same is also true when a preverbal element is not present. Returning to hypothesis (3d), although there were not enough examples in just the HLVC dataset to build a model in Rbrul, across the corpora we do not see phrases with reflexive pronouns favouring null subjects. Thus, the hypothesis is not supported.

As we have seen in this section, PREVERBAL ELEMENT was only a significant variable for the HLVC results, and the range of effect was small (8). Also, the category of preverbal element is not significant. Finally, the results from hypothesis (3d) do not support the functional hypothesis as there is not a significant difference for distinctive reflexive pronouns, but instead for other preverbal elements.

In the following section, I examine ambiguity broadly as it relates to the results presented in this thesis.

7.4. Ambiguity

The linguistic feature that motivated this dissertation the most was how ambiguous verbal forms may relate to subject pronoun realization. Hochberg (1986:618), in her study of /s/ deletion in Puerto Rican Spanish, which renders 2SG regular verb forms identical to the 3SG, found "pronoun usage is highest with those verb forms rendered ambiguous by /s/ deletion". This finding is often referenced in studies of pro-drop in Romance languages, and ambiguous verbal morphology has become a commonly investigated variable in pro-drop studies. While ambiguity of verb forms is only a significant variable for the AIS data, in this section I present an overview of the variable, how it presented in each of the corpora, and how it relates to the findings of this dissertation.

Rate of ambiguous VPs in comparable studies

Study	%Ambi	%Ø
CHILS	34%	72%
AIS	21%	76%
Lastra & Butragueño, 2015	17%	78%
Carvalho & Child, 2011	14%	65%
Limerick, 2019	11%	73%
Padilla, 2020	10%	81%
HLVC	7%	80%

Table 75. Ambiguous verbal inflection rates across Spanish varieties and this thesis' results alongside their respective null rates.

Table **75** shows the rate of ambiguous verb forms in the datasets presented in this study, alongside four studies of Spanish varieties. None of the Spanish varieties has an ambiguous rate as high as the Ciociaro corpora, but they all have higher rates than found in Calabrian Italian. From the Spanish varieties, it does not appear that rates of ambiguity and null subjects are inversely related, although the difference in ambiguous rates is much smaller than between the three Italian varieties.

As discussed in **3.1.2**, Spanish varieties have less syncretism across verbal paradigms than Ciociaro. For studies that examine verbal ambiguity of Spanish varieties, there is a general pattern of ambiguous paradigms as described by Limerick (2019:255), "first and third person singular verbs in the imperfect, subjunctive (present or past), conditional, or pluperfect". This contrasts with Ciociaro, where ambiguous forms are possible across all singular subjects and, though less common, some plural subjects as well.

The CHILS data have a high rate of ambiguous verbal forms (34%). But this ambiguity is not evenly distributed across subjects. Table **76** shows how common ambiguous verbal forms are for each subject. The first three columns show how many tokens have an ambiguous inflection for each subject (*n*), their percentage of the total number of tokens in the dataset (*%Ambi*), and the null rate of those ambiguous tokens. The final two columns show the total number of tokens for each subject in the dataset (*total*), and the overall null rate of that subject.

Ambiguous verbal forms by SUBJECT for CHILS

SUBJECT	n	%Ambi	%∅	total	%∅
1SG	363	70%	72%	518	68%
2SG	23	45%	57%	51	66%
3SG.M	148	32%	64%	464	64%
3SG.F	39	34%	77%	116	77%
3PL.M	17	5%	59%	374	93%
3PL.F	1	5%	0%	20	75%
1PL	0	0%	---	193	57%
Total	**591**	**34%**	**69%**	**1,736**	**72%**

Table 76. Ambiguous verb forms in the CHILS data and their null rate as a proportion of the total number of tokens for each subject.

The expected result is that the null rate for ambiguous tokens would be lower than the overall null rate for a given subject. This is the case for 2SG, which has a null rate of 57% for ambiguous phrases and an overall null rate of 66%, as well as for 3PL.M (59% vs. 93%). The other subjects, however, do not have this result; 3SG.M and 3SG.F have identical null rates for ambiguous phrases as in their overall results, and 1SG subject actually have a higher null rate in ambiguous phrases (72%) than the overall null rate (68%). In other words, if an ambiguous verb form rendered the subject unrecoverable for the listener, we might expect a much lower rate of null subjects, but my results do not show that for 3SG.M and 3SG.F.

Despite the frequency of ambiguous verb forms in the CHILS dataset (34% of all tokens), it is not a significant variable. This is something that has also been found in some Spanish dialects (e.g., Carvalho and Child 2011:23). On the other hand, many studies do find ambiguity to be a significant variable for pro-drop. A potential reason for why it is not significant in the CHILS data is that ambiguity is not entirely dependent on a verbal inflection but is part of the larger context of the narrative. This is also noted by Travis (2007:118), "cases of true ambiguity are rare in natural discourse, as even with unexpressed subjects the morphological ambiguity is generally resolved by context". The priming variables, one of which was significant in each of the analyses, were a means of testing the importance of that context. As shown in **7.3.2**, there is

an inverse relationship between the significance of SWITCH REFERENCE and the rate of AMBIGUITY.

This context, or lack of it, can also explain why ambiguous verb forms were found to be a significant factor in the AIS dataset, but not in the CHILS or HLVC analyses. For the elicitation of the AIS data there is no context for each phrase; each one is not part of a larger narrative from which the subject can be interpreted. Nagy et al. (2011:137) describe the use of overt subject pronouns as a "'last resort' strategy where recoverability of the subject is not possible through either agreement morphology or a topic antecedent". For the CHILS and HLVC, topic antecedent applies, but it does not for the AIS.

Therefore, this artificial environment of the AIS data collection, absent from natural discourse, means true ambiguity is more common, and thus morphological ambiguity cannot be resolved from context. However, in the CHILS corpus, ambiguity of verbal forms is only tangentially linked to pro-drop. Returning to the functional hypothesis, I do not think that a narrow interpretation of it, whereby an ambiguous inflection is correlated with an overt subject pronoun, is supported by these results. Throughout this chapter, there have been examples that hint at the functional hypothesis, but often once scrutinized, it simply is not supported. That is not to say that it is not a useful idea, only that for the CHILS corpus to fully support a strict interpretation of it, subject pronouns would be overt much more frequently when occurring with an ambiguously inflected verb.

As this section has shown, verb-final vowels are frequently realized as [ə] or entirely absent in Ciociaro, and there is also a much higher rate of ambiguous verb forms than in the AIS and HLVC dataset. Yet, the ambiguity of these forms is not a significant variable in pronoun

realization for either the CHILS or the HLVC analyses. Despite this lack of significance, the

more syncretic verbal paradigms of Ciociaro may interact with other variables, such as SWITCH

REFERENCE. Future research and analysis of verbal ambiguity and its interactions is necessary to

better understand its role in pro-drop.

8. Conclusion

In this concluding chapter, I situate the results of this study within the wider field of variationist research, and I discuss how the work presented here can be used by researchers and communities more broadly. In **8.1**, I provide a summary of the linguistic contributions of this dissertation to the study of pro-drop in Romance languages and beyond. **8.2** presents this dissertation within the field of heritage language research, and how the CHILS corpus can provide a model for scholarly access to varieties that may not exist anymore in their homelands. **8.3** is a discussion of future directions for this research; ideas and areas of study that fall outside the scope of this thesis but would further our understanding of the constraints on pro-drop, of the Ciociaro language, and of heritage languages.

8.1. Pro-drop across Romance languages

Pro-drop is one of the most-studied features of Romance languages, both in formal syntactic and in variationist linguistics. The variables that were included in this dissertation have been found to be significant in methodologically comparable studies (see **3.3**). This background of research provided a solid foundation upon which I built the methodology of this work.

There is little previous research on Italian varieties within the variationist framework (*cf.* Baird et al., 2021; Cristiano & Nagy, 2024; Dal Negro & Vietti, 2006; Filiaci, 2010; Heap, 1997; Nagy & Celata, 2022). This dissertation provides an opportunity to test and modify the variables that have been significant in Spanish varieties to accommodate the linguistic differences of Italian and Italo-Romance varieties. While pro-drop designates the same phenomena in any language, the variable realization of a subject pronoun, there is no reason to assume that the variables which constrain or predict pro-drop in Spanish varieties would necessarily work the same, or to

216

the same degree, in Calabrian or Ciociaro. This is also true of the heritage varieties. Many of the pro-drop studies of heritage Spanish use corpora collected from communities in the United States (Bayley & Pease-Alvarez, 1997; Erker & Guy, 2012; Otheguy et al., 2007; Travis, 2007; inter alia). While these heritage communities vary in their makeup and size, it is important to note that 42 million Americans speak Spanish at home (U.S. Census Bureau, 2022), making heritage Spanish in the United States very different from heritage Calabrian Italian or Ciociaro in Canada. Therefore, the present study of understudied Italian languages expands our understanding of pro-drop in Romance languages, both as homeland and heritage varieties.

To understand how pro-drop compares in these languages to Spanish varieties, it is important to also look at similarities and differences in the significant variables. Priming variables are significant in both Spanish and Italian varieties. Specifically SWITCH REFERENCE is a significant variable for Calabrian Italian as well as for many Spanish varieties (see Torres Cacoullos & Travis, 2019:20 for an overview). They also share the direction of effect: the same referent is more likely to be null, and a switched referent is more likely to have an overt pronoun. The SUBJECT variable, often the most significant variable in analyses of Spanish varieties, is also significant for the Italian varieties presented here, and the variable with the biggest range of effect for the CHILS analysis (see **7.3.1**). However, the hierarchy of variants is different for Spanish varieties than for the analyses presented in this dissertation. The established pattern in Spanish varieties is plural subjects are more likely to be null than singular subjects, so that number is the most significant aspect of subject in these analyses. While this is mostly true for the HLVC analysis (with the exception of 3SG.M), for the CHILS analysis shows that 3rd person subjects favour null subjects (with the exception of 3SG.M), and 1st and 2nd person subjects disfavour null subjects, so PERSON is the most important conditioning factor of subject in this

corpus. This division in the variant hierarchy for subject is also reported by Nagy et al. (2018) for Faetar. This reinforces the importance of examining less-studied varieties. SUBJECT is a significant variable for pro-drop, potentially as a pan-Romance effect, but how PERSON and NUMBER affect pronoun realization can vary in Italian (and Francoprovençal) when compared to Spanish varieties.

The potential effects of the functional hypothesis on pro-drop, however, is the central research question of this dissertation. I wanted to test if verbal ambiguity would affect the use of null subjects, as well as what other independent variables may affect the functional hypothesis. The ambiguity of verbal morphology has been included as a variable for pro-drop analysis in some Spanish varieties, but I believe this dissertation is the first time it has been brought to the study of an Italian dialetto.

In the previous chapter (**7.4**), I described how the high rate of morphological ambiguity in the verbal paradigm is correlated with a lower null-subject rate in Ciociaro than in Calabrian Italian. For example, SUBJECT is a significant variable in all three analyses. There is a pattern in the verbal paradigm for ambiguous inflections: 1SG inflections are often ambiguous for subject and 3PL inflections often are not. At the same time, 1SG subjects disfavour null subjects and 3PL favours null subjects across all the corpora studied here.

Despite this overall result, it is not as simple as an increase in the rate of ambiguous verb forms leading directly to a decrease in the use of null subjects. A listener can recover the subject of a VP with an ambiguous verb form and a null subject from the context, priming variables, and other features of the discourse. This may explain, in part, why ambiguity is only a significant variable in the AIS analysis, despite CHILS having a higher rate of ambiguous forms and a lower

rate of null subjects. The AIS corpus is a series of disconnected prompts and responses, as such the verbal inflection is often the only indication of subject when a subject pronoun is not realized. However, as shown in the results of the AIS analysis (**6.2.3**), ambiguity is not a binary parameter that flips based on the verbal inflection. Unlike the HLVC and CHILS participants, the AIS participants are not having a conversation and do not have the same communicative needs. The participants are translating prompts that are given to them, so an unambiguous subject for each VP is less important than in natural discourse. So, while AMBIGUOUS VERB FORMS was a significant variable in the AIS analysis, it had the smallest range of effect. Further, it was not a significant variable in either the CHILS or HLVC analyses. Therefore, the functional hypothesis, in the narrow sense of an ambiguous verbal inflection triggering an overt subject pronoun, is only partially supported in this dissertation.

Comparison of null and ambiguous verb rates across corpora

	Calabrian (HLVC)	AIS	CHILS
%∅	80%	76%	72%
%Ambi	7%	21%	34%

Table 77. Rates of null subjects and of ambiguous verbal forms in the HLVC, AIS, and CHILS corpora.

Despite this, Table **77** does point to a relationship between ambiguity and pro-drop. This supports a broader interpretation of the functional hypothesis. When a verb is ambiguously inflected, as is often the case in Ciociaro, that does not mean the subject of the phrase is necessarily ambiguous. And, if the subject *is* ambiguous, it can be clarified in many ways within a conversational context, one of which is an overt subject pronoun. The question is not whether the verb is ambiguous for subject, it is whether a subject pronoun is required to disambiguate it. Instead, as I have shown in my analyses, other aspects of the VP, such as TENSE, SUBJECT, or the preceding phrase play a more significant role in predicting subject pronoun usage. Taken together, Table **77** does point to a relationship between higher rates of ambiguity and lower rates of null subjects, though the two

are not directly correlated. Compared to the HLVC, CHILS has a 27% higher rate of ambiguous inflections, but only an 8% lower rate of null subjects. For this higher rate of ambiguous subjects one of the ways to recover the subject is by using overt subject pronouns, but there are other solutions to ambiguous inflections. This is a more nuanced approach to the functional hypothesis: one of the ways a subject is disambiguated is by distinct verbal inflections, and another is by overt subject pronouns. Yet, as I have shown in this dissertation, there are other ways speakers respond to ambiguity—either other means of recovering the subject (e.g., priming variables), or by simply tolerating the ambiguity.

The functional hypothesis, as discussed in **3.1.3**, proposes that the function of language is to transmit information efficiently to the listener. For pro-drop, this means that subject pronouns should be used when the subject is otherwise ambiguous, and that the subject pronoun is not necessary when the phrase's subject is unambiguously indicated in some other way. The direct correlation between ambiguous verbal inflections and overt subject pronouns is not supported in this dissertation. That being said, a more nuanced approach that looks elsewhere in the discourse for other independent variables that may transmit the information of the subject is supported. This broader approach relies less on a direct connection between subject pronouns and verbal ambiguity, and instead takes a more holistic look at the larger discourse in which the ambiguity occurs. I suggest this is a more useful avenue of research for pro-drop studies, both within the subfield of Romance languages and beyond. By extending the scope of the functional hypothesis, we do not assume every occurrence of subject ambiguity should be resolved within the phrase by an overt subject pronoun. This encourages researchers to seek other explanations rather than an overreliance on the strength of the functional hypothesis (see Labov, 1987). This also allows us to extend our findings and hypotheses beyond Romance languages, to languages that do not

show verbal agreement, tolerate heavily syncretic verbal paradigms, or use pro-drop in entirely different ways (Huang, 1984).

8.2. Heritage language research

Research on a heritage language provides insights into a language community that is separated from its homeland. Linguists use corpora like Nagy's HLVC corpus (2011, 2024) to analyse these homeland and heritage varieties to examine how they evolve, how (or indeed if) they diverge, and what can account for these differences.

In this dissertation, I examine pro-drop in both Calabrian Italian and Ciociaro. In **6.4** and **7.1**, I show the similarities and differences between the homeland and heritage varieties. The comparison of the homeland and heritage varieties of Calabrian Italian shows almost no difference between the two varieties as it relates to pro-drop. Similarly, pro-drop appears to function similarly in the AIS and CHILS analyses, allowing for the methodological differences. Comparing homeland and heritage varieties allows linguists to better understand contact effects with dominant languages. For the Calabrian Italian community in the GTA and the Ciociaro community of Sarnia, contact with English has not led to an increased rate of subject pronoun realization. There is, though, more to studying heritage languages than just the possible effects of contact with a dominant language.

Through the process of chain migration, heritage-language communities are often tightly knit, and can provide access to a variety that would be difficult to research in its homeland, as in the case of Ciociaro in Sarnia. These communities, in some cases, may hold the only remaining speakers of a variety. This is particularly true of Italian varieties, as noted by Haller (1987:393), "[w]hile the use of dialects in Italy was gradually limited to home and family, or abandoned in

221

favor of the standard language, the local dialects remained for decades the only form of speech among Italian emigrant communities."

Thus, the CHILS corpus provides a record of a dialetto that may already be lost in Italy. As discussed in **2.1**, Italy's linguistic landscape has changed a great deal since the period of mass emigration following World War II. Most Ciociari in Sarnia emigrated from Italy in the 1950s, when schooling in standard Italian was nearly non-existent, and illiteracy was more than 20% (Mariani, 2021:8). Further, Ciociaro is the majority Italian variety of Sarnia (Di Cocco, 1991:147). This majority status in Sarnia's Italian community has allowed Ciociaro to be preserved, as evidenced by its similarities to the AIS data for the same region. This is in contrast to what is found by both Alfonzetti (2005) and Haller (1987) in New York City, which has a much more diverse and larger Italian community than Sarnia. They both found that a speaker's variety was used with parents, and (quasi-)standard Italian ("a non-standard lingua franca" according to Haller, 1987:393) was used to communicate across communities. So, while there are Ciociaro communities elsewhere (e.g., in Windsor, Ontario), it is not the variety of the majority, and thus the larger Italian community is likely to have developed its own non-standard lingua franca. Therefore, while difficult to say with certainty, Ciociaro, as it exists among Sarnia's heritage community, may provide our best or only access to a dialetto that no longer exists anywhere else.

Every heritage community is different, but they all provide access to varieties worthy of study. This, of course, does not just apply to how pro-drop works in Calabrian Italian and Ciociaro. By studying these heritage varieties, we expand our understanding of Romance languages. And for

any variety, finding and engaging with these communities may be easier for North American researchers than travelling to Italy to conduct fieldwork.

By engaging with the community, I was able to create an online archive (www.italiansinlambton.ca) in partnership with Lambton County to preserve the community's voices, stories, images, letters, and other ephemera for posterity. These materials provide a wealth of knowledge for researchers interested in immigrant communities, Italians in Canada, and for Italian Canadians hoping to learn more about their heritage. While Sarnia's Ciociari may not find the analysis of pro-drop especially interesting, the CHILS corpus and archive do contribute meaningfully to the community. The stories the participants kindly shared with me of early life in Italy, their immigration to Canada, and building a life in Sarnia are important chapters in the histories of many of Sarnia's families. These are important, both for their contribution to our understanding of Romance languages, but also as part of the story of our community.

The fieldwork that was conducted with the Ciociaro community also provides a potential template to future researchers hoping to engage with immigrant communities. By partnering with Western University and local community leaders, I was able to provide a means of preserving the Ciociaro variety in Sarnia. I am also a part of the Italian Canadian Archives Project (ICAP) and from my personal experiences with this research I have created toolkits and materials so that communities can create their own archives and preserve their histories (https://icap.ca/toolkit/).

For this dissertation, sociolinguistic interviews were conducted with members of Sarnia's Ciociaro community. However, the research already conducted by the community, such as Di Cocco (1991), provided an understanding of the community that would have been impossible to

collect on my own. This wealth of demographic information about the makeup of the community, its history, and its place in the city of Sarnia provides incredibly valuable insights into the community for researchers in other disciplines, including anthropological linguists.

Sarnia's Ciociaro community will likely not engage with a dissertation about subject pronoun realization. However, by creating a digital archive; digitizing photos, documents, and videos; and creating a toolkit on how to conduct fieldwork like this, I hope to help other communities do the same. This is both a means of giving back to the community which made this thesis possible, but also a means of helping to ensure that other heritage languages are preserved for future study by linguists.

8.3. Future research

There are many features of Ciociaro that I believe are of linguistic interest but are outside the scope of this study. In this section, I first propose features of Ciociaro that are related to the VP and (potentially) to pro-drop. This is followed by motivating study of other aspects of the language that are not part of the verb phrase.

First, the possibility of metaphony occurring in Ciociaro is a potentially fruitful avenue of investigation. It did not appear that metaphony was productive in the AIS or CHILS datasets. Ciociaro is also said to be just outside the regions where metaphony is common (Merlo, 1919; Merlo & Vignoli, 1920). However, some verbal paradigms from the AIS appear to show limited evidence of metaphony. This is not predictable based on location, verbal paradigm, or phoneme. But there are some examples, such as the participant from Scanno who produces metaphony for the verb *venire* ('to come'), contrasting the 2SG subject (/vinːə/) with 1SG and 3SG (/vɛnːə/) (Jaberg & Jud, 1928:1688). If metaphony is productive in Ciociaro, this would mean that in some

224

cases a listener could recover the subject of a verb because the inflection is not on the final vowel, but instead on the stressed vowel. Metaphony is also described as productive in the dialetto spoken in the neighbouring province of Isernia (Iannacito, 2000). Investigating metaphony would require a careful acoustic examination of the tokens of the CHILS data, but an examination of its frequency in the AIS dataset, both for VPs and more broadly, would be a useful first step toward determining whether metaphony is productive in Ciociaro, or used to be.

Second, an analysis that looks more closely at priming effects in the corpora included in this dissertation could yield interesting results. A more fine-grained analysis of SWITCH REFERENCE may lead to a better understanding of this priming effect on subject pronoun realization. For example, it would be interesting to include cases when the referent may be introduced as an object in the preceding clause (e.g., 'I talked to my grandma$_j$, [she$_j$] is awesome'). In my analysis, the subject of this VP would be coded as a switch in referent, but the referent is nonetheless present in the preceding context. It would be interesting to see how these types of phrases compared to other variants. Another approach, as proposed by Torres Cacoullos & Travis (2019:11), is to further divide SAME REFERENT into those that share "syntactic linking" and "prosodic linking" to those that do not. Syntactically linked phrases are those that fall within the same phrase, coordinated with a conjunction (e.g., "Chris$_j$ called and [she$_j$] told us the story"). Similarly, prosodically linked phrases fall into the same prosodic contour (e.g., "Julie$_j$ came over, [she$_j$] brought thyme and lavender"). When coding for SWITCH REFERENCE, Torres Cacoullos & Travis found that for VPs that were both prosodically linked and syntactically linked to the preceding VP, the subjects were more likely to be null than if only linked in one way, and those subjects of VPs linked in one way were more likely to be null than if not linked

in either way. This may shine more light on the priming effect of SWITCH REFERENCE on subject pronoun realization in both Calabrian and Ciociaro.

Third, following from the framework of Erker & Guy (2012), it would be very interesting to see what a more thorough examination of the most frequent verbs in the HLVC and CHILS datasets could yield. The results of the analysis of lexical frequency presented in **6.3.5** were not conclusive. However, a more thorough analysis could provide useful insights into patterns of pro-drop in Calabrian Italian and Ciociaro. There are also several other ways to analyze verbs and their relation to pro-drop. As well as coding VPs for verb and verb class, Erker & Guy examined lexical frequency as a variable, coding not by verb, but by verb form. So, instead of coding all the VPs that have the verb *essere*, I could separate the verb into its paradigm, examining the most frequent forms. This would be a novel way to compare, for example, *è* for a 3SG subject, as well as the ambiguity of *sono*, which is the verb form for both 1SG and 3PL subjects. They also coded verbs that were conjugated with regular morphology from the verbs with irregular morphology. For example, *comprare*, *pensare*, and *mangiare* ('to buy', 'to think', and 'to eat') are all conjugated in the present tense with *-are* removed and *-o* (1SG), *-i* (2SG), and *-a* (3SG) added; however, the equivalent forms of *fare* ('to do') are *faccio*, *fai*, and *fa.* The hypothesis of Erker & Guy (2012) is that more frequently occurring verb forms and irregularly conjugated verb forms both favour null subjects. It would be interesting to examine this in both Calabrian, which typically has distinct forms across the paradigm, and Ciociaro, which does not, to see how the results compare. They also propose a distinction between what they term "constraints that are systemic to the verb forms themselves" and "discourse-level constraints" (Erker & Guy, 2012:553). Put simply, the first category are the factors that are inherent to the verb form and cannot vary from one occurrence of that form to the next (e.g., PERSON, NUMBER,

TENSE, etc.). The other category, including factors such as SWITCH REFERENCE and TENSE, are not tied to verbal selection, and operate at the higher-level of the larger discourse. In summary, understanding which variables should co-vary at the verb-form level, and which variables operate at the discourse level will provide novel ways to examine interactions between pro-drop and verb use.

Fourth, my analysis included only human subjects (see **3.1.1**). As a result, phrases such as the second VP in (**61**) were not included in my analysis. In this example, Frank L says that his wife only speaks to him in English when there are problems (meaning he is in trouble).

61) [kwandə paɹlə ɪŋlɛz so pɹoblem]
     When   speak.PR.3SG English   is.PR.3PL   problems
     Quando parla inglese, sono dei problemi
     'When (she) speaks English, (there) are problems'
     (CHILS, Frank L, 17:22)

More may be revealed about pro-drop in Ciociaro by comparing 3rd person subjects that are human (e.g., 'he is rambling', for a person), non-human animate subjects (e.g., 'she is running', for a dog), meteorological (e.g., 'it is raining', for the weather), and impersonal (e.g., 'it is revealing', for a topic). While most 3rd person subjects from the CHILS interviews are for human or impersonal subjects, the AIS questionnaire contains a good balance of prompts from each of these categories. Examining if these different categories of subject have different rates of null subjects may provide insights into how the subject pronoun functions in Ciociaro. Meteorological verbs cannot vary by subject, regardless of inflection, so we might expect these phrases to occur at a higher rate of null subjects. For example, the AIS interview includes prompt #396 *tuona* ('it's thundering'). In this case, a subject pronoun does not disambiguate, and the six

AIS locations included in this dissertation do not include a pronoun either. This type of analysis could also add to research that has been done on other Gallo-Romance varieties, such as Heap's analysis of northern Italian and Occitan varieties (1997), and Diémoz's analysis of Francoprovençal varieties (2007).

Finally, there are several features in Ciociaro that merit investigation related to nominal phrases. Commonly, possessives are postnominal. For example, *la nonna mia* instead of *la mia nonna* for 'my grandma' (*Antonia,* 08:18). This is found in many varieties throughout Italy (see Cardinaletti & Giusti, 2019 for an overview of the phenomenon). There is also a reduction in the distinctiveness of articles. Often in both the CHILS recordings and the AIS corpus definite articles are pronounced as /lə/ or /lu/, regardless of gender or number. Further, many nouns also end in /ə/ or a consonant, with no inflection for grammatical gender or number. There are also words in Ciociaro that do not have an obvious cognate in Italian, such as /wajonə/ for both *bambino* and *bambini* ('child' and 'children'). As a result, phrases like (**62**) are common:

62)   [a   nu   lə   wajonə]
            to   us   the-PL child-PL
            a noi, i bambini
            'to us, the children'
            (CHILS, Antonia, 46:48)

This dissertation is the first description of pro-drop in Ciociaro. However, it is also, I believe, the first description of Ciociaro in more than a century. As such, there are many interesting aspects of the grammar that merit further investigation and research.

8.4. Summary

Pro-drop is a well-studied area of Romance languages, and in my dissertation, I focussed on how it may relate to the functional hypothesis. I used three corpora to investigate pro-drop in a regional Italian and an Italian dialetto. Nagy's HLVC corpus established a framework for my research project (2011), providing both homeland and heritage varieties of Calabrian Italian. Heap's examination of Italo-Romance dialetti introduced me to the AIS (1997), from which I created my homeland Ciociaro corpus (Jaberg & Jud, 1928). I created the third corpus, the heritage Ciociaro corpus, by interviewing members of Sarnia's heritage Ciociaro community. I sought to examine pro-drop within each of these language communities, but also to compare them: homeland to heritage, Calabrian to Ciociaro, and both to pro-drop research in other languages.

Once I had the corpora and framework, the next step was defining an appropriate methodology. To the best of my knowledge, the only studies of the Ciociaro dialetto were conducted more than a century ago (Merlo, 1919; Merlo & Vignoli, 1920). The AIS data was recorded during this same period (Jaberg & Jud, 1928). Most research on pro-drop in Romance languages, and especially in heritage communities, is conducted on Spanish varieties. While there is a body of research to draw on when preparing a methodology to examine pro-drop in Spanish varieties, I cast a wide net in choosing which variables to include in my analyses. While many, in fact most, of these variables were not significant, I believe they all strengthened my understanding of pro-drop in Ciociaro, and of the language more broadly.

The analyses were conducted using Rbrul, with the overall null rate being 80% for both heritage and homeland Calabrian Italian, 76% in the AIS, and 72% in the CHILS. The analyses also

found that three independent variables were significant in each corpus. SUBJECT was a significant variable for each corpus, and although there were outliers, plural and 3rd person subjects broadly favour null subjects, while singular and 1st and 2nd person subjects broadly disfavour them. A priming variable was significant in each corpus as well, with SWITCH REFERENCE being significant in the analysis of Calabrian Italian, PRONOUN IN PROMPT being significant in the AIS, and PREVIOUS REALIZATION in CHILS. Each corpus had a third variable that was also significant, PREVERBAL ELEMENTS for the HLVC analysis, VERBAL AMBIGUITY for the AIS, and TENSE for the CHILS.

Hypothesis	HLVC	CHILS	AIS
(1) Heritage and homeland			
(a) Similar null subject rate	✓	✗	✗
(b) Similar significant variables	✓	✳	✳
(2) Null subject rates			
(a) Ciociaro has a lower null rate than Calabrian		✓	
(b) Lower null rate correlated with higher rate of ambiguity	✳	✳	✓
(c) AIS has lower null rate than CHILS, similar rate of ambiguity		✗	✗
(d) i. Higher null rates than French and Faetar	✓	✓	✓
(d) ii. Lower null rates than Italian	✓	✓	✓
(d) iii. Similar null rates to Spanish	✳	✓	✓
(3) Significant variables			
(a) i. Subject is a significant variable	✓	✓	✓
(a) ii. Ambiguous inflections disfavour null subjects	✳	✳	✓
(b) Switch reference / pronoun in prompt is significant	✓	✗	✓
(c) Tense is significant for AIS			✗
(d) VPs with reflexive pronouns favour null subjects	✗	✗	✗

Table 78. Table summarizing the hypotheses that were supported (green), refuted (red), and mixed (blue).

These results were then compared to my hypotheses, which are summarized in Table **78** (and is repeated from Chapter **7**).

The first group of hypotheses sought to compare the homeland to the heritage varieties. The Calabrian varieties had nearly identical null subject rates (1a), and similar significant variables (1b). This was not the case for the CHILS and AIS analyses. The heritage and homeland varieties of Ciociaro have different null subject rates (1a). In terms of significant variables (1b) they share SUBJECT as a significant variable, but otherwise have different significant variables. Also, the heritage variety does not show evidence of contact effects with English, and in fact has a lower null-subject rate than the homeland variety. There is also no evidence of contact effects with English in the heritage Calabrian from the HLVC analysis.

Ciociaro does have a lower null rate than Calabrian (2a), and it does have a higher rate of ambiguous verbal inflections (2b). However, although they are correlated, Rbrul analyses do not show a significant effect of ambiguity on subject pronoun realization—except for the AIS. Although the AIS had a higher rate of null subjects, and a lower rate of ambiguity (2c), ambiguity is a significant factor in subject pronoun realization. When these results are compared to other pro-drop studies with similar methodologies, the results are largely as expected (2d). Ciociaro and Calabrian Italian have an overall null rate that is higher than French, lower than Italian, and similar to many Spanish varieties. Calabrian Italian, as a regional Italian, was expected to have a higher null rate than all varieties of Spanish. While it is higher than most reported null subject rates, Padilla (2020) reports a slightly higher overall rate in her study of Equatoguinean Spanish.

The significant variables were next considered. First, SUBJECT was a significant variable for each analysis (3a). I hypothesized this would be due to ambiguous inflections, which are more common for certain subjects, leading to certain subjects disfavouring null subjects. That was not

borne out by the analysis of the HLVC or CHILS corpora, but it was true of the AIS. I propose

this is because of the AIS's disconnected prompts, which provide no context from which a VP's

subject can be recovered. For the priming variables, SWITCH REFERENCE was significant in the

HLVC analysis, and PRONOUN IN PROMPT was significant in the AIS analysis (3b). While it was

not a significant variable for the CHILS analysis, another priming variable, PREVIOUS

REALIZATION, was significant. I also predicted that TENSE would be significant for the AIS (3c),

and the variants would be ranked according to their relative ambiguity. Tense was not a

significant factor for the AIS analysis, and while it was for the CHILS analysis, their relative

ambiguity is not related to their ranking. Finally, reflexive pronouns, which help to disambiguate

some subjects, were predicted to favour null subjects (3d). However, this was not the case. While

PREVERBAL ELEMENTS was a significant factor for the HLVC dataset, reflexive pronouns did not

favour null subjects more than negative particles or object pronouns.

Overall, the results of my analyses from the HLVC and CHILS corpora do not consistently

support the functional hypothesis. However, particularly with hypotheses (2b) and (3a), it is

supported by the AIS results. As discussed in **7.4**, I believe this can be explained by context in

the discourse. Both the CHILS and HLVC are interviews; participants are talking for minutes at

a time telling a particular story or discussing a certain topic. In this context, an ambiguous

inflection simply does not matter as much as I had hypothesized it would. However, the AIS is a

collection of disconnected phrases, devoid of conversational context. In this atypical

environment ambiguity plays a more significant role in subject pronoun realization.

The linguistic question that steered the methodological approach of this dissertation was: could

the functional hypothesis explain variation in subject pronoun usage? While ambiguity in the

verbal morphology is a significant variable for the AIS data, it is not significant for either the CHILS or the HLVC. This does not mean that the functional hypothesis does not account for any of the variation in subject pronoun usage, only that it cannot explain all of it.

This dissertation was motivated by a desire to better understand pro-drop, Ciociaro, and heritage language communities more broadly. This dissertation, and the fieldwork that produced the CHILS corpus, add an under-described Italian variety to the body of research on pro-drop in Romance languages. This study also contributes to our understanding of how heritage languages relate to their homeland varieties, as well as to their broader, "dominant language" communities. It has contributed to our understanding of pro-drop in Romance languages, and the methodological approaches that can be brought to its study. This dissertation is also a meaningful contribution to the study of Ciociaro, and, to the best of my knowledge, the first in more than a century. Finally, the creation of the CHILS corpus and archive contribute a permanent resource for, and record of, Sarnia's heritage Ciociaro community, without whom this dissertation would not have been possible.

9. References

Abreu, L. (2009). Spanish subject personal pronoun use by monolinguals, bilinguals and second language learners [Ph.D., University of Florida]. In *ProQuest Dissertations and Theses* (304885870). ProQuest Dissertations & Theses Global. https://www.lib.uwo.ca/cgi-bin/ezpauthn.cgi?url=http://search.proquest.com/dissertations-theses/spanish-subject-personal-pronoun-use-monolinguals/docview/304885870/se-2?accountid=15115

Abreu, L. (2012). Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, 1–8.

Alfonzetti, G. (2005). Intergenerational variation in code switching. Some remarks. *Italian Journal of Linguistics*, *17*(1), 20.

Avila-Jimenez, B. (1996). Subject pronoun expression in Puerto Rican Spanish: A sociolinguistic, morphological, and discourse analysis [PhD Thesis, Cornell University]. In *ProQuest Dissertations and Theses*. https://proxy.library.brocku.ca/login?url=https://www.proquest.com/dissertations-theses/subject-pronoun-expression-puerto-rican-spanish/docview/304287797/se-2?accountid=9744

Avolio, F. (2000). "Ma nuje comme parlamme?" Problemi di descrizione e classificazione dello spazio dialettale "campano." *Romance Philology*, *54*(1), 1–28. https://doi.org/10.1484/J.RPH.2.304360

Baird, A., Cristiano, A., & Nagy, N. (2021). Apocope in Heritage Italian. *Languages*, *6*(3), 120. https://doi.org/10.3390/languages6030120

Bayley, R., & Pease-Alvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, *9*(3), 349–371. https://doi.org/10.1017/S0954394500001964

Berruto, G. (2005). Dialect/standard convergence, mixing, and models of language contact: The case of Italy. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change* (pp. 81–95). Cambridge University Press. https://doi.org/10.1017/CBO9780511486623.005

Berruto, G. (2017). What is changing in Italian today? Phenomena of restandardization in syntax and morphology: an overview. In M. Cerruti, C. Crocco, & S. Marzo (Eds.), *Towards a New Standard*. De Gruyter. https://doi.org/10.1515/9781614518839-002

Bianchi, B. (1888). *Il dialetto e la etnografia di Città di Castello, con raffronti e considerazioni storiche; memoria*. S. Lapi. https://ia902604-us-archive-org.proxy1.lib.uwo.ca/26/items/ildialettoelaetn00bianuoft/ildialettoelaetn00bianuoft.pdf

Biberauer, T. (2018). Pro-drop and emergent parameter hierarchies. In *Null subjects in generative grammar: A synchronic and diachronic perspective*. Oxford University Press. https://doi.org/10.1093/oso/9780198815853.003.0005

Bonifazi, C., & Heins, F. (2000). Long-term trends of internal migration in Italy. *International Journal of Population Geography*, *6*(2), 111–131. https://doi.org/10.1002/(SICI)1099-1220(200003/04)6:2<111::AID-IJPG172>3.0.CO;2-L

Calabrese, A. (2011). Investigations on markedness, syncretism and zero exponence in morphology. *Morphology*, *21*(2), 283–325. https://doi.org/10/bw4gqd

Calabrese, A. (2016). On the morphophonology of metaphonic alternations in Altamurano. In F. Torres-Tamarit, K. Linke, & M. V. Oostendorp (Eds.), *Approaches to Metaphony in the*

Languages of Italy (pp. 89–126). De Gruyter Mouton. https://doi.org/10.1515/9783110366310-005

Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, *5*(3), 305–334. https://doi.org/10/db5sp3

Cappellaro, C. (2017). The semantic specialization of third person pronoun 'esso' as (−human) in standard Italian. *Revue Romane. Langue et Littérature. International Journal of Romance Languages and Literatures*, *52*(2), 113–136. https://doi.org/10.1075/rro.52.2.01cap

Cardinaletti, A. (2018). On different types of postverbal subjects in Italian. *Italian Journal of Linguistics*, *30*(2), 79–106.

Cardinaletti, A., & Giusti, G. (2019). Micro-variation in the possessive systems of Italian dialects. *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistics Colloquium*, 137–154.

Carvalho, A. M., & Child, M. (2011). Subject pronoun expression in a variety of Spanish in contact with Portuguese. *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*, 14–25.

Carvalho, A. M., Orozco, R., & Shin, N. L. (Eds.). (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown University Press.

Casalicchio, J., & Cordin, P. (2020). *Grammar of central Trentino: A romance dialect from North-East Italy*. Brill.

Cerruti, M. (2011). Regional varieties of Italian in the linguistic repertoire. *International Journal of the Sociology of Language*, *2011*(210), 9–28. https://doi.org/10.1515/ijsl.2011.028

Cerruti, M. (2016). L'italianizzazione dei dialetti: Una rassegna. *Quaderns d'italià*, *21*, 0063–0074. https://doi.org/10.5565/rev/qdi.15

Cerruti, M., & Regis, R. (2014). Standardization patterns and dialect/standard convergence: A northwestern Italian perspective. *Language in Society*, *43*(01), 83–111. https://doi.org/10.1017/S0047404513000882

Colucci, M., & Sanfilippo, M. (2006). L'emigrazione dal Lazio: Il dibattito storiografico. In E. Franzina & M. Sanfilippo (Eds.), *Modelli di emigrazione regionale dall'Italia centro-settentrionale* (pp. 107–123). Sette Città.

Comrie, B., Haspelmath, M., & Bickel, B. (2015, May 31). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Max Planck Institute of Evolutionary Anthropology. https://www.eva.mpg.de/lingua/resources/glossing-rules.php

Cristiano, A., & Nagy, N. (2024). Toronto Heritage Italian (r): Maintaining homeland patterns. *Bergen Language and Linguistic Studies*, *14*(1), 37–48.

Cugno, F. (2008). Dialetti e lingue minoritarie nell'Italia contemporanea. *Italia contemporanea*, 30.

Dal Negro, S., & Vietti, A. (2006). The interplay of dialect and the standard in anonymous street dialogues: Patterns of variation in northern Italy. *Language Variation and Change*, *18*(02). https://doi.org/10.1017/S0954394506060108

Dal Negro, S., & Vietti, A. (2011). Italian and Italo-Romance dialects. *International Journal of the Sociology of Language*, *2011*(210). https://doi.org/10.1515/ijsl.2011.031

Daniele, V. (2021). L'emigrazione italiana dall'unità ad oggi. In L. Sinisi (Ed.), *Emigrazioni e immigrazioni: Una sfida della storia e della contemporaneità* (pp. 51–74). Pacini giuridica.

De Mauro, T. (1979). *Storia linguistica dell'Italia unita* (10. ed). Laterza.

De Pascale, S., & Marzo, S. (2016). Gli italiani regionali. Atteggiamenti linguistici verso le varietà geografiche dell'italiano. *Incontri. Rivista europea di studi italiani*, *31*(1), 61. https://doi.org/10.18352/incontri.10151

Del Boca, D., & Venturini, A. (2003). Italian Migration. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.475021

Devoto, G., & Giacomelli, G. (1972). *I Dialetti delle regioni d'Italia*. Sansoni. https://books.google.ca/books?id=G3cIAQAAIAAJ

Di Cocco, C. (1991). *One by one: History of the Italian community in Sarnia-Lambton, 1870-1990 = passo dopo passo, storia della comunità italiana di Sarnia-Lambton*. Italo-Canadian Cultural Club of Sarnia.

Di Domenico, E., Baroncini, I., & Capotorti, A. (2020). Null and overt subject pronouns in topic continuity and topic shift: An investigation of the narrative productions of Italian Natives, Greek Natives and near-native second language speakers of Italian with Greek as a first language. *Glossa: A Journal of General Linguistics*, *5*(1). https://doi.org/10.5334/gjgl.1009

Diémoz, F. (2007). *Morphologie et syntaxe des pronoms personnels sujets dans les parlers francoprovençaux de la Vallée d'Aoste*. A. Francke.

Dota, F., & Panaccione, D. (2017). *L'Uso della lingua italiana, dei dialetti e delle lingue straniere* (Communicato Stampa 207961; Lingua italiana, dialetti, lingue straniere, p. 22). Instituto Nazionale di Statistica. https://www.istat.it/it/files//2017/12/Report_Uso-italiano_dialetti_altrelingue_2015.pdf

ELAN (6.0). (2020). [Computer software]. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, *88*(3), 526–557. https://doi.org/10.1353/lan.2012.0050

Erker, D., & Shin, N. L. (2015). The Emergence of Structured Variability in Morphosyntax: Childhood Acquisition of Spanish Subject Pronouns. In A. M. Carvalho, N. L. Shin, & R. Orozco (Eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 171–191). Georgetown University Press. https://www.tandfonline.com/doi/full/10.1080/23247797.2017.1325125

Filiaci, F. (2010). Null and overt subject biases in Spanish and Italian: A cross-linguistic comparison. *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 171–182.

Germani, A. (2014). *Il tipo dialettale del Lazio meridionale: Alcuni fenomeni linguistici caratteristici*. 31.

Ghimenton, A. (2013). Quantitative Approaches to Italian / Dialect Family Interactions: Considerations of Methodology and Language Transmission in a Contact Situation. *Journal of Language Contact*, *6*(1), 106–133. https://doi.org/10.1163/19552629-006001010

Givón, T. (2013). On the Intellectual Roots of Functionalism in Linguistics. In S. Bischoff & C. Jany (Eds.), *Functional Approaches to Language* (pp. 9–30). DE GRUYTER. https://doi.org/10.1515/9783110285321.9

Google. (n.d.-a). *AIS locations in this dissertation* [Map]. Google My Maps. Retrieved February 25, 2024, from https://www.google.com/maps/d/u/0/edit?mid=17gcLWeEorKBLe5u4-MH5wi3x42pSc1c

Google. (n.d.-b). *AIS locations included in this dissertation, compared with hometowns of majority of Sarnia's Ciociari* [Map]. Google My Maps. Retrieved March 16, 2024, from https://www.google.com/maps/d/u/0/edit?mid=1Xxulc61vYKgXeoJcyZRhURjRe6pMX BuJ&cid=mp&cv=1DrBfMdOFkc.en_US.

Gramellini, F. (2008). Il dialetto del nuovo millennio: Usi, parlanti, apprendenti/El dialecte del nou mil· lenni: usos, parlants, aprenents. *Ianua: Revista Philologica Romanica*, *8*, 181– 201.

Haller, H. W. (1987). Italian Speech Varieties in the United States and the Italian-American Lingua Franca. *Italica*, *64*(3), 393–409. https://doi.org/10.2307/478498

Harrison, B. (2000). Passing on the language: Heritage language diversity in Canada. *Canadian Social Trends*, *58*, 14–19.

Heap, D. (1997). *La variation grammaticale en géolinguistique, les pronoms sujet en roman central* [PhD Thesis]. University of Toronto.

Hernández Constantin, A. (2021). *Variable Subject Pronoun Expression in the Spanish of Londombia: A study of language contact in Canada*. Western University.

Hochberg, J. G. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language*, *62*(3), 609–621. https://doi.org/10/gh7ffs

Holmberg, A. (2010). Null subject parameters. *Parametric Variation: Null Subjects in Minimalist Theory*, 88–124.

Huang, C.-T. J. (1984). On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry*, *15*(4), 531–574.

Iannacito, R. (2000). *Il dialetto molisano in Villa San Michele* [University of Toronto]. https://www.collectionscanada.ca/obj/s4/f2/dsk2/ftp02/NQ53771.pdf

Iannozzi, M. (2017). *The Good [w]ita in Canada* [Master's Thesis]. University of Western Ontario.

Instituto dell'Atlante Linguistico Italiano. (2018). *Segni e simboli usate nelle carte*. Istituto dell'Atlante Linguistico Italiano. https://www.atlantelinguistico.it/images/ali/documenti/segni_e_simboli_ALI.pdf

IStat. (2020). *National Accounts regional main aggregates: Per capita values*. Istituto Nazionale Di Statistica. http://dati.istat.it/?lang=en&SubSessionId=d7346671-4bc3-4bf8-a16a-0d9cd55e9ce9&themetreeid=-200

Jaberg, K., & Jud, J. (1928). *Sprach-und Sachatlas Italiens und der Südschweiz* (Vol. 1). Ringier.

Jaberg, K., & Jud, J. (1987). *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale. 2: Scelta di carte commentate* (G. Sanga, Ed.; S. Baggio, Trans.; 1. ed. italiana, Vol. 4). Ed. Unicopli.

Johnson, D. E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass*, *3*(1), 359–383. https://doi.org/10/dnqx2d

King, R., Mortimer, J., & Strachan, A. (1984). Return Migration and Tertiary Development: A Calabrian Case-Study. *Anthropological Quarterly*, *57*(3), 112. https://doi.org/10.2307/3317926

Labov, W. (1972). *Sociolinguistic patterns*. Univ. of Pennsylvania Press.

Labov, W. (1987). The overestimation of functionalism. In R. Dirven & V. Fried (Eds.), *Linguistic and Literary Studies in Eastern Europe* (1st ed., Vol. 20, pp. 311–332). John Benjamins Publishing Company. https://doi.org/10.1075/llsee.20.16lab

Lastra, Y., & Butragueño, P. M. (2015). Subject pronoun expression in oral Mexican Spanish. *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective*, *39*, 57.

Ledgeway, A. (2010). *Lingua italiana in bocca calabra*: Italian in Calabria. *The Italianist*, *30*(sup2), 95–120. https://doi.org/10.1080/02614340.2010.11917480

Legendre, G. (2010). A formal typology of person-based auxiliary selection in Italo-Romance. *Syntactic Variation: The Dialects of Italy*, 186–200.

Lepschy, A. L., & Lepschy, G. C. (1988). *The Italian language today* (2nd ed). New Amsterdam.

Leroux, M., & Jarmasz, L.-G. (2006). A study about nothing: Null subjects as a diagnostic of convergence between English and French. *University of Pennsylvania Working Papers in Linguistics*, *12*(2), 16.

Limerick, P. (2019). Subject expression in a Southeastern U.S. Mexican community. *Borealis – An International Journal of Hispanic Linguistics*, *8*(2), 243–273. https://doi.org/10/gqfxjv

Loporcaro, M. (2009). *Profilo linguistico dei dialetti italiani* (Vol. 275). Laterza.

Loporcaro, M. (2016). Metaphony and diphthongization in Southern Italy: Reconstructive implications for sound change in early Romance. In F. Torres-Tamarit, K. Linke, & M. van Oostendorp (Eds.), *Approaches to Metaphony in the Languages of Italy* (pp. 55–87). De Gruyter. https://doi.org/10.1515/9783110366310-004

Loporcaro, M., & Paciaroni, T. (2016). The dialects of central Italy. In A. Ledgeway & M. Maiden (Eds.), *The Oxfrod Guide to the Romance Languages* (pp. 228–245). Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199677108.001.0001/acprof-9780199677108-chapter-15

Lorusso, P., Caprin, C., & Guasti, M. T. (2005). Overt Subject Distribution in Early Italian Children. *A Supplement to the Proceedings of the 29th Annual Boston University Conference on Language Development*, 7.

MacDonald, J. S., & MacDonald, L. D. (1964). Chain Migration Ethnic Neighborhood Formation and Social Networks. *The Milbank Memorial Fund Quarterly*, *42*(1), 82–97. https://doi.org/10.2307/3348581

Mariani, A. M. (2021). *Rilettura della evoluzione dei modelli antropologici in Ciociaria dagli anni cinquanta ad oggi*. Tofani Editore.

Martín-Butragueño, P. (2020). An approach to subject pronoun expression patterns in data from the "Project for the Sociolinguistic Study of Spanish in Spain and America." *Spanish in Context*, *17*(2), 294–316. https://doi.org/10/gqfxjt

Merlo, C. (1919). *Fonologia del dialetto di Sora*. F. Mariotti.

Merlo, C., & Vignoli, C. (1920). *I dialetti di Roma e del Lazio; studi e documenti pub. In memoria di Ernesto Monaci sotto il patrocinio del Comune di Roma*. Roma, Società Filologica Romana. http://archive.org/details/idialettidiromae00merluoft

Microsoft Corporation. (2021). *Microsoft Excel* (2021 (16.1)) [Computer software]. https://office.microsoft.com/excel

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, *7*(2), 125–142. https://doi.org/10/bnkbfj

Montrul, S., & Polinsky, M. (Eds.). (2021). *The Cambridge Handbook of Heritage Languages and Linguistics* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108766340

Nagy, N. (2000). *Faetar*. LINCOM Europa.

Nagy, N. (2011). A Multilingual Corpus to Explore Variation in Language Contact Situations. *Rassegna Italiana Di Linguistica Applicata*, *43*(1–2), 65–84.

Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua*, *164*, 309–327. https://doi.org/10/gf7kft

Nagy, N. (2024). *Heritage languages: Extending variationist approaches*. Cambridge University Press.

Nagy, N., Aghdasi, N., Denis, D., & Motut, A. (2011). Null Subjects in Heritage Languages: Contact Effects in a Cross-linguistic Context. *University of Pennsylvania Working Papers in Linguistics*, *17*(2), 16.

Nagy, N., & Celata, C. (2022). Un corpus per lo studio della variazione sociolinguistica dell'italiano in contesto migratorio. *Corpora e Studi Linguistici*, *6*, 223–237. https://doi.org/10.17469/O2106SLI000015

Nagy, N., Iannozzi, M., & Heap, D. (2018). Faetar null subjects: A variationist study of a heritage language in contact. *International Journal of the Sociology of Language*, *2018*(249), 31–47. https://doi.org/10.1515/ijsl-2017-0040

Orozco, R. (2015). Pronominal Variation in Colombian Costeño Spanish. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 17–37). Georgetown University Press.

Orozco, R., & Hurtado, L. M. (2020). A Variationist Study of Subject Pronoun Expression in Medellín, Colombia. *Languages*, *6*(1), 5. https://doi.org/10/gqfxjs

Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and Dialect Contact in Spanish in New York: Toward the Formation of a Speech Community. *Language*, *83*(4), 770–802. https://doi.org/10.1353/lan.2008.0019

Padilla, L. V. K. (2020). *The Case of Equatorial Guinea*. Arizona State University.

Pier21.ca. (n.d.). *History of Italians in Canada*. Retrieved September 13, 2022, from https://pier21.ca/culture-trunks/italy/history

Posio, P. (2011). Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics*, *43*(3), 777–798. https://doi.org/10/dv3svz

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Ramat, A. G. (1995). Code-switching in the context of dialect/standard language relations. In L. Milroy & P. Muysken (Eds.), *One Speaker, Two Languages* (1st ed., pp. 45–67). Cambridge University Press. https://doi.org/10.1017/CBO9780511620867.003

Rizzi, L. (1986). Null Objects in Italian and the Theory of pro. *Linguistic Inquiry*, *17*(3), 501–557.

Schmitz, K., Di Venanzio, L., & Scherger, A.-L. (2016). Null and overt subjects in Italian and Spanish heritage speakers in Germany. *Lingua*, *180*, 101–123. https://doi.org/10/gg9dmb

Schmitz, K., & Müller, N. (2008). Strong and clitic pronouns in monolingual and bilingual acquisition of French and Italian. *Bilingualism: Language and Cognition*, *11*(1), 19–41. https://doi.org/10.1017/S1366728907003197

Sornicola, R. (2006). Dialetto e processi di italianizzazione in un habitat del Sud d'Italia. In *Lingua e dialetto nell'Italia del duemila* (p. 44).

Statistiche IStat. (2022a). *Popolazione residente per singole età, sesso e cittadinanza—Regioni*. https://esploradati.censimentopopolazione.istat.it/databrowser/#/it/censtest/categories/RE

GPROV/REGPROV_POP/REGPROV_POP_DEM_CHAR/IT1,DF_DCSS_POP_DEMC
ITMIG_SETA_TV_1_REGIO,1.0

Statistiche IStat. (2022b, July 3). *Risultati del censimento permanente della popolazione, 2021—
Lazio*.
https://esploradati.censimentopopolazione.istat.it/databrowser/#/it/censtest/categories/ITE
4/IT1,DF_DCSS_POP_DEMCITMIG_SETA_REG_13_ITE4,1.0

Statistics Canada (Ed.). (2017, February 8). *Census Profile, 2016 Census—Sarnia, City [Census
subdivision], Ontario and Canada [Country]*. http://www12.statcan.ca/census-
recensement/2016/dp-
pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3538030&Geo2=PR&Code2=0
1&Data=Count&SearchText=Canada&SearchType=Begins&SearchPR=01&TABID=1&
B1=All

Statistics Canada. (2023a). *Census Profile of Sarnia* [Census Profile; Table]. Statistics Canada.
https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

Statistics Canada. (2023b, March 29). *Census Profile of Toronto, ON*. Statistics Canada.
https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

Taylor, S. M. (1986). *The Impact of Standard Italian on the Central Italian Dialect of Moricone*.
University of Michigan.

Telmon, T. (1993). Varietà regionali. In A. Sobrero, P. Benincà, & G. Berruto (Eds.),
Introduzione all'italiano contemporaneo (Vol. 1, pp. 93–149). Laterza.

Tisato, G. (2009). *NavigAIS* [Computer software]. Università di Padova.
https://www3.pd.istc.cnr.it/navigais/

Tomasi, L. F. (1977). The Italian community in Toronto: A demographic profile. *International
Migration Review*, *11*(4), 486–513.

Torres Cacoullos, R., & Travis, C. E. (2019). Variationist typology: Shared probabilistic
constraints across (non-)null subject languages. *Linguistics*, *57*(3), 653–692.
https://doi.org/10.1515/ling-2019-0011

Tosi, A. (2001). *Language and society in a changing Italy*. Multilingual Matters.

Tosi, A. (2004). The Language Situation in Italy. *Current Issues in Language Planning*, *5*(3),
247–335. https://doi.org/10/dg5q2r

Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and
conversation. *Language Variation and Change*, *19*(2), 101–135.
https://doi.org/10.1017/S0954394507070081

TUBS. (2011a). *Province of Frosinone, Lazio, Italy* [Map]. Wikimedia Commons.
https://en.wikipedia.org/wiki/File:Frosinone_in_Italy.svg

TUBS. (2011b). *Region of Calabria, Italy* [Map]. Wikimedia Commons.
https://commons.wikimedia.org/wiki/File:Calabria_in_Italy.svg

U.S. Census Bureau. (2022). *American Community Survey, ACS 1-Year Estimates Subject
Tables, Table S1601: Language Spoken at Home*. U.S. Census Bureau.
https://data.census.gov/table/ACSST1Y2022.S1601?q=Language Spoken at Home

Vignuzzi, U. (1997). Lazio, Umbria and the Marche. In M. Parry & M. Maiden (Eds.), *The
Dialects of Italy* (pp. 311–320). Routledge.

Wilcox, V. (2011). Encountering Italy: Military service and national identity during the First
World War. *Bulletin of Italian Politics*, *3*(2), 283–302.

Wolfram, W., Reaser, J., & Vaughn, C. (2008). Operationalizing Linguistic Gratuity: From Principle to Practice. *Language and Linguistics Compass*, *2*(6), 1109–1134. https://doi.org/10.1111/j.1749-818X.2008.00092.x

Yeh, H. (2020). *Chinese Verb Phrases: Continuum of Patterns with Different Lexical Statuses* [Ph.D., Stanford University]. https://www.proquest.com/docview/2456896890/abstract/DF37F547D8B3400FPQ/1

Appendices

A. Ethics Approval (2016) and renewal (2023)

Western Research

Research Ethics

Western University Non-Medical Research Ethics Board
NMREB Delegated Initial Approval Notice

Principal Investigator: David Heap
Department & Institution: Arts and Humanities\French,Western University

NMREB File Number: 108230
Study Title: The CHILS (Ciociaria Heritage Italian Language of Sarnia) Corpus

NMREB Initial Approval Date: August 09, 2016
NMREB Expiry Date: August 09, 2017

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Western University Protocol	Received August 8, 2016	
Letter of Information & Consent		2016/08/03
Recruitment Items	Recruitment Email - Received July 26, 2016	
Recruitment Items	Telephone Script	2016/06/29
Advertisement	Newspaper	2016/06/29
Advertisement	Bulletin	2016/06/29
Advertisement	25 word ad	2016/06/29
Instruments	CHILS Questionnaire	2016/06/30

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Date: 8 August 2023

To: David Heap

Project ID: 108230

Study Title: The SOHL (Southwestern Ontario Heritage Languages) Corpus

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Date Approval Issued: 08/Aug/2023 07:01

REB Approval Expiry Date: 09/Aug/2024

_____

Dear David Heap,

The Western University Non-Medical Research Ethics Board has reviewed this application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Electronically signed by:

Mr. Joshua Hatherley, Ethics Coordinator on behalf of Dr. Isha DeCoito, NMREB Chair 08/Aug/2023 07:01

Reason: I am approving this document

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

B. CHILS Questionnaire

Interview Guide

All questions will be asked in English (as written), but answers will be given in Italian. The interviewer will only repeat what was said in English in Italian if the participant cannot understand the question in English in order to avoid influencing the dialect.

All questions depend on the context and appropriateness. No single individual will be asked all the questions. The only parts that will be asked of every person will be the 'basic demographic' section, 'wordlist' and 'final questions'. The interview will not take longer than 1.5 hours.

For example, if someone has never been married, or is the child of immigrants, then those questions about marriage or life in Italy won't apply, and will be dropped, or questions will be asked about their perceptions of those subjects.

Basic Demographic

1. What is your name?
2. Where were you born?
3. How old are you?
4. Do/did you have any brothers and sisters?
5. Where do/did they live?
6. What did your parents do?

Wordlist

7. Please give me the Italian translation of the following words (*The first 17 are to prompt words with 'v' occurring inside the word, the second 13 at the beginning of the word)

1)	Egg	16)	Grape
2)	Eggs	17)	Grapes
3)	Horse	1)	Where
4)	Table	2)	Voice
5)	Cloud	3)	Worm
6)	Bad	4)	Wine
7)	A ship	5)	Old
8)	Please	6)	Near
9)	Cauliflower	7)	Green
10)	Snow	8)	To see
11)	Deer	9)	Wind
12)	Spring	10)	Clothes
13)	Winter	11)	Twenty-six
14)	Sink	12)	A flight
15)	nineteen	13)	City

Social Network

8. Who is your wife/husband?
9. Where is he/she from?
10. Where do your brothers and sisters live?
11. Do you talk with them often?
12. Is anyone else from your family in Sarnia? Cousins, aunts, uncles…?
13. Where does the rest of your family live?
14. Did brothers and sisters and family leave Italy too, or are they still there?
15. Who do you spend your time with?
16. Do you speak to them (the people you spend time with) in Italian?
17. In a given week, how many people would you say you speak English with?
18. And how many do you speak Italian with?
19. Who are your friends?
20. When you (used to) play Scopa who would come over?
21. Do you talk on the phone?
22. Who do you talk on the phone with?
23. Who did you talk on the phone within the last week?

Family in Italy

24. Who were your parents?
25. Where were they from?
26. How much education did they have?
27. What did they do for a living?
28. How did your parents meet?
29. Were they from the same area?
30. Did they ever tell you about their wedding?
31. Growing up, where did your family live?
32. What did your parents do for a living? Did they always have the same jobs?
33. When your parents got home from work, what would you do in the evenings?
34. When you were a kid what did your family do on weekends?
35. What were the big holidays?
36. What would you do?

Childhood

37. Who were your friends when you were a child?
38. What was is like to be a kid back then?
39. What were meals like back in Italy?
40. What kind of food did you eat?
41. What was your favourite supper (meal?) growing up?
42. How important was food to your family?
43. Did your family grow anything?
44. What kind of farming did they do?

45. Did your family make olive oil or wine?
46. How did they do that?
47. Did your mum do most of the cooking when you were a kid?
48. What kind of things did she cook?
49. Did you often help her?
50. Who taught you to cook?
51. What else did he/she teach you?
52. Did you have any festivals or fairs in your hometown?
53. Who went to them?
54. What were they like?

School

55. Did you go to school?
56. What did you like to learn about?
57. How young were you when you left school?
58. When you were young who were your friends?
59. How did you get to school and back each day?
60. Do you remember any games you played with your school friends?
61. What did you love to do when you got home/weren't at school?
62. Did you have chores to do when you came home each night? (what did you do / like to do / not like to do outside of school?)

Pre-emigration

63. What was it like in Italy before you moved?
64. Why did you choose to move from Italy?
65. Why did you choose Canada?
66. How did you get to Canada?
67. Who were you meeting in Canada once you arrived?
68. What did you know about Canada before you arrived?
69. Did a lot of people leave your hometown? Why?
70. Did they all go to the same places or different places?
71. Are you in touch with anyone who went elsewhere?

Emigration

72. What did you hope would happen in Canada?
73. Why couldn't it happen in Italy?
74. What was your journey like to Canada?
75. Did you travel alone? Who else did you travel with, if anyone?
76. What did you pack? What did you have with you on the trip?
77. What did you think when you arrived?
78. Did you know/speak any English?
79. What did you think about Sarnia when you arrived?
80. Who met you/how did you get from the port to Sarnia?

Settling in

81. Where did you live first?
82. Who did you live with at the start?
83. What did you do first for work?
84. Did you know many people when you arrived?
85. What were the other Italians like when you arrived?
86. What were the Canadians like to you when you arrived?
87. How did you learn English?
88. What did you like best about the first couple years?
89. Did you think about leaving and going back to Italy in the first 5 years?
90. What was it like to become a Canadian citizen? How long did it take?
91. Were you surprised by the oath of citizenship?

Marriage

92. Who did you get married to?
93. When did you get married?
94. Where were you married? In Italy or in Canada?
95. What was your wedding like?
96. Was that a normal wedding in your hometown?
97. Did people often have photographers, or what things did you have as a keepsake?
98. What did you do after the wedding ceremony?
99. Who came to your wedding?

Work

100. What was your first job here in Canada?
101. How did you get your first job?
102. What other jobs have you done?
103. Did you work mostly with Italians or with lots of other Canadians?
104. Do you remember something you really wanted to buy when you starting working?
105. What do you remember being a big thing that everyone wanted to buy when they arrived from Italy?
106. What do you remember about your first home?
107. Where did you live first?
108. did you pick that area to live in for a reason?
109. Do you/did you drive?
110. what was it like getting a driver's licence?
111. had you driven when you lived in Italy?

Social Life

112. Did you spend time with other Italians?
113. Were the majority of your friends Italian or Canadian?

247

114. What would you do on a typical weekend?
115. Did you make wine? Or did someone you know?
116. Did you help them?
117. Where did you/they get the grapes? What about the press?
118. When did you normally do the winemaking?
119. When did you first drink that year's wine (how long after making it)?
120. What were your typical weeks like?
121. What did you miss most about Italy? Least?
122. Where could you buy Italian foods?
123. Did you know the people who ran the shop?
124. Was there a strong Italian community?

Italian Community in Sarnia

125. Was there an Italian community when you arrived?
126. what about now?
127. What do you think of the Italian community In Sarnia today?
128. What has changed?
129. What do you miss about Italy?
130. Did you take part in community events with the Italian community?

Community

San Rocco Festival

131. Did you go to the San Rocco festival?
132. Who was San Rocco?
133. Do you go to the festival anymore?
134. What do you like about It?
135. Did you ever take part in the festival planning?

St. Peter's Catholic Church

136. Did you go to St. Peter's Church?
137. How did the church get started?
138. How many people there were Italian? Most? All?
139. What kind of events would take place there?
140. Now that it's gone, do you think it hurt the community to lose it?

Dante Club of Italian Canadians

141. Did you go to the Dante Club?
142. Who started the Dante Club?
143. Do you play bocce?
144. Did you play in a league there?
145. What were the people at the Dante Club like?
146. What kind of events took place there?

Contact with Italy

147. Did you have a phone?
148. Did you call Italy?
149. How often?
150. Did you ever go back? (for a visit?)
151. What did you think going back?
152. How many times have you gone back?
153. When did you first call home?
154. Do you talk to people in Italy often?
155. What do you talk about?
156. What do they think about your life in Canada?
157. What do they think about Canada generally?

Here

158. Is it important for you to retain your Italian heritage in Canada?
159. What has living in Sarnia been like?
160. How is it different than if you had stayed in Italy?
161. What do you think of Canadian people?
162. What do you like best about living here?
163. Do you think your childhood was different from the childhood of kid's today?
164. When you came to Canada did you think children were treated differently by Canadians?
165. How has your neighbourhood changed in the last couple years/decades?

Foods and Cooking

166. Did you make wine here in Canada?
167. How did you do it?
168. Did you or your friends do any canning?
169. What sort of things did you guys can?
170. Where did you get the (pears, tomatoes, peaches, etc.) that you canned?
171. Who taught you how to can/prepare sausages/bake bread/make pasta?
172. what was it like making them? Who came over to help, how long did it take?
173. Did you have a garden? What sort of things did you grow?
174. Did you do gardening back in Italy? How did you learn how to do it?
175. Did you have a fig tree?
176. How did you keep it over winter?
177. What do you think of food in Canada?
178. What did you think when you got here?
179. Do you think eating and supper is different in Canada?
180. Did you/do you make supper every night?
181. What kind of things did you like to cook?
182. Did you ever make pasta or bread?
183. Who taught you how to do that?

184. Is the food you cook for supper similar to what you ate in Italy?
185. What kind of things did you eat mostly in Italy?
186. Do you like to cook?
187. Do you like to go out to restaurants?
188. What's your favourite type of food?
189. Where's your favourite restaurant?

Language

190. Do you think it is important for people to learn other languages?
191. How hard was it to learn English?
192. Who helped you?
193. Do all the Italians in Sarnia sound the same to you?
194. Who sounds different if anything?
195. Do your kids speak Italian the same way you do?
196. Was it important for you that your children/you learn Italian?
197. When you speak to your children/parents do you use Italian or English?
198. Do your grandchildren/children/you speak Italian?
199. Do you wish they did? Why?
200. Do you watch TV? (in English / Italian?)
201. What kind of shows do you like?
202. Do the people speaking Italian on TV sound like you do?
203. Why or why not?
204. Who did you talk to yesterday/last week? In English / Italian?
205. Do you think all the Italians in Sarnia are the same, or are there differences based on where people are from in Italy?
206. What about the Italians in Windsor?
207. Do you know Italians in Windsor?
208. What are they like?
209. Do you know Italians in Toronto?
210. What are they like?
211. Do you think you are a Canadian?
212. Why or why not?

Last Questions

213. What do you think now about your choice (or your parent's choice) to emigrate from Italy?
214. How did immigrating to Canada influence your family and their lives?
215. Do you have anything you would like to add about any of these questions?

Image and document collection (If consented to before interview begins)

216. Do you have any photographs or documents you'd like to share with me?

C. Coding schema of linguistic variables

Verb

Verb Tense

Code	Description
AP	Absolute Past
CP	Compound Past (Verb has an auxiliary and past participle)
Fᴜᴛ	Future
Iᴍᴘ	Imperfect
Pʀ	Present
OTH	Other verb Tenses (Conditional, subjunctive, etc.)

Auxiliary for compound past VPs

Code	Description
Aᴠᴇʀᴇ	*Avere* is the past participle
Eꜱꜱᴇʀᴇ	*Essere* is the past participle
Sᴛᴀʀᴇ	*Stare* is the past participle
N/A	The VP does not have a past participle

Final segment of verb

Code	Description
[ə]	Final vowel of verb is realized as [ə]
Ø	The final segment of the verb is a consonant, not a vowel
V	The final segment of the conjugated verb is the same final vowel as the equivalent in standard Italian

Ambiguous

Code	Description
1ꜱɢ/2ꜱɢ/3ꜱɢ	Verb has ambiguous form shared with 1st, 2nd, and 3rd singular subjects
1ꜱɢ/3ꜱɢ	Verb has ambiguous form shared for verb with 1st and 3rd singular subjects
ꜱᴏɴᴏ	Verb is a form of *sono*, either as present or auxiliary verb, past participle lacks disambiguating agreement
N/A	Subject is clear from the verb form

Subject

Person

Code	Description
1ST	The verb's subject is in the 1st person
2ND	The verb's subject is in the 2nd person
3RD	The verb's subject is in 3rd person

Number

Code	Description
SING	Subject of the verb is singular
PL	Subject of the verb is plural

Grammatical gender of 3rd person subject

Code	Description
MASC	Masculine subject
FEM	Feminine subject
N/A	1st or 2nd person VP

Subject of VP

Code	Description
1SG	The verb's subject is 1st person and singular
2SG	The verb's subject is 2nd person and singular
3SG.M	The verb's subject is 3rd person singular and masculine
3SG.F	The verb's subject is 3rd person singular and feminine
1PL	The verb's subject is 1st person and plural
2PL	The verb's subject is 2nd person and plural
3PL.M	The verb's subject is 3rd person plural and masculine
3PL.F	The verb's subject is 3rd person plural and feminine

Priming variables – HLVC and CHILS

Switch reference

Code	Description
SAME.REF	Same referent as previous VP
SWITCH.REF	Different referent from previous VP
NEW	New Topic/Turn

Switch tense

Code	Description
SAME.TENS	Same tense as previous VP
DIFF.TENS	Different tense from previous VP
NEW	New Topic/Turn

Previous Realization

Code	Description
OVERT	Yes, the previous VP had a realized subject pronoun
NULL	No, the previous VP had a null subject
NEW	New Topic/Turn
NOUN	The previous VP had an NP subject

Priming variable – AIS

Pronoun in prompt

Code	Description
PRESENT	The prompt contains an overt pronoun
ABSENT	The prompt does not contain an overt pronoun

Phrase structure – HLVC and CHILS

Clause type

Code	Description
EMBED	Embedded clause
MATRIX	Matrix clause
SIMPLE	Simple phrase

Dependent variable – HLVC and CHILS

Subject pronoun realization

Code	*Description*
Full	Subject pronoun is realized
NULL	Subject pronoun is null
Noun	Subject pronoun is a common noun

Dependent variable – AIS

Subject pronoun realization

Code	*Description*
Prompt	Subject pronoun is realized, and was also in the prompt
Add	Subject pronoun is realized, but was not in the prompt
Null	Subject pronoun is null, and was also null in the prompt
Remove	Subject pronoun is null, but was realized in the prompt

D. Participants

HLVC

Homeland participants					Heritage participants				
Speaker	*Gender*	*Age*	*Duration*	*Tokens*	*Speaker*	*Gender*	*Age*	*Duration*	*Tokens*
IXF35A	F	35	24:34	87	I2F57A	F	57	30:06	95
IXF38A	F	38	33:41	66	I1F61A	F	61	1:18:31	98
IXF51A	F	51	44:38	101	I1F65A	F	65	52:51	86
IXF61A	F	61	33:41	79	I1F71A	F	71	54:53	96
IXF94A	F	94	1:15:04	102	I1F73A	F	73	49:29	100
IXM35A	M	35	50:47	93	I1M60A	M	60	1:16:09	112
IXM47A	M	47	56:32	94	I1M61A	M	61	39:42	109
IXM52A	M	52	56:27	100	I1M61B	M	61	57:29	106
IXM61A	M	61	32:29	95	I1M62A	M	62	57:22	102
IXM64A	M	64	1:10:21	103	I1M75A	M	75	45:45	99
Average		**54**	**47:49**	**92**	**Average**		**65**	**47:56**	**100**

Table 79. Table showing the speakers from the HLVC corpus used in this study.

AIS

AIS participants

AIS Point	Location	Gender	Parents' P.O.B.	Work	Date	Age	Tokens
654	Serrone	M	Serrone	Sharecropper	18-25 Sept., 1924	51	99
664	Veroli	M	Father: Santa Francesca Mother: Isola Liri	Sharecropper	27-30 Sept., 1924	56	105
701	San Donato	M	San Donato	Municipal usher	12-15 Sept., 1924	60	72
710	Ausonia	M	Ausonia	Sharecropper	9-12 Oct., 1924	51	101
656	Scanno	M	Scanno	Sharecropper	13-16 Sept., 1923	44	101
682	Sonnino	M	Sonnino	Homemaker	2-6 Oct., 1924	56	105

Table 80. Summary of demographic information from the AIS points included in this study (Jaberg & Jud, 1987:143-148).

CHILS

CHILS Participants

Name[39]	Gender	Place of birth (POB)	Partner's POB	Age	Age of arrival	Duration	n
Annita[40]	F	Fontechiari	Fontechiari	58	20	41:32	91
Antonia	F	Fontechiari	Canada	69	9	49:17	83
Antonietta	F	Posta Fibreno	Posta Fibreno	89	26	31:21	88
Arcangela	F	Casalvieri	Casalvieri	82	18	24:34	86
Assunta	F	Alvito	Alvito	86	24	1:26:26	84
Caroline	F	Fontechiari	Fontechiari	66	6	35:01	85
Clara	F	Castelliri	Castelliri	82	17	28:13	80
Joanna[c]	F	Fontechiari	Fontechiari	57	8	30:36	85
Nunciata	F	Fontechiari	Fontechiari	85	27	33:30	77
Palma	F	Fontechiari	Fontechiari	84	21	52:10	90
Angelo F.	M	Vicalvi	Castelliri	80	17	1:20:52	90
Ennio	M	Veroli	Veroli	84	20	25:42	85
Ernie	M	Fontechiari	Fontechiari	58	10	29:33	84
Frank D'A.	M	Casalvieri	Casalvieri	78	19	43:50	91
Frank L.	M	Fontechiari	Fontechiari	62	18	28:04	89
Isy	M	Casalvieri	Canada	65	10	40:48	87
Nello	M	Casalvieri	Canada	74	10	43:00	93
Raffaele	M	Alvito	Alvito	95	32	56:58	94
Renzo	M	Alvito	Alvito	81	19	1:04:16	100
Sante	M	Posta Fibreno	Posta Fibreno	89	28	48:07	74
Average				**76**	**18**	**43:42**	**87**

Table 81. Participants who were included in this study from the CHILS corpus.

---

[39] First names have not been changed. Participants consented to their recordings being shared on a publicly available archive, which is open online (www.italiansinlambton.ca). Those with an initial of their surname share a first name with another participant who may included in this study (e.g., Frank L. and Frank D'A.), or not (e.g., Angelo F.)

[40] The following participants were married and interviewed together: Annita and Frank L, Antonietta and Sante, and Joanna and Ernie. However, all the tokens included in this dissertation for Antonietta are from her interview with her sister-in-law, Nunciata.

E. Phonetic Inventory of Ciociaro from AIS

A phonetic inventory of Ciociaro has not been described since Merlo described the dialetto as spoken in the town of Sora (1919). Before coding the CHILS recordings, I created a phonetic inventory of Ciociaro based on the transcriptions of the AIS responses. I hope it is useful to other researchers of the dialetto. Phones are organized by groups, with locations divided by columns. Phones that only occur in some of the locations are placed last in each group. I give the total in paratheses for those phones that occur fewer than 20 times for at least one of the locations.

Note that this inventory only includes the prompts that have a tensed verb, and not the complete list of transcriptions from the AIS.

	Serrone (654)	Veroli (664)	San Donato (701)	Ausonia (710)	Scanno (656)	Sonnino (682)
stops	p / b	p / b	p / b	p / b	p / b	p / b
	t / d	t / d	t / d	t / d	t / d	t / d
	k / g	k / g	k / g	k / g	k / g	k / g

	Serrone	Veroli	San Donato	Ausonia	Scanno	Sonnino
fricatives	f / v	f / v	f / v	f / v	f / v	f / v
	s / z	s / z	s / z	s / z	s / z	s / z
	l	l	l	l	l	l
	r / ɹ	r / ɹ	r	r	r	r / ɹ
	ʃ	ʃ	ʃ	ʃ	ʃ	ʃ
	ʎ	ʎ	ʎ	ʎ	ʎ	ʎ
	------	ç (1)	ç (1)	ç (2)	ç (4)	ç (1)
	------	------	------	β	------	β
	------	------	ð	------	------	------
	------	------	------	------	ɣ	------

	Serrone	Veroli	San Donato	Ausonia	Scanno	Sonnino
affricates	ʧ / ʤ	ʧ / ʤ	ʧ / ʤ	ʧ / ʤ	ʧ / ʤ	ʧ / ʤ
	tɕ (7) / dʑ (3)	tɕ (12) / dʑ (1)	tɕ (11)	tɕ (45) / dʑ (6)	tɕ (2) / dʑ (1)	tɕ (5)
	ts (4) / dz (6)	ts (1)	ts (2)	ts (2)	ts (5)	ts (8)

	Serrone	Veroli	San Donato	Ausonia	Scanno	Sonnino
nasals	m	m	m	m	m	m
	n	n	n	n	n	n
	ɲ	ɲ	ɲ	ɲ	ɲ	ɲ
	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ

	Serrone	Veroli	San Donato	Ausonia	Scanno	Sonnino
vowels	i	i	i	i	i	i
	e	e	e	e	e	e
	ɛ	ɛ	ɛ	ɛ	ɛ	ɛ
	a	a	a	a	a	a
	u	u	u	u	u	u
	o	o	o	o	o	o
	ɔ	ɔ	ɔ	ɔ	ɔ	ɔ
	------	------	------	æ	æ	------
	ɐ (6)	------	------	------	ɐ (5)	------
	ɪ (2)	------	------	------	ɪ (6)	------
	------	------	------	------	ʊ (3)	------
	ə (2)	ə (5)	ə (159)	ə (2)	ə (220)	ə (3)

	Serrone	Veroli	San Donato	Ausonia	Scanno	Sonnino
glides	j	j	j	j	j	j
	w	w	w	w	w	w

F. AIS correspondences

The AIS is transcribed using a modified Romanist phonetic notation. In order to convert the transcriptions into IPA, I created a list of correspondences using the index included in the Italian translation reprint of the atlas (Jaberg & Jud, 1987).[41] The primary difference between the AIS phonetic alphabet and the IPA is the differences in how the two notations use diacritics. Prosodic stress is marked with diacritics above vowels in the AIS transcriptions. While most of the diacritics above vowels describe prosodic stress, there are some that do describe vowel quality (such as length), and these can both occur above a vowel.[42] These are summarized in Table **82**.

Symbol	Example	Diacritic	Description
́◌	á	Acute accent	Marks primary stress
̀◌	à	Grave accent	Marks secondary stress
̋◌	a̋	Double acute accent	Marks strong stress (*accento forte*)
̄◌	ā á	Macron accent	Vowel is long (or, as in the second example, long and stressed)

Table 82. Diacritics used above vowels in the AIS transcriptions.

There are also diacritics that occur below vowels. These are specific to vowel quality and distinguish one phoneme from another. Consonants are transcribed with fewer diacritics than the

---

[41] A summary is provided at this link: https://www3.pd.istc.cnr.it/navigais-web/AIS_symbols.htm

[42] Only the diacritics that apply to the transcriptions of the regions included in this dissertation are described here.

vowels, and always distinguish one phoneme from another. These are described below in the complete list of correspondences (Table **83**).

Figure 14. Transcription of a geminate in the AIS.
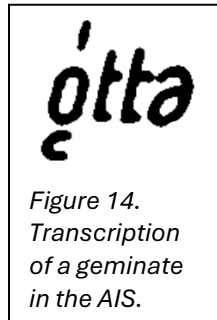
Finally, the AIS transcriptions repeat consonants to describe geminates. Figure **14** shows the differences between the AIS phonetic alphabet and the IPA. The participant from San Donato gives this response for prompt #287: *otto* ('eight') is transcribed as [ˈɔtːə] (Jaberg & Jud, 1928:287).

Below are the correspondences for the phonetic alphabet:

Code	Symbol	IPA equivalent
A1		[a]
A2		[æ]
A3		[ɐ]
B1		[b]
B2		[β]
C1		[tɕ]
C2		[ʧ]
D1		[d]
D2		[ð]
E1		[e]
E2		[ə]

E3	$\hat{e}$	[ɛ]
F1	$\int$	[f]
G1	g	[g]
G2	$\acute{g}$	[ʤ]
G3	$\tilde{g}$	[dz̧]
I1	i	[i]
I2	i	[ɪ]
K1	k	[k]
L1	l	[l̦]
L2	t	[ʎ]
M1	m	[m]
N1	n	[n]
N2	$\tilde{n}$	[ɲ]
N3	η	[ŋ]
O1	o	[o]
O2	ϱ	[ɔ]
P1	p	[p]
R1	r	[r]

S1	**S**	[s]
S2	**š**	[ʃ]
S3	**š̬**	[ʧ] (used by Rohlfs in lieu of C2 "il segno è stato usato solo da Rohlfs" [Jaberg & Jud, 1987:43])
T1	**t**	[t]
U1	**u**	[u]
U2	**u̯ɔ**	[ʊ]
V1	**ν**	[v]
W1	**ω**	[w]
Y1	**y**	[j]
Z1	**z**	[z]
X1	**X**	[ç] Voiceless palatal fricative ("suono del ted. *ich* [spirante sorda palatale]" [Jaberg & Jud, 1987:43])
ɣ1	**ɣ**	[ɣ] Voiced velar fricative ("suono sonoro corrispondente al precedente [spirante velare sonoro]: g in sp. *Fuego*" [Jaberg & Jud, 1987:43])

Table 83. Complete list of correspondences between the AIS phonetic alphabet and the IPA.

G. AIS prompts

As I explain in my dissertation (see **5.1.2**), I used an online version of the AIS maps for this research (Tisato, 2009). There are nearly 2,000 maps included in the AIS. For my project, I identified 258 prompts containing an inflected verb. From these 74 were excluded according to my criteria for tensed verbs which *could* have an overt subject pronoun and had a human subject, in keeping with the methodology for the HLVC and CHILS corpora (see **5.2**). This left 184 prompts, which I include below for reference.

Prompt #	Prompt	Translation
11	io manderò	I will send
41	è già battezzata?	is [she] baptized?
50	quanti anni hai?	how old are [you]?
52	non vedi	[you] don't see
53	che sei	that [you] are
65	lui l'ama molto	he loves her very much
69	non vi sposate?	[you] are not married?
74	è gravida	[she] is pregnant
76	sono morti	[they] are dead
80	si chiama	[he/she] is named
112	perché hai?	why do [you] have?
168	soffia il naso	[he/she] blows their nose
189	è guercio	[he] is cross-eyed
192	zoppica	[he/she] limps
194	tartaglia	[he/she] stutters
199	è un mestiere	[he] is a craftsman
263	te la prova	[he/she] shows it to you
359	vieni?	[you] coming?
384	egli dice	he says
385	che ha freddo?	who is cold?
388	ho le mani	I have hands
390	ho visto	I have seen
394	è cascato	[he] did fall
512	ho preso	[I] took
519	va a caccia	[he] goes hunting
522	è andato	[he] has gone
543	si tagliano	[they] (refl.) cut themselves
592	portano	[they] bring
643	ho sonno	[I] am tired
646	è andata	[she] has gone
649	dormi già?	are [you] sleeping already?
650	non dormo mai	[I] never sleep

651	egli non dorme mai	[he] never sleeps
653	non dormirò	[I] will not sleep
654	egli russa	[he] snores
656	si sveglia	[he] wakes up
659	è coricato	[he] is lying down
660	ci leviamo	we (refl.) get up
661	se voi vi levate	if 2PL (refl.) get up
_2_667	bisogna **che si vestano**	It's needed **that they (refl.) get dressed**
668	è vestito	[he] is dressed
671	era tutta nuda	she was totally naked
672	si pettina	[he] (refl.) combs himself
676	potrebbe ferirsi	he could hurt himself
694	mi sono raffreddato	I got cold
695	eppure, sto	yet I remain
697	ho la febbre	I have a fever
698	ho la voce rauca	I have a hoarse voice
699	sto per svenire	I'm going to faint
703	soffro assai	I suffer a lot
704	andava spesso	he went often
707	è guarita	she is healed
709	fu ben curata	she was well cared for
712	sei un bugiardo	you are a liar
713	non mento mai	I never lie
730	perché lo fai piangere?	why do you make him cry?
761	egli fuma	he smokes
768	egli legge	he reads
770	l'hai letto?	have you read it?
813	mi ha dato il malocchio	he gave me the evil eye
821	dove vai?	where are you going?
822	vo a comprare	I'm going to buy
826	che ho comprati	that I had bought
835	vogliono	they want
836	che io ve li venda	that I'm selling it to 2PL
837	mi offrono solamente	they only offer me
887	ho messo	I had put
900	si è nascosto	he is hidden
913	l'ho accesa	I turned it on
921	spegnere la fiamma	he/she puts out the flame
936	soffia sul fuoco	he/she blows on the fire
946	si rigovernano / si lavano	they (refl.) wash themselves
949	poi si asciugano	then they (refl.) dry
952	cuocere la carne	cook the meat
954	mondare le patate	peel the potatoes
994	bruciare l'arrosto	burn the roast
1013	sono digiuno	I am fasting

1015	poiché hai fame!	because you're hungry!
1016	mangeresti	you would eat
1017	se avessi fame?	if you were hungry?
1019	ne mangerei	I would eat it
1027	sono sazio	I'm full
1032	quando si ha sete	when he/she is thirsty
1033	si ha la gola secca	if he/she has a dry throat
1035	beverei	I will drink
1044	comprerebbero	they would buy
1052	vogliamo ingrassare	we want to fatten
1075	tosare le pecore	shear the sheep
1085	spellare un capretto	skin the goat
1086	voglio attaccarla	I want to attack it
1107	l'ho ricevuto	I've received it
1110	**te lo darò**, se lo vuoi	**I'll give it to you**, if you want it
_2_1110	te lo darò, **se lo vuoi**	I'll give it to you, **if you want it**
1111	ce l'ha dato	he gave it to us
1112	se io ve lo dessi	if I gave it to you
1113	cosa ne fareste?	what would 2PL do with it
1143	bada che le galline	Minds the chickens
1145	hai venduto le uova?	did you sell the eggs?
1146	le venderò domani	I'll sell them tomorrow
1248	abbiamo avuto	we had
1250	avreste dovuto vedere	2PL should have seen
1264	hanno già cominciato	they have already started
1278	se non mangiamo	if we don't eat
1346	bada! **Tu versi il vino**	Careful! **You're spilling the wine**
1519	vorrei di questa qui	I would like this one here
1533	dove tu cuci adesso	where you now sew
1534	cantava sempre	he always sang
1537	hai cucito bene	you sewed well
1586	perché hai scelto	why did you choose?
1589	pagatemi quello	2PL pay me that
1593	è una buona serva	she's a good servant
1594	non vuol rimanere	she doesn't want to stay
1597	non lo trovo in nessun luogo	I can't find it anywhere
1598	siete venuta senza niente	You (formal) came with nothing
1599	avete guadagnato qualchecosa	you have gained something
1600	perché taci?	why are you silent?
1601	lo dico di nuovo	I say it again
1602	il nostro padrone è cattivo	our master is mean
1606	non ha mai fretta	he's never in a hurry
1607	ma noi ci sgrida continuamente	but we chide ourselves always
1613	lavoreremmo di più	we would work more
1614	se fossimo pagati meglio	if we were paid better

1619	sono tutto stordito	I am all stunned
1622	cade supino	he falls on his back
1623	cade bocconi	he falls on his face
1624	voi altri chiudete la porta	2PL others close the door
1627	gli parlerei io	I would talk to him
1628	se lo trovassi	if I found it
1629	se tu lo trovassi	if you found it
1630	non sarebbe contento	he would not be happy
1632	c'è stato qualcheduno?	has anyone been there?
1633	voi lo trovereste in qualche luogo	2PL would find it someplace
1634	se lo cercaste	if 2PL were looking for it
1635	abbiamo cercato insieme dappertutto	we searched everywhere together
1637	che viene da noi	that he comes to us
1638	**volete** che ci vada io	**2PL want** that I go there
_2_1638	volete **che ci vada io**	2PL want **that I go there**
1639	che ci mandi qualcheduno?	that you send it to someone?
1640	andiamo senza voltarci indietro	we go without looking back
1641	mi rincresceva che non la trovassimo	he apologized to me that it was not found
1645	sento un rumore	I hear a noise
1646	siamo arrivato poco fa	we arrived a little bit ago
1649	**hai dimenticato** che dovevi condurci in cima?	**did you forget** that you have to take us to the top?
_2_1649	hai dimenticato **che dovevi condurci in cima**?	did you forget **that you have to take us to the top?**
1650	**credi** che lo troviamo?	**do you think** that we'll find it?
_2_1650	credi **che lo troviamo?**	do you think **that we'll find it?**
1651	**mi meraviglio** che non lo troviate	**I am surprised** that you didn't find it
_2_1651	mi meraviglio **che non lo troviate**	I am surprised **that you didn't find it**
1652	l'ho saputo troppo tardi	I knew it too late
1653	**voglio** che tu finisca	**I want** that you finish
_2_1653	voglio **che tu finisca**	I want **that you finish**
1654	e che ci dica tutto	and tell us everything
1656	se no, me ne vado	if not, I will leave
1658	non capisco	I don't understand
1661	faranno ciò che vorranno	they will do what they want
1664	l'afferrò per il collo	he grabbed it by the neck
1667	l'hanno cacciato	they kicked him out
1669	io non potevo andare	I could not go
1673	m'ha tirato un sasso	he threw me a stone
1677	ho i piedi bagnati	I have wet feet

Table 84. Complete list of AIS prompts used in my analysis, with English translation.

H. Tables and analyses

H. 1. Combined corpora – HLVC and AIS

One-level analysis of pro-drop across HLVC and AIS

Tokens (n):	2,173	%Ø	79%
Subject			
	n	*%Ø*	*FW*
2PL	(44)	(66%)	N/A
3PL	319	91%	0.67
3SG.F	129	87%	0.61
1PL	349	84%	0.53
3SG.M	308	81%	0.47
1SG	898	74%	0.38
2SG	170	71%	0.35
Range			**32**
Corpus			
	n	*%Ø*	*FW*
HLVC	1,634	80%	0.52
AIS	539	77%	0.48
Range			**4**
Speaker (random)			
Std. dev.			0.34

Table 85. One-level analysis using tokens from the CHILS and AIS datasets with the variables SUBJECT and CORPUS.

Variation in pronoun usage across HLVC and AIS participants

	n	*%Ø*	*FW*
IXM47A	69	90%	0.60
IXM47A	69	90%	0.60
IXF38A	61	93%	0.59
I1M61A	95	91%	0.59
IXF38A	61	93%	0.59
I1M61A	95	91%	0.59
I1F61A	93	88%	0.57
IXF35A	70	87%	0.57
I1F61A	93	88%	0.57
IXF35A	70	87%	0.57
710	93	83%	0.55
710	93	83%	0.55
I1M75A	86	85%	0.54
I1M75A	86	85%	0.54
I1M62A	80	83%	0.53
664	97	80%	0.53
I1F65A	80	84%	0.53
I1M62A	80	83%	0.53
701	67	79%	0.52
682	97	79%	0.52
IXM64A	89	82%	0.52
I1F71A	89	82%	0.51
I2F57A	89	80%	0.51
IXM52A	83	78%	0.49
I1M60A	97	77%	0.47
IXF61A	63	78%	0.47
IXF94A	78	77%	0.46
654	92	73%	0.46
IXM35A	85	72%	0.44
I1M61B	90	73%	0.43
IXF51A	82	74%	0.43
IXM61A	68	71%	0.42
656	93	68%	0.42
I1F73A	87	61%	0.33
Range			**27**

Table 86. Null subject rates by participant, combining the CHILS and AIS datasets.

H. 2. Combined corpora – HLVC and CHILS

One-level analysis of pro-drop across HLVC and CHILS

Tokens (n):	3,382	%Ø	76%

Subject

	n	%Ø	FW
2PL	(16)	(100%)	N/A
3PL	646	91%	0.77
3SG.F	207	79%	0.54
3SG.M	690	73%	0.46
1PL	491	73%	0.43
1SG	1,176	71%	0.40
2SG	172	69%	0.37
	Range		**40**

Corpus

	n	%Ø	FW
HLVC	1,634	80%	0.58
CHILS	1,748	72%	0.42
	Range		**16**

Speaker (random)

Std. dev.			0.38

Table 87. One-level analysis using tokens from the HLVC and CHILS datasets with the variables subject and corpus.

Variation in pronoun usage across HLVC and CHILS participants

	n	%Ø	FW
Angelo F	91	87%	0.65
IXF38A	61	93%	0.62
Rafaelle	95	82%	0.62
IXM47A	69	90%	0.61
I1M61A	95	91%	0.60
I1F61A	93	88%	0.58
Frank D'A	91	79%	0.58
IXF35A	70	87%	0.56
Isy	88	78%	0.56
Caroline	86	78%	0.56
Sante	76	76%	0.56
I1M75A	86	85%	0.55
I1F65A	80	84%	0.54
I1M62A	80	83%	0.54
Palma	91	74%	0.53
I1F71A	89	82%	0.52
Antonietta	89	78%	0.52
Nunciata	78	76%	0.52
Annita	92	75%	0.51
IXM64A	89	82%	0.51
Nello	93	72%	0.51
Joanna	86	71%	0.50
I2F57A	89	80%	0.50
Frank L	89	70%	0.49
IXM52A	83	78%	0.48
Assunta	85	72%	0.47
I1M60A	97	77%	0.47
IXF61A	63	78%	0.47
Clara	80	69%	0.47
IXF94A	78	77%	0.45
Ennio	85	67%	0.44
I1M61B	90	73%	0.43
IXF51A	82	74%	0.43
IXM35A	85	72%	0.43
Renzo	100	62%	0.42
IXM61A	68	71%	0.41
Arcangela	86	59%	0.39

Ernie	84	60%	0.39
Antonia	83	53%	0.35
I1F73A	87	61%	0.32
Range			**33**

Table 88. Null subject rates by participant, combining the HLVC and CHILS datasets.

H. 3. Combined corpora – AIS and CHILS

One-level analysis of pro-drop in the AIS and CHILS corpora

Tokens (n):	2,287	%Ø	73%
Subject			

	n	%Ø	FW
2PL	(54)	(72%)	N/A
3PL	461	92%	0.81
3SG.F	154	82%	0.60
1SG	758	71%	0.43
2SG	120	68%	0.39
1PL	248	64%	0.38
3SG.M	546	63%	0.36
Range			**45**
Corpus			

	n	%Ø	FW
AIS	539	77%	0.55
CHILS	1,748	72%	0.45
Range			**10**
Speaker (random)			

Std. dev.			0.33

Table 89. Analysis of pro-drop in both CHILS and AIS corpora with subject and corpus as variables.

Variation in pronoun usage across AIS and CHILS participants

	n	%Ø	FW
Angelo F	91	87%	0.63
710	93	83%	0.60
Rafaelle	95	82%	0.59
664	97	80%	0.58
Frank D'A	91	79%	0.57
682	97	79%	0.57
701	67	79%	0.56
Isy	88	78%	0.54
Caroline	86	78%	0.53
Sante	76	76%	0.53
Palma	91	74%	0.52
Joanna	86	71%	0.51
Nunciata	78	76%	0.51
Annita	92	75%	0.51
654	92	73%	0.51
Antonietta	89	78%	0.50
Nello	93	72%	0.49
656	93	68%	0.47
Assunta	85	72%	0.47
Frank L	89	70%	0.46
Clara	80	69%	0.45
Ennio	85	67%	0.43
Renzo	100	62%	0.40
Ernie	84	60%	0.38
Arcangela	86	59%	0.38
Antonia	83	53%	0.33
Range			**30**

Table 90. Null subject rates by participant, combining the CHILS and AIS datasets.

H. 4. Combined corpora – HLVC, AIS, and CHILS

One-level analysis of pro-drop across HLVC, AIS, and CHILS

Tokens (n):	3,921	%∅	76%
Subject			

	n	%∅	FW
2PL	(60)	(75%)	N/A
3PL	713	91%	0.77
3SG.F	245	82%	0.58
1PL	544	75%	0.44
3SG.M	772	71%	0.42
1SG	1,416	72%	0.40
2SG	231	69%	0.37
	Range		**40**
Corpus			

	n	%∅	FW
HLVC	1,634	80%	0.56
AIS	539	77%	0.52
CHILS	1,748	72%	0.42
	Range		**14**
Speaker (random)			
Std. dev.			0.36

Table 91. One-level analysis using tokens from the HLVC, AIS, and CHILS datasets with the variables subject and corpus.

Variation in pronoun usage across HLVC, AIS, and CHILS participants

	n	%∅	FW
Angelo F	91	87%	0.64
IXF38A	61	93%	0.61
Rafaelle	95	82%	0.61
IXM47A	69	90%	0.60
I1M61A	95	91%	0.59
Frank D'A	91	79%	0.58
I1F61A	93	88%	0.58
IXF35A	70	87%	0.56
710	93	83%	0.55
Isy	88	78%	0.55
Caroline	86	78%	0.55
Sante	76	76%	0.55
I1M75A	86	85%	0.54
I1M62A	80	83%	0.54
664	97	80%	0.53
I1F65A	80	84%	0.53
Palma	91	74%	0.53
701	67	79%	0.52
I1F71A	89	82%	0.52
682	97	79%	0.52
Antonietta	89	78%	0.52
Annita	92	75%	0.52
Nunciata	78	76%	0.51
IXM64A	89	82%	0.51
Joanna	86	71%	0.51
Nello	93	72%	0.51
I2F57A	89	80%	0.50
Frank L	89	70%	0.49
IXM52A	83	78%	0.48
Assunta	85	72%	0.47
I1M60A	97	77%	0.47
IXF61A	63	78%	0.47
Clara	80	69%	0.46
654	92	73%	0.46
IXF94A	78	77%	0.45
Ennio	85	67%	0.45
IXF51A	82	74%	0.44

I1M61B	90	73%	0.44
IXM35A	85	72%	0.43
Renzo	100	62%	0.42
656	93	68%	0.42
IXM61A	68	71%	0.42
Arcangela	86	59%	0.39
Ernie	84	60%	0.39
Antonia	83	53%	0.35
I1F73A	87	61%	0.32
Range			**33**

Table 92. Null subject rates by participant, combining the HLVC and CHILS datasets.

273

H. 5. Combined corpora – HLVC homeland and heritage

One-level analysis of pro-drop in the HLVC homeland and heritage corpora

n:	1,634	%Ø:	80%
Switch reference			

	n	%Ø	FW
SAME REF.	591	89%	0.69
SWITCH REF.	647	79%	0.46
NEW TURN	396	69%	0.35
	Range		**34**

Subject

	n	%Ø	FW
(2PL)	(6)	(100%)	N/A
3SG.M	226	91%	0.66
3PL	252	90%	0.65
1PL	296	83%	0.52
3SG.F	91	82%	0.49
1SG	658	73%	0.34
2SG	111	70%	0.34
	Range		**32**

Preverbal Element

	n	%Ø	FW
YES	617	83%	0.54
NO	1,017	79%	0.46
	Range		**8**

Subcorpus

	n	%Ø	FW
Homeland	748	80%	(0.50)
Heritage	886	80%	(0.50)
	Range		**0**

Speaker (Random)

Std. Dev.		0.43

Table 93. Significant variables of pro-drop realization in the HLVC corpus comparing homeland and heritage speakers.

Variation in pronoun usage across HLVC participants

	n	%Ø	FW
IXM47A	69	90%	0.66
I1F65A	80	84%	0.59
I1F61A	93	88%	0.59
IXF38A	61	93%	0.59
I1M61A	95	91%	0.59
IXF35A	70	87%	0.58
I1M75A	86	85%	0.55
I2F57A	89	80%	0.51
IXM52A	83	78%	0.51
IXM64A	89	82%	0.51
I1M62A	80	83%	0.51
I1F71A	89	82%	0.51
I1M60A	97	77%	0.47
IXF61A	63	78%	0.46
IXF94A	78	77%	0.46
IXM35A	85	72%	0.44
IXM61A	68	71%	0.40
IXF51A	82	74%	0.40
I1M61B	90	73%	0.39
I1F73A	87	61%	0.31
Range			**33**

Table 94. Inter-speaker variation in the HLVC analysis of homeland and heritage speakers.

274

Curriculum Vitae

Focus

Heritage Languages, Rural dialects, Italian languages, Language documentation and archiving, Language variation and change

Education

Western University | London, Ontario   2017-2024
PhD French Linguistics
Dissertation:
        The imbalanced interaction of verbal ambiguity and pro-drop:
        The functional hypothesis in homeland and heritage varieties of Calabrian Italian and Ciociaro

Western University | London, Ontario   2015 – 2017
M.A. Linguistics
Major Research Project:
        Variable /v/ realization as [w] in Ciociaro

University of Toronto | Toronto, Ontario   2010 – 2015
B.A. French Linguistics
        Minors in Linguistics and Spanish

Research Projects

Co-Investigator | CHILS Project   2016 – Present
        Organize and conduct sociolinguistic interviews in Sarnia, Ontario
        Investigate the heritage Ciociaro spoken in Sarnia from Frosinone, Lazio

Co-Investigator | SWORE Project   2016 – Present
        Organize and conduct sociolinguistic interviews
        Investigate the regional English of Southwestern Ontario

Publications

Nagy, N., Iannozzi, M., & Heap, D. (2018). Faetar null subjects: A variationist study of a heritage
        language in contact. *International Journal of the Sociology of Language* 249:31-47.

Iannozzi, M. (2017). The Good [w]ita in Canada [Master's Thesis]. University of Western Ontario.

Iannozzi, M. (2016). Pro-drop in Faetar in Canada: A Study of a Heritage Language in Contact.
        Western Papers in Linguistics 2(2).

Conferences

Iannozzi, M. (2020, February). *An emigrated Italian dialect and its variable realization of /v/ as
        [w].* Paper presented at Southwest Popular Culture Association, Albuquerque, NM.

Iannozzi, M. (2017, June). *A Seat at "La Tavola".* Paper presented at Canadian-Italian
        Cooperation at 150 and 156, Toronto, ON.