Reference Impersonals in Texas Spoken Spanish: A Corpus-Based Analysis of
Sociopragmatic Variation

by
Allison Lynn Milner

A dissertation submitted to the Department of Hispanic Studies,
College of Liberal Arts and Social Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Spanish

Chair of Committee: Dr. Manuel Gutiérrez

Committee Member: Dr. Mata Fairclough

Committee Member: Dr. C. E. Goodin-Mayeda

Committee Member: Dr. Antonio Moreno Sandoval

University of Houston
May 2022

ACKNOWLEDGEMENTS

ABSTRACT

Reference impersonals comprise a subtype of impersonal expression under the agent-defocusing approach to impersonality (Siewierska, 2011). These forms, which employ *man*-type pronouns or subject pronouns used anaphorically, hold a variety of pragmatic functions that, to date, have been underexamined (Akinṛẹmi, 2013). In an effort to analyze Spanish reference impersonals under a comprehensive pragmatic framework, the current study examines the use of impersonal *se*, indefinite *uno*, and reference impersonal forms of second-person singular *tú*, first-person plural *nosotros*, and third person-plural *ellos/ellas* in corpus data collected from the Spanish In Texas Corpus. Specifically, this study investigates the potential relation between the use of these impersonal forms and specific sociolinguistic (gender, generation, age, age range, education level) and linguistic (verb tense, verb mood, presence of adverbs of time, overt pronouns, or conditional *if* clauses) variables through a series of chi-square tests of association. Qualitative analysis also explores how the aforementioned reference impersonals serve as mechanisms for generalization, inclusive defocalization, and speaker concealment. Results indicate that U.S. Spanish speakers employ a host of reference impersonals in achieving distinct pragmatic intentions, including focalized generalization and rhetorical usage in addition to the three uses listed above. Additionally, statistically significant differences in reference-impersonal use based on speaker gender and verb tense point to a complete, developed reference-impersonal framework. This study is innovative in its consideration of reference-impersonal forms in addition to traditional impersonality mechanisms and demonstrates the need for a comprehensive framework for reference-impersonality that takes all reference-impersonal forms into account.

Keywords: reference-impersonals, impersonality, deictic shift, U.S. Spanish, corpus linguistics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Impersonal expressions represent an interesting phenomenon from a syntactic, pragmatic, and sociolinguistic point of view. While its title lends a degree of elucidation, one of the most interesting facets of impersonal expressions is that there is not a comprehensive classification to define or characterize them (Siewierska, 2008a; Speshilova and Gulyás, 2014). This may be for several reasons: in several languages, they share surface similarity with passives, which have received more attention and analysis (Blevins, 2003; Malchukov & Ogawa, 2011); availability or use of impersonal expressions differs between languages and even dialects of the same language (Light & Wallenberg, 2015; Malchukov & Siewierksa, 2011; Malchukov & Ogawa, 2011); much of the research concerning impersonal expressions has only focused on Indo-European languages (Siewierska, 2008a); finally, depending on the area of study, considerations of what is or is not an impersonal differ—for instance, forms considered impersonal in pragmatic analyses may not qualify as impersonals from strictly syntactic or semantic points of view. As a result, some aspects of impersonality have received ample attention while others, such as the pragmatic, have not (Akinrẹmi, 2013). Despite the differing concepts of what constitutes an impersonal expression, the use and structure of these forms are a key aspect of language use.

1. 1 Classifications of impersonal expressions
1.1.1 Subject-centered vs agent-defocusing

From the ample work conducted on impersonal forms, two chief conceptualizations have emerged: a subject-centered view and an agent-defocusing view (Siewierska, 2008a). Syntactic considerations partially define each category of impersonal expression. While the subject-centered approach is also couched in semantic consideration, the agent-defocusing account approaches impersonality from a pragmatic

point of view. These two conceptualizations of impersonal expressions will be discussed in more detail below.

    1.1.1.1 Subject-centered approach. Strictly speaking, the subject-centered (also referred to as "structure-based") account classifies an impersonal expression as one that lacks a canonical subject (Suñer, 1976). Siewierska (2008a) provides a succinct description of the subject-centered approach:

> In terms of this subject-based approach, constructions which
> have been viewed as impersonal include: (a) those with a
> subject which is not fully referential, (b) those with a subject
> which does not display canonical subject properties, (c) those
> with a subject which is not a verbal argument but merely a
> place filler manifesting no semantic or referential properties,
> i.e. an expletive subject[1], and (d) those with no overt subject at
> all. (p. 116)

The following illustrate examples of impersonal expressions that fit each category referenced above:

1 (a) They always start events late.

  (b) Ñuka-ta-ka   uma-ta   nana-wa-n-mi.

     Me-ACC-TOP   *head*-ACC   *hurt*-OBJ-PRES-3-VAL

     'My head hurts me.'   (Quechua; from Hermon, (2001), p.

151)

  (c) There   is   no   time   to   wait.

  (d) Wolno   (nam)   wracać   do domu.

     Allow   *(we:DAT)*   *return:INF*   *to home.*

---

[1] This is a common occurrence in non-pro-drop languages, such as English.

'One is allowed to return home.' / 'We are allowed to

return home.'   (Polish; from Siewierska, 2008a, p. 120)

Although questions arise regarding the total applicability of these subdivisions—

some, such as the last, are highly theoretical—a structure-based account represents a

useful tool for exploring a typology of impersonal expressions across languages

(Malchukov & Siewierksa, 2011). Indeed, much of the existing research on impersonal

expressions focuses on creating a broad comparison of forms across languages.

1.1.1.2 Agent-defocusing approach. On the other hand, the agent-defocusing

view considers discourse-level linguistic features. Under this framework, impersonal

expressions are those that involve agent-defocusing or diminishing the prominence of the

instigator (Siewierska, 2011), hence its title. Referring to agent-defocusing as a

communicative functional view, Malchukov and Ogawa (2011) state that

"impersonalization is viewed as under-elaboration or demotion of the agent/instigator"

(p. 20). Because it is based in part on a pragmatic sense of diminishing the importance of

an agent in comparison to the lexical meaning of its governing verb (Fried, 2006, p. 85),

the agent-defocusing approach to impersonality is broader in scope than its subject-

centered counterpart. In their analysis of agent-defocusing impersonals, Malchukov &

Ogawa (2011) highlight five prominent subtypes and argue that the type of subject

properties lost (based on Keenan's (1985) characterization of subject) will determine the

specific type of impersonal form that may arise. This framework distinguishes between

impersonal expressions with subjects that (a) have no referential argument (b) are

indefinite (c) are non-topical (d) are inanimate and (e) are non-agentive (p. 24).

Furthermore, they demonstrate how different agent-defocusing strategies rely on different

types of linguistic coding, as shown below:

2 (a) Llueve.

    Rain:3SG.

    'It's raining.'   (Spanish)

(b) En este negocio, uno   nunca   sabe.

    *In   this business one:*indef *never   know*:3SG

    'You never know in this business.'   (Spanish)

(c) Une   femme   viendra.

    INDEF.SGF   woman.SG   come.FUT.A3SG

    'A woman will come.'   (French) (from Malchukov & Ogawa, (2011,

    p. 30))

(d) Sneg   zanes   dorog-u.

    snow   cover.PAST.3SG.M   road.ACC

    'The snow covered the road.'   (Russian)   (from Malchukov & Ogawa

    (2011, p. 33))

(e) U   menja slomal-sja   zub.

    to   me   broke.3SG-REFL   tooth.NOM

    'I have a tooth broken.' (Russian)   (from Malchukov & Ogawa (2011,

    p. 35))

Importantly, the agent-defocusing approach views impersonality as a continuum, wherein instigators may indicate their level of involvement or distance in the actions in question through different forms of impersonal expression. This accounts for a broader scope of impersonality under the agent-defocusing viewpoint. For instance, Malchukov and Siewierksa (2011) further consider forms such as anticausatives, manifesting in Spanish as accidental *se*, as in *Se quebró el vidrio*, and action nominalizations (*The circling of the camp* (from Malchukov & Siewierksa, (2011), p. 2) as impersonal under

the agent-defocusing approach as well, given their shift of focus from instigator to event.

Though this focus has received less attention than the subject-centered approach

(Akinrẹmi, 2013), forays into this conceptualization of impersonal expressions have

yielded promising results at the pragmatic and semantic levels (Hurtado, 2015; Malamud,

2012).

1.1.2 Syntactic vs Semantic Impersonality

Similar in its acknowledgment of the broad quantity of forms encompassed in the

idea of an *impersonal expression*, the framework of syntactic versus semantic

impersonality offers an additional perspective on how to differentiate between impersonal

forms. In simple terms, a syntactically impersonal form lacks a lexical-syntactic subject,

while a semantic impersonal conceals or diminishes the verbal Agent. Motivations for

concealment or diminishment can include "an indetermination, generalization, or

intentional concealing" (Lamas, 2015, p. 101, translation mine). This distinction was

explored extensively by Gómez Torrego (2013) in his analysis of Spanish impersonality.

While the purely syntactic and purely semantic impersonal forms do exist, he also argued

that a great number of forms considered impersonal were semantic-syntactic in nature.

Below are examples, extracted from Gómez Torrego (2013), that reflect purely syntactic,

purely semantic, and semantic-syntactic impersonals:

    3. (a) Syntactic impersonal: ***Está nevando*** *en la sierra.*   (p. 9)

       (b) Semantic impersonal: ***Se dice*** *que no hay dinero.*   (p. 13)

       (c) Semantic-syntactic impersonal: ***Se vive*** *bien en España.*   (p. 20)

In 3(a), there is no lexical-syntactic subject with the meteorological verb "*nevar*,"

and the expression is thereby impersonal from a syntactic point of view. In 3(b), the

phrase "que no hay dinero" functions as syntactic subject; however, there is no semantic

Agent present to fulfill the verbal argument. Therefore, it is semantically impersonal. The

semantic-syntactic interpretation of 3(c) presents in that there is an impersonality marker

"se" present, but this particle does not function as a syntactic subject[2]. At the same time, the phrase suggests the existence of an Actor or Agent, but this entity is not mentioned, lending semantic impersonality (Gómez Torrego, 2013, p. 20).

As will be reviewed in more detail in Chapter 2, many syntactically impersonal expressions in Spanish relate to necessarily Agentless actions, communicated through the use of meteorological verbs and process verbs such as *haber, hacer, ser,* and *parecer* (Gómez Torrego, 2013). Semantic and semantic-syntactic impersonals employ the use of distinct lexical-subject forms and pronouns while maintaining impersonality based on semantic and discourse-level features.

1.1.3 Reference-impersonals

Reference-impersonals (R-impersonals) are a specific subtype of impersonal expression under the agent-defocusing approach. These forms use indefinite pronouns (French *on*, Spanish *uno*, English *one*) or personal pronouns with no contextual antecedent (non-anaphoric use, such as English *you* (Berry, 2009)), and they fulfill the syntactic role of subject in transitive sentences (Barberà & Cabredo Hofherr, 2017; Siewierska & Papastahi, 2011). This subtype was presented by Malchukov and Ogawa (2011) in complementary fashion with two other subtypes that together compose a tentative cross-linguistic typology of impersonal expression. They state that "one could distinguish between the domains of A-impersonals (sensitive to agentivity/animacy properties), T-impersonals (sensitive to topicality), and R-impersonals (sensitive to referentiality/definiteness properties)." (p. 44) Furthermore, R-impersonals cover constructions with a notional subject deficient in referential properties, with a further subdivision into constructions with a non-referential subject of meteo-verbs (meteorological expressions) as opposed to *man*-impersonals (Barberà & Cabredo

---

[2] See Gómez Torrego (2013: pp. 21-25) for an argument against "se" as a lexical-syntactic subject in impersonals.

Hofherr, 2017, p. 4). In this way, R-impersonals fit the category of semantic or semantic-syntactic impersonal forms as has been characterized by Gómez Torrego (2013), Lamas (2015), and others.

As previously mentioned, null subject pronouns used in R-impersonals are non-anaphoric, meaning that there is no previous referent in the expression by which they gain meaning. This distinction is also made by a deficiency in the phi-features carried by pronoun use—whereas anaphoric pronoun expressions carry a fully-specified set of phi-features[3], impersonal usage lacks this full specification (Cabredo Hofherr, 2003: p. 88). A second defining characteristic of R-impersonals which separates them between the meteo-verb classification and *man*-impersonals is the [number] and [human] features carried by each form. Null third person singular impersonals (3sg), termed "quasi-arguments," differ from third person plurals (3pl) in the [number] feature. Third person plurals have an underspecified [number] feature that lends itself to a count noun feature. This also allows for a [+human] interpretation. Conversely, the third person singular has no number feature whatsoever, and this lack of number slot causes mass noun and inanimate interpretations that correspond to a [-human] interpretation. Meteo-verbs are one of the most common examples of null singular third person impersonals under this framework; the Spanish "Llueve," (*It's raining*) takes a third person singular, [-human] interpretation. This also holds for non-pro-drop languages, such as English, wherein expressions such as "It's raining" or "It's snowing" also adopt a 3sg, [-human] interpretation.

Man-impersonals under the R-impersonal framework may be categorized into five distinct interpretations, as noted below (examples taken from Cabredo Hofherr, 2003, p. 83):

---

[3] phi-features (φ-features) are agreement features relating to traits such as person, number, gender, or class (den Dikken, 2011; Kerstens, 1993).

4 (a) specific existential reading: ***Tocan*** *a la puerta.*

   (They) knock.3pl at the door. (=someone is knocking…)

(b) vague existential reading: ***Han encontrado*** *una motocicleta en el patio.*

 (They) have.3pl found a motorbike in the courtyard.

(c) inferred existential reading: *Aquí **han comido** mariscos.*

   Here, (they) have.3pl eaten seafood. (=someone)

(d) corporate reading (Kärde 1943, cited in Cabredo Hofherr, 2003, p. 83):

Volvieron *a aumentar el IVA.*

   (They) raised the VAT again.

(e) universal Reading*: En España **hablan** español.*

   In Spain (they) speak.3pl Spanish.

The three distinct existential interpretations differ from corporate and universal readings based on the trait of a unique maximal group: whereas types 4a-4c permit replacement of the null pronoun with the indefinite *someone*, corporate and universal readings require delineation of a unique maximal group that must exclude the speaker and addressee, and therefore cannot accept *someone* as a possible replacement (Cabredo Hofherr 2003, p. 90). The corporate and universal readings, while similar in necessitating a unique maximal group, differ in the means of specifying the group: corporate readings are interpreted via specification by their predicate, whereas universal reading is achieved via a locative expression. With respect to corporate readings, the context of the predicate implies that a particular group is responsible, though the members of that group are

unidentified (thereby allowing the utterance to retain an impersonal interpretation). Not every predicate can yield a corporate reading, as is shown in Example 5:

5 (a) ***Firmaron*** *una nueva Constitución.*

(They) sign.3p.PST a new Constitution.

(b) ***Firmaron*** *sus nombres.*

(They) sign.3pl.PST their names.

Though the referents in 5(a) are not explicitly named, the predicate portion "una nueva Constitución" narrows the scope of potential candidates to those involved in government. On the other hand, the same verb "firmaron" with the argument "sus nombres" in 5(b) does not have the same power of suggesting a collective group of referents, given that signing one's name is an activity not limited in scope to a particular group. Identification of a universal impersonal via locative expression is given in 4(e) above. Siewierska and Papastahi (2011) have deemed corporate and universal expressions "semi-impersonals," given that the referents for corporate and universal readings, while still unknown, are partially conceivable (p. 581).

Existential impersonals, on the other hand, do not have such restrictions on their potential referents. Rather, the distinction between specific, vague, or inferred comes from other features: specific existentials are anchored in time, vague existentials "only imply that an event of the type described has taken place" (Cabredo Hofherr, 2003, p. 83), and inferred impersonals involve an event deducible by a perceivable result. Using the examples provided in 4 above, (a) demonstrates an anchoring in time—in this instance, the present. In 4(b), there is no reference to time but rather a result. Finally, the inferred impersonal in 4(c) takes its interpretation from the perceptive quality (smell) that led the speaker to deduce that seafood had been eaten. Generally, inferred impersonal

expressions take form based on arguments related to touch, smell, or taste. The following

table, adapted from Siewierska and Papastahi (2011), represents the differing features of

the five *man*-impersonals described above for third-person plural impersonals:

Table 1: Man Impersonals

Property	Universal	Corporate	Vague existential	Inferred existential	Specific existential
Necessarily group interaction	+	+	-	-	-
Delimitation of group by locative	+	-	-	-	-
Delimitation of group by predicate	-	+	-	-	-
Anchoring in time	-	+	+	+	+
Specific time reference	-	-	-	-	+
Inferred by result	-	-	-	+	-

In analyzing R-impersonals in English, Kitagawa and Lehrer (1990) employ a

similar, yet more streamlined distinction between what they term impersonal and vague

uses of personal pronouns:

> An 'impersonal' use of a pronoun applies to anyone and/or everyone. A 'vague'
>
> use applies to specific individuals, but they are not identified, or identifiable, by
>
> the speaker. Whereas impersonal uses are akin to an interpretation with the
>
> universal quantifier (often subject to restricted quantification), the vague use
>
> requires the existential quantifier. (p. 742)

Comparing the frameworks laid by Cabredo Hofherr and Kitagawa and Lehrer,

then, we notice that the possibility of a universal quantifier such as *someone, anyone,*

todos, or *todo el mundo* serves as a dividing line between different types of impersonals:

for Cabredo Hofherr, this marks the potential for a universal R-impersonal, and for

Kitagawa and Lehrer it delineates an impersonal use. While Kitagawa and Lehrer

maintain a simplified system of all other unidentified or unidentifiable non-universal

referents under the "vague" interpretation, Cabredo Hofherr and those expanding on her

framework, such as Siewierska and Papastahi (2011) elaborate on other semantic or

lexical properties to yield the corporate, vague existential, inferred existential, and specific existential interpretations.

1.2 Conclusion

The term *impersonal expression* encompasses a wide range of forms that generally serve to de-emphasize the syntactic or pragmatic subject of an expression. However, the sheer number of ways in which impersonality may be expressed discourages, for the time being, a more comprehensive definition. Various frameworks for impersonals highlight different characteristics; the subject-centered and agent-defocusing approaches represent two chief conceptualizations of impersonality. R-impersonals represent a specific subclass of agent-defocusing impersonal forms. These impersonals, which include either null third-person meteorological expressions or null or overt non-anaphoric pronouns, are prevalent forms of de-focusing the agent or instigator of an activity.

Given the broad scope of impersonal forms under the agent-defocusing view, many studies are limited to specific subtypes of impersonal expressions. However, other impersonal expressions utilizing non-anaphoric third, second, and first-person forms, as well as indefinites are also present in many languages. The current investigation seeks to determine the frequency and pragmatic function of various *man* R-impersonal forms in Spanish, including second and third-person forms. By examining their usage alongside the more traditional impersonal *se* and impersonal expressions with the indefinite pronoun *uno*, a more comprehensive view of how impersonality is articulated in oral Spanish will be achieved. This represents a novel advance in studies on impersonality and lends itself to current advances in sociolinguistics, pragmatic analysis, and Spanish language pedagogy.

The dissertation is divided into five chapters, including this first introductory portion. Chapter Two provides the theoretical framework relevant to the current

investigation, including a discussion of the research and function of R-impersonals in Spanish language and pedagogy. The research questions are also presented in Chapter Two. Chapter Three, which details methodology, reviews the role that corpus analysis has placed in Hispanic linguistics; details relevant demographic information for the corpus used for the current study; introduces the variables and data coding strategy used in the study; and expounds on the quantitative and qualitative analysis that was conducted. In Chapter Four, the results and analysis of the research study are shared. This includes pertinent findings of both the qualitative and quantitative analysis, as well as a discussion of the most salient results. Finally, Chapter Five offers concluding remarks on the significance of the study, its limitations, and future avenues for investigation.

Chapter 2: Theoretical Framework

2.1 Impersonal Expressions in Spanish

There are various ways of expressing impersonal ideas in Spanish. As mentioned in

Chapter 1, these forms may be divided based on whether they are syntactically,

semantically, or semantically-syntactically impersonal. In terms of semantic and

semantic-syntactic impersonals—which includes R-impersonals—prevalent forms

include the use of impersonal *se*, the indefinite pronoun *uno*, and the second-person

pronominal forms *tú* or *usted.* Certain dialects also utilize the first-person plural pronoun

and/or verb conjugation (Hurtado, 2015; Posio, 2012) and third person plural forms

(Lapidus & Otheguy, 2005; Siewierska & Papastahi, 2011), though their use has received

less academic attention than the aforementioned forms. The choice to employ an R-

impersonal over an expression with impersonal *se* is a notable language choice that, as

will be discussed in detail, serves an important pragmatic role. While impersonal *se*

expressions were previously considered the standard choice for agent-defocusing

impersonality (Cabañas Maya, 2006), we must note that the use of other pronouns,

especially *uno* and *tú*, is growing (Fernández, 2013; Guirado, 2011; Hurtado, 2015; Vila,

1987).

In addition to specific pronominal forms, other linguistic elements, deemed

markers of genericity, have been identified as co-occurring in phrases with impersonal

expressions. These elements motivate and facilitate an impersonal interpretation (Serrano,

2013) and include active voice, adverbial phrases of time (especially expressions with

"cuando"), conditional or hypothetical "if" clauses, and present tense (Lamas, 2015, p.

103; Serrano, 2013, p. 186). Additional consideration has been given to the register and

modality; investigations of oral speech by Hurtado (2015), Ibáñez (2018), Lapidus &

Otheguy (2005), and written documents by Dahl (2017), Fernández (2008), and Munuera

(2018) (among others) have convincingly shown that impersonality mechanisms differ

based on physical presence of a direct interlocutor, formal or informal register, and

written or oral modality.

Analyzing impersonal expressions through a pragmatic lens can assist in

exploring the functions of these various forms. This idea has been recognized due to

distinctions in impersonality based on either syntactic or semantic divisions (Gómez

Torrego, 2013). In general, R-impersonals can serve as "un juego inclusivo-

exclusivo…respecto de la situación comunicativa, es decir, a un mayor o menor

distanciamiento entre los interlocutores" (Hidalgo Navarro, 1996, p. 171). The subsequent

sections explore the distinct uses of specific R-impersonals in the overarching theme of

speaker-interlocutor distancing or message generalization. Pragmatic functions of each

form are highlighted, and a tentative cohesive framework is presented.

2.1.1 Se

The relevant literature suggests various uses for distinct pronouns in impersonal

expressions. In fact, various research investigations specifically examine the frequencies

and stylistic or pragmatic use of different pronouns (Fernández, 2008; Guirado, 2011;

Hurtado, 2015; Pulido Astorga & Rivadeneira Valenzuela, 2017). Traditionally,

impersonal *se* has received much attention as a primary form for generic impersonal

expressions in Spanish and is considered an impersonality marker (Serrano, 2018). In

contrast to other forms included in the present analysis, s*e* functions as an option for

impersonality in both colloquial and formal registers. In fact, analyses of language use in

specific contexts, such as judicial proceedings, have shown that *se* is an attractive choice

for maximum impersonality or heightened objectivity of a message (Lamas, 2015).

Grammarians focus on the use of *se* in passive-reflexive (*pasiva-refleja*) and

impersonal forms. According to Moreno Castrillón (2013), passive-reflexive *se* is

employed when the nominal argument of a verb is an object or a person that is not

preceded by the personal preposition *a*. On the other hand, a nominal argument preceded

by personal *a* in conjunction with *se* yields an impersonal interpretation:

6 (a) ***Se buscan*** *especialistas.*

(b) ***Se abrieron*** *las puertas*. (from Westphal, 1999, p. 156)

(c) ***Se saludó*** *a los nuevos empleados.*

6a and 6b contain nominal arguments—*especialistas* and *puertas*—that function as

syntactic subjects and are therefore considered passive-reflexive. However, the personal *a*

in 6c places the nominal argument, *los nuevos empleados*, in the role of complement,

impeding its ability to serve as syntactic subject. This results in the lack of a syntactic *and*

semantic subject and represents an instance of impersonal *se*. All three examples are

categorized as semantically impersonal because the agent for the verb is absent. In

analyzing forms of impersonal *se*, some authors, such as Westphal (1999) also suggest the

existence of a medium reflexive (*media refleja*), characterized as not implying the

participation of any agent (p.156). Accidental *se*, as in "***Se me rompieron*** *las lentes"*, is

the choice form cited for a medium reflexive.

Semantic and syntactic analyses have centered around whether or not the particle

se functions as a subject in impersonal sentences (Gómez Torrego, 2013; Mendikoetxea,

2008; Oca, 1914; Romero & Ormazabal, 2019). This form has also been the primary

emphasis of second-language pedagogy on impersonal expressions (see Section 2.3 for

further detail on this point), in many cases representing the only explicitly taught form for

non-native speakers. Example 7 shows a prototypical use of impersonal *se*.

7 ***Se trabaja*** *mucho en esta oficina.*

One works a lot in this office. (Cited in Giordano (2015, p. 151))

This example is one of semantic-syntactic impersonality with an intransitive verb; there is

no nominal clause to act as syntactic subject, nor is there a nominal clause to fulfill the

role of semantic Agent. This contrasts with a purely semantic impersonal form, shared

initially in Chapter 1 and reproduced here:

      8 ***Se dice*** *que no hay dinero.*

       They say there's no money.   (From Gómez-Torrego (2013, p. 13))

In this phrase, the clause *que no hay dinero* serves as a syntactic subject, but there is no

semantic Agent corresponding to the verbal argument. In this instance, it is unclear *who*

makes the claim, and this is enhanced by the use of impersonal *se*.

As demonstrated in the interpretations of the above examples, impersonal *se*

serves a pragmatic role of agent de-emphasis, removing reference to a specific Agent

(assumed to be human) and yielding a generic or generalized interpretation. This may in

part be aided by *se* being an unmarked form for gender, number, case, or reflexivity

(Gervasi, 2007, p. 350). As De Cock (2021) asserts in her contrastive analysis of

impersonal *se* and *uno* in oral discourse, impersonal *se* provides maximum agent-

defocusing compared to other forms. Her corpus analysis considered discourse-pragmatic

features in the prevalence of the two impersonal forms and focused on the communicative

genre (conversation or TV debate), the co-occurrence of other pronominal forms in the

dialogue, and restrictions on space, time, or miscellaneous. Results showed that

restrictions on time and in miscellaneous functions highly correlated with the use of

impersonal *se*. In addition, impersonal *se* was preferred to *uno* in the genre of TV debate

and other professional discourse genres. This further suggests that impersonal *se* serves a

pragmatic function of maximum distancing or impartiality compared to other R-

impersonals; it also contextualizes the frequency of adverbial restrictors with the form:

"the lack of an agent or experiencer in *se*-constructions seems partially compensated by

the presence of other information that restricts the scope of the construction" (p. 103).

The following example, taken from De Cock (2021), illustrates the use of impersonal *se*

for maximum agent de-focusing in the context of a formal debate:

> 9   *Pero, por supuesto, es imprescindible ampliar este régimen de garantías,*
>
> *dotarle de más flexibilidad, de forma que el usuario pueda, en el momento en*
>
> *el que se produzca el problema, dentro de unos plazos de esta garantía que se*
>
> *establezcan el exigir una indemnización y una reparación inmediata, que*
>
> *luego **se solventará entre el resto de los agentes** que intervienen quién es el*
>
> *responsable(…)*
>
> (CORLEC debate)
>
> But, of course, it is indispensable to amplify this system of guarantees, to give
>
> it more flexibility, so that the user can, in the moment that the problem occurs,
>
> within some delays of this guarantee which are establish [sic], demand a
>
> compensation and an immediate redress, which later on **will be settled among**
>
> **the rest of the agents** who intervene who is responsible(…) (from De Cock
>
> (2021, p. 104)).

Here, when reviewed through a discourse-pragmatics focus, we see that the emphasis of

the message is on the recompense offered to users. The exact people responsible for this

recompense is more or less irrelevant (to the users), yet still sufficiently specified through

the clause "*entre el resto de los agentes*." In assessing the pragmatic function of

impersonal *se* and other R-impersonals, considering larger linguistic context and

extralinguistic features such as genre are important. The present analysis reviews the

frequency and function of impersonal *se* in both its semantic and syntactic-semantic

forms to consider how its use contributes to pragmatic impersonality.

2.1.2 Uno

The pronoun *uno* has received particular attention, given its status as an

"indefinite / impersonal pronoun" (Real Academia Española, 2011). While prior iterations

of descriptive grammar manuals such as the Real Academia Española (RAE) did not offer

more information about the status of *uno* apart from its indefinite quality (Gelabert-

Desnoyer, 2008), it is now acknowledged that the use of *uno* may convey various

meanings based on context, including that of referencing an unknown or indefinite

subject, expressing a universal impersonal (Díaz Blanca, 2005), or "*con referencia al yo*

que habla" (Real Academia Española, 2011; Gómez Torrego, 2013). The existence of

these uses, mentioned on the RAE's *Diccionario panhispánico de dudas*, point to a

continuum of meanings for *uno* based on context. One of the most recognized uses is

associated with the generalization or de-emphasis of the agent, extending the referent

from one subject (the speaker) to a more general population (De Cock, 2021; Guirado,

2011). For this reason, *uno* and *se* are often presented as complementary impersonal

forms; in many impersonal sentences with *se*, the clitic could be replaced with the

indefinite *uno* and the meaning of the utterance would hold (Gómez Torrego, 2013, pp.

16-17). Observe Example 10, which illustrates how 2 above may be re-formulated with

uno:

    10 (a) ***Se***   ***trabaja***   *mucho en*   *esta*   *oficina*. (Giordano (2015, p. 151))

      One works a lot in this office.

     (b) ***Uno***   ***trabaja***   *mucho*   *en*   *esta*   *oficina.*

       One works a lot in this office.

In this example, use of *uno* conveys a generality—the particular identity of an agent(s)

working is not of primary importance. Rather, the general message—that there is a lot of

work accomplished—takes the forefront. At the same time, we note that the phrase "en

esta oficina" places a certain limit on the potential referents. In other cases, impersonal

uno is employed to express a more universal statement, as shown in 11.

    11 *...****uno no sabe****, ¿verdad?, cómo se va a morir.* (Lope Blanch (1990, p. 109))

One never knows, right? how you're going to die.

However, it is important to note that the function of *uno* differs depending on such factors as context and dialect—for example, investigations by Hurtado (2015) and Gelabert-Desnoyer (2008) revealed a high frequency of *uno* in contexts in which the speaker referred to themselves. Hurtado's analysis of impersonality in spoken Spanish in Bogotá revealed that *uno* was the most frequently employed impersonal form, followed by impersonal *se,* second person informal *tú,* and the second person formal *usted.* In the majority of the instances quantified, *uno* was used in contexts in which speakers were assumed to be referring to themselves, as exemplified in 12.

> 12 *Ahora cuando fui a Ibagué estuve muy contenta, la vida mía fue muy sabrosa, sino que el cambio a veces de la como…a **uno neura**. Pensar **uno** que no es lo mismo, ¿no?*
>
> Now when I went to Ibagué I was very happy, my life was very rich, but sometimes the change of how…makes you nervous. One thinks that one's not the same, right? (from Hurtado (2015, pp. 132-133))

Here, the switch-reference between first and third person, or the co-occurrence of the indefinite pronoun *uno* with first person verb conjugations (*fui* and *estuve*) and adjectives (*mía*), signals *uno*'s use as a reference to the speaker. This use of *uno* for nonspecific first-person reference contrasts with the use of impersonal *se* in pragmatic contexts intended to exclude the speaker found in other varieties of Spanish (Hurtado, 2015). It is worth noting that the use of the impersonal form is dependent, among other factors, on dialect; a sociolinguistic analysis of impersonality in the spoken Spanish of Salamanca revealed that speakers employed the second person singular impersonal form far more frequently than *uno* (Posio, 2017). Additional features, including syntactic considerations

like the presence of the clitic *se* in a reflexive form, may also encourage the use of

indefinite *uno* in discourse to avoid repetition (De Cock, 2021; Fernández, 2008). This

data emphasizes the importance of incorporating not only the semantic, but also syntactic,

pragmatic, and sociolinguistic factors into consideration when examining the use of

impersonal expressions.

2.1.3 Second person singular / Tú

Despite its omission from pedagogical materials, the Spanish singular second-

person verb conjugation (2SG) and its accompanying pronoun *tú* represents a common

form for impersonal expressions (Hidalgo Navarro, 1996; Vila, 1987). There are multiple

labels for this construction, including generic, generalized, impersonal, unspecific, or

non-deictic second-person (Fernández, 2008; Kitagawa & Lehrer, 1990; Serrano, 2013;

Vila, 1987). Though each title carries a specific interpretation of the form, the common

trait shared between them is that the referent of the second person pronoun and/or verb

conjugation is understood *not* to be explicitly and exclusively the interlocutor. This use

appears to counter the linguistic assumption that personal pronouns always involve the

notion of *hearer* or *interlocutor* (Aijon Oliva & Serrano, 2013, pp. 68-69), yet serves an

important stylistic function in discourse, both written and oral (Tolchinsky & Rosado,

2005).

Use of the impersonal 2SG has been found in several varieties of Spanish and has

attracted attention from those interested in analyzing its sociolinguistic and pragmatic

features. Broadly speaking, use of impersonal 2SG over impersonal *se* or indefinite *uno* is

seen to characterize, in a way more closely relatable to the interlocutor, a generalized

situation that could ostensibly apply to anybody (Fernández, 2013, p. 90). Similarly,

Serrano (2013) proposes an objectivizing *tú*, a specific use of impersonal 2SG aimed at

centering the speaker's ideas or opinions in a less subjective frame. Her study correlated

the use of the null pronoun with a closer speaker-interlocutor relationship in which the

interlocutor is assumed to either have the same discourse information as the speaker or is assumed to accept the speaker's ideas or opinions as generalizations without question, thereby allowing for pronoun omission. This use of null pronoun in lieu of the overt form with second person conforms with Accessibility Theory (Ariel, 2001), which posits that highly-accessible forms in discourse need not appear in overt structure. However, null pronoun usage is also seen as a strategy to de-emphasize the agent[4]. The following 2SG impersonal represents a generalizing use:

13 *La verdad es   que (tú) nunca* **sabe***s cuando vas a morir.*

The truth is you never know when you're going to die. (from Gómez Torrego, (2013, p. 15))

In this well-cited example, the verbal desinences in *sabes* and *vas* have syntactic concordance to the second-person singular form, yet do not semantically refer to the interlocutor. Rather, the use of 2SG in this impersonal generalizes the broad idea of the inability to know the limits of one's mortality. This idea is presumably accessible or understandable, in a general sense, to most people.

Drawing from Serrano's classification of objectivizing *tú,* Pulido Astorga and Rivadeneira Valenzuela (2017) analyzed impersonal 2SG in Chilean Spanish. The authors found ample use of the impersonal 2SG form and divided its usage among the pragmatic functions of generalization, focalized generalization, inclusive defocalization, and concealment of the speaker (*encubrimiento del yo*) (p. 26). When used for generalization, the speaker makes a general reference that could be expanded to any person in question, not specifically to the interlocutor. The use of focalized generalization is "at the same time generic and personalized…it could refer to a specific group" (Pulido Astorga &

---

[4] Posio (2013) attests to a pragmatic distinction between the overt and null use of the first person singular pronoun with specific verbs such as *creer* or *pensar* in Peninuslar Spanish; whereas the overt pronoun usage emphasizes the speaker's subjective perspective on the matter, null pronoun usage serves a pragmatic role of presenting the information in a more objective light.

Rivadeneira Valenzuela, 2017, p.207, translation mine*). The authors explain that focalized generalization places the semantic Agent within a broad group of referents, akin to a corporate reading of impersonal third person forms. However, the agent is never specifically referenced. Through inclusive defocalization, the speaker tells the interlocutor about a personal experience that could also, in theory, apply to the listener. In other words, this use shows an empathetic or inclusive style toward the interlocutor. Finally, the use of *tú* for concealment of the speaker replaces the use of the first-person pronoun *yo* in discourse. In this case, the use of second person aims to conceal the identity of the speaker in instances in which the speaker is referring to themselves but does not wish to implicate themselves directly in the actions described for one reason or another. This use coincides with the idea of *indefinición del sujeto*, (Seco (1973), as cited in Hidalgo Navarro, (1996)) as a strategy by which the speaker "en la conversación como recurso con que la modestia o la timidez eluden le mención del YO bajo un disfraz de <cualquiera>, de manera que el caso individual tiende a diluirse" (p. 169). Each of these interpretations, while context-dependent and at times difficult to identify in discourse, can function as unique pragmatic operators in the use of *tú* impersonal expressions. The following examples demonstrate the four classifications of impersonal 2SG according to Pulido Astorga and Rivadeneira Valenzuela (2017):

> 14 (a) Generalization: *…podía ser un, un benefit buenísimo para todos los gentes aquí, no solo los gente que tiene hijos en esos daycare, pero también los XX que **tú tienes** XX porque si **tú tienes** gente trabajando y haciendo, avanzando su carrera…*
>
> (b) Focalized generalization: *Si me ha tocado experimentos... experimentar cuando yo vivía en México a los dieciocho años, pues uno tiene que manejar*

*obviamente (…) pues los policías te paraban para que no **te sirvas** una multa te están pidiendo dinero.*

(c) Inclusive defocalization: *…encontré unos cuadernos de, que yo tenía cuando era pequeño, cuando yo tenía como 8 años, y tenía una imaginación increíble, XX mirando los dibujos, no podía creer que era yo que los dibujé. Porque, no sé si.,. Es cuando **tú lees** tu diario y **dices** '¿Quién es esta persona?'...me sentí igual, entonces*

(d) Concealment of the speaker: *aprendí a valerme por mí misma, a tomar mis propias decisiones, mis padres ya no influían en mí…tuve más libertad de casada ya. Entonces, cuando tú **te vuelves** a juntar con tu pareja, te quieren volver a llevar ellos la rienda y es cuando uno vuelve a chocar un poco porque **aprendiste** a valerte sola, ver por ti sola. Si **tienes** hijos, **aprendes** a ver tú por ellos sin contar con la imagen del hombre en la casa. Y este, entonces como mi esposo se vino para acá, a una mujer sola ya no se te acercan con buenas intenciones. No **puedes** atener…no **puedes** tener amistad con hombres porque **eres** casada*

These examples of stylistic use of 2SG impersonals are discourse-dependent, but nevertheless show an agent-defocusing approach to impersonality. In 14(a), the speaker proposes the idea of free daycare as an appealing notion, using the 2SG form to suggest that this would be beneficial to anyone. The corporate reading accessible through the use of impersonal 2SG in 14 (b) roughly corresponds to "people in Mexico." In 14 (c), the 2SG forms serve to make the situation discussed more accessible to the interlocutor, in turn increasing speaker-interlocutor empathy. Finally, the speaker in 14 (d) discusses an experience that could be considered emotionally undesirable or shameful. In an attempt to de-emphasize her own involvement in the emotions and activities, she engages in switch-

reference, transitioning from the use of first-person pronouns and verb forms at the beginning to 2SG impersonal forms at the end. Presence of an impersonal 2SG with the overt pronoun in a switch reference context in which the first person *yo* was previously used further suggests that the second person aims to extend the frame of reference outward from the speaker (Fernández, 2008).

The use of 2SG, with both overt and null pronouns, represents a popular choice for R-impersonals in Spanish. Pragmatic uses for 2SG range from efforts to make the message appear more objective, to attempting to create speaker-interlocutor empathy, to a pragmatic attempt to diminish the Agent's prominence.

2.1.4 First person plural / nosotros

Though less common in usage than the second-person singular, the first-person plural (1PL) in Spanish has been documented in R-impersonal forms. Regarding 1PL as in impersonal form, the *Nueva Gramática de la lengua española* states:

> La referencia de las demás personas incluidas por *nosotros* es imprecisa y depende de factores discursivos: una pareja, todos los miembros de una comunidad, de un país, de un continente, del planeta, etc. Ello da pie a los usos llamados genéricos, en los que la forma de plural adquiere un sentido cercano al de 'cualquiera, la gente en general', o al que se manifiesta en las pasivas reflejas o en las impersonales (Real Academia Española, 2011, p. 304)

One potential pragmatic use is that of implying a universal nature, in which the 1PL form could be substituted with a general "anyone," in line with the impersonal use of personal pronouns proposed by Kitagawa & Lehrer (1990). This is the primary characterization of 1PL impersonals documented by Posio (2012) in his analysis of Peninsular Spanish and European Portuguese first person plural pronoun usage. His corpus study showed that between hearer-inclusive, hearer-exclusive, and impersonal interpretations, impersonals

represented the smallest portion of 1PL forms (8.2% of all first-person plural forms in the

data). A typical universal 1PL impersonal is shown in Example 15:

15 *No, desnudos* ***somos*** *todos   iguales.*   (from Posio (2012, p. 353))

No, naked we are all equal.

On the other hand, 1PL can also serve as indicating hearer inclusivity or exclusivity in the

actions in question. When used in a hearer-inclusive manner, 1PL impersonals

"symbolically includes the addressee" as a potential referent, though the addressee may

not actually perform the action in question (Posio, 2012, p. 352). Hearer-exclusive 1PL

impersonals do not include the addressee as a potential referent. The following examples,

extracted from Ajión Oliva (2020), provide instances of hearer-inclusive and hearer

exclusive 1PL impersonals:

16 (a): ***nos vamos*** *a:: otras cuestiones / enseguida* ***oímos*** *un poquito de música /*

y: ***nos: iremos*** *al cine dentro de muy poquito*

Let *us* move on to some other issues. In a minute *we*'ll listen to a little bit of

music and *we*'ll be going to the movies right after.   (Ajión Oliva

(2020, p. 6)

(b): ***hemos tenido*** *estas dificultades / en algunas / por- / por impericia nuestra*

/ por-porque ***hemos metido*** *la pata…*

We've experienced these difficulties, some of them due- due to our own lack

of skill, be-because we've messed up…   (Ajión Oliva (2020, p. 9)

In 16(a) the speaker, a radio talk host, utilizes a 1PL impersonal in a hearer-inclusive

sense, blending her actions into the experience of her listeners. On the other hand, 16(b) is

a quote spoken by a local politician providing an explanation. The use of the 1PL forms

hemos tenido and *hemos metido* are not meant to infer that the listeners in the discourse

have participated in these actions with the politician, therefore demonstrating how 1PL

may be used in a hearer-exclusive way. Interestingly, 16(b) shows how 1PL may be used

with a pragmatic function of the speaker distancing themselves from a potentially

negative implication, encompassing the *encubrimiento del yo* presented in 2SG forms. In

an analysis conducted by Posio (2012) on the prevalence of impersonal, hearer-inclusive,

and hearer-exclusive 1PL impersonal forms in Peninsular Spanish and European

Portuguese, oral corpus data indicated that hearer-exclusive forms almost always included

use of the pronominal *nosotros* and referred to specified groups of people "depicted as a

single agent involved in a positively evaluated joint activities" (Posio, 2012, p. 349). An

example provided in the study was that of a surgeon, discussing their responsibilities in

transplant operations. Though the context does not explicitly reference other medical

staff, the surgeon employs a 1PL impersonal in diminishing their own prominence in the

endeavor in favor of generalizing the responsibility among a group. In this context,

hearer-excusive IPL impersonal forms defocus the individual agent but retain some

definable referent properties through group evaluation.

      In her genre-specific pragmatic analysis of Spanish legal proceedings, Ibáñez

(2018) highlights an additional value, that of attenuator, for 1PL impersonals. This use

creates a "*yo dilatado* que presenta contornos vagos y que, incluso, puede referirse a otras

personas…permite alterar el eje deíctico…en favor de la adquisición de valores

pragmáticos" (pp. 1060-61). When speakers employ a so-called "modesty use" of 1PL,

they imply that others are involved in their mentioned actions, therefore defocalizing

themselves as agent or diminishing their prominence. This can serve as an attenuating

factor in making the speaker's ideas appear more objective or less biased, harking back to

Serrano (2013)'s objectivizing *tú*. Finally, several scholars have noted the use of 1PL in

this way to serve as a hearer-dominant form (De Cock, 2011; Ibáñez, 2018). The hearer-

dominant form has been characterized as patronizing, empathetic, or polite, depending on

context. Finally, using 1PL in lieu of first-person singular to evoke the concept of group

involvement as a means of obscuring personal attachment or relation to the activity, as is

the case with the 2G *encubrimiento del yo*, is also a possibility. The following example,

also from Posio (2012), exemplifies this use:

> 17: *Yo creo que…bueno, que cada uno de nosotros…pues **tenemos** unas…cosas*
>
> *que nos pasan en la vida pero creo que **tenemos que pensar** siempre en el*
>
> *futuro.*
>
> I think that…well, that each one of us…well we have some…things that
>
> happen to us in life, but I think that we have to think always about the future.
>
> (p. 354)

In this instance, a woman was asked about a difficult instance from her past. While she

begins her answer with the first person singular, making a clear deictic link that situates

herself as the center of the action, she switches her reference to the 1PL as though to

deflect attention or discourage an emphasis on herself. Instead, she uses the 1PL pronoun

and the pronominal and verb "nosotros…temenos" to encourage the concept that the idea

she's describing is something shared between herself and a wider, though undefined,

group.

Additional pragmatic analyses confirm that the use of 1PL functions similarly to

the use of 2SG in generalizing personal experiences (Fernández, 2013). By opting for the

first person plural in making generalized statements, the speaker has the option of

implicitly including themselves as a potential referent, as shown in the following example

from Fernández (2013):

> 18 *¡Qué mal   **jugamos**   ayer   contra   Alemania!* (p. 98)
>
>   how   bad   play.1pl.pas   yesterday   against   Germany!

We played so badly yesterday against Germany!

This type of use is labeled by Borthen (2010) as a representative element reading. The condition laid out by the speaker need not necessarily apply to the speaker so long as it is true of a large enough representation of the group. So, although the speaker in 18 could be a spectator of a sports match and did not personally play the German team, the use of 1PL here symbolically links the spectator to his/her team. Posio (2012) proposes an empathetic motivation for this, defined in terms of a temporary perspective switch and manifested as a symbolic inclusion of the speaker in the referential range through the use of a first-person plural form (p. 343). Along the same lines, Ajión Oliva (2020) points to a hearer-inclusive use of 1PL impersonals in such a way as to create a "blending or fusion" of the speaker and interlocutor's viewpoints (p. 6). He contrasts this use of impersonal 1PL with impersonal 2SG. While both forms seek to extend the perspective of the speaker outward toward the interlocutor, the use of second person is seen as placing more distance and stronger "desubjectification" (Ajión Oliva, 2020) between the speaker and interlocutor than the use of 1PL. This perspective aligns with the overall concept of using 2SG and other personal pronoun forms in "non-standard" ways as a means to develop a stronger connection to the interlocutor as a means to facilitate communication.

2.1.5 Third person plural / ellos

Though less discussed than other R-impersonals, the third person plural (3PL) does appear in Spanish in this context. These impersonal expressions reference non-specific human entities (Cabredo Hofherr, 2006), and though they take the syntactic plural, may refer to singular referents at the semantic level. While many analyses neglect to consider 3PL, it is important to note that context and dialect may play an important role in its frequency. For instance, a qualitative analysis by Díaz Blanca (2005) found that 3PL forms were the most frequent case of impersonal expression found in a study of spoken Spanish in Mérida, Venezuela. On the other hand, an analysis of impersonal expressions

in the Spanish of monolingual residents of Mexico City by Gervasi (2007) found a

relatively low frequency for 3PL at the expense of impersonal *se.*

Pragmatically, Fernández (2013) situates third person plural impersonal forms in

an "indefinite" space. Others assert that 3PL adopts a necessarily exclusive reference in

relation to the speaker when used impersonally. The analysis of impersonal expressions

by Gervasi (2007), mentioned above, also included oral corpus data from bilingual

speakers in the United States. Her study analyzes the frequency of various impersonal

forms in oral contexts, and also examines the frequency of the forms based on the

discourse task (looking at whether the impersonal verbs appear in the habitual past, non-

habitual task, or present). This is one of the few studies that examines impersonal

expressions from a pragmatic perspective, and it is all the rarer in that it does so in the

context of United States Spanish. In contrast to the preference of the monolingual

speakers in Mexico City, bilingual speakers in the United States showed a strong

preference for 3PL in exclusive impersonal forms. Table 2 below demonstrates the

difference in speaker preference based on region:

Table 2 *Exclusive se and Third Person Plural Impersonals (Adapted from Gervasi (2007, pp. 349-350))*

	Exclusive *se* (count)	Exclusive *se* (%)	Third person plural (-n) (count)	Third person plural (-n) (%)	Total
Speaker group					
Bilingual	33	23.9%	157	79.3%	190
Monolingual	105	76.1%	41	20.7%	146
Total	138	100%	198	100%	336

Gervasi's study is significant because it shows the prevalence of 3PL impersonal

forms in United States Spanish. She demonstrates that this form appears to be a more

popular choice for expressing third person, exclusive impersonals, occupying a space

previously held by impersonal *se*. Unfortunately, the study lacks detailed information

about the nature of the bilingual speakers (whether or not they are heritage speakers or second/foreign language learners) as well as other demographic information (age, educational status, etc.). Regardless, the effort to analyze impersonal forms with attention to pragmatic factors represents an important step in profiling impersonality in United States Spanish. Prevailing scholarship asserts that third person expressions may only be impersonal when there is a null subject (Cabredo Hofherr, 2006; Fernández, 2013; Jaeggli, 1986; Siewierska & Papastahi, 2011), as shown in 19.

19 ***Llamaron*** *del banco.* (Example from Lapidus & Otheguy (2005, p. 157))

They called from the bank.

However, other research has found examples of R-impersonals with overt third-person pronouns (*ellos*). In their analysis of spoken Spanish in New York City, Lapidus and Otheguy (2005) found that the overt *ellos* was used in R-impersonals in recent arrivals, established immigrants, and Spanish speakers born and raised in New York City, as illustrated in the following example:

20 *Disco, disco… ya está arde. A mí me gusta más rock, pero me gustan muchos diferentes estilos de música, pues… a mí me gusta mucho la música pero, no pienso que tengo, no sé si se dice así en español, un buen oído para la música. Entonces, después de un tiempo me suena todo igual, pero me gusta, ¿tú sabes? A mí me gusta mucho el rock, me gusta diferente tip… típico de música americana, de que, no sé,* ***ellos, ellos*** *lo llamarían diferentes cosas, pero yo lo llamo todo rock ¿tú sabes? Como* ***ellos*** *tienen, like ballads,* ***ellos*** *tienen funky,* ***ellos*** *tienen rock, y tienen heavy metal, y tienen new wave, para mí esto es todo bien mezclado.* (013U)

Disco, disco is already dead. I like rock better. I like a lot of different styles of music. I like music a lot, but I don't think that I have, I don't

know if you say it like this in Spanish, a good ear for music. So, after a

while, everything sounds the same to me, but I like it, you know? I like

rock a lot, I like different types of American music, of which, I don't

know, they, they would call it different things, but I call it all rock. You

know? Like they have ballads, they have funky, they have rock, (they)

have heavy metal, and (they) have new wave. For me, this is all mixed.

From Lapidus & Otheguy (2005, p. 163)

In this example, there is no explicit referent for the bolded instances of *ellos*/they; while a reader or listener may intuit that the speaker is referring to, for instance, American music listeners, this context is absent from the dialogue. The use of overt *ellos* with no clear referent can therefore function as an impersonal expression. As indicated by Fernández (2008), the use of third person plural over other impersonal forms narrows the potential frame of referents, excluding the speaker and interlocutor.

The sample of overt *ellos* in Lapidus and Otheguy (2005) was marginal, yet present, in comparison to null pronoun usage in impersonal forms, constituting only 4% (107 of 2,834) of the analyzed tokens. The authors argue that the relatively low frequency of the form has led to its absence in sociolinguistic literature. At the same time, their data indicate that this form is not necessarily a linguistic innovation born from contact with English, given that it was found in the speech of recent arrivals from six Latin American countries; rather, the language contact situation between Spanish and English present in New York City relaxed pragmatic rules that disfavor the overt use of *ellos* and encourage its use, already present in the pre-contact variety, in the contact environment (p. 166). This viewpoint expands on prior claims by Morales (1995), who asserts that contact with English has resulted in the loss of Spanish impersonal *se* in favor of other forms. In her study of Spanish impersonal forms in Puerto Rican speakers, she found that those who

had resided in the continental United States disfavored impersonal *se* when compared to speakers who had exclusively lived in Puerto Rico. Similarly, Gervasi (2007) suggests that increased use of 3PL impersonals among bilingual Spanish speakers in the United States in comparison to their monolingual Mexican counterparts could be attributed to contact with English; specifically, she suggests that the lack of direct translation for Spanish *se* leads to activation of the 3PL form, given its similarity to the English impersonal *they*. However, Lapidus an Otheguy show that 3PL impersonals are present as much in the speech patterns of recent arrivals as they are in long term residents. Their perspective on language contact as an opportunity to relax pragmatic rules and permit alternate forms accomplishes a similar task as the perspectives of Morales (1995) and Gervasi (2005) in that it acknowledges the effect of Spanish in contact with English. At the same time, it goes one step farther in seeking an explanation. By comparing the patterns of speaker groups with different amounts of exposure to English, they show that 3PL impersonal use is not the simple result of English "replacing" or "erasing" a Spanish form, but rather providing a context for alteration.

Finally, Fernández (2013, 2008) details the presence of third-person singular and plural forms and conjugations for impersonal expressions in oral discourse and periodicals in Latin America. She utilizes various examples of the third person in impersonal expressions to denote her realms of generic, group, and indefinite membership. This conceptualization places different impersonal forms in distinct realms based on their potential referentiality: the generic area includes forms meant to generalize repeated, non-specific situations; the membership area loosely implicates a specific membership group through lexical and discourse clues; and the indefinite group contains forms with unspecified human agents with no discernable antecedents. Figure 1 below is a reproduction of this concept of the impersonality continuum:

Figure 1: *Personal Pronouns and the Impersonal Space (from Fernández (2013, p. 92))*
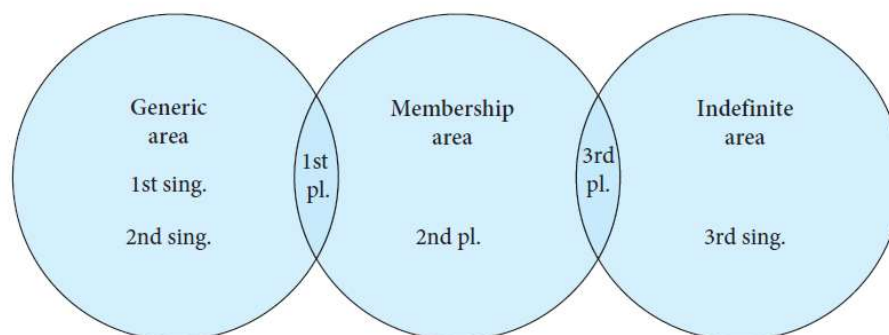
Fernández's framework for personal pronouns used impersonally in Spanish in the indefinite area partially overlaps with the framework for third-person R-impersonals presented by Cabredo Hofherr (2006, 2003). However, whereas Cabredo Hofherr's R-impersonal paradigm was designed specifically to accommodate third person forms, the generic-membership-indefinite continuum considers all pronominal forms used impersonally in Spanish.

To summarize, 3PL impersonals represent an impersonality option to designate speaker-exclusive groups. As is the case with other R-impersonals, its use differs based on context, register, or dialect. Though there are less studies analyzing 3PL impersonals in comparison to other forms, important efforts by Gervasi (2007) and Lapidus& Otheguy (2005) demonstrate its use in United States Spanish as an attractive option for impersonality.

2.1.6 Additional forms: generic, vague, or indefinite uses

The forms reviewed up to this point are those traditionally included in analyses of Spanish R-impersonals. In considering the broad category of impersonal forms, some authors have elected to expand their considerations toward inclusion of additional forms that may be better described as generic, vague, or indefinite. For instance, Dahl (2017)

and Díaz Blanca (2005) include the nominal phrase "la gente" (*people*) in their qualitative analyses of impersonality. This phrase serves as a generic and universal referent or could be used to emphasize disinterest in focus on the Agent (Díaz Blanca, 2005, p. 44). An example of generic "la gente" is as follows:

21 *...me veo en veces americana por mi color, pero cuando hablo en español **la gente está** muy sorprendida.*

...I see myself as American sometimes because of my color, but when I speak in Spanish people are very surprised.

Additional nominal elements like "todos" (*everyone*) are also used in generic statements, serving to implicate the widespread or generic property of an action without naming specific referents. While the scope of the current analysis does not include these forms, it is interesting to note that Spanish speakers have multiple options inside impersonal, generic, vague, or indefinite realms to express themselves. Additional analyses into the pragmatic nature of these nominal forms, in addition to impersonality markers, would be of use in further defining impersonal matrices in Spanish.

2.1.7 Pragmatic profile of Spanish impersonal forms

R-impersonal forms in Spanish are best conceptualized as a continuum that constitute varying degrees of speaker or listener involvement (Aijon Oliva & Serrano, 2013; Ajión Oliva, 2020; Fernández, 2013; Kluge, 2016). Due to the variety of functions and expressions, many studies only examine one or a small number of these forms, thus leaving an absence of a comprehensive framework for Spanish R-impersonals. Nevertheless, the analysis reviewed at present demonstrates that impersonal *se*, indefinite *uno*, 2SG, 1PL, and 3PL forms share certain degrees of overlap and distinction that allow for a tentative general framework of impersonality.

In terms of pragmatic functions, R-impersonal forms have been found to create generalizations, increase the odds of an objective interpretation of a message, or promote

agent de-emphasis or defocusing through concealment. Some forms, such as indefinite

uno, 2SG, and 1PL, may include the speaker as a potential referent; others, such as 3PL

forms and *se*, are speaker exclusive. In the same vein, 2SG and 1PL have been analyzed

for their capacity to include or exclude the interlocutor as an intended referent in

discourse. Table 3 provides a general schema of the distinct R-impersonals included in

the present study alongside their potential pragmatic functions.

Table 3: *R-impersonals and Pragmatic Uses*

	Generalization	Inclusive Defocalization	Speaker Concealment
Impersonal *se*	X		
Uno	X		X
2SG	X	X	X
1PL	X	X	X
3PL	X		

The current investigation seeks to confirm the pragmatic uses of various R-

impersonal forms in spoken US Spanish. In addition to other pertinent linguistic and

extralinguistic variables (see Chapter 3), the pragmatic use(s) of each form will be

considered an important aspect in characterizing their function in oral discourse.

2.2 Modality and Genre Considerations in Spanish R-impersonals

In considering patterns of language use, taking such factors as modality and genre

into account are useful in contextualizing which forms speakers choose to employ (Biber

et al., 1988; Biber, 1992). In this case, *modality* refers to the communicative nature, either

written or spoken, of information, while *genre* refers to a more extensive concept related

to "conventionalized communicative events (written and spoken) recognized as such by

the professional communities which adopt them" (Stewart, 1999, p. 135). In other words,

different genres are defined in part by the intended audience and communicative goals,

and linguistic choices will differ based on genre (Yasuda, 2011). In truth, the term

"genre" encompasses a large quantity of different oral or textual materials. For the

purposes of the present investigation, considering genre will be done in the context of formality.

Prior research shows that the distribution of R-impersonals in Spanish differs based on modality and genre (De Cock, 2011; Tolchinsky & Rosado, 2005). Oral corpus data in Spanish represents colloquial genres (conversations), formal genres (political debates, judicial processes, television newscasts), and communicative encounters that may straddle the line between the two (such as sociolinguistic interviews or radio broadcast media). As Biber (1988) indicates in his seminal work on register and modality analysis, the two features are often conflated as historically, the written and oral modalities lent themselves to specific types of genre, i.e., formal genres in written documents and informal genres in oral speech. However, it is important to recall that modality and genre are distinct features, and each modality may carry different genres. In the oral mode, colloquial communicative acts are characterized as more spontaneous and subjective than their formal counterparts (Ajión Oliva, 2020). By virtue of their being composed with anticipation and often to a nonspecific audience, written discourse is viewed as more objective and idea focused. However, it is important to remember that these are not absolute characterizations, and different genres within oral or written modality may differ.

An extensive multi-dimensional analysis of Spanish by Biber et al. (2006) grouped nineteen different registers into distinct dimensions, or sets, based on co-occurring linguistic features. Their analysis of 146 different linguistic features across 8 written and 11 oral registers yielded six different dimensions. Of interest to the current investigation is Dimension 1, the "stereotypical oral" dimension. The linguistic characteristics most associated with Dimension 1 are as follows:

Table 4: *Characteristics of Dimension 1, from (Biber et al. (2006, p. 13))*

Dimension	Positive Linguistic Features
1: Stereotypical oral	indicative mood, causal subordinate clauses, time adverbs, first person pronouns, copula *SER,* demonstrative pronouns, specific single-word conjuncts. first person pro-drop, copula *ESTAR.* mental verbs, place adverbs, existential *haber. que* verb complement clauses (indicative), tag questions, present tense, future *ira,* perfect aspect, communication verbs, third person pronouns, progressive aspect, *el que* clauses, yes-no questions, *que* relative clauses (indie.)., manner adverbs, augmentatives, quantifiers, *CU* verb complement clauses, premodifying demonstratives, conditional subordinate clauses, *nu listed,* desire verbs, general single-word conjuncts. verbs of facilitation, simple occurrence verbs

The registers with the highest average scores for Dimension 1 were business telephone conversations, casual conversations, contests, political debates, drama, and sociolinguistic interviews—all oral registers. The registers scoring highest—business telephone conversation and casual conversation—represent registers in which communicative acts are less planned, focus on actions or occurrences in the present tense, and are person-focused. This is shown through the salience of first, second, and third-person pronouns, indicative mood, and perfective clauses. These traits, along with a high use of time and other adverbs, characterize several of the "genericity triggers" found with R-impersonals, thus we may anticipate that some of these linguistic features will co-occur with the impersonal expressions analyzed in the present study.

Research in the distribution and use of Spanish R-impersonals by mode and genre have demonstrated the unique sociopragmatic properties of each form. A corpus analysis of impersonality mechanisms in published academic writing in Spanish and English by

Alonso Alonso (2011) found that impersonal *se* represented the most popular impersonality mechanism of all forms considered. Given that academic texts seek to impart information in an objective manner and contain formal writing features, high usage of the objectivizing *se* is not surprising. This study also found ample use of the 1PL impersonal in the so-called "modesty form," used in a way to "crear un sentido de solidaridad que invita al lector a compartir un punto de vista incluyéndolo de esta forma en el discurso" (p. 34).

Analysis of impersonality and agency alternation in Tolchinsky & Rosado (2005) looked at the use of distinct impersonal forms in writing from an age perspective, looking at development and utilization of the forms in grade school, junior high school, high school, and university students in Spain. This study analyzed the use of distinct agent-defocusing or distancing constructions in oral and written modalities, in expository and narrative texts, among different groups of literate Spanish students at the grade school, junior high, high school, and college levels. Focusing primarily on syntactic-semantic constructions of defocusing or impersonality, the study also acknowledges the importance of assessing modality, genre, and thematic content in a consideration of the distribution of the forms. Though passive *se* had been traditionally considered common in the Spanish tradition, results from this study showed a preference for active-voice constructions implying the second-person singular impersonal *tú* and impersonal *se*.

Combining a focus of specialized genre—legal proceedings—with the genre of newspaper articles, Lamas (2015) analyzed impersonality in cartel laws and edicts in Venezuelan newspapers. Ten different impersonality mechanisms, including personal pronominals, impersonal and passive *se* constructions, and constructions of *hay que + infinitivo* were included in a corpus study of twenty published newspaper articles. Impersonal *se* surfaced as the most frequent impersonality mechanism, reinforcing the

form's pragmatic usage as an objectivizing tool in formal registers. At the same time, the use of 3PL impersonals was acknowledged as significant in presenting a neutral or objective manner and providing distance between the edict-issuer and potential readers (p. 104). Lamas' study is also significant in that it identified other textual features that frequently appeared alongside impersonal forms, such as nominal abstraction, nominal precision, and co-occurrence of infinitival forms. These structures appear to reinforce the formality of the register but may also serve as potential impersonality triggers. At present, analysis of the written modality centers around formal genres such as academic writing and journalistic writing. These studies reveal a high use of impersonal *se* in formal registers, with use of 2SG and 3PL increasing as a means of stylistic objectivity.

Studies of R-impersonals in oral Spanish corpus data, especially with respect to genre, illustrate a rich variety of forms used and stylistic functions for each form. Results regarding the use and distribution of 1PL and 2SG in oral (radio) and written (press) media discourse in Ajión Oliva (2020) point to a preference for the first person plural over second person. In this article, he indicates that audience-inclusive 1PL impersonals serve as a means of "viewpoint blending or fusion," extending personal views or experiences of the speaker to a wider audience (p. 7). As such, this form serves to build a collaborative sense with the listener or audience, especially when there is no such specific referent present, as is the case in media discourse. Ibáñez (2018) also analyzed the use of 1PL impersonals, specifically in the context of oral court cases. Though speaker-inclusive 1PL was less common than the speaker-exclusive use of the form (21.4% inclusive vs 40.1% exclusive), its presence alongside "modesty" uses (impersonal) combined for 57.3% of sample data. As discussed above, analyses of oral corpus data composed of sociolinguistic interviews by Guirado (2011), Hurtado (2015) Posio (2013), and Pulido Astorga & Rivadeneira Valenzuela (2017), to name a few, provide a more complete

picture of the ample use of R-impersonals employed in oral contexts. Owing to the

spontaneous nature of speech, more informal nature, and opportunity for variable

thematic content during oral discourse, oral modality shows the potential to allow for

many different types of R-impersonals. Underlining the heterogeneity of uses for each R-

impersonal is the importance of considering modality, (Ajión Oliva, 2020; Biber, 1988;

De Cock, 2021), theme (Gervasi, 2007), genre (Ajión Oliva, 2018; Ibáñez, 2018), and

dialect (Guirado, 2011; Hurtado, 2015). The current investigation examines R-impersonal

distribution and frequency in oral corpus data, focusing on speech collected via

sociolinguistic interviews.

2.3 Impersonal Expressions in Spanish Language Pedagogy
2.3.1 Research on impersonal expressions and pedagogy

Compared to other topics of instruction, there are relatively few studies on

acquisition or use of impersonal expressions in second language learners or heritage

speakers of Spanish (Fernández, 2013). Earlier work focuses on one particular structure

deemed impersonal without elaborating on the properties of the structure involved. For

instance, Chandler (1996) and Feeny (1978) each offer brief reviews on learning

strategies for the acquisition of impersonal existential phrases requiring the subjunctive

mood (*Ser* + adj + *que* _____) but do not offer much in the way of theoretical explanation

of the structures themselves or how they constitute expressions of impersonality.

More often, the impersonal *se* appears in pedagogical literature in opposition with

the passive *se*, as found in Hernández's (2012) discussion of teaching constructions with

inverted subject positions. Interestingly, Hernández refers to the impersonal *se*

constructions as "impersonal passives." Giordano (2015) elaborates on the ambiguity

present when phrases with *se* could be interpreted as either passive or impersonal based

on the transitive/intransitive interpretation of the verb, as in constructions like "*se terminó*

a tiempo" (p. 152). In other cases, nominal arguments that function as patients match

their verbs in number, and it is unclear whether they function as subject or direct object. This leads to possible interpretations of either a passive or impersonal construction and can be challenging for Spanish language learners. Note the example presented in 22:

22 (a) *Ya **se pagó** la cuenta$_{DO}$. (impersonal)*

(b)*Ya **se pagó** la cuenta$_{SUJ}$. (passive)* (From Chaves (2006, p. 198))

In 22(a), *la cuenta* serves syntactically as the direct object and semantically as the patient of the action-process verb *pagar*. This gives the sentence an impersonal interpretation, leaving the agent of the paying unknown. On the other hand, 22(b) presents an alternative interpretation in which the nominal *cuenta* represents the syntactic subject in a passive construction that could alternately be expressed as *La cuenta fue pagada*.

In general, literature surrounding the teaching of these distinct *se* constructions in Spanish as a second or foreign language stresses that non-native speakers may have difficulty in acquiring them as distinct forms (Espinoza, 1997; Tremblay, 2005). In an effort to address this concern, Cardona (2015) argues in favor of dissolving the distinction of passive and impersonal *se* constructions. Her alternative conceptualization of the qualities in impersonal and passive *se* forms provide an opportunity to merge the two categories for the sake of teaching and acquisition. She then offers corpus-based and form-focused instruction pedagogy for teaching these constructions in Spanish. In the same vein, contrastive analysis has also been suggested as a useful means to assist students in consciously identifying how impersonality is expressed, especially in the case of students whose native Romance language partially parallels Spanish (Giordano, 2015).

2.3.2 Textbook presentation of impersonal expressions

Calls for the inclusion of impersonal expressions in second-language pedagogy go as far back as Hidalgo Navarro (1996), whose investigation of impersonal second person indicates its importance as a topic of colloquial language or informal registers. His article points to the widespread usage of impersonal *tú* in spoken Spanish and insists that in

order to gain communicative competence, second and foreign-language learners must be

exposed to stylistic linguistic elements commonplace in spoken, informal registers (p.

167). In response to this call for inclusion of impersonality in pedagogy, several

university level Spanish textbooks were examined. The contents of the books were

examined to see how information on impersonal expressions was presented both in terms

of explanation of the function of impersonal forms and the different types of impersonal

forms presented. Table 5 shows basic information about the textbooks included in the

analysis:

Table 5: *Impersonal Expressions in Spanish Language Textbooks*

Textbook Name & Author / Editor	Publication Date	Publisher	Level	Type of analysis
Portales 1, José Blanco	2016	Vista Higher Learning	Beginner	Impersonal *se*
Portales 2, José Blanco	2018	Vista Higher Learning	Intermediate	Expletive impersonals w/subjunctive Impersonal *se*
Repase y escriba, Sexta edición, María Canteli Dominicis	2014	John Wiley & Sons	Advanced	Expletive impersonals w/subjunctive; impersonal *se*; indefinite third person plural
Exploraciones, Mary Ann Blitt, Margarita Casas & Mary T. Copple (Eds.)	2015	Cengage	Intermediate	Expletive impersonals w/subjunctive

2.3.2.1 Portales 1. *Portales 1* is a second-language textbook that encompasses oral, written, and sociolinguistic Spanish-language competency. It is produced in conjunction with an interactive online platform and is designed for beginner-level students. An index search for the term "impersonal" yields results for two distinct grammar modules. The first, in Chapter 10, concerns impersonal *se* constructions. In the preliminary description for impersonal *se*, the book instructs that "**Se** can also be used to form constructions in which the person performing the action is not expressed or is de-emphasized" (Blanco, 2016, p. 350). A note on the sidebar of the textbook further states "In English, the passive voice or indefinite subjects (*you, they, one*) are used where Spanish uses impersonal constructions with **se.**" (p. 350). Examples of impersonal expressions with *se* are then provided, along with an additional note, shown in Figure 2:

Figure 2: *Portales 1 Uses of se*

(from Blanco (2016, p. 350))

In this resource, impersonal *se* constructions receive an explanation in terms of their pragmatic function: to de-emphasize the speaker or indicate their anonymity. *Portales 1* also seeks to make connections between this form and its English parallels by introducing the idea that it is used when English uses passive voice or indefinite expressions. Curiously, it lists indefinite *you* as an English impersonal form without explaining how *you* is used non-deictically in English. It is also interesting to note that the second image portrayed in Figure 2, of a sign that reads "*Se necesitan programadores*," displays a use of *passive se*, not *impersonal se*.

The other index result for "impersonal" in *Portales 1* is tied to the use of the subjunctive with impersonal expressions that utilize the existential impersonal. In an explanation of the general uses of the subjunctive, the book provides a list of phrases with *ser + adjective*, such as *Es bueno que…, es urgente que…, es importante que* and notes "These impersonal expressions are always followed by clauses in the subjunctive" (Blanco, 2016, p. 424).

Finally, the glossary of grammatical terms, located in the appendix, contains an entry for impersonal expressions, defining them as: "A third-person expression with no expressed or specific subject" (Blanco, 2016, p. A6). Three example sentences are listed: an existential, "Es muy importante," a null-subject meteorological "Llueve mucho," and an impersonal *se* form, "Aquí se habla español."

Overall, *Portales 1* provides a broad introduction to impersonal expressions by explaining that impersonal *se* can be used to deemphasize the subject of a sentence. It also conforms to longstanding grammar pedagogy tradition in introducing the terminology for impersonal expressions in use with *ser + adjective* phrases for the subjunctive. However, this resource lacks a more nuanced approach to impersonal expressions and does not mention R-impersonal phrases with the indefinite pronoun *uno* or with first person plural, second person, or third person forms.

2.3.2.2 Portales 2. The next textbook in this Spanish language textbook series is *Portales 2,* which takes a similar functional-communicative approach as *Portales 1* and is designed for intermediate-level students.

An index search for the term "impersonal" yields results for a learning module on the subjunctive with impersonal expressions. This learning module, located in the appendix, instructs on the use of the present subjunctive with forms comprised of *ser + adjective*, such as *es bueno* and *es necesario*. The other item indexed under "impersonal" is a grammar

module on the uses of *se*, in which a discussion of impersonal *se* is introduced. The

instructional text on impersonal *se* is copied in Figure 3 (Blanco, 2018, p. 408):

Figure 3*: Portales 2 Uses of se*

This pedagogic material presents that *se* may be used with either transitive or

intransitive verbs to form impersonal expressions, and the impersonal *se* form is presented

alongside other uses of *se*, such as the passive and accidental *se.* One point of attention is a

side note presented alongside these uses, as reproduced in Figure 4:

Figure 4: *Portales 2 Se Side note*

From Blanco, (2018, p. 408)

This note explains that *passive se* is often used on signs and warnings. However, as shown in Figure 2, copied from *Portales 1,* students are also advised that signage in Spanish typically uses *impersonal se*. Both textbooks, which come in the same language series, differentiate in their explanation of impersonal and passive *se*, yet contain conflicting explanations for the language used on signs or printed material. This unfortunate discrepancy could be a source of confusion for language students attempting to acquire these distinct forms based on grammar categorizations.

Finally, while *Portales* 2 includes sample phrases with *se*, an explanation of why an impersonal *se* expression could be used over a more direct phrase is lacking. Additionally, the use of other forms, such as first, second, or third person impersonals, is omitted.

2.3.2.3 Repase y escriba. *Repase y escriba, Séptima edición,* is a grammar-focused textbook for advanced second-language learners intended to address literacy development. Impersonal expressions are featured in two places in the textbook: firstly, in Chapter 5, in the context of "The subjunctive with impersonal expressions" (Dominicis, 2014). The instructional text regarding impersonal expressions is as follows: "Most impersonal expressions fall into one of the categories that call for the subjunctive (wish, doubt, emotion, unreality, etc.) and, therefore, require the subjunctive when there is a change of subject." (p. 123). A vocabulary list of expressions comprised of *ser* + adjective is then presented along with example phrases and writing activities.

Impersonal expressions are later featured in Chapter 12, under the heading "Impersonal use of *se*," though this information may only be found by referencing *se* in the index, not by looking at index references for impersonal expressions. Here, the text instructs that *se* may be used with intransitive verbs: "**Se** is found with the third-person singular of the verb (used intransitively) to mean *one, they, people, you* (indefinite)*. This construction is similar to the reflexive substitute for the passive discussed on pages 327-328 but is much less

common." (p. 323). The asterisk in the text signals a footnote that reads "It should be noted

that the indefinite or impersonal English *you* is sometimes expressed in Spanish by **tú,**

especially in the spoken language. Occasionally **usted** is also used in this way." (p. 323).

One point of interest in the discussion of impersonal *se* as presented in *Repsase y*

escriba is that it is placed in opposition to the purportedly more-popular passive *se* form, not

presented as a unique language feature. Secondly, this explanation contextualizes impersonal

se as a form used with intransitive verbs; however, as demonstrated in cases like *Ya se pagó*

la cuenta, impersonal *se* expressions can take transitive verbs as well. Finally, while this

textbook acknowledges the existence of impersonal second- and third-person constructions,

they are placed in a footnote and their usage is de-emphasized. Given that impersonal

expressions with *tú* are commonplace, it would be more beneficial to incorporate this

information in the primary text.

One further relevant structure in *Repase y escriba* is third-person impersonal forms.

However, the textbook does not include these forms in their discussion of impersonality, but

rather in a grammatical section on passive voice, as shown in Figure 5:

Figure 5: *Repase y Escribe Uses of Third-Person Impersonal*

⊚ THE INDEFINITE THIRD-PERSON PLURAL OF THE ACTIVE VERB

When the agent is not expressed or strongly implied in a passive voice sentence, a very common equivalent of the passive voice is the indefinite third-person plural of the active verb.* In this construction, the subject in Spanish is not **ellos** or **ellas** but an unexpressed indefinite *they.*

SECCIÓN GRAMATICAL **327**

The English subject becomes the direct object in Spanish.

Traen sacos de dormir al campamento.	*Sleeping bags are brought (They are bringing sleeping bags) to the camp.*
Hace años instalaron telescopios en el desierto.	*Telescopes were installed (They installed telescopes) in the desert years ago.*
No han publicado ese libro todavía.	*That book hasn't been published yet. (They haven't published that book yet.)*

(Dominicis, 2014, pp. 326-327)

The three example sentences in the grammar module shown above employ non-overt third person impersonal forms: *traen, instalaron,* and *han publicado*. These examples align pragmatically with the third-person R-impersonals examined by Cabredo Hofherr (2006, 2003) and Siewierska & Papastahi (2011) in that they de-emphasize the nominal agent through employing an impersonal, non-overt form. These example sentences also clearly employ the active voice, yet the textbook places these structures in a unit about passive verbs and states that they function as an equivalent form of the passive. This may be confusing for students—while it appears that this structure was placed in the module on passives because the two forms serve similar de-emphasizing pragmatic roles, the textbook is primarily organized by grammar, not pragmatics. Another curious point is the final sentence in the explanation: "In this construction, the subject in Spanish is not **ellos** or **ellas** but an unexpressed indefinite *they*" (p. 327). The intention of this sentence may be to explain that the agent involved in the following examples are undefined or uncertain. However, stating that the subject is not "**ellos** or **ellas** but an unexpressed indefinite *they*" is misleading; the subject *in Spanish* would not be *they*, as *they* is an English pronoun. If an overt subject pronoun were to surface in these constructions, it would almost certainly be *ellos*, and this overt pronoun would be used *non-anaphorically*.

Overall, the discrepancies presented in *Repase y escriba* stem from a common source: there is no specific pedagogic unit on impersonal expressions. Rather, the various forms are divided along form-focused, grammatical lines, leading to an incomplete understanding of impersonality. This is not for lack of knowledge regarding the function of impersonals—the textbook describes an agent-defocusing view of impersonality while discussing the "pasiva refleja" form of *se* as "when the subject of a sentence is either inanimate, animal, or a de-individualized person or group of people" (Dominicis, 2014, p. 327). This explanation serves as a succinct description of agent-defocusing and could be used to present impersonal *se*,

forms with *tú* or *usted*, the indefinite *uno*, and other impersonal forms. It is possible that this

textbook, which is designed for strengthening written proficiency and Spanish language

literacy, adopts its current stance based on forms most likely to appear in written Spanish.

However, by presenting a more coherent framework of R-impersonals, it would allow

learners to acquire these forms in oral and written communication.

     2.3.2.4 Exploraciones. This book is described as an intermediate-level resource for

all students and focuses on reading, listening, speaking, an writing proficiency (Blitt et al.,

2015, pp. AIE-12-AIE-13). Similar to the classifications found in the *Portales* series, an

index search for impersonal expressions carries reference to a unit on the present subjunctive

with impersonal expressions using the verb *ser.* There is also a listing in the index for

"impersonal *se*." Chapter 9 contains a grammatical unit that presents passive *se*, impersonal

se, and accidental *se*. Of all the textbooks referenced in this section, *Exploraciones* appears to

be the only one that addresses the use of impersonal *se* from a communicative or pragmatic

standpoint. The idea of impersonality with *se* is first proposed in the introduction to passive

se: "The pronoun **se** is used when the person or thing performing an action is either unknown

or unimportant" (Blitt et al., 2015). The idea of de-emphasizing the agent is then extended to

impersonal *se*, as shown in Figure 6.

Figure 6: *Exploraciones Treatment of Se*

A comprobar

Los usos de *se*

El *se* pasivo

1. The pronoun **se** is used when the person or thing performing an action is either unknown or unimportant and the object affected by the action is used as the subject. This is known as a passive sentence. The verb is conjugated in the third person form to agree with the object. The singular form is used with singular nouns and the plural form with plural nouns. Notice that the subject can either precede or follow the verb.

   Se escucha el jazz aquí.
   *Jazz **is listened** to here.*

   Los discos compactos ya no **se compran** mucho.
   *Compact discs **are** not **bought** much anymore.*

2. When using an auxiliary verb such as **deber, poder,** or **necesitar** that is followed by an infinitive and a noun, the verb is conjugated in agreement with the noun because it is the subject.

   No **se puede entender** la letra.
   *The lyrics **can't be understood**.*

Las canciones **se deben bajar** de este sitio.
*The songs **should be downloaded** from this site.*

Se impersonal

3. Similar to the passive **se**, the impersonal se is also used when the subject is unknown, unimportant, or not specified; however, the impersonal **se** is not used with a noun. As a result, the verb is always conjugated in the singular form. The pronoun **se** translates to *one, you,* or *they* in English.

   No **se puede** entrar tarde al concierto.
   You can't get in late to the concert.

   Se dice que es un buen concierto.
   They say it's a good concert.

4. When the noun receiving the action is a specific person or persons, it becomes the direct object and must have a personal **a**. The verb is then conjugated in the singular form (impersonal **se**). Otherwise, the personal **a** is not necessary, and the noun acts as a subject (passive **se**); the verb must then be conjugated in agreement with the noun.

(Blitt et al., 2015, p. 286)

In *Exploraciones*, students are instructed that impersonal *se* should be used when the emphasis of a phrase is not placed on the subject. Note the seemingly contradictory information, however, in the grammar explanations listed above: we are told that impersonal *se* "is not used with a noun" in the first rule presented for impersonal *se*, and then are subsequently instructed that "When the noun receiving the action is a specific person or persons," it receives the personal *a* and receives a conjugation via impersonal *se* (Blitt et al, 2015, p. 286). Does the reference to a noun in 3 shown above relate to a noun in subject position? While this may be the case, the ambiguous explanation here could cause confusion for students.

The content in *Exploraciones* about impersonal expressions appears to be limited to the use of impersonal and passive *se* constructions—there are no references to the use of other subject or indefinite pronouns in impersonal expressions. One potential note regarding this resource is that while there is a distinction made between passive and impersonal *se*, the

distinction is referenced from a purely grammatical point of view—students learn that passive

se constructions contain an object that functions as subject, while the impersonal *se* sentences

do not. However, since the idea of agent de-emphasis is provided as a rationale for both

forms, students may not fully appreciate the communicative or pragmatic differences between

forms based off the explanation presented, or they may have trouble distinguishing between

forms.

2.3.3 Research and pedagogy conclusions

From this brief review, it is noteworthy that there is an apparent split between theory

and practice: research stresses that impersonal and passive forms are separate (Blevins, 2003;

Suñer, 1976), while applied pedagogy suggests either an ambiguity between the forms in

some cases (Espinoza, 1997; Palacios & Olivares, 2015) or a lack of functional distinction

(Blitt et al., 2015; Cardona, 2015) in others. In addition, it appears that there is very little

peer-reviewed research on the teaching or acquisition of R-impersonals or impersonal

expressions other than impersonal *se* or *ser* and *hacer* in existentials. Allusions to the idea

that pronouns or forms other than *se* are not considered in language instruction echo in

Hernández (2012): "Recipes can be employed to practice the impersonal passive. In cooking

instructions, the use of the first-person plural to provide directions sounds informal in

Spanish. Instead, the impersonal passive is stylistically preferred" (p. 325). Given that native

Spanish speakers utilize a variety of forms in impersonal expressions and that pragmatic and

communicative competence is an integral part of language learning, it is important to expand

extant research on the pragmatic implications of impersonal expressions in Spanish language

learning.

Acquiring R-impersonal expressions in Spanish will allow language learners to

articulate more abstract, generalized ideas and expand their communicative capacity, and

appropriate pedagogical materials would be beneficial in this process. As will be explored in

more detail below, one objective of the current investigation is to provide informed

recommendations on the instruction of R-impersonals in Spanish for heritage and second language learners based on patterns of usage found in oral corpus data. In this way, linguistic theory and pedagogical practice may be more closely linked in an effort to provide a contextualized language learning experience.

2.4 Research Questions

The present study seeks to analyze R-impersonal expressions in oral Spanish of the United States based on the following research questions:

R1: What is the frequency and distribution of different forms of impersonal expressions? Is there a statistically significant difference in the use of particular forms of impersonal expressions by gender of speaker, age range of the speaker, generation of speaker, or educational level of speaker?

R2: What is the relationship between the use of different impersonal expressions and their respective pragmatic functions, as evidenced in the corpus?

R3: Is there a statistically significant difference in the use of particular forms of impersonal expressions based on the tense/mood of verbal forms, the presence of hypothetical *if* clauses, or adverbs of time (genericity triggers)?

R4: How may the findings of this analysis inform current pedagogy for teaching impersonal forms in both second language and heritage Spanish?

Chapter 3: Methodology

3.1 Corpus analysis in Hispanic Linguistics

Many analyses of impersonal expressions in Spanish use language samples obtained

through a set of sociolinguistic interviews compiled by the researchers in question (Hurtado

2001, Posio 2016, Morales 1995, Pulido Astorga & Rivadeneira Valenzuela 2017, among

others). This technique allows for the opportunity to conduct in-depth qualitative analyses of

impersonal expressions in specific dialects or populations. At the same time, relying upon a

limited sample size limits the generalizability of findings and the ability to conduct

quantitative analysis. For this reason, it is useful to consider corpus analysis as a useful tool

to investigate the frequency and use of impersonal expressions. Corpus linguistics, while

considered by some as a distinct branch of study, can also be considered a specific tool or

methodology to be incorporated in the analysis of various linguistic phenomena (Anderson,

2010; Szmrecsanyi, 2017). Drawing on the tradition of text analysis utilized in historical

linguistics, corpus linguistics encompasses the analysis of frequencies and features in

collected corpora that range in scope, size, or purpose (Nieuwenhuijsen, 2016; Togini

Bonelli, 2010). While corpus linguistics has been applied extensively to material in English

(Biber, 1992; Deutschmann, 2006; McCarthy, 2002), there is a lack of comparable

comprehensive corpus analysis in Spanish (Parodi, 2007). This is surprising, given that

several Spanish corpora containing both written and oral data are available for analysis.

The present investigation will utilize a Spanish language corpus to assess the

frequency and use of reference impersonal expressions in spoken Spanish. While corpus

linguistics has received some pushback from more generativist schools, there are several

benefits of adopting a corpus-oriented approach. First, corpus linguistics presents an

opportunity to expand past relying on intuition for testing linguistic hypotheses. Hunston

(2002) highlights the usefulness of corpus analysis over speaker intuition in particular when

considering collocations, frequency, pragmatic meaning, and phraseology. Corpus linguistics

uses the investigator's intuition as a base and tests it through the data compiled in the corpus

to explore and identify salient features, thereby allowing for the incorporation of intuition

into a quantitative analytic framework (Viana et al., 2011). Additionally, adequate corpora

have been shown to contain enough data to identify trends in language use. For investigations

interested in examining real language use, corpora are a useful tool (Baker, 2010). By

combining quantitative analysis with language variation analysis, corpus linguistics can

identify larger patterns present in language use, going beyond simple frequency counts and

toward a framework for explaining language use in broader contexts (Biber, 1988;

Szmrecsanyi, 2017). In other words, the sample size present in modern corpora allow for

large-scale data analysis and stronger arguments to be built based on parametric analysis. In

fact, corpus linguistics has been successfully applied in areas such as discourse analysis

(Mautner, 2012), semantics (Gil-Vallejo et al., 2018), pragmatics (Jørgensen, 2013), cross-

language analysis (Corpas Pastor, 2021), and sociolinguistics (Szmrecsanyi, 2017). Largely

benefitting from advances in computational technology, data transfer, and data storage,

corpus linguistics represents a promising and viable tool for linguistic analysis.

   Just as there are benefits in using corpora, there are points of potential concern that

must be addressed. While a corpus may seek to collect language samples that are

representative of real language use, not all corpora are broad enough as to be generalizable to

an entire speech population. Corpora may be general or specialized based on register or

genre; they may be synchronic or diachronic; or, they may aim to collect instance of language

as a lingua franca or be centered on language learner output (Hunston, 2002). Simply put,

corpora contents differ based on the needs of the researchers who compile them. That said,

the selection or compilation of corpora for linguistic analysis should consider the design and

contents of the corpora in question. Many corpora available for free access online contain

detailed descriptions of the corpora content and may even provide the methodology used in compilation, aiding in the linguist's assessment of its validity.

Another cited concern regarding the use of corpora, particularly in sociolinguistic or pragmatic research, is the limited scope of information provided: namely, the lack of paralinguistic information that accompanies oral speech acts (Baker, 2010). However, more and more frequently, linguistic corpora are compiled with this concern in mind and contain multi-level tagging (i.e., phonetic, semantic, and pragmatic tags), metadata, and video and audio files in addition to written transcriptions (Anderson, 2010). Since the popularization of corpus linguistics in the 1990s, important technological advances have allowed for more advanced and detailed corpus annotation. This, in turn, has alleviated concerns about the scope of information available for analysis and allowed for successful pragmatic research based in corpus analysis (García, 2007).

A final concern to be addressed relates to the quantitative nature inherent in corpus analysis. Corpus analysis lends itself to large-scale data analysis, which some have argued denies the ability for detailed analysis of individual features or loss of notice of nuance (Enrique-Arias, 2016). Some fields, such as sociolinguistics, have been reticent to adopt these quantitative methodologies in part because of a traditional preference for qualitative analysis (Baker, 2010). While it is true that multivariate quantitative analyses are useful in interpreting data obtained from corpora, corpus analysis does not reject qualitative analysis. Indeed, as Biber (1988, 1992) stresses in his seminal work on language variation across registers in English, corpus analysis allows for qualitative and quantitative analysis—or macroscopic and microscopic analyses (Biber 1988, pp. 61-63)—to work in a complementary fashion. Given that collocation is a primary analytical tool in corpus analysis, researchers have the opportunity to select particular cases of the phenomenon under study and analyze its specific use or application. A balanced approach that incorporates both qualitative and quantitative

analysis allows for identification of larger trends or patterns unavailable to smaller case

studies and, at the same time, provides opportunity for detailed analysis of specific linguistic

features.

Corpus-based analysis presents no serious problems to the linguist in terms of

inappropriate or lacking methodology. Rather, it requires the same concern for methodology,

consideration of research questions and objectives, and appropriate analytical knowledge

called for with any research tool or framework.

The current study seeks to expand on existing research on R-impersonals in Spanish,

augmenting existing data by examining the phenomenon on a larger scale. Basing the current

study in corpus research will allow for a more detailed study of R-impersonal usage than

found in smaller case studies. In this way, the study examines if the high or low frequencies

of specific forms are borne out in the larger trends of language use, helping to identify

contemporary use of reference impersonals in spoken Spanish. Additionally, the use of a

corpus based in oral data from the United States represents an innovative effort to expand

existing literature on R-impersonal expressions in Spanish to an understudied variety. This

objective also speaks to the aim of widening the use of corpus analysis in Spanish linguistics

in general; while corpora studies in Spanish do exist, they are widely outnumbered by studies

centered on English and other European languages (Parodi, 2007). Given Spanish's

widespread use in the United States and worldwide, it is important to show how corpus-

driven methodologies are useful and pertinent tools in the field.

3.2 Study corpus

3.2.1 Spanish in Texas (SIT)

The present investigation analyzes R-impersonal forms present in the Spanish in

Texas Corpus (SIT). A description of this corpus is provided below in Table 6:

Table 6: *Spanish in Texas Corpus Information*

Corpus	Total Number of Speakers (Male / Female)	Word Count (full corpus)
Spanish in Texas	96 (36 / 60)	507,544

The Spanish in Texas Corpus (SIT), developed through the Center for Open Educational Resources and Language Learning at the University of Texas at Austin, contains data from 96 bilingual English/Spanish speakers living in Texas (Toribio & Bullock, 2012). The corpus contains video files of sociolinguistic interviews conducted with participants and corresponding audio files, transcripts, and part-of-speech (POS) annotations. Data for the corpus was collected between 2011 and 2012 in various Texas cities, and participants were between the ages of 18 and 86 at the time of interview with a median age of 41. Mexico was the most frequent country of origin for the participants—79 of the 96 participants listed Mexico as their own birth country, at least one parent's birth country, or at least one parent's country of origin. Other Spanish-speaking countries of origin include Argentina, Colombia, El Salvador, Peru, Uruguay, and Venezuela. One participant reported that their mother is of Korean descent and their father from the United States and of Mexican heritage. Upon further review of demographic information, it was determined that this participant was not a heritage speaker of Spanish but rather a proficient second-language Spanish speaker.

Of the 96 interviews in SIT, 94 were included for study in the current investigation. One interview was excluded due to the interviewee being a proficient second-language learner of Spanish and not a heritage speaker. The second interview was excluded because the recorded data was one continuous monologue and not a turn-taking sociolinguistic interview. For the sake of consistency in genre, only sociolinguistic interview data was included in the current study. The final data set included for study included interviews from 59 women and 35 men.

Data on the highest level of education (achieved or in progress) was available as part of demographic data with the SIT database; this information, along with generation information[5], is shown in Table 7.

Table 7: *Educational Attainment and Generational Distribution, Spanish in Texas Corpus*

		Total (%)	Total (n)
Generation	**1**		
Elementary		14%	5
High School / GED		30%	11
Associates / Technical		5%	2
Bachelors		49%	18
Postgraduate (MA, JD, MD, etc.)		3%	1
Total		**100%**	**37**
	2		
Elementary		0%	0
High School / GED		15%	7
Associates / Technical		4%	2
Bachelors		71%	34
Postgraduate (MA, JD, MD, etc.)		10%	5
Total		**1**	**48**
	3		
Elementary		0%	0
High School / GED		33%	2
Associates / Technical		0%	0
Bachelors		67%	4
Postgraduate (MA, JD, MD, etc.)		0%	0
Total		**100%**	**6**

---

[5] To define the parameters for each generation in this study, the structure proposed by Silva-Corvalán (1994) was used: individuals in the first generation were those that were born in a Spanish-speaking country (in this case, Mexico or El Salvador) and moved to the United States after eleven years of age. Individuals were considered second generation if they were born in the United States (and were children of first-generation parents) or if they moved to the United States from a Spanish-speaking country before turning eleven. Finally, third-generation speakers are classified as those born in the United States to at least one second-generation parent. Classifications of generation provided by the corpora compilers differed in some instances from the generation distinctions made based on Silva Corvalán's classification.

Table 7 (continued)

	4+	
Elementary	0%	0
High School / GED	0%	0
Associates / Technical	0%	0
Bachelors	100%	3
Postgraduate (MA, JD, MD, etc.)	0%	0
Total	**100%**	**3**

The SIT data included in this study includes data from 37 first generation speakers, 48 second generation speakers, 6 third generation speakers, and 3 speakers who identified as fourth or subsequent generation speakers. The majority of speakers held or were working on at least a bachelor's degree in each generation breakdown. Note that the only participants whose education threshold was elementary school were first generation speakers who immigrated to the United States after childhood.

3.2.2 Representativity of the Sample

Data for the current investigation has been compiled from sociolinguistic interviews in the SIT corpus. A total of 94 interviews will be analyzed for the distribution and use of R-impersonals. The SIT corpus was selected for inclusion based on its abundant availability of oral speaker data in the same genre. Though conducted by different researchers, the data was compiled in a similar format during sociolinguistic interviews, such that similar patterns of language and thematic content may be expected. Data collected from a prior pilot study on R-impersonals adopted this same approach by analyzing a collection of sociolinguistic interviews conducted by different researchers in the same Hispanic Studies department at a U.S. university. Although the interviews used in the pilot study were conducted by different researchers, the same general sociolinguistic interview pattern was followed and the resulting interviews were similar enough for comparison and analysis. Thus, it was determined that

analyzing sociolinguistic interviews from SIT collected by different researchers would not cause undue issues with data sample representativity.

Since the objectives of this study are to analyze the frequency and distribution of a particular linguistic feature in a semi-spontaneous setting, a data set with only one genre is appropriate. Additionally, the data included in this investigation was collected in a community in which Spanish as a minority language thrives; Texas represents one of the geographic areas of the United States that contained a population of 1 million or more Hispanic residents as of 2019 (Alonzo, 2020). While other states, such as California and Florida, also contain high populations of Spanish-speaking Hispanic/Latino populations, limitations of relevant data availability prevented inclusion of linguistic data from these areas. In the future, comprehensive corpus collection and analysis of oral Spanish from all states with Spanish-speaking populations should be addressed; the scope of the current investigation is limited to oral Spanish sociolinguistic corpora from Texas.

With 37 first generation and 48 second generation speakers, these two speaker groups are relatively balanced in their representation in the data set. Representation of third generation speakers is lesser, with six speakers. Only three speakers came from the fourth or subsequent speaker generations. Given that some trends in heritage language maintenance and use indicate that heritage language use decreases by the third generation (see Fishman (1964) or Solé (1990) for more details regarding transitional bilingualism in minority language communities), it is not entirely unexpected to see a lower representation of fourth, fifth, or sixth generation heritage speakers in this data compared to their first or second generation counterparts. Nevertheless, the speaker data for the 4+ generation speakers will be included in analysis.

In terms of educational level, the majority of the participants had received or were actively pursuing a bachelor's degree (63%). Furthermore, 21% of participants earned a high

school diploma or GED, 4% a technical or associates degree, and 6% a postgraduate degree, suggesting that the overwhelming majority of speakers possessed formal education. Only 5% of the participants had only completed all or some elementary school; these five participants were all first-generation speakers. In general, then, the speaking population represented in the current study data is an educated group consisting primarily of individuals with at least a high school and college education. This representation is in line with overall increased education rates for Hispanic/Latino students at all levels of education in the past decade as reported by the U.S. Census Bureau (Bauman, 2017); therefore, we may say that the data included in the current study presents a representative profile of oral Spanish spoken in the United States in terms of educational level of its speakers.

The majority of the speakers identified Mexico as their country of origin, either as their birth country or the birth country of their parents or grandparents. While there were some participants from Spanish-speaking countries in Central and South America, these speaking populations were less represented than their Mexican counterparts, as represented in Table 8. Finally, distribution of speakers by gender is 59 females and 38 males, a 61% to 39% split.

Table 8: *Speaker Country of Origin*

Country	Participants (#)	Participants (%)
Mexico	81	86.17%
Venezuela	4	4.26%
Colombia	3	3.19%
El Salvador	2	2.13%
Argentina	1	1.06%
Spain	1	1.06%
Peru	1	1.06%
Uruguay	1	1.06%
Total	**94**	**100%[6]**

---

[6] Participant percentage may not sum to 100 due to rounding.

Of the 94 interviews comprising the data set, 81, or about 86%, come from speakers of Mexican descent. Meanwhile, the total number of speakers from all other Spanish-speaking countries represented in the data set only comprises about 13% of the data (13 speakers). Due to homogeneity of the data set in terms of country of origin, this feature will not be considered as a sociolinguistic factor for analysis. Interviews were also divided into age ranges based on the age of the interviewee at the time the interview was conducted. The three categories were young adult (18-35), adult (36-54), and older adult (55+). The complete information for age ranges is available in Table 9:

Table 9: *Speakers Classified by Age Range*

Age Range (Yrs)	Speakers(#)	Speakers (%)
18-35	40	42.5%
36-54	39	41.5%
55+	15	16.0%
Total	**94**	**100%**

The distribution of young and middle-aged adults is comparable, with speakers between the ages of 18 and 35 comprising 42.5% of the sample (40 of 94) and speakers between the ages of 36 and 54 accounting for 41.5% (39 of 94). Finally, a total of 15 adults aged 55 or more comprised 16% of the study data.

In a pilot study of the frequency and use of Spanish R-impersonals conducted prior to the current investigation, only speaker generation and gender were considered as sociolinguistic variables. However, it was noticed that speaker generation is not necessarily indicative of a speaker's age but rather their relative time of personal and/or familial residence in the United States. Failing to consider age as a separate factor could obscure patterns of language use (Silva-Corvalán & Enrique-Arias, 2017, pp. 50-51). In order to examine the distribution and usage of R-impersonals in a more complete, transparent fashion, speaker's age was added alongside generational status in the current investigation.

Educational status of the speakers is an additional variable absent in the pilot study that will

be included in the current investigation. The justification for considering a speaker's

educational status in relation to their linguistic patterns comes from Silva-Corvalán and

Enrique-Arias (2017), who indicate that among sociolinguistic studies of immigrants and

bilingual populations:

> Los hijos y nietos de estos inmigrantes…tienen un nivel alto de educación (doce o
>
> más años de escolaridad) impartida en inglés (la lengua mayoritaria), pero han
>
> adquirido el español de manera informal y a menudo no leen ni escriben en esta
>
> lengua. Esto significa que, de todos los parámetros que definen nivel socioeconómico,
>
> el nivel educacional es de gran importancia (p. 50).

In order to better capture nuances present in bilingual and heritage Spanish-speaking

communities in the United States, considering educational level is valid. By including

educational status as an independent variable, the current investigation can assess whether

there is a relationship between educational attainment and use of R-impersonals in the

heritage language. In general, then, preliminary data from the pilot study highlighted the

importance of considering all pertinent sociolinguistic variables in an analysis of

impersonality.

Data from the Spanish in Texas Corpus is included for analysis in the current study.

Sociolinguistic interviews of bilingual English-Spanish speakers ranging from first to fifth

generation will be analyzed for the frequency and distribution of R-impersonal forms. The

data from the SIT corpus constitutes an adequately sized sample of oral linguistic data for

qualitative and quantitative analysis. In the following sections, the methodology for data

coding and analysis will be discussed.

3.3 Data Coding and Analysis

3.3.1 Variables

The current study utilizes frequency and distribution data in a linguistic corpus to analyze the use of R-impersonal forms in spoken Spanish. The dependent variable will be the type of impersonal form used; specifically, the study will examine the use of impersonal *se*, indefinite *uno*, the 2SG form, 1PL form, and 3PL form. In the case of 2SG, 1PL, and 3PL, the presence or absence of an overt subject pronoun will also be examined. This dependent variable is categorical and is measured at the nominal form.

The independent variables are divided among linguistic and sociolinguistic considerations. Linguistic variables that will be considered are impersonality or genericity triggers, including 1) verb tense and 2) verb mood; 3) presence of adverbs of time, 4) presence of an overt pronoun and 5) presence of hypothetical *if* clauses (cláusulas *si*). As mentioned previously, studies by Lamas (2015) and Serrano (2013) found that present tense, indicative mood, and active voice favored the use of impersonal forms. The Lamas (2015) study, which analyzed written text about judicial matters in newspapers, focused on nominal traits and co-occurrence of infinitival forms. Given the genre and modality distinction between that study and the present one, it is possible that nominal abstraction and infinitival forms may not be present in the oral study data. The sociolinguistic variables that will be considered are 6) speaker gender 7) speaker generation 8) speaker age range and 9) speaker educational level. Given that the majority of the participants included in the corpora are from Mexican descent, there is not enough dialectal diversity present to consider country of origin or dialect as an influencing factor. Additionally, it should be noted that considering speaker generation and speaker age as separate sociolinguistic variables should not be considered redundant; speaker generation refers to the speaker's familial and residential history in the United States and can indicate, generally, the degree of bilingualism, whereas age range indicates the overall life period of an individual.

As with the dependent variable, all nine independent variables are categorical. Verb type, tense/mood, the presence of adverbs of time, overt pronouns, and *if* clauses, as well as speaker gender, are nominal variables. Speaker generation, speaker age range, and speaker educational level are measured at the ordinal level. Given that the dependent variable is categorical and not continuous, non-parametric statistical tests are appropriate for quantitative analysis (Abu-Bader, 2016). The following sub-sections describe the programs used for data analysis, how the corpus data was cleaned for accuracy, and the different types of qualitative and quantitative analysis that will be conducted.

Each impersonal form that was analyzed through WordSmith Tools represents a sample, and each sample was coded for analysis based on classification of its variables. Looking at the dependent variable of impersonal type, the following coding was used:

Table 10*: Data Coding for Dependent Variables*

Impersonal Form	Code
Impersonal *se*	1
Uno	2
2SG	3
1PL	4
3PL	5

The full coding information for the samples, including codes used for each independent variable, is included in the Appendix A.

3.3.2 Programs used for analysis

Various tools were used to collect, clean, and analyze the data. Access to the SIT was freely available online. A simple web scraper was designed to select and download interviews from the SIT website. Python and its Natural Language Toolkit (NLTK) were utilized for cleaning the interview data and removing interviewer dialogue (Bird et al., 2009). The TagAnt corpus linguistics software was used to assign Spanish-language tags to the corpus data, and these tags were further structured and organized with a second Python function. WordSmith Tools, a software piece designed for corpus linguistics, was used to analyze the corpus data selected for study. Python programs to identify the verb forms accompany

impersonal *se* and indefinite uno, as well as the presence of adverbs of time, conditional "if,"

and overt programs were also utilized at this stage. Finally, SPSS, a statistical software

package for social science research, was used for statistical analysis. The Python codes used

for data collection and cleaning are available in Appendix B.

3.3.3 Mutual Information Score & T-Score

The corpus analysis for the study data and much of the descriptive analysis took place

in WordSmith Tools. WordSmith Tools possesses a concordance program, which allows the

user to search for contextualized uses of a specified form. In this case, the different

impersonal forms were loaded separately into the concordance tool so that their uses could be

quantified and observed. In addition to the potential for returning each use of a specified

form, it is possible to return a sampling of the form; this may be useful when a form is highly

productive. As Hunston (2002) indicates in her overview of corpus analysis, concordance

features are useful for allowing researchers to examine the contextual use of a given form,

which may differ from the researcher's intuition.

Additionally, the concordance tool was also used to look for collocates, or series of

frequently co-occurring words. The collocation function will serve to identify the potential

effect of linguistic variables like genericity triggers or verb types on the type of impersonal

form used. When collocates with impersonal expressions were identified, their Mutual

Information scores and T-scores were calculated. The Mutual Information score is a standard

measure utilized in corpus analysis that calculates the strength of the collocation, or the given

likelihood of the series of words appearing together in a given span of words (Biber et al.,

1998; Hunston, 2002). Mutual Information scores range from 0 upward, with a score of 0

indicating that there is no relationship between the words' co-occurrences. The farther a score

is from 0, the stronger the relationship between the words that occur in the collocate and the

likelihood that their co-occurrence is not chance. Note that the Mutual Information score is

influenced by the overall frequency of a word in a collocate set—for instance, a highly

productive word like *and* would lend a low Mutual Information Score to a collocate in which

it forms a part since it is also likely to appear without its collocate pairs in other contexts. For

this reason, it also useful to consider a collocation's T-score. The T-score shows "a statistical

measure (in standard deviations) of words that are more likely to appear as collocates of one

word rather than of another word" (Biber et al., 1998: p. 267). This measure can be useful to

show if different words are in complementary distribution with each other, based on the T-

score of their collocate pairs. For example, Biber et al. (1998) illustrates how T-scores for the

words *big* and *large* in the Longman-Lancaster Corpus demonstrate the distinct uses for each

term:

Table 11: *T-scores for Big and Large (from Biber et al. (1998), p. 267)*

Big	
Collocate Pair	**T-score**
Man	5.06
Deal	4.17
Toe	3.93
Boy	3.24
Dog	3.13
House	3.06
Number	-11.38
Numbers	-9.35
Proportion	-5.96
Amounts	-5.43
Quantities	-4.74
Amount	-4.64

Words with a positive T-score are those that are likely to appear alongside *big* in a

collocate pair, and the larger the number, the stronger the likelihood. Thus, words such as

man and *deal* were highly likely to form a collocate pair with *big* as compared to the

likelihood that they would appear with the counterpart word *large*. Negative t-scores, on the

other hand, signify that the given term is unlikely to appear as a collocate word with *big*; the

larger the negative score, the less likely.

By looking at T-scores for collocate pairs, we can examine whether or not forms are in complementary distribution. In the context of the current investigation, it will be useful to see which, if any, linguistic factors favor or disfavor collocation with impersonal forms.

Using the concordance and collocate tools available in WordSmith Tools allows an observation of the contextualized use of the distinct impersonal forms analyzed. This descriptive feature is complemented by the consideration of the Mutual Information and T-scores, two measures that explain a collocation's strength in the corpus data.

3.3.5 Qualitative Analysis of Data

The pragmatic function of the R-impersonal forms was analyzed. Using contextual information and discourse analysis, the role of the 2SG, 1PL, and 3PL R-impersonals for generalization, inclusive defocalization, or concealment of the speaker was determined. Patterns of usage across discourse topic and examples that highlight the aforementioned pragmatic functions were explored.

3.3.6 Quantitative Analysis of Data

A quantitative analysis of the corpus data was conducted to examine the potential relationships between the dependent variable of R-impersonal expression and the nine independent variables previously described. As previously mentioned, the dependent variable is categorical in nature, as are all nine independent variables. Given the categorical nature of the dependent variable, non-parametric statistical tests are appropriate (Abu-Bader, 2016, p. 284). In this case, several chi-square tests of association will be conducted. The chi-square test of association examines the potential relationship between the dependent variable and one or more independent variables, looking to see if the observed frequency of the independent variable on the dependent is significantly different than its expected frequency in the general population (Baker, 2010, p. 38).

First, contingency tables for the levels of the dependent and independent variables will be compiled. Next, the expected frequencies for each intersection of dependent and

independent variable will be calculated based on the following formula, taken from Abu-

Bader (2010):

$$E_i = \frac{R_i * C_i}{N} \qquad (1)$$

In Equation (1), E_i representes the expected frequency for the ith cell; R_i equals the

row margin for the ith cell; C_i equals the column margin for the ith cell, and N represents the

sample size.

Assuming that the contingency tables reveal a difference between the observed and

expected frequencies, the chi-square test of association will be conducted against a null

hypothesis. The assumption for sample size will be evaluated. As long as the assumption is

met, the chi-square value and p value will be analyzed. In order to control for the non-chance

outcome of a potential relationship, a p value of $p \leq 0.05$ will be used. The Cramer's V and

Cramer's V^2 coefficients will also be assessed to consider the strength of the potential

relationship between the dependent and independent variables. Squaring the Cramer's V

coefficient will provide the variance in the dependent variable that is accounted for by the

independent variable—this can indicate a weak, moderate, or strong relationship, based on its

value. Table 12, adapted from Abu-Bader (2016), shows the various interpretations for

Cramer's V coefficients:

Table 12: *Correlation Coefficients (from Abu-Bader (2016, p. 165)):*

Interpretation	$|r|$ (abs value of r)	r^2
Very strong	$r \geq 0.91$	$r^2 \geq 0.2$
Strong	$0.71 \leq r \leq 0.90$	$0.50 \leq r^2 \leq 0.81$
Moderate	$0.51 \leq r \leq 0.70$	$0.26 \leq r^2 \leq 0.49$
Weak	$0.31 \leq r \leq 0.50$	$0.10 \leq r^2 \leq 0.25$
Very weak	$r \leq 0.30$	$r^2 \leq 0.09$

The results of the chi-square tests of association will determine whether or not the null

hypotheses are accepted or rejected. Given that the variables under consideration in the

current investigation are non-continuous in nature, the chi-square test is the most appropriate.

This non-parametric statistical test will explore whether or not there is a relationship between the dependent variable of R-impersonal expression and the independent variables of verb type, tense/aspect/mood, adverbs of time, hypothetical *if* clauses, speaker gender, speaker age range, speaker generation, and speaker educational level.

3.4 Conclusion

Data collected from the SIT, which encompasses oral sociolinguistic interview data from U.S. Spanish speakers, will be analyzed for the distribution, frequency, and pragmatic usage of R-impersonal forms. Through the application of qualitative and quantitative analysis, as well as the use of word frequency scores developed specifically for corpus analysis, the current study seeks to explore the use of R-impersonals in spoken Spanish and draw findings that will inform contemporary language pedagogy. Results of the analysis and a discussion of salient findings are presented in the next chapter.

Chapter 4: Results and Analysis

The corpus data was analyzed, and R-impersonal forms expressed through impersonal *se*, indefinite *uno,* 2SG, 1PL, and 3PL forms were extracted. This chapter provides the results of the linguistic analysis of these forms, including details about their frequency, distribution, and pragmatic use.

Regarding the identification of impersonal *se* forms, it should be noted that reflexive verbs were excluded from analysis. In addition, forms of *se* with conjugations of the verbs *llamar* (to call) and *decir* (to say) were also excluded. In phrases such as "que se llama / ¿Cómo se llama?" and "¿Cómo se dice?," the *se* morphemes are associated semantically in a very close manner, and the forms are therefore considered lexicalized. By excluding these forms, the analysis is better able to assess the frequency and use of impersonal *se* in non-lexicalized contexts. are highly common, lexicalized phrases in Spanish[7].

4.1 Quantitative Analysis

A total of 4,989 R-impersonal forms were identified in the SIT data. For the uses of impersonal *se* and *uno*, these forms were only counted if they appeared alongside a conjugated verb at the clausal level. The distribution of the impersonal forms by type are displayed below in Table 13:

Table 13: *Distribution of R-impersonals: Spanish in Texas*

Form	Cases (#)	Cases (%)
2SG	1,614	32.35
1PL	1,145	22.95
Se	930	18.64
3PL	960	19.24
Uno	340	6.81
Total	**4,989**	**100.00**

The 2SG form appeared most frequently in corpus data, constituting just over one-third (32.35%) of the total impersonal cases registered. Following 2SG, 1PL was the most

---

[7] Removal of uses of lexicalized *se* with *decir* and *llamar* was context-dependent; in some cases, the forms were maintained when used in a non-lexicalized impersonal manner. For example, an utterance such as "Aha/ ¿Así **se dice** , canjil _FS?" was discarded, whereas "Nunca pensé que **se dijeran** las barbaridades que **se dicen**" was kept.

frequent form, constituting 22.95% of sample data. Use of the 3PL and impersonal *se* were

comparable at 19.24% and 18.64%, respectively. Finally, impersonal *uno* was the least-used

R-impersonal, comprising only 6.81% of sample data.

4.1.1 Sociodemographic Variables

4.1.1.1 Gender. When analyzing the distribution of impersonal forms by gender, we

see that 1,793 of the 4,989 forms (35.94%) were made by men, while the remaining 3,196

(64.06%) were made by women. It is important to note that the distribution of participants by

gender was not equal in the SIT corpus sample; the sample consisted of data from 35 males

and 59 females. As such, it is expected that there were more impersonal expressions produced

by female speakers based on raw production. Table 14 shows the distribution of impersonal

forms by gender:

Table 14: *Distribution of R-impersonals in SIT by Gender*

Form	Male expressions (raw #)	Male expressions (%)	Female expressions (raw #)	Female expressions (%)
Se	381	21.25	549	17.18
Uno	161	8.98	179	5.60
2SG	604	33.69	1,010	31.60
IPL	308	17.18	837	26.19
3PL	339	18.91	621	19.43
Total	**1,793**	**100.00**	**3,196**	**100.00**

The 2SG was the most popular impersonal form for both males and females; these

forms constituted at least a third of all impersonal forms for each gender. However, males

preferred impersonal *se* as the second most popular choice, whereas the second most-used

form for females was the 1PL impersonal form. 3PL forms ranked as the third most popular

choice for both genders, followed by 1PL for males and impersonal *se* for females. Finally,

impersonal *uno* ranked as the least-used R-impersonal for both genders.

A chi-square test of association was used to explore the potential association between

impersonal form and gender. There was no violation of the assumption of sample size noted

from reviewing the 2 x 5 contingency table. Furthermore, the results of the chi-square test

show a statistically significant association between form and gender ($\chi^2_{(df=4)}$= 71.799, p <

0.01). The results of the chi-square test of association are reported in Table 15:

Table 15: *Chi-square Test of Association: Form x Gender*

Form	Male		Female		Total		χ^2	p^a
	n	% of total	n	% of total	n	% of total		
Se	381[b]	7.64	549[e]	11.00	930	18.64	71.799	<.01
Uno	161[c]	3.23	179[f]	3.59	340	6.81		
2SG	604	12.11	1010	20.24	1614	32.35		
1PL	308[d]	6.17	837[g]	16.78	1145	22.95		
3PL	339	6.79	621	12.45	960	19.24		
Total	1,793	35.94	3,196	64.06	4,989	100.00		

[a] Two-tailed alpha
[b] Adjusted residual = 3.5
[c] Adjusted residual = 4.5
[d] Adjusted residual = -7.3
[e] Adjusted residual = -3.5
[f] Adjusted residual = -4.5
[g] Adjusted residual = 7.3

Positive adjusted residuals indicate that a given case is overrepresented in the sample,

whereas negative adjusted residuals indicate underrepresentation. Following the standard that

an adjusted residual either greater than 1.96 or smaller than -1.96 bears significance (Abu-

Bader, 2016, p. 306), we may note a statistically significant difference in the use of

impersonal forms by gender for impersonal *se, uno,* and 1PL for both males and females. In

the case of 1PL impersonals, there is a positive association between the form's use and

gender, with women more likely to use the form. On the other hand, there was a negative

association among women with the use of impersonal *se* and *uno*, with an adjusted residual of

-3.5 for *se* and -4.5 for *uno*. Accordingly, men showed a statistically significant positive

association with impersonal *se* and *uno* (Adjusted residual for *se* = 3.5 and *uno* = 4.5). In

other words, women showed a greater tendency toward expressing impersonality through a

1PL collective lens, whereas men trended toward the use of singular, more generalized forms.

These results coincide with similar findings by Posio (2016), whose analysis of Peninsular

Spanish revealed a similar preference for 1PL in females. The author notes that female

speakers may feel more inclined to express impersonality through a collective means or

engage in more subjective conversation, while males rely upon third-person anaphoric impersonality and engage less in Experiencer-role dialogue (p. 13). Though the difference in use of forms by gender was also confirmed through a Cramer's V value of 0.120 ($p < .01$), gender accounted for a relatively low percentage of variance overall, only around 1.4% (Cramer's $V^2 = .0144$), thus indicating a very weak relationship.

4.1.1.2 Generation. Generation, age range, and educational level were also considered as independent variables factoring into the usage of R-impersonals. There was an uneven distribution of speakers by generation, and this impacted the raw frequencies of impersonal forms accordingly. Figure 7 show the distribution of forms by generation:

Figure 7 *Distribution of Impersonals by Generation Impersonals by Generation (%)*
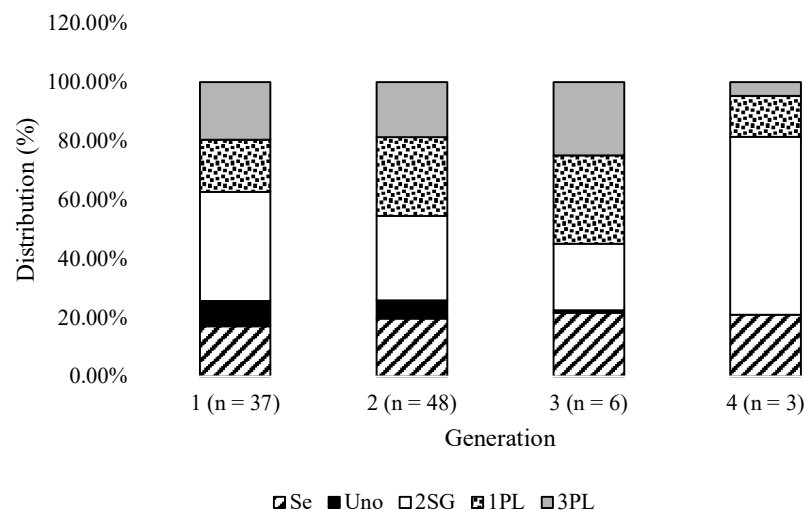
First generation speakers, numbering 37 in total, produced 2,177 R-impersonal forms. Of these, 2SG were the most numerous, constituting roughly 37.12% of all utterances (808/2,177). 3PL was the second most popular form, representing 19.47% of utterances (424/2,177), followed closely by 1PL (17.82%, or 388/2,177) and impersonal *se* (17.04%, or 371/2,177). Impersonal *uno* has the smallest representation, with its 186 cases representing roughly 8.54% of first-generation R-impersonals.

In the second generation, which included data from 48 speakers, 2SG is also the most popular form, representing 28.77% of uses (717/2,492). However, 1PL forms are the second

most popular form, and their usage is close to that of 2SG; 1PL forms constituted 26.81% of

the second-generation R-impersonal uses (668/2,492). Impersonal *se*, ranking as the third

most popular form, represents about 19.66% of R-impersonals (490/2,492). 3PL forms

constituted 18.66% of the sample, and impersonal *uno* only 6.10% of forms. As can be noted,

the difference in distribution of forms is more pronounced in first generation speakers—2SG

represents the most popular form (37.12%), and the next most-used form, 3PL, has a

frequency of only 19.48%. On the other hand, R-impersonal usage in the second generation is

more evenly spread, with four of the five forms' frequency falling between 18 and 28

percent.

The samples of third and fourth generation speakers were significantly smaller than those

for first and second-generation speakers—data from the third generation came from 6

speakers and there were only 3 speakers in the fourth generation. The third-generation

speakers produced 277 R-impersonal forms. Of these, the 1PL forms were the most popular

(29.96%), followed by 3PL (24.91%), 2SG (22.74%), and impersonal *se* (21.66%). There

were only 2 uses of impersonal *uno* (0.72%) in the third-generation data. In the data collected

from fourth generation speakers, 2SG forms showed an overwhelmingly majority in

popularity (60.47%), followed *by se* (20.93%), 1PL (13.95%), and finally 3PL (4.65%).

There were no registered uses of impersonal *uno* in the fourth generation.

It is interesting to note similar patterns in the popularity of the various R-impersonal

forms in all generations save the third, in which case 1PL ranked as the most popular choice.

The 2SG form, which was most used in the first, second, and fourth generations, ranked as

the second-to-last most used form for third generation speakers. In a chi-square test of

association analyzing the potential relationship between generation and R-impersonal form,

significant variance between expected and actual forms by generation was detected. Table 16

shows the results of the 4 x 5 contingency table:

Table 16: *Chi-square Test of Association: Form x Generation*

Form	Generation										χ^2	p^a
	1		2		3		4		Total			
	n	% of total	n	% of total	n	% of total	n	% of total	n	%	137.265	<0.01
Se	371[b]	7.44	490	9.82	60	1.20	9	0.18	930	18.64		
Impersonal *uno*	186[c]	3.73	152[d]	3.05	2[e]	0.04	0	0.00	340	6.81		
2SG	808[f]	16.20	717[g]	14.37	63[h]	1.26	26[i]	0.52	1614	32.35		
1PL	388[j]	7.78	668[k]	13.39	83[l]	1.66	6	0.12	1145	22.95		
3PL	424	8.50	465	9.32	69[m]	1.38	2[n]	0.04	960	19.24		
Total	2,177	43.64	2,492	49.95	277	5.55	43	0.86	4,989	100.00		

[a] Two-tailed alpha
[b] Adjusted residual = -2.6
[c] Adjusted residual = 4.3
[d] Adjusted residual = -2.0
[e] Adjusted residual = -4.1
[f] Adjusted residual = 6.3
[g] Adjusted residual = -5.4
[h] Adjusted residual = -3.5
[i] Adjusted residual = 4.0
[j] Adjusted residual = -7.6
[k] Adjusted residual = 6.5
[l] Adjusted residual = 2.9
[m] Adjusted residual = 2.5
[n] Adjusted residual = -2.4

The chi-square test of association confirmed a statistically significant variance in the use of forms based on generation ($\chi^2_{(df=12)}$= 137.265, p < 0.01). In addition, multiple instances of statistically significant variance were detected within generations based on the R-impersonal form in question. When looking at the use of *se* in the first generation, we note an adjusted residual of -2.6. This indicates a statistically significant negative relationship between use of *se* and first-generation speakers compared to other generational groups. For the impersonal *uno* form, there is a statistically significant positive association with first-generation speakers (Adjusted residual = 4.3) and a statistically significant negative association with second and third generation populations (Adjusted residuals of -2.0 and -4.1, respectively).

Use of the 2SG form is statistically significant for all generations, showing a positive association in first and fourth generations. This positive association was stronger in the first generation (Adjusted residual = 6.3) than in the fourth (Adjusted residual = 4.0). A statistically significant negative association was found between use of 2SG in second and third generations (Adjusted residuals of -5.4 for second generation and -3.5 for third generation).

For the 1PL form, there is a statistically significant negative association with first generation speakers (Adjusted residual = -7.6) and positive associations with second-generation (Adjusted residual = 6.5) and third generation (Adjusted residual = 2.9) speakers. Finally, for the 3PL form, there is a statistically significant positive association between its use in the third generation (Adjusted residual = 2.5), yet a negative association in the fourth generation (Adjusted residual = -2.4). No other statistically significant relationships between form and generation were detected through the chi-square analysis. The relationship between form and generation is very weak, as indicated by a Cramer's V^2 value of 0.009.

4.1.1.3 Age and Age Range. When submitted to a chi-square test of association, age range was also identified as a salient variable. A statistically significant variance between

REFERENCE IMPERSONALS IN TX SPANISH

expected and actual counts was identified ($\chi^2_{(df=8)}$ = 122.777, p < 0.01). Table 17 displays the

contingency table with adjusted residuals for the three age ranges:

Table 17: *Chi-square Test of Association: Form x Age Range*

Age Range	18-34		35-54		55+		Total		χ^2	p^a
Form	n	% of total	n	% of total	n	% of total	n	%	122.777	<0.01
Se	**403[b]**	8.08	**403 [c]**	8.08	**124[d]**	2.49	930	18.64		
Impersonal *uno*	**90[e]**	1.80	**193[f]**	3.87	57	1.14	340	6.81		
2SG	**688[g]**	13.79	762	15.27	**164[h]**	3.29	1614	32.35		
1PL	**401[i]**	8.04	**493[j]**	9.88	**251[k]**	5.03	1145	22.95		
3PL	**312[l]**	6.25	464	9.30	**184[m]**	3.69	960	18.58		
Total	1,894	37.96	2,315	46.40	780	15.63	4,989	100.00		

[a] Two-tailed alpha
[b] Adjusted residual = 3.7
[c] Adjusted residual = -2.1
[d] Adjusted residual = -2.1
[e] Adjusted residual = -4.5
[f] Adjusted residual = 4.0
[g] Adjusted residual = 4.7
[h] Adjusted residual = -7.4
[i] Adjusted residual = -2.3
[j] Adjusted residual = -2.6
[k] Adjusted residual = 6.7
[l] Adjusted residual = -3.9
[m] Adjusted residual = 3.4

The three age ranges included in the present analysis span distinct periods of adulthood: 18-35 (young adult) (40/94), 36-54 (mid adult) (39/94), and 55+ (older adult) (15/94). The chi-square test of association found that there was a statistically significant difference in the use of impersonal forms based on age range.

Looking at the use of *se*, the young adult group showed a statistically significant positive relationship (Adjusted residual = 3.7), while the mid and older adult groups showed a statistically significant negative relationship (Adjusted residuals of -2.1 for both age ranges). On the other hand, young adults were less likely to use impersonal *uno* (Adjusted residual = -4.5). There was a statistically significant positive relationship between use of impersonal *uno* and middle-aged adults (Adjusted residual = 4.0); this was the only age range for which a statistically significant positive association was formed. Use of 2SG had a statistically significant positive relationship with young adults (Adjusted residual = 4.7) and a negative relationship with adults aged 55+ (Adjusted residual = -7.4). A statistically significant negative relationship was also found with the use of 1PL forms and the young and middle-aged adults (Adjusted residuals = -2.3 and -2.6, respectively), while a positive association is shown in the older adult group (Adjusted residual = 6.7). Finally, young adults showed a statistically significant negative relationship with 3PL usage (Adjusted residual = -3.9), whereas older adults were more likely to use this form (Adjusted residual = 3.4).

The large number of statistically significant variances between R-impersonal and generation and age range show emerging speech profile patterns for these groups. In terms of generation, first generation speakers differ from second and third generations speakers in the use of all five R-impersonals (when statistically significant); compared to second and third generation speakers, first generation speakers show decreased uses of impersonal *se* and 1PL and an increased use of impersonal *uno* and 2SG. Accordingly, second and third generation speakers pattern similarly in showing an increased likelihood of using 1PL and a decreased

tendency toward impersonal *uno* and 2SG. The only forms for which fourth generation speakers showed a statistically significant difference in usage were 1) 2SG, in which their positive association patterns with second and third generation speakers, and 2) 3PL, which demonstrated a negative association. It should be noted that the fourth generation was the only speaker generation to show a statistically significant negative association with 3PL forms, but this may be partially due to the extremely low sample count overall from fourth generation speakers.

In terms of age range, there are additional associations that speak to demographic differences in the use of R-impersonal forms. There were statistically significant differences across all three age ranges for *se* and 1PL forms. Regarding *se*, younger adults showed a positive association with the form, while middle and older adults showed a negative association. Prior studies have suggested that increased contact with English among Spanish-English bilinguals in the United States has contributed to a decrease in the vitality of impersonal *se* (Morales, 1995), yet this finding is not borne out in the SIT data. In fact, results of the chi-square test of association demonstrate that first generation speakers (those who would presumably have the least contact with English out of all generation groups studied) and the oldest speakers (who might demonstrate less assimilation / use of non-native forms) are those that show a decreased use of impersonal *se*.

For the 1PL impersonals, the older adults showed a positive association, in direct contrast to young and middle-aged adults. This finding is in accordance with similar results in Posio (2012, 2016), which suggest that older speakers use plural impersonals as a signal of group identification more so than young adults. In lieu of 1PL impersonals, young adults showed a positive significant use of impersonal *se* and 2SG forms. Middle-aged adults' negative association with 1PL impersonals could potentially be countered by their positive association with impersonal *uno.* It is interesting to note that while younger adults also

showed a positive association with other non-1PL forms, middle-aged adults did not; of the

three statistically significant forms for middle-aged adults, impersonal *uno* is the only

positive. This indicates that middle-aged speakers may use a wider variety of R-impersonal

forms to less dramatic extents, or more evenly, than their older or younger counterparts.

For the remaining R-impersonal forms, impersonal *uno*, 2SG, and 3PL, there were

only statistically significant variances between two of the three age ranges. There was an

opposite relationship with impersonal *uno* between young adults (negative) and middle-aged

adults (positive); the same is true for the relationships with 2SG, in which young adults

showed a positive association for the form and older adults a negative one. For 3PL usage,

young adults showed a negative association in contrast to older adults, who were more likely

to utilize the form.

4.1.1.4 Educational Status. Finally, speakers' educational attainment was also

examined in relation to the use of distinct R-impersonal forms. As mentioned in Chapter 3,

the majority of the speakers in the SIT data set (61.1%) had earned or were actively earning a

bachelor's degree at the time their interviews were conducted; the next largest portion of

speakers (21.7%) held a high school diploma or GED. Those with postgraduate degrees

composed 6.2% of the study data, those with only elementary education accounted for 5.8%,

and those with technical or associates degrees 5.3%. A 5 x 5 contingency table revealed no

violations of sample size; no cells had expected counts of less than five. Additionally, a chi-

square test of association revealed a statistically significant difference in the expected versus

actual use of R-impersonal form by educational level, $\chi^2_{(df=16)} = 414.676$, p < 0.01. Table 18

shows the contingency table with the associations that were statistically significant based on

adjusted residuals:

Table 18: *Chi-square Test of Association: Form x Educational Status*

Form	Elementary n	Elementary % of total	High School n	High School % of total	Associates / Technical n	Associates / Technical % of total	Bachelors n	Bachelors % of total	Postgraduate n	Postgraduate % of total	Total n	Total % of total	χ^2	p^a
Se	55	1.10	191	3.83	47	0.94	**523[b]**	10.48	**114[c]**	2.29	930	18.64	360.919	<0.01
Impersonal uno	**70[d]**	1.40	**89[e]**	1.78	**10[f]**	0.20	**162[g]**	3.25	**9[h]**	0.18	340	6.81		
2SG	**70[i]**	1.40	**249[j]**	4.99	**122[k]**	2.45	**1072[l]**	21.49	101	2.02	1614	32.35		
1PL	**39[m]**	0.78	**278[n]**	5.57	**27[o]**	0.54	**754[p]**	15.11	**47[q]**	0.94	1145	22.95		
3PL	62	1.24	**279[r]**	5.59	55	1.10	**536[s]**	10.74	**28[t]**	0.56	960	19.24		
Total	296	5.93	1,086	21.77	261	5.23	3,047	61.07	299	5.99	4,989	100.00		

[a] Two-tailed alpha
[b] Adjusted residual = -3.4
[c] Adjusted residual = 8.9
[d] Adjusted residual = 11.8
[e] Adjusted residual = 2.0

[f] Adjusted residual = -2.0
[g] Adjusted residual = -5.3
[h] Adjusted residual = -2.7
[i] Adjusted residual = -3.3
[j] Adjusted residual = -7.5

[k] Adjusted residual = 5.1
[l] Adjusted residual = 5.4
[m] Adjusted residual = -4.1
[n] Adjusted residual = 2.3
[o] Adjusted residual = -5.0

[p] Adjusted residual = 3.8
[q] Adjusted residual = -3.1
[r] Adjusted residual = 6.1
[s] Adjusted residual = -3.7
[t] Adjusted residual = -4.5

There were statistically significant variances in usage across all educational levels for impersonal *uno* and 1PL. Regarding impersonal *uno*, speakers with an elementary education showed a significantly stronger inclination for its usage (Adjusted residual = 11.8); this increased chance of usage was also found, to a smaller degree, in speakers with a high school education (AR = 2.0). Speakers with higher educational attainment all showed a negative association with impersonal *uno*. When looking at the association between education and 1PL R-impersonals, we note a larger variation; there was a negative association between 1PL usage and speakers with elementary, associates or technical, and postgraduate education AR = -4.1, -5.0, and -3.1, respectively). On the other hand, speakers with a high school education or bachelor's degree showed an increased likelihood for using 1PL forms (AR = 2.3 and 3.8).

For the three other R-impersonal forms, statistically significant usage patterns were only identified for certain educational levels. Speakers with a bachelor's degree showed a negative association toward *se* (Adjusted residual = -3.4), whereas there was a positive association between the form and postgraduate speakers (Adjusted residual = 8.9). Example 23 demonstrates a typical use of impersonal *se* from a postgraduate speaker:

23: *Estacionarse y no parquearse. So yo he oído eso aquí en el sur de Texas. No sé si **se***

***utiliza** esa palabra, que a la mejor no es formal, en otros lugares*.   SIT 48

The use of impersonal *se* in this context serves to generalize the speaking habits of those in the south of Texas, avoiding the need for a direct referent. On the other hand, a similar function may be achieved through the use of a 3PL impersonal. Example 24 spoken by a participant with a high-school education, demonstrates this usage:

24: *instead of, voy a estacionar el carro o el coche. Ah... voy a mirar. No, **usan** la*

palabra wátchalo, que es like watch, looking at, mirar.   SIT 46

In 24, SIT 46 uses the 3PL impersonal "usan" to describe slang used by specific speaker populations. This use takes the place of an impersonal *se* phrase and fulfills a generalizing

use. Given that impersonal *se* is considered generally a more formal form of impersonality, it is possible that speakers with a postgraduate education have been exposed to its usage in a formal register more often and are thus more inclined toward its use. Furthermore, the use of 3PL impersonals was found to have a statistically significant positive association with high school-level speakers but was disfavored among speakers with bachelors or postgraduate degrees. Speakers with a maximum elementary or high school education showed a negative association to 2SG forms, while those with an associates or bachelor's degree showed a positive association. These usage tendencies, along with those for 1PL, are less consistent than that found for impersonal *uno*. In the case of impersonal *uno*, there was an increased likelihood of use among speakers from lower educational levels, and this likelihood decreased as education increased. Prior studies into the use of impersonal *uno* have focused primarily on its use in distinct dialects of Spanish, as in the cases of Hurtado (2001) and Hurtado (2015). Given the tendency toward impersonal *uno* based on educational status found here, it would be fruitful to expand on analysis of the form and the factors that may condition its use.

4.1.2 Linguistic Variables

There were five linguistic factors considered as independent variables in the present analysis of R-impersonals: 1) verb tense and 2) mood; 3) presence of adverbs of time, 4) presence of an overt pronoun and 5) presence of hypothetical *if* clauses (cláusulas *si*). This section reviews the findings regarding these variables and their association with R-impersonals.

4.1.2.1 Verb Tense. In terms of verb tense, the majority of the R-impersonal forms were in the present tense—3,614 of the 4.989, or 72.44% of the sample. Given that the SIT sample is composed of data from sociolinguistic interviews, this is not surprising; speakers were most likely to be speaking about topics related to their current states, opinions, or situations. The second-most used verb tense was the imperfect, constituting 18.40% of the

sample, or 918 of the 4,989 cases. All other verb tenses recorded had low frequencies in comparison to present and imperfect: the simple preterit accounted for 6.37% of forms (318/4,989); the preterit perfect for 1.24% (62/4,989); and the conditional for 0.90% of forms (45/4,989). There were 31 cases of future tense (0.62% of the overall sample); as well as one instance of pluperfect.

Just as the high frequency of forms in the present tense may be contextualized by considering the genre of the sample data, the increased use of the imperfect in comparison to the preterit may also be justified. Consider that R-impersonals are often used to generalize actions, and the imperfect tense is used in Spanish for descriptions or ongoing actions in the past. The imperfect tense fits as an appropriate choice for the pragmatic use of R-impersonals to generalize or emphasize this generic, ongoing event. One example of the impersonal for such a use is demonstrated in Example 25:

25　*La, la comida era diferente allá y **se cocinaba** diferente. Vi a mi abuela hacer un... de elote, si... los granitos del elote los metía en una... es como algo así...*　　(SIT 31)

Here, the impersonal "se cocinaba" takes the impersonal indicative, along with the other verbs in the phrase, according to its function as a habitual action in the past. Although it could be inferred from the context of the interview that the speaker is referring to culinary customs in their home country or in their own family, they choose an impersonal, generalized, "*Se cocinaba*" to characterize the action.

A chi-square test of association was used to examine the potential variation in R-impersonal form based on verb tense. Given the relatively small frequency of forms in the preterit perfect, pluperfect, future, and conditional, these forms were excluded from the analysis; the present, preterit, and imperfect were analyzed (4,850 total cases). Overall, a statistically significant variance between the expected and actual uses of form were found

based on verb tense ($\chi^2_{(df=8)}$=542.752, p < 0.05). The results of the 3 x 5 contingency table for

the present, preterit, and impersonal are displayed in Table 19:

Table 19: *R-impersonals x Verb Tense*

Form	Present		Preterit		Imperfect		Total		χ^2	p^a
	n	% total	n	% total	n	% total	n	% total		
Se	**700[b]**	14.43	71	1.46	**131[c]**	2.70	902	18.60	542.752	<.05
Impersonal *uno*	**295[d]**	6.08	**7[e]**	0.14	**37[f]**	0.76	339	6.99		
2SG	**1426[g]**	29.40	**26[h]**	0.54	**136[i]**	2.80	1588	32.74		
1PL	**664[j]**	13.69	77	1.59	**344[k]**	7.09	1085	22.37		
3PL	**529[l]**	10.91	**137[m]**	2.82	**270[n]**	5.57	918	18.93		
Total	3,614	74.52	318	6.56	918	18.93	4,850	100.00		

a Two-tailed alpha
[b] Adjusted residual = 2.4
[c] Adjusted residual = -3.7
[d] Adjusted residual = 5.5
[e] Adjusted residual = -3.5
[f] Adjusted residual = -3.9
[g] Adjusted residual = 17.0
[h] Adjusted residual = -9.7
[i] Adjusted residual = -12.9
[j] Adjusted residual = -11.4
[k] Adjusted residual = 12.2
[l] Adjusted residual = -14.1
[m] Adjusted residual = 11.1
[n] Adjusted residual = 8.6

As can be noted from the data provided in Table 19, significant variance in R-impersonal use across the present, preterit, and imperfect tenses was found for impersonal *uno,* 2SG, and 3PL forms. Impersonal *uno* showed an increased use in the present tense (Adjusted residual = 5.5) but was disfavored for both forms of the past tense (Adjusted residual = -3.5 for the preterit and -3.9 for the imperfect). The 2SG form showed a statistically significant stronger use in the present (Adjusted residual = 17.0) and was also disfavored for the past, with an adjusted residual of -9.3 for the preterit and -12.9 for the imperfect. For the 3PL form, the preterit showed a stronger likelihood for use (Adjusted residual = 11.1), as did the impersonal (Adjusted residual = 8.6), while the present showed a lesser tendency for use (Adjusted residual = -14.1). For other R-impersonal forms, there were statistically significant associations found with select verb tenses. For instance, *se* was less likely to be used in the imperfect tense (Adjusted residual = -3.7), but more likely in the present (Adjusted residual = 2.4). On the other hand, the 1PL showed a significantly higher use in the imperfect (Adjusted residual = 12.2) and was underrepresented in the present (Adjusted residual = -11.4).

In noting the different positive and negative associations between tense and R-impersonal, several interesting patterns occur. As an indefinite subject pronoun, *uno* showed a stronger likelihood for use in the present than in the past. Impersonal *se* also trended toward use in the present over the past. The underrepresentation of *uno* in the preterit could be explained by considering that preterit is used to specify completed actions in the past, in which case speakers are likely to reference specific events with definite referents. In these instances, the use of an indefinite pronoun *uno* in an R-impersonal would not match the definite properties advanced by the preterit tense. Impersonal *uno* also showed an underrepresentation in the imperfect tense, as did impersonal *se* and the 2SG form. These three forms all have the capacity to reference singular agents, while 1PL and 3PL are plural

impersonals. Both 1PL and 3PL forms showed an overrepresentation in the imperfect, thus

demonstrating an opposition in the imperfect between singular and plural impersonality. This

opposition between singular and plural impersonals also manifests in an analysis of the

present tense: impersonal *uno* and 2SG forms were overrepresented, while 1PL and 3PL

forms were underrepresented. This suggests that while recounting events from recent

occurrences or sharing thoughts and opinions—in the present tense—speakers were more

likely to utilize singular impersonal forms. These forms, while still lacking definite referents,

perhaps reinforced that speakers were focusing on their own, singular life perspectives, as

shown in Example 26:

> 26 *mi papá es uno de los que me enseñó a...portarse **uno** y... ser un buen hombre, no*
>
> *de... así que **uno** tiene que trabajar para ganar la vida, y ser honesto.*   SIT 36

Here, speaker SIT36 is referring to lessons learned from his father, and the use of impersonal

uno clearly represents a substitution for the first-person personal form. This substitution of the

yo form for *uno* has been documented in prior impersonal studies and is not unusual. When

recounting events from the past, however, the same speaker utilizes the first-person *yo* form.

R-impersonals, when they occur, take the form of 3PL to generalize:

> 27 *voy para nueve años ahí, y más, entré que era warehouse, la bodega. Ahí empecé y*
>
> *después de tiempo me **cambiaron** a... que le dicen tester*   SIT 36

In this example, the speaker explains their current occupational status and uses the 3PL

"*cambiaron*" in an impersonal sense when explaining how their role changed. In this instance,

the emphasis is not on who reassigned SIT 36, hence the use of a referentless impersonal;

rather, the speaker aims to emphasize their new role as a tester. The 3PL impersonal form

shown here serves a generalizing and distancing function, and this is partly achieved through

the vagueness connected with the plural properties. Increased use of 1PL in the imperfect

tense also patterned with speakers' descriptions of past habitual or routine activities in which

they declined to mention specific referents, instead blending their own action into a collective, as shown in 28:

> 28 *Lo que es diferente a como están creciendo mis hijos…es que en la mañana en el verano, en la mañana **desayunábamos** y **salíamos** a jugar afuera. Mi mamá no sabía dónde estaba yo. Hasta la hora de comer, de comer, mi mamá y todas las vecinas de la calle eran a comer. Y ya **regresábamos** a comer, **estábamos** adentro un ratito y otra vez **salíamos** a jugar eh... hasta la noche que nos hablaban para cenar y luego **nos juntábamos** afuera y **jugábamos** encantado, **jugábamos** beisbol, **jugábamos** kickball, este, **nos subíamos** a todos los techos y árboles que había en la cuadra. Y **éramos** muy libres y muy sanos.*   SIT 29

In this example, the speaker contrasts their upbringing with that of their children's. Early on in the response, they switch from a current analysis of these differences into a recollection of their childhood, employing several 1PL R-impersonals to describe their habits. At no point does the speaker mention potential referents (siblings, neighbors, classmates, etc.). At the same time, the use of a 1PL R-impersonal emphasizes the collectivity of this action, that the speaker was not engaging in these activities alone. The impersonality present obscures the potential for these referents to be identified, most likely because context provides a somewhat clear idea of whom the speaker could most likely be engaging in these activities with. This example is one of many in the SIT corpus data in which speakers share recollections from their past with 1PL R-impersonals to express a collective action. While Section 4.3 further explores pragmatic roles of R-impersonals, Examples 26-28 here illustrate how singular and plural R-impersonal lend themselves to distinct temporal realms. It should be noted that there was a relatively weak relationship found between verb tense and R-impersonal use, based on a Cramer's V^2 of 0.052. In other words, verb tense accounted for only about 5% of the variance in R-impersonal use in the SIT data.

REFERENCE IMPERSONALS IN TX SPANISH

4.1.2.2 Verb Mood. Verbal mood was also included among the linguistic variables analyzed. The overwhelming majority of the verbs took the indicative mood (93.27%). There was a small representation of subjunctive mood, constituting 6.73% of the sample. A chi-square test of association found statistically significant variances in actual and expected frequencies, though the statistical significance was rather weak in comparison to that of the other independent variables analyzed ($\chi^2_{(df=4)}$=20.050, p <0.05). This statistical significance was only found for 1PL and impersonal *se*. For 1PL R-impersonals, there was an adjusted residual of 3.7 for the subjunctive mood, suggesting an overrepresentation. Of the 105 subjunctive 1PL impersonal forms, there were 63 instances of the word "*digamos*" (*let's say*) used in a generalizing or rhetorical manner[8]. One speaker in particular (SIT 92) used "*digamos*" in this fashion 43 times. This rhetorical overuse of "*digamos*" explains its overrepresentation in the overall data sample, the resulting positive adjusted residual for the subjunctive mood, and a negative adjusted residual of -3.7 for the indicative mood. For impersonal *se*, the chi-square test of association showed a positive adjusted residual (3.3) for the indicative mood and a negative adjusted residual (-3.3) for the subjunctive mood. This significance appears to reflect the overall tendency for indicative over subjunctive mood in the sample data. There were no other statistically significant variances in mood found for the other R-impersonal forms. Overall, there was a very weak relationship found between mood and R-impersonal usage, confirmed through a Cramer's V^2 of 0.003.

4.1.2.3 Adverbs of Time and If Clauses. The presence of adverbs of time or *if* clauses were also included as linguistic variables. The data was examined at the clausal level for whether or not any of forty-two specific adverbs of time were embedded (a complete list of adverbs included is available in Appendix C). Results show that 85.3% of the R-impersonals

---

[8] An example of *diagmos* used rhetorically: "y también tenía, **digamos**, buenas conexiones aquí desde que llegué como a los dos meses empecé a trabajar (SIT 92).

were not accompanied by adverbs of time, while 14.7% were. Table 20 shows the distribution

of adverbs of time by R-impersonal.

Table 20: *Adverbs of Time Present with R-Impersonals*

Adverbs present	Yes		No		Total (n)
Form	n	% of form total	n	% of form total	
Se	100	10.75	830	89.25	**930**
Impersonal *uno*	61	17.94	279	82.06	**340**
2SG	276	17.10	1,338	82.90	**1,614**
1PL	179	15.63	966	84.37	**1,145**
3PL	118	12.29	842	87.71	**960**
Total	**734**	**14.71**	**4,255**	**85.29**	**4,989**

The distribution of adverbs of time by R-impersonal form is relatively even across all

five forms, with a range of 82-90% of all forms showing a non-use for temporal adverbs.

Impersonal *uno* showed the highest rate of adverbs of time, with these forms present in

17.94% of the cases. 2SG showed the next-highest use of adverbs of time, with 17.10% of

cases. 2SG showed a statistically significant positive association, shown through the chi-

square test of association ($\chi^2_{(df=4)}$=27.038 p<0.05). All other forms either showed significant

underrepresentation in the sample for use of adverbs of time (*se*, 3PL) or no significant

association (impersonal *uno,* 1PL). Although the chi-square test shows an association, this

association is very weak and accounts for less than 1% of the variation in the data (Cramer's

V^2=0.005).

Further data analysis shows that *if* clauses were less frequent in the sample than

adverbs of time; there were 363 instances of *if* clauses and 4,626 cases without. In other

words, of all instances of R-impersonals, only 7.28% contained the conditional *if* clause. Table

21 further clarifies the limited use of *if* clauses with R-impersonals.

Table 21: *If-clauses Present with R-Impersonals*

If clause present	Yes		No		Total (n)
Form	n	% of form total	n	% of form total	
Se	17	1.83	913	98.17	**930**
Impersonal *uno*	25	7.35	315	92.65	**340**
2SG	226	14.00	1,388	86.00	**1,614**
1PL	46	4.02	1,099	95.98	**1,145**
3PL	49	5.10	911	94.90	**960**
	Total	**% of sample total**	**Total**	**% of sample total**	
	363	**7.28**	**4,626**	**92.72**	**4,989**

Of all R-impersonal forms, 2SG favored *if* clauses the most, as they appeared in 14%

of all 2SG samples. For all other R-impersonals, *if* clauses had a representation rate of less

than 10%. Since *if* clauses are restricted in their discourse function in the sense that they are

specifically used to present a hypothetical or conditional, it is not extremely surprising that

their overall representation in the data sample is low. In many cases, speakers chose the 2SG

form for a conditional *if* clauses with the goal of generalization, situating the hypothetical

situation in a generalized sense, as seen in 29:

> 29 *...apenas abres la boca sabes... o al menos muy, muy seguro **si** eres de Colombia,*
>
> *o **si** eres de Argentina, o **si** eres de México.*   SIT 0

In this case, the speaker presents their hypothetical analysis of a Spanish-speaker's accent,

stating that they are table to tell *if* you (impersonally) are from Colombia, Argentina, or

Mexico. In addition to generalization, *if* clauses were also present when used for inclusive

defocalization, or including the interlocutor more closely in the discussion, as in 30:

> 30 *es lo único que necesitaba para cross over porque ahoritita soy sargento. So no*
>
> *puedes hacer uhm... officer **si** no tienes bachelor 's degree.*   SIT 27

Here, SIT 27 explains the education needed to rank in the Army. There are two 2SG

impersonal forms here, *puedes* and *tienes*, and the corresponding *if* clause "*si no tienes*

bachelor's degree" presents a hypothetical situation. In this instance, the 2SG impersonal

serves as a discourse feature to include the interlocutor—who may be less familiar with the

standards required in the armed forces—more closely in the discussion. Examples 29 and 30

demonstrate how *if* clauses are embedded into impersonality with 2SG R-impersonal forms.

It should be noted that there was a very weak relationship established between the

presence/absence of *if* clauses and the five different R-impersonal forms (Cramer's V^2=0.032,

$\chi^2_{(df=4)}$=173.891, p<0.05).

4.1.2.4 Overt Subject Pronouns. Finally, the presence of an overt subject pronoun

for 2SG, 1PL, and 3PL R-impersonals was often considered. While prior analyses have

suggested that the appearance of an overt subject pronoun removes an impersonal

interpretation, further studies by Lapidus and Otheguy (2005), Posio (2012), and others have

demonstrated that the two features—impersonality and explicit subject pronouns—are not

mutually exclusive. In the current investigation, 15.41% of the R-impersonal cases, or 573 of

the 3,719 samples, contained an overt subject pronoun. The distribution of pronoun by R-

impersonal form is shown in Table 22:

Table 22: *Overt Subject Pronouns Present in R-impersonals*

Form	Overt pronoun cases (n)	Overt pronoun cases (%)
2SG	276	48.17
1PL	179	31.24
3PL	118	20.59
Total	**573**	**100.00**

Use of the 2SG pronoun *tú* accounted for the highest use of overt pronouns in R-

impersonals, followed by the 1PL *nosotros* and the 3PL *ellos/ellas*. The chi-square test of

association found a mild statistically significant difference between expected and actual

values for overt pronoun usage in R-impersonals ($\chi^2_{(df=2)}$=37.244 p<0.05, Cramer's

V^2=0.01). In this case, there was a positive association between overt pronoun usage with

2SG (Adjusted residual = 4.8) and a negative association between overt pronoun usage and

3PL (Adjusted residual = 5.7).

These results appear to support an observation made by Posio (2012), who suggested

that plural impersonal forms may be more likely to produce overt pronouns due to their

decreased referential accessibility in discourse (p. 340). 1PL R-impersonals composed over

one-third of the overt pronoun cases in the sample data. The 118 instances of 3PL R-

impersonals with overt subject pronouns is also notable; as Lapidus and Otheguy (2005)

indicated in their sociolinguistic analysis, 3PL R-impersonals are generally less frequent than

their 2SG and 1PL counterparts and are as such often overlooked in corpus data. One case of

R-impersonal usage with overt 3PL pronoun usage may be seen in 31:

31 *Pues, yo pienso que aunque la gente no lo quiera decir hay un poco de racismo. Y,*

*este, se nota como trabajo en construcción con mi papá. Cuando **ellos** nos **ven**, este,*

*haciendo trabajo y les molesta **a ellos**, este, nos **hablan** así como con una... con coraje, pues,*

*porque **dicen** que el... lo que estamos haciendo no... no es importante para ellos. **Ellos se***

***enfocan** a lo de ellos y no... no tienen respeto a uno.*   SIT 79

 Example 31 demonstrates how 3PL R-impersonals may be employed with overt pronoun

usage. The speaker employs the pronoun *ellos* in subject position twice (and an additional

time as an indirect object) without assigning or alluding to a potential referent. The current

analysis has demonstrated that 3PL R-impersonals are not only frequent in oral discourse

(constituting over 18% of all R-impersonals, in this case), but their appearance with an overt

pronoun is also existent. More common, however, was the use of the overt 2SG pronoun *tú* in

a non-referential capacity, such as that demonstrated in 32:

32 (a) *Y yo creo que tienes amistades en Texas para toda la vida. Mientras que **tú***

   ***seas** amigo, vas a tener amigos.*   SIT 80

    (b) *Pero tienes que estar feliz. Porque si **tú** no **te puedes** ser feliz, nadie te puede*

   *poner ahí. **Tú** solamente **eres** la que te puedes hacer más feliz.*   SIT 19

Here, note that speakers SIT 80 and 19 each employ 2SG R-impersonal forms with overt

pronouns. In both instances, the insertion of the 2SG pronoun *tú* comes after a phrase with no

overt pronoun, suggesting that the overt pronouns may serve emphatic uses. At the same

time, the overt usage of *tú* retains an impersonal reading in each of these instances, given that

the speaker is not employing the forms in direct reference to the interlocutor. Each of the

phrases in Example 32 express generic or generalized readings, hence the lack of specificity

of the interlocutor as referent.

While use of overt subject pronouns with 1PL, 2SG, and 3PL R-impersonals was

present in the SIT, it was not the predominant form. Presence of an overt pronoun in an R-

impersonal counters previous assertions that state that it is impossible to glean an impersonal

interpretation in such cases. However, in support of observations made by Lapidus &

Otheguy (2005), the usage of overt pronouns in R-impersonals is possible and prevalent.

4.1.3 Quantitative Analysis Conclusion

Nine independent variables were evaluated for a potential effect on the use of

impersonal *se,* impersonal *uno*, 2SG, 1PL, and 3PL forms in the SIT data. Statistically

significant effects were found for the use of different impersonal forms based on the

sociodemographic variables of gender, generation, age range, and educational level. In

particular, notable statistically significant differences were found in the use of

1PLimpersonals in male and female speakers, with females showing a stronger inclination

toward plural forms and males the singular. The use of plural forms was also stronger among

older speakers in comparison to younger and middle-aged adults, suggesting that group

membership or deictic reference may shift with age or life experience.

Results from distinct chi-square tests of association show a statistically significant

association between the independent variables and use of the five distinct R-impersonals,

though the overall amount of variance attributed to the independent variables is low, given

very weak relationships revealed through the Cramer's V^2 figures. Regardless, data from SIT

confirms previous findings from Posio (2012) regarding differences in the use of 1PL R-impersonals based on gender and speaker generation. The present analysis also supports assertions made by Lapidus and Otheguy (2005) regarding the infrequent, though existent, use of 3PL impersonal forms with overt subject pronouns, as well as more general findings in Gervasi (2007) and Morales (1995) regarding the frequent use of R-impersonals in U.S. Spanish.

The quantitative results reviewed in this section also present new findings regarding the use of R-impersonals. Educational status of speakers, previously underexplored in prior analyses, appears to impact, to some degree, the choice of impersonal employed in oral discourse. In particular, the 2SG form was less used among speakers with less formal education but was overrepresented in speakers with some college schooling. This statistical significance, however, was not confirmed in speakers with a postgraduate level of education, indicating that overall variation in R-impersonal usage may be caused by a variety of inter and extra-linguistic factors.

In a quantitative analysis of the linguistic variables of verb tense, verb mood, and presence of adverbs of time, *if* clauses, and overt pronouns, we note a significant association among verb tense and R-impersonal. Specifically, singular forms appeared more likely to be used in the present tense, whereas plural forms lent themselves to the impersonal past. While the topic of conversation or discourse is certain to impact the verb tense and mood employed by speakers, this finding nevertheless suggests that variance in R-impersonals can assist in creating profiles of genericity or specificity in oral discourse. In addition to examining the associations between R-impersonal distribution and distinct linguistic and extralinguistic factors, this section also engages in a qualitative analysis that allows for examination of salient examples of R-impersonal usage. This information is presented below.

4.2 Qualitative Analysis

As detailed in Chapter 2, R-impersonals serve a variety of pragmatic functions including generalization, inclusive defocalization, and concealment of the speaker. While prior studies have centered on individual R-impersonals or traditional impersonality mechanisms, there is little to no comprehensive research that compares distinct R-impersonals' chief pragmatic uses. In this section, qualitative and pragmatic analyses of the 2SG, 1PL, and 3PL forms will explore how these more nuanced R-impersonals serve as stylistic tools in oral discourse.

4.2.1 Si MI & T-Scores

An analysis of the collocate list provided for the 2SG, 1PL, and 3PL R-impersonal forms in WS Tools showed that "si" appeared as a collocate for several different verb forms within each R-impersonal. As such, the Mutual Information (MI) and T-scores for collocates with "si" were calculated. Though quantitative analysis in Section 4.2.1.2 showed a weak relationship between the appearance of "si" and R-impersonals in general, studying the collocates of "si" with R-impersonals can still provide insight on which verb(s) are most likely to accompany the form. As mentioned previously in Chapter 3, calculating both the MI and T-scores for collocate pairs is useful in truly determining a collocate pair's uniqueness in a corpus sample (Hunston, 2002). The MI and T-Scores were calculated based on formulas provided in Th. Gries (2009):

$$\text{MI: } \log_2 \frac{observed\ frequency}{expected\ frequency} \quad (2)$$

$$\text{T-Score: } \log_2 \frac{observed\ frequency - expected\ frequency}{\sqrt{expected\ frequency}} \quad (3)$$

Table 23 shows the results of the MI and T-Score calculation.

Table 23: *Mutual Information and T-Scores for "Si" Node Collocate Pairs*

2SG Form	MI Score	T-Score
puedes	4.056	4.231
vas	4.056	4.231
quieres	4.056	4.023

hablas	4.056	3.789
eres	4.056	3.342
tienes	4.056	3.231

Table 23
(cont)

1PL Form	MI Score	T-Score
estamos	4.056	3.842
vamos	4.056	3.438

3PL Form	MI Score	T-Score
están	4.056	3.842
hablan	4.056	3.668
pueden	4.056	3.438
eran	4.056	3.231
van	4.056	3.099
tienen	4.056	3.099
quieren	4.056	3.099

The determination of collocate pairs with *si* was made based on information returned through the "collocates" function of WS Tools. We may note that, when acting as a node, *si* had more collocate pairs with 3PL forms than the other two R-impersonal forms, though the MI and T-Scores for collocate pairs with 3PL forms are not the highest in comparison to 2SG or 3PL collocate pairs.

The MI score expresses the difference between the observed and expected frequencies for a given collocate pair. Hence, a higher MI score will indicate that the actual occurrence of a collocate pair exceeds its expected appearance. It is important to note that MI score is very sensitive to the frequency of the node and collocate words that compose the collocate pair; when the frequency of the collocate pair is low overall in the corpus sample, the MI score may not be the best sole indicator of the pair's frequency. For this reason, considering the T-Score is also important. The T-Score addresses the likelihood that the association of the node and collocate in a collocate pair is not one made by chance. In the data provided in Table 23, we note that the forms with the highest T-scores are the 2SG forms *puedes* and *vas;* of all R-impersonals included for analysis, these two R-impersonal forms were most likely to form a significant collocate pair with the conditional *si.* All of the T-scores for R-impersonal

collocate pairs with *si* fell between 3.099 and 4.231, which indicates that there was not significant variance between the distinct verbs in terms of their likelihood of appearing alongside *si*.

Although there does not appear to be wide variation *between* the T-scores for the R-impersonals that formed collocate pairs with *si,* it is important to note the verbs that appeared frequently enough with *si* to form a collocate pair. Iterations of the verb *ir* appeared in *si* collocates with 2SG, 1PL, and 3PL R-impersonals. This was the only verb to appear among all three R-impersonal types. Given its high frequency overall in language use, its appearance among all three R-impersonals is not surprising.

The verbs *estar, querer, poder, ser, tener,* and *hablar* appeared in *si* collocate pairs for two of the three R-impersonal forms analyzed. All of these verbs are considered to be frequently-used words that form part of contemporary Spanish language usage (Fratini et al., 2014), hence their appearance in collocate pairs reflects their overall usage in Spanish language in general. The appearance of *hablar* as an R-impersonal collocate with *si* relates to the thematic matter of the sample; throughout the sociolinguistic interviews, speakers were asked about the linguistic habits of their communities. Collocate pairs of *si* and forms of *hablar*, then, indicate participants' willingness to speculate on speech conditions, as shown in sample data in 31:

31 *Pero además con eso en el trabajo se me hace que tienes más oportunidad **si***

*hablas los dos idiomas. Aquí en particular **si** hablas el español…*   SIT 48

In these and similar cases, participants discussed the potential advantages of bilingualism and speech habits of their community (i.e. English vs. Spanish dominance, Spanglish usage).

An examination of collocate pairs with the conditional "si" showed that frequently-used verbs, such as the copulatives *ser* and *estar* and other highly frequent verbs like *querer* or *poder* were potential candidates for collocation. At the same time, a calculation of the MI

and T-Scores for these collocate pairs failed to reveal a significantly strong association

between any verb form and the conditional *if*. Considering a sample with a larger overall

frequency of conditional *if* statements (which may also implicate widening the consideration

of genre and text type to include in the overall sample) may in the future prove fruitful to

further explore the relationship between conditionals and R-impersonals.

4.2.2 Pragmatic Analysis of 2SG and 1PL R-Impersonals

As discussed in Chapter 2, R-impersonal forms have distinct pragmatic applications

that include generalization, inclusive defocalization, and concealment of the speaker. When

used to generalize, R-impersonals serve to make the message more objective or universal.

Through inclusive defocalization, R-impersonals engage in agent de-emphasis that also seeks

to include the interlocutor's experience in the dialogue, building speaker-interlocutor

empathy. Finally, R-impersonals may also serve a function of total agent de-emphasis as a

means to distance the agent from the action, usually in the context of an undesirable situation.

Of the five impersonality mechanisms included in the present analysis, the 2SG and

1PL R-impersonal forms are the only ones that have the potential to act in all three pragmatic

capacities described above. In order to explore which pragmatic functions lend themselves to

the 2SG or 1PL R-impersonal form, a random sampling from the SIT data was drawn. Fifty

instances of impersonal 2SG and 1PL were selected at random from pools of male and female

participants and were classified according to whether they were used for generalization,

inclusive defocalization, or concealment of the speaker (100 instances total of 2SG + 100

instances total of 1PL = 200 R-impersonal forms total analyzed). The results of this analysis

are shown below.

Table 24: *2SG and 1PL R-impersonals: Pragmatic Analysis*

Pragmatic Function	1PL (n)	2SG (n)	Total (n)
Generalization	35	56	91 (45.5%)
Inclusive Defocalization	8	29	37 (18.5%)

Focalized Generalization	30	0	30 (15.00%)
Concealment	13	14	27 (13.50%)
Rhetorical	14	1	15 (7.50%)
Total	**100**	**100**	**200 (100.00%)**

The remainder of this section is divided into sub-sections that analyze the use of 2SG and 1PL R-impersonals for the aforementioned pragmatic purposes of generalization, inclusive defocalization, and speaker concealment. Additional pragmatic purposes for these forms are also discussed.

4.2.2.1 Generalization. Generalization was the most-used pragmatic application for both 1PL and 2SG R-impersonal forms. These instances constituted nearly half (45.5%) of the sample, though 2SG was the favored impersonality form for generalization (61.5% of all generalization uses were 2SG) over 1PL (38.5%). The second most-used pragmatic application was inclusive defocalization, though the 2SG form was used more for this purpose than the 1PL by a significant margin. R-impersonal usage as a means for concealment of the speaker occurred in 13.50% of the cases and was distributed nearly equally between the 2SG and 1PL forms.

When used for generalizations, R-impersonals present a speaker's perspective in a more objective light. While impersonal *se* is considered the impersonality mechanism best suited for generalization, 1PL and 2SG forms are also employed, as seen in 33:

33 (a) *Spanglish es hablar el inglés y el español mixto, como... cada otra palabra... Si* **estás hablando** *en inglés, cada otra palabra, cada tercera o cuarta palabra, podría ser en español. O decir una frase en inglés y la siguiente en español, la siguiente en inglés. No, no* **continúas** *la conversación en solamente un lenguaje…*   SIT 25

　(b) **nosotros** *aquí* **metemos** *mucho la palabra en inglés y la* **hacemos** *en español, como troca…*   SIT 3

Both of the examples in 33 show how R-impersonals may be used to generalize. In 33(a), Speaker 25 describes Spanglish and uses 2SG forms; in this context, these 2SG forms could

be substituted with impersonal *se* or other grammatical constructions that would also denote impersonality, such as:

> 34 *Spanglish es hablar el inglés y el español mixto, como... cada otra palabra... Si **se***
>
> ***habla** en inglés, cada otra palabra, cada tercera o cuarta palabra, podría ser en*
>
> *español. O decir una frase en inglés y la siguiente en español, la siguiente en inglés.*
>
> *No, no **se continúa** la conversación en solamente un lenguaje…*

While impersonal *se* could serve as a viable choice for describing speech patterns, speaker SIT 25 instead employs 2SG R-impersonal forms to generalize. These forms do not engage the interlocutor as a direct referent, but rather indicate how people who use Spanglish speak. Similarly, in 33(b), the 1PL impersonal phrases ***nosotros** aquí **metemos** mucho la palabra* and *la **hacemos** en español* describe general habits of the speech community. While the use of 1PL implies that the speaker (and perhaps the interlocutor) is considered among those who use Spanglish, the potential referents extend past just these individuals. The inclusivity included in the impersonality of this 1PL form is contrasted in cases with impersonal *se*:

> 35 (a) *cerca de... del puente **se habla** más o menos como **se habla** en Texas…SIT* 27
>
> (b) *es muy diferente el español que **se habla** en México al que **se habla** aquí, sí es*
>
> *diferente.*   SIT 65
>
> (c) *en Uruguay **se habla** un español uhm... también diferente a otras partes del*
>
> *mundo, ni siquiera a España, ¿verdad?*   SIT 84

All examples in 33 and 35 discuss speaking habits, yet they differ by impersonality mechanism. The forms in 35 employ impersonal *se*, which creates a greater degree of impersonality or genericity than the 2SG and 1PL forms in 33. This greater degree of genericity in 35 could be due to the fact that the speakers are discussing speech habits or activities outside of their physical or epistemic proximity; in 35(a), use of impersonal *se* over a 2SG or 1PL form suggests that speaker does not consider themselves close (either

physically or epistemically) to the bridge; in 35 (b) and (c), speakers reflect on habits of speakers in other countries, Mexico and Uruguay. The distance that the speakers perceive between their own experiences and those of Spanish speakers in other countries is reflected through the use of impersonal *se,* whereas the R-impersonal examples in 33 serve to reflect speakers' perceived inclusion in generic references.

4.2.2.2 Inclusive Defocalization. Instances of R-impersonals for inclusive defocalization were also present in the sample data. Though inclusive focalization was more prominently expressed through 2SG, instances of 1PL R-impersonals with this function also surfaced:

36 (a) *En ese lugar hay casas muy antiguas hechas de piedra, de lodo, materiales raros para **uno** cuando **estás** niño que **conoces** de la ciudad casas de ladrillo y de cemento…* SIT 60

   (b) *Yo tengo mucha fe. Y este, cuando yo le rezo a Dios, siento que Él me escucha, y, si siento que hay resultados de cuando yo le pido algo o le rezo por alguien, también creo que las cosas pasan por algo y que Dios tiene que ver mucho en eso. Hace las cosas para que **nosotros aprendamos**, ya sea por algo bueno o por algo malo, pero no es por lastimarnos de ninguna manera es razón por algo, **deberíamos aceptarlo**. Y, pues, sí, yo creo en los milagros. Si, si **crees**... si **tienes** fe, yo, yo creo que sí existen los milagros.* SIT 25

 In 36(a), the speaker describes a personal experience and switches from a more generic indefinite *uno* to 2SG for inclusive defocalization. This switch to 2SG retains the overall impersonality of the utterance but also serves as an attempt to engage the interlocutor more fully in the scenario being described. Similarly, in 36 (b), speaker SIT 25 utilizes both 2SG and 1PL R-impersonals for inclusive defocalization in describing spiritual beliefs. The focus on their own experience in the first sentence, demonstrated by first-person singular forms

such as "**Yo tengo** mucha fe…cuando **yo** le **rezo** a Dios…**siento** que hay resultados de cuando **yo** le **pido** algo o le **rezo** por alguien, también **creo** que las cosas pasan por algo," is broadened to a more indefinite potential referent through 2SG and 1PL R-impersonal forms. Here, SIT 25 expresses their personal beliefs through a communal lens of "para que **nosotros aprendamos**" and extends the experience of having faith out to the interlocutor through 2SG forms such as "crees" and "tienes." At the same time, these forms retain impersonality in that SIT 25 is not literally insisting that the interviewer believe and have faith; rather, this form carries a generic, impersonal interpretation that may extend to anyone.

The agent-defocusing view of impersonality presents impersonality as a continuum with varying degrees of referentiality. When comparing the two examples provided in 36 (a) and (b), we may note how the use of 1PL holds a larger degree of generalization or objectivity and the 2SG retains a more subjective degree. This extension of an experience from the speaker's perspective to that of a wider set of referents via inclusive defocalization is often notable through the existence of switch reference, or concurrent use of different pronominal forms or verb conjugations signaling a potential deictic change:

> 37 *…eso nunca yo creo que lo **sabes**, pero como **te enamoras**, **crees** que **estás** con la persona correcta, y simplemente, yo sabía que era una persona buena. Ah... pues **conoces**, que es... no lo **conoces**, sino que te **empiezas** a dar cuenta que es una persona de buenos sentimientos y yo pienso que eso es lo principal…*   SIT 49

In this instance, speaker SIT 49 utilizes 2SG R-impersonal forms alongside the first person "yo sabía" to describe the process of finding a romantic partner. Though the speaker is obviously speaking from personal experience, she employs 2SG forms through the means of inclusive defocalization in order to make her experience more relatable to her interlocutor and appear more objective in her judgments.

Inclusive defocalization serves to increase speaker-interlocutor empathy and facilitate communication of the message. In the sample analyzed from SIT, we note that both 2SG and 1PL R-impersonals are used to this end.

4.2.2.3 Speaker Concealment. Thirdly, uses of R-impersonals for speaker concealment were prevalent in the sample. Identification of an R-impersonal for speaker concealment is highly context-dependent, given that this usage shares some similar properties with inclusive defocalization. Speaker concealment and inclusive defocalization are both employed when a speaker discusses something that they themselves experienced through an impersonal lens; in both instances, first-person singular forms would be appropriate. However, their core purposes differ. With inclusive defocalization, speakers aim to make their experiences more accessible or understandable to the listener, whereas concealment of the speaker is intended to distance the speaker from the action expressed. Speaker concealment is generally used to describe undesirable or uncomfortable situations and has less to do with involvement of the interlocutor than it does de-emphasis of the speaker as agent. Select examples of R-impersonals for speaker concealment are shown in 38:

38 (a) *y allá es puro español so, y luego lo tienen todo perfecto. Cause como* **nosotros**, *unos* **batallamos** *a ser, a decir la ere o like los acentos y todo eso...* SIT 2

(b) *cuando era niño, a cierta edad todavía no me daba cuenta de muchas cosas, de muchas injusticias, de la pobreza en la que... en la que nos... en la que* **vivíamos**...SIT 75

(c) *lo peor que me ha pasado es este... una vez que* **intentamos** *pasar como tres veces. Tres veces y este nos agarraban. Y nos regresaban para atrás...SIT* 81

The three situations described in 38 are uncomfortable for the speakers in different ways. In 38 (a), the speaker recalls difficulty in communicating exclusively in Spanish. The speaker in

38 (b) reflects on the poverty they experienced in their youth, and the speaker in 38 (c)

recounts their experiences with deportation. In each instance, the speakers make no prior

reference to other individuals to which the 1PL forms could refer; hence, these 1PL R-

impersonals serve to de-emphasize the speaker as an individual and demote their prominence.

In 38 (c) it is also interesting to note the use of "nos agarraban" and "nos regresaban," two

3PL impersonal forms. Although the potential referents for these verbs is somewhat

accessible based on the context of the utterance, these forms retain a degree of impersonality,

perhaps as a means of de-emphasis or distancing, given the undesirability of the situation.

These three examples show concealment communicated through 1PL impersonals, but further

data in the sample show that 2SG may also be used for concealment:

> 39 (a) *yo necesito que me den consejos, que me digan, 'ay estás haciendo mal esto,*
>
> *hazlo de este modo,' pero con buenos modos, ¿me entiende? no que te humillen*
>
> *porque no **sabes** expresarte, primeramente, por decir si **tú** no **tuviste** mucha escuela,*
>
> *no **te sabes expresar** con una persona que tuvo más escuela que uno, se siente uno,*
>
> *bueno yo digo de mi parte…*   SIT 50
>
>   (b) *es un pueblo bien chiquito. Ah… pues, todo el mundo se conoce a todo el*
>
> *mundo. Allá no **puedes** hacer nada malo porque al ratito tu mamá se enteraba. Lo*
>
> ***hacías** en la mañana y en la... en la tarde ya tu mamá lo sabía que **habías hecho**. No*
>
> ***podrías tener** novio. No **podías andar pintándote** en la escuela. No **hacías** nada malo*
>
> *porque en la mañana lo **hacías** y en la tarde ya tu mamá sabía…*   SIT 59
>
>   (c) *cuando no sabía hacer alguna tarea y me sentía desesperada que no la podía*
>
> *lograr hacer…porque, o sea... por ejemplo, matemáticas, si **te llegabas** a atrasar en*
>
> *algún, en alguna área, y no la **llegabas** a entender y la clase avanzaba, pues este...*
>
> *pues sí **te desesperabas** porque **tenías** que estar al nivel, ¿verdad? De la clase...*   SIT
>
> 68

In each of these situations, speakers switch from first-person to second-person references, engaging with 2SG impersonality mechanisms as a way to de-emphasize their attachment to the situations in question. For instances, the speaker in 39 (a) admits that they would like advice with certain areas of their speaking, using first-person forms "**yo necesito** que **me** den consejos" to express this. However, to express the sentiment that they wish to avoid criticism, they utilize 2SG R-impersonals, changing the referentiality from themselves in the first person to a less definite form. This switch from first to second person also presents in 39 (c), when the speaker admits that they struggled with homework using the first-person "no **sabía** hacer alguna tarea y **me sentía** desesperada." In further elaborating on their experience, they switch from first to second person, distancing themselves from the negative emotions that result from the feeling of falling behind in school. The experience relayed by the speaker of 39 (b) is autobiographical yet solely employs 2SG R-impersonals as a distancing mechanism. It may be that the speaker feels strong negative emotions about the restrictions she faced while growing up in her small town and wishes to take emphasis off of the idea that it was something she personally experienced. The use of 2SG R-impersonals in this case conveys a secondary idea of generality while fulfilling a primary focus of agent de-emphasis. R-impersonal usage for concealment of the speaker was the least-used pragmatic function in the sample when compared to generalization and inclusive defocalization, but this may be due in part to the conversation topics elicited during the sociolinguistic interviews. In other words, speakers probably spent more time speaking about topics that would not naturally elicit speaker concealment. However, the representation for this pragmatic function present in the sample taken from SIT demonstrates that both 2SG and 1PL forms may be used when speakers wish to distance their role as agent from a given action.

       4.2.2.4 Focalized Generalization and Rhetorical Uses. In addition to generalization, inclusive defocalization, and concealment, note that there are two additional uses displayed in

Table 28: "focalized generalization" and "rhetorical." Upon analyzing the samples of 1PL

and 2SG impersonals, it was determined that there were several instances, chiefly among 1PL

forms, that could not cleanly be considered as generalization, inclusive defocalization, or

concealment. In these instances, speakers employed a 1PL R-impersonal with no mentioned

referent, though the context of the usage indicated that there was an implied group. This

usage of an R-impersonal aligns with the *generalización focalizada* use of R-impersonality

referenced in Pulido Astorga and Rivadeneira Valenzuela (2017), who characterize the form

as "a la vez genérico y personalizador…puede estar refiriéndose a un grupo en específico" (p.

27). Though their analysis identifies focalized generalization as a usage for 2SG R-

impersonals, there were several instances of this usage for 1PLs in the current analysis, as

shown in Example 40:

> 40 (a) *en el trabajo en que estoy ahorita* **hacemos** *libros con plástico, con el wire de*
>
> *plástico, con el wire de alambre,* **ponemos** *glue,* **ponemos** *las cover, las pastas…* SIT
>
> 60
>
> (b) *Fue en recreo. Ya* **íbamos**... *se había acabado.* **Andábamo***s todos sudados.*
>
> **Fuimos** *al... ya nos* **íbamos** *a cambiar y todo y a ver el... el clima estaba muy*
>
> *frío…*SIT 67
>
> (c) *mi pasatiempo favorito es el... es la pesca… y disfruto mucho la música...*
>
> *tocar guitarra también…Hay varios, varios, varios conciertos que han sido muy, muy*
>
> *especiales. Por la razón de que los conciertos que... que* **nosotros ejercemos** *son...*
>
> *son, este... son religiosos. O sea* **tocamos** *música religiosa. Entonces todo... todo los...*
>
> *ah... los eventos que* **nosotros tenemos** *siempre son muy especiales…*   SIT 80

In40 (a)-(c), we note the use of 1PLs in impersonal contexts, given that there is no referent

mentioned in the dialogue. In 40 (a) and (c), the speakers transition from speaking in

reference to themselves, with first-person singular forms, to a first-person plural. This switch

reference from singular to plural acts as a defocalization mechanism yet does not fall under

the category of concealment, since the speakers are not necessarily attempting to distance

themselves from the actions. Rather, the switch to a 1PL R-impersonal serves as a mechanism

to generalize, to de-emphasize the sole participation of the speaker and imply the

participation of other (unknown) agents. Though the interlocutor may be able to roughly infer

a potential referent set for each of these 1PL forms—40 (a) references a workplace, (b) a

school setting, and (c) a musical group—there is still a certain ambiguity and impersonality.

For example, in 40 (c), there are a few different potential referents for "*nosotros ejercemos*"

and "*nosotros tenemos*," including the group of musicians that the speaker is part of *or* the

larger community that attends the event. This category was not initially included in the

consideration for pragmatic applications of 2SG and 1PL R-impersonals; given that it

constitutes roughly one third of the 1PL R-impersonal uses in this sample, however, it is

considered a valid and vital impersonality mechanism.

   The 1PL R-impersonals included in this sample also deviated from the framework of

pragmatic functions provided in that there were several instances of rhetorical, lexicalized

forms. These instances, all of which save one corresponded to the verb *decir* to express a

sentiment of "let's say," serve a purely rhetorical function in dialogue:

41 (a) *Pues nací aquí, pero yo sé que, eh, parte de mí, bueno,* ***diremos*** *mis*

antepasados son mexicanos…   SIT 33

   (b) *lo más difícil fue aprender el idioma, pero no, no tardé mucho tiempo, o sea,*

ya como en seis meses me podía comunicar y también tenía, ***digamos****, buenas*

conexiones aquí…   SIT 92

   (c) *no hay muchas oportunidades allá. Allá lo que... lo que reina más... o lo* ***vamos***

a decir *lo que... no lo que reina más, sino lo que se trabaja más en el área el es*

petróleo…   SIT 80

In each of these examples, verb phrases with conjugations of *decir* serve as a negotiation tactic put forth by the speaker to the interlocutor. Iterations of "let's say" are common, lexicalized forms both in English and Spanish; pragmatically, they may be used to signal overtly that a Gricean maxim will be violated in the enunciation (Cogen Herrmann, 1975). These forms may be considered impersonal in that they are not directly referring to the speaker and interlocutor in the use of a 1PL form; at the same time, they are not specifically used to generalize, defocalize, or conceal the speaker. Fourteen of the 15 rhetorical impersonals used the 1PL form. The sole use of a rhetorical 2SG form appeared in the lexicalized phrase *tú sabes* ("you know") as a discourse marker for narrative conclusion (Said-Mohand, 2007). A full analysis of the pragmatic functions for "let's say" or "you know" fall outside of the scope of the present analysis; however, mention of their usage warranted inclusion in the current analysis given their surfacing in the sample.

4.2.2.5 1PL and 2SG Usage by Gender. As was discussed previously in 4.1, speaker gender was a statically significant variable in the use of R-impersonals, and women were more likely to employ 1PL forms than men. In order to assess whether there was a gender effect in the pragmatic use of R-impersonals, the 2SG and 1PL forms that were categorized by pragmatic use were also classified by gender, as shown in Table 25:

Table 25: 2SG and 1PL R-Impersonals: Pragmatic Analysis by Gender

Pragmatic Function	1PL		2SG	
	Male (n)	Female (n)	Male (n)	Female (n)
Generalization Inclusive	15 (30%)	20 (40%)	32 (64%)	24 (48%)
Defocalization Focalized	6 (12%)	2 (4%)	11 (22%)	18 (36%)
Generalization	12 (24%)	18 (36%)	0 (0%)	0 (0%)
Concealment	6 (12%)	7 (14%)	7 (14%)	7 (14%)
Rhetorical	11 (22%)	3 (6%)	0 (0%)	1 (2%)
Total	50 (100%)	50 (100%)	50 (100%)	50 (100%)

Among both genders, generalization was the most popular pragmatic application for both 1PL and 2SG R-impersonals. For 1PL forms, focalized generalization was the second

most popular form. Concealment through 1PL usage was more popular among women by

men, whereas men employed 1PLs in rhetorical uses more often than women. Inclusive

focalization was used by both genders, though men used this form more than women.

Regarding the use of 2SG R-impersonals, both genders favored generalization and

inclusive defocalization as the most popular pragmatic applications. However, use of 2SG to

generalize was more popular among men (64%) than women (48%); in terms of inclusive

defocalization, women employed 2SG forms more than men. There were no uses of 2SG

focalized generalization and only one use of rhetorical 2SG. For concealment, the distribution

of 2SG forms was equal among men and women (14% of the sample for each gender) and

was nearly equally distributed across the 1PL and 2SG forms.

Prior analyses have posited that women tend to use R-impersonals as a means to

extend their own experiences into a generic realm more often than men, whose primary focus

for R-impersonality usage is generalization (Posio, 2016; Serrano & Ajión Oliva, 2014).

Under this framework, women's use of 2SG for inclusive defocalization indicates a

preference toward maintaining something of an epistemic sense while still employing

impersonality. In choosing 2SG over 1PL, women opted for a less (to some degree) generic

impersonal reference Men, on the other hand, opted to express inclusive defocalization

through the 1PL form, and the plurality expressed through 1PL widens the generic

interpretation. Note how both 2SG and 1PL may be used for inclusive defocalization:

42 (a) *¿Ciudadanía? Pues es un derecho que **tenemos** aquí en el lugar en que, en que*

***vivimos**, ¿verdad? Ser ciudadano, ser parte de la comunidad…*   SIT 71

(b) *estuviera tratando de hacerse... no hacerse, pero, sino aprender el idioma, de,*

*de aprender los modos americano porque ya **estás** aquí. Ya **estás** en el Estados*

Unidos. Y, sí, nunca se te va a olvidar tu…   SIT 30

In 42 (a), a male speaker expresses his beliefs on citizenship through an 1PL R-impersonal. The use of "*tenemos*" and "*vivimos*" here can be interpreted as hearer-inclusive since the speaker conceives of the interviewer as a US citizen. At the same time, the use of a 1PL form is meant to generalize even beyond the speaker-interlocutor space and apply to an indefinite group, potentially to anyone holding citizenship. The use of 2SG R-impersonals in 42 (b) has a similar application for inclusive defocalization. In a similar conversational context, female speaker SIT 30 discusses acculturation and utilizes the 2SG form "*estás*" to broaden the scope of reference from herself to a wider set of referents. In this instance, the second-person singular form invites the listener to identify with the statements being made and extends the experience of being in the United States but is not restricted solely to the interlocutor.

      4.2.2.6 1PL and 2SG Conclusion. Generalization, inclusive defocalization, and concealment of the speaker are three chief pragmatic applications for R-impersonal forms. 1PL and 2SG R-impersonals each hold the ability to execute these pragmatic functions, and a random sampling of 1PL and 2SG forms from the SIT data shows that both forms carry out all three functions. In general, 1PL and 2SG forms were employed for generalizing means, with applications toward inclusive defocalization and concealment present to a lesser degree. In addition to these three functions, 1PL impersonals were also used for focalized generalization and in rhetorical contexts. An analysis of the random sample by gender showed that women used 2SG for inclusive defocalization more than men, who opted for the 1PL in similar circumstances. This increased use of 2SG inclusive defocalization among women coincides with similar results found by Serrano & Ajión Oliva (2014) and Posio (2016) that show a trend toward generalization for men and more subjective expressions by women. Consideration to the use of 1PL R-impersonals for focalized generalization warrant future consideration so that these forms may also be considered in a comprehensive R-impersonal framework.

4.2.3 Pragmatic Analysis of 3PL R-Impersonals

In addition to examining the pragmatic applications of 2SG and 1PL R-impersonals, 3PL R-impersonals from SIT data were also analyzed. 3PL R-impersonals operate in a different deictic space from 2SG and 1PL forms in that they necessarily exclude the speaker and interlocutor from potential reference; therefore, the pragmatic frame of generalization, inclusive defocalization, focalized generalization, or speaker concealment is not appropriate. Instead, we may consider the use of 3PL R-impersonals in the context of generalization—expressing a generic statement that emphasizes objectivity of the message—or denoting an unspecified group. In the case of an unspecified group, 3PL forms may be used to indicate presence of an unknown actor or agent, either singular or plural. Their participation or engagement in the action is not the focal point of the message, hence the use of an impersonality mechanism for backgrounding means. This sub-section further discusses the use of 3PL R-impersonal forms for generalization or denotation of an unspecified group.

4.2.3.1 Generalization. A sample of 200 3Pl R-impersonals was selected randomly from the SIT data. One hundred cases were taken from female speakers, and the other hundred from male speakers. Each case was classified according to whether it was used for generalization or for denoting an unspecified group, and these results are shown in Table 26:

Table 26: *3PL R-Impersonals: Pragmatic Analysis by Gender*

Pragmatic Function	Male	Female	Total
Generalization	72	60	**132**
Unspecified Group	28	40	**68**
Total	**100**	**100**	**200**

Generalization proved the most popular use for 3PL R-impersonals in the sample, constituting 132 of the 200 instances (66%). However, there is a slight difference in the use of generalization based on gender: of the 100 male utterances, 72 were generalized; in the same raw number of utterances for females, there were 60 generalizations. Accordingly, female speakers used 3PL to refer to unspecified groups more than men (40% for females versus 28% for males). Example 42 below shows an example of 3PL for generalizing means:

43 >>i[9]: *¿Así que nunca has estado en México?*

>>s: *No. He estado cerca, en McAllen.*

>>i: *McAllen. ¿Y cómo era el español allí?*

>>s: *Ah... pues lo **hablan** mejor porque lo **hablan** más eh... más seguido. So, you know, **tienen**, **pueden** hablar en, en oraciones completas y hablar bien y...*

SIT 88

In this example, we see how speaker SIT 88 utilizes 3PL R-impersonals to describe speaking habits in McAllen, Texas. The lack of referent is clear from the preceding dialogue; the interviewer's questions make no reference to humans and merely ask how "the Spanish is there." It is interesting to note the potential for impersonal *se* in this response—an answer of "***se habla** mejor porque **se habla** más eh…más seguido*" would be appropriate. In fact, similar impersonal *se* structures were used in 33, demonstrating that either choice is available for speakers in generic contexts. An additional example of generalized 3PL usage is shown in 44:

44 *algo que es no sé si es no más de... de la parte de que soy de México, o grande, o todo México, pero este... uhm... **celebran**, este... el día de Santiaguito…*   SIT 38

A 3PL R-impersonal use of *celebrar* fits within a generalized usage to express that people, in general, celebrate a given holiday. In this case, there is no definite referent, though intuition from context of the utterance indicates that those celebrating are residents of a certain part of Mexico. Use of a 3PL R-impersonal over impersonal *se* in this context retains the impersonality of the message but may also be seen as a less indefinite form.

Since both impersonal *se* and 3PL R-impersonals can be employed for means of generalization, it is interesting to see if this a preference in form based on verb. When looking

---

[9] >>i refers to the interviewer, and >>s to the speaker.

REFERENCE IMPERSONALS IN TX SPANISH

at the most-used verbs for impersonal *se* and 3PL employed to generalize, we note the

following frequencies:

Table 27: *Impersonal se and 3PL Generalization by Verb Frequency*

Infinitive	Uses with impersonal *se*	Uses with 3PL	Total
Hacer	**251**	44	295
Poder	**102**	10	112
Decir	15	**93**	108
Hablar	**55**	27	82
Tener	15	**58**	73
Ver	**51**	6	57
Usar	**33**	14	47
Ir	**30**	16	46
Estar	7	**25**	32
Dar	4	**27**	31
Total	563	320	883

Bolded numerals indicate the class of R-impersonal with greater representation in the corpus data.

In six of the 10 cases for most frequently appearing generalized verbs, impersonal *se*

outranked 3PL in terms of usage; in five of these six cases (all save with *ir*), use of

impersonal *se* topped 66%. There were four verbs among the top ten most used that had more

conjugation in 3PL than impersonal *se*: *decir, tener, estar,* and *dar*. It is important to recall

that many instances of impersonal *se* with *decir* were discarded from analysis due to their

appearance in highly lexicalized uses such as "*cómo se dice*." What was unexpected,

however, was the potential usage of 3PL R-impersonals with *decir* to communicate a similar

sentiment. Nevertheless, this generalized usage with *decir* appeared with 3PL R-impersonals,

as shown in 45:

45 (a) *me han dicho que el negocio de ellos ya es el aguacate. El aguacate, el*

*avocado que le **dicen** aquí.*   SIT 81

(b) *me gustaba uhm... vacilar con ellos y salir con ellos a algún lado. Pienso que*

*yo era normal, no era muy, ¿cómo **dicen**? El bomb, no era el bomb del party.*   SIT

36

(c) *Pero... para empezar aquí a El Paso le **dicen** el Chuco. No sé. No sé por*

*qué…hay palabras... por ejemplo…en vez de empujar le **dicen** púshale. En vez de*

*decir, un... un car, **dicen** carro.*   SIT 74

These three examples of 3PL impersonal "decir" function in parallel to impersonal *se* with

decir to express what things are called. For instance, 45 (a) could be re-phrased "*El aguacate,*

*que **se dice** aquí el avocado*." Similarly, utterances in 45 (b) and (c) could also employ

impersonal *se*. The use of 3PL R-impersonals with *decir* for a form typically considered

standard with impersonal *se* coincides with the idea that 3PL R-impersonals may be in a

process of overtaking impersonal *se* in U.S. Spanish (Gervasi, 2007). At the same time, this

phenomenon may be limited in scope to specific verb structures; data from Table 27, as well

as the overall frequency of impersonal *se* in the SIT corpus, demonstrates that the form is still

actively used in oral discourse by U.S. Spanish speakers. Future investigation into the use of

the verb *decir* with 3PL R-impersonals, especially through a diachronic lens, would be useful

in evaluating whether there is a linguistic change in progress from preference of impersonal

se to 3PL R-impersonals in this communicative context. In examining a different verb form

with more 3PL uses than impersonal *se*, such as *dar*, it is possible to see how 3PL R-

impersonals lend a more subjective lens to a generalization:

46 (a) *quizá la educación que **se ha dado** en Reynosa no haya estado totalmente a la*

par con los fenómenos del progreso en Reynosa…   SIT 45

(b) *Sí, ahí nos **daban** una clase diaria de también…Todo era español en México,*

todo era español.   SIT 71

Each utterance in 46 discusses education. While the use of impersonal *se* through the *pasiva*

refleja in 46 (a) adds a degree of removal from the education and those who provided it, the

3PL R-impersonal in 46 (b) contrasts this complete objectivity. The 3PL "*daban*" in lieu of

"*se daba*" expresses less overall indefiniteness through its use of active voice with a plural conjugation.

    4.2.3.2 Unspecified Group Denotation. Use of 3PL R-impersonals for generalization outranked its use for denoting an unspecified group, yet this secondary function still constituted a healthy presence in the limited sample of data analyzed. Of the 200 analyzed sample, 68, or 34%, were utilized to denote an unspecified group of people. Used in a non-anaphoric sense, these R-impersonals functioned to background the agents of these actions while still implying their presence. Though no referents are explicitly mentioned, potential referents may be vaguely inferred from context. Example 46 shows a few examples of 3Pl R-impersonals used for unspecified groups:

    47 (a) *No llegué, me faltaba medio bloque para llegar, cuando me paró…'¿para dónde vas?' Pues voy a mi casa. '¿Pues dónde vives?' Ahí, vivo en el tres, quince y medio, Valverde. 'No, no, súbete.' Ah okay. Me* **llevaron** *a la cárcel, me* **tomaron** *las digital... las huellas, las huellas digitales y luego dice el Willy 'hijo, Raúl.' Ya iba allí el Willy, ya lo* **llevaban***. Le digo Willy 'te dije, vamos'. Anda, a pues para un taxi alcanza uno Raúl anda. 'Bueno, ándale pues. Ahora vamos a la cárcel, carnal'. 'Pues sí', dice 'y a mí me* **van** *a castigar feo ahora, verás…'*   SIT 34

    (b) *en San Luis, que es una ciudad tan tradicionalista y que las costumbres, muchas veces toman el lugar y de la creencia…***ellos** *te* **fuerzan** *a creer algo que no sientes o que no quieres y aquí tienes la decisión de descubrir lo que de verdad quieres…*   SIT 55

    (c) *y un señor de una estación de gasolina nos dijo, bueno si ustedes muchachos quieren ir allí al hospital, allá les reciben sangre y les pagan diez dólares por cada pinta, y nosotros con ese hambre vendimos sangre, el argentino se desmayó, a mí me*

> ***dieron*** *una copa de, de coñac o de vino, no sé qué era, para restablecerme porque,*
>
> *porque era muy difícil…* SIT 93

In 47 (a), there are several examples of 3PL R-impersonals used non-anaphorically. Although the context of these forms suggests that those responsible for stopping the speaker and taking him to jail were some sort of law enforcement, these agents are never explicitly mentioned. Rather, their presence is that of an unspecified group. This unspecified group sentiment is also present in 47 (b) through the use of 3PL "*fuerzan*." We are not certain if those responsible for forcing beliefs are the entire community of San Luis, members of the speaker's social circle, or a different group of referents altogether. At the same time, the scope of potential referents is limited somewhat in that they are known to the speaker. The utterance in 47 (b) also exemplifies how 3 PL R-impersonals may be used alongside overt use of the subject pronoun *ellos* as attested previously in Lapidus & Otheguy (2005). Although the frequency of overt *ellos* with 3Pl impersonals was very low in the SIT corpus (less than 10 occurrences total), it is present. Finally, the 3PL R-impersonal in 47 (c) loosely implies a referent of someone at the hospital, though their identity is not certain. This usage of impersonal "*dieron*" is interesting in that it coincides alongside other non-impersonal 3PL verbs; this example demonstrates how the identification of 3PL R-impersonals for unspecified group denotation is highly context-specific and requires great attention to detail for accurate identification. That these forms were less frequent than generalization 3PL R-impersonals is not surprising—speakers may prefer to use specific referents in narration to ensure that their message is clearly understood. Utilizing 3PL forms as reference impersonals, in turn, is a mechanism that can serve to background the agents responsible for the action and imply de-emphasis or less importance.

4.2.3.3 3PL R-impersonals, Impersonal se, and Verb Tense. Alternation between 3PL R-impersonals and impersonal *se* may be due to some degree to the extent to which

speakers wish to de-emphasize agenthood. Another factor that may influence selection of

3PL forms or impersonal *se* is verb tense (Gervasi, 2007; Morales, 1995). In her analysis of

monolingual Mexican and bilingual U.S. Spanish, Gervasi (2007) found an increased use of

3PL R-impersonals over impersonal *se* across the present, habitual past (imperfect) and non-

habitual past (preterit) for bilingual speakers. This difference was more notable in the two

past tenses, and Gervasi explains the preference for impersonal *se* in the present by proposing

that this form, neutral in gender and number, serves to better express a generality that may

also reflect speakers' personal opinions or experiences (p. 350). The overall increased use of

3PL R-impersonals, in turn, is proposed to be a result of contact with English and

convergence of impersonal *se* and 3PL forms onto the English impersonal "they" (Gervasi,

2007, pp. 350-351). Addressing Gervasi's call for additional analysis of the relationship

between 3PL R-impersonals and impersonal *se* with respect to verb tense, the 3PL and

impersonal *se* forms in SIT were coded for verb tense. Table 28 shows the distribution of

both forms across tense[10].

Table 28: *3PL R-impersonals and Impersonal se by Tense*

	Present	**Preterit**	**Imperfect**	**Total**
Impersonal *se*	700	71	131	902
3PL	529	137	270	936
Total	1229	208	401	1838

When looking at the present tense, impersonal *se* was used more frequently,

constituting roughly 60% of all present tense forms. However, 3PL R-impersonals outrank

impersonal *se* for both forms of the past; 3PLs constitute 65.8% of preterit forms and 67.3%

of impersonal forms. These results coincide with the findings in Gervasi (2007) in

demonstrating that 3PL R-impersonals are preferred by U.S. Spanish speakers when using

impersonality in the past. At the same time, the results of the current analysis show a less

---

[10] Only forms in the indicative mood were included in this sub-analysis.

drastic preference for 3PL than those of Gervasi (2007)—her analysis for bilingual U.S.

Spanish speakers indicated a 79.6% use of 3PL in the preterit and a 78.4% use of 3PL in the

imperfect in comparison to impersonal *se* (p. 349). However, the sample size of the data in

Gervasi's study was significantly lower than that of the current investigation in that she only

recorded 54 total uses of impersonal *se* and 3PL in the preterit and 37 in the imperfect. Given

a larger sample size, it is possible that the distribution of 3PL and impersonal *se* forms in the

past would approach a more normal distribution. While the verb tense analysis of 3PL and

impersonal *se* in the current analysis generally support Gervasi's findings regarding use of

3PL R-impersonals in U.S. Spanish, there is not full support for the idea that this form is

definitively replacing impersonal *se* or restricting its use (Morales, 1995). Rather, the two

forms are both actively used in U.S. Spanish as is recorded in the SIT, and further analysis is

needed to more completely explore their function and distribution.

  4.2.3.4 3PL Analysis and Conclusion. 3PL R-impersonal forms differ from their

2SG and 1PL counterparts in that they may not be used as a substitute for first-person

singular forms; these forms, by their nature, exclude the speaker and interlocutor (Cabredo

Hofherr, 2003; Fernández, 2008). Regardless, linguistic analysis has shown that these forms

may still be employed impersonally when they are not used anaphorically. Two chief

functions for 3PL R-impersonals proposed by Fernández (2013) are 1) operating in a

"membership area" (p. 98) to denote an unspecified agent whose affiliations may loosely be

inferred from context[11] or 2) working in an indefinite area in which the human agents are

completely unspecified and there are no discernable antecedents. The membership and

indefinite areas are distinct from a third space, a generic space, in which only first and

second-person R-impersonals may operate as a means to create generalizations.

---

[11] Fernández's membership area shows partial overlap with the corporate reading of R-impersonals proposed by Cabredo Hofherr (2003).

However, sample data from SIT shows various examples of 3PL R-impersonals that may take a generic, or generalized, interpretation, as demonstrated in 42-43. These utterances use 3PL R-impersonals in contexts that could also accept impersonal *se*. These results counter the proposition regarding the exclusion of 3PL R-impersonals from generalization in Fernández (2013) and raise the question of whether a pragmatic framework for R-impersonals should also include 3PL under mechanisms for generalization. It should be noted that much of the corpus data cited by Fernández (2013) comes from written sources, such as newspaper and magazine articles. As such, it is possible that these sources reflect a deference to impersonal *se* for generalization over 3PL based on use of written modality and a more formal genre. This would suggest that the pragmatic usage of R-impersonals and impersonal *se* may vary based on modality and genre. While the scope of the current investigation does not allow for consideration of multiple modalities and genres, these factors should be taken into consideration in future analyses of reference impersonality.

The following chapter will provide a summary of the current investigation and contextualize its relevance to the field. In addition, limitations of the study will be discussed, and future research opportunities examined.

Chapter 5: Conclusion

5.1 Introduction

The linguistic concept of impersonality encompasses a broad variety of forms and structures that serve, to varying degrees, a purpose of distancing or de-emphasis of the subject/agent in relation to the verbal action. Given the broad range of forms and functions, especially cross-linguistically, there is not currently a comprehensive definition for impersonality (Siewierska, 2008b), yet two primary characterizations have been placed forefront in the discussion. The subject-centered approach to impersonality emphasizes lack of or decreased properties of a syntactic subject. On the other hand, the agent-defocusing approach considers impersonal forms as those that may either reflect syntactic or semantic impersonality. This latter conceptualization lends itself to pragmatic and discourse analysis, and has received overall less scholarly attention than the subject-centered approach (Akinrẹmi, 2013). However, the agent-defocusing approach to impersonality serves as a vital reference point for considering reference impersonals (R-impersonals), or impersonal forms that employ *man*-type indefinite pronouns or subject pronouns in non-anaphoric capacities.

Analysis for agent-defocusing impersonality in Spanish has traditionally centered on the use of impersonal *se* (often referred to as *pasiva refleja*) or the indefinite pronoun *uno*. However, more recent analyses have shown that second-person singular, first-person plural, and even third-person plural pronouns and verbal forms have been employed impersonally in both oral and written capacities. Nevertheless, these studies are few in number and capacity of forms analyzed, and there is a lack of a comprehensive framework that considers the pragmatic function of these different R-impersonals. Additionally, much of the study of R-impersonals in Spanish centers on the Peninsular variety. In an effort to identify and examine the sociopragmatic role of R-impersonals in United States Spanish, the current study analyzed corpus data collected from Spanish speakers in Texas. This section will briefly summarize the results of the study, which addressed the following research questions:

1) **R1:** What is the frequency and distribution of different forms of impersonal expressions? Is there a statistically significant difference in the use of particular forms of impersonal expressions by gender of speaker, age range of the speaker, generation of speaker, or educational level of speaker?

2) **R2:** What is the relationship between the use of different impersonal expressions and their respective pragmatic functions, as evidenced in the corpus?

3) **R3:** Is there a statistically significant difference in the use of particular forms of impersonal expressions based on the tense/aspect/mood of verbal forms, the presence of hypothetical *if* clauses, or adverbs of time (genericity triggers)?

4) **R4:** How may the findings of this analysis inform current pedagogy for teaching impersonal forms in both second language and heritage Spanish?

In terms of addressing Research Question R1, data from the SIT corpus yielded 4,989 R-impersonal forms encompassing second-person singular, first-person plural, and third-person plural, as well as impersonal *se* and indefinite *uno*. These forms were submitted to quantitative and qualitative analysis. Significant associations between the use of distinct R-impersonal forms and the sociolinguistic variables of speaker gender, generation, age range, and educational status, were confirmed through chi-square analysis. In particular, women showed a greater tendency toward first-person plural R-impersonals than men, reflecting data found in studies of Peninsular Spanish (Posio, 2012). Men, on the other hand, showed a greater tendency toward stronger impersonality mechanisms such as indefinite *uno* and impersonal *se*, suggesting a difference in R-impersonal usage based on gender. Patterns in the use of first-person plural forms based on speaker generation and age range also indicate that older adults are more likely to utilize these forms, potentially reflecting speaker perception of community status, in comparison to younger adults. Despite prior claims regarding a decrease

in the use of impersonal *se* in United States Spanish (Morales, 1995), data collected in the

present analysis demonstrate its usage and firm presence in an oral context.

In addition to the examination of relationships between R-impersonal usage and

sociolinguistic variables, linguistic variables of verb tense and verb mood were also

examined (Research Question R1). Chi-square analysis confirmed statistically significant

relationships between R-impersonal usage based on verb tense and mood. Specifically, all

singular R-impersonal forms showed a greater likelihood of use in the present (both habitual

and non-habitual), whereas plural R-impersonals showed a greater likelihood of expression in

the past. While conversation topic most likely had some degree of influence over the R-

impersonal forms that speakers employed, the clear distinction in likelihood of singular and

plural R-impersonal based on the present/past distinction is significant. Weaker relationships

between R-impersonal usage and the presence of adverbs of time*, if* clauses, and overt

pronouns were also identified. Given that prior analyses suggesting a link between R-

impersonal usage and adverbs of time and *if* clauses were established based on analysis of

written text, modality differences are a potential explanation for differing results in the

current study.

The distinctions in use between the different R-impersonal forms were further

examined through a qualitative lens in which pragmatic functions of generalization, inclusive

defocalization, and speaker concealment were used (Research Question R3). In addition, the

rhetorical use and defocalized generalized use of first-person plural R-impersonals was

revealed in the data, suggesting an amplification of this form's utility in a pragmatic sense.

Furthermore, SIT data established that third-person plural forms were used to generalize, a

stance not supported in prior literature (Fernández, 2013). Taken together, these findings

demonstrate that second-person singular, first-person plural, and third-person plural R-

impersonal forms constitute vibrant mechanisms for impersonality in oral discourse of U.S. Spanish.

The remainder of this chapter is divided into four sections. The following two sections will discuss this study's relevance to the field of sociolinguistics, pragmatics, and pedagogy (this section addresses Research Question R4). After that, study limitations will be discussed. Finally, future research opportunities based on this study's results will be proposed.

5.2 Implications to the Field

As was briefly mentioned in 5.1, there is a dearth of research regarding the agent-defocusing approach to impersonality and a pragmatic analysis of R-impersonals in particular. While there is extant literature on agent-defocusing impersonality in Spanish, much of this research focuses on the Peninsular variety, and the forms included for analysis are the more traditional impersonal *se* and indefinite *uno*.

To date, the only other popularly-referenced scholarly literature on R-impersonals in U.S. Spanish lack important detail for full sociolinguistic consideration; both Gervasi (2007) and Morales (1995) include U.S. Spanish speakers in their analyses but do not provide relevant sociodemographic details such as speaker generation or language preference. Important work by Lapidus and Otheguy (2005) provides more complete speaker demographics but is also limited in scope in that it only examines third-person R-impersonals in a specific context.

This study, in turn, represents the first comprehensive analysis of reference impersonality, including in its scope not only impersonal *se* and indefinite *uno*, but second-person singular, first-person plural, and third-person plural forms. In addition, this study analyzes how all of these forms are employed in the Spanish of the United States, a language variety under-explored in discussions of impersonality. That this analysis has shown the widespread, grammatical use of innovative reference impersonal forms is an unarguably significant contribution to the study of impersonality. Whereas prior literature is hesitant to

accept the usage of certain forms, such as third-person plural with overt subjects (Cabredo Hofherr, 2006; Jaeggli, 1986), the current study confirms their usage in an impersonal capacity. Additionally, the use of a linguistic corpus for sociolinguistic study provides quantitative data supporting the validity of R-impersonal usage in U.S. Spanish. As technological innovation has made corpus research and analysis increasingly more accessible, this study responds to the call of Parodi (2007) for increased corpus research into Spanish, demonstrating the usefulness of this methodology in consideration of sociolinguistic variation.

Prior analyses of R-impersonals, such as those encompassed in Fernández (2013) or Cabredo Hofherr (2003) have asserted different pragmatic functions for the forms based on singular and plural properties. Fernández (2013) proposes a framework of generalization, group membership, or indefiniteness and divides R-impersonals into these spaces, with singular forms serving to generalize or show group membership and plural forms fulfilling group membership or indefinite roles. Under this framework, plural forms are not considered able to generalize. However, data from third-person plural R-impersonals in the current study show how these forms can actually be used to generalize, similarly to how generality is conveyed through impersonal *se*.

These findings call for a reconsideration of a comprehensive framework for R-impersonals. Based on the data from the current study, pragmatic functions of generalization, inclusive defocalization, focalized generalization, and speaker concealment should be considered for second person-singular, first-person plural, and third-person plural R-impersonals. Prior frameworks have been created for specific R-impersonals, such as that proposed for second-person singular by Pulido Astorga & Rivadeneira Valenzuela (2017) and third-person plural by Cabredo Hofherr (2006) and Siewierska (2011). While these frameworks may be useful for language typologies or cross-linguistic comparison, a single

pragmatic framework for all available R-impersonality mechanisms is vital for a more complete understanding of how and why speakers employ these forms. The results of the current analysis support the adoption of such a framework for classifying the pragmatic functions of Spanish R-impersonals. Considering the pragmatic categories of generalization, inclusive defocalization, focalized generalization, and concealment serves to aid in analysis of R-impersonal forms in both written and oral modalities, as well as a variety of text genres. In addition to aiding research, a comprehensive consideration of R-impersonal pragmatics would also serve language learners and bridge the gap in pedagogic knowledge currently in place.

5.3 Pedagogical Implications

The field of applied linguistics has acknowledged the role of pragmatics in both language use and language acquisition. Regarding language learning, a body of growing scholarly literature confirms the importance of pragmatic interpretation (sometimes considered as part of sociocultural competence) in second or heritage language acquisition (Alcón & Martínez-Flor, 2008; Canale & Swain, 1980; Liddicoat, 2014). Given that different R-impersonal forms may be used for distinct pragmatic purposes, such as communicating generalizations, building speaker-interlocutor empathy, or agent de-emphasis, inclusion of these forms in Spanish language pedagogy is critical. However, at present, discussions of impersonality in beginner and intermediate language texts fall short in introducing learners to different R-impersonals. A review of contemporary pedagogical materials for the second and mixed-language Spanish classroom has shown a focus on impersonal *se* with occasional reference to indefinite *uno* and possible the second-person singular; however, first and third-person R-impersonals are absent. Similarly, research on acquisition or use of these forms shows a preference for impersonal *se*.

The results of the current study show that U.S. Spanish speakers employ a variety of forms outside of impersonal *se* to express impersonality. As such, pedagogical materials and

classroom instruction should incorporate R-impersonal usage. In particular, Spanish language classrooms should include explicit lessons on how second-person singular, first-person plural, and third-person plural may be employed impersonally. Given that these three impersonal forms take the same surface forms as in the deictic uses learned by students, it would be simple to incorporate lessons on how they may be used impersonally alongside grammatical instruction. Instruction on the various uses of the clitic *se*, including its use for passive or impersonal means, is often included at the elementary and intermediate language level (Blanco, 2016, 2018; Blitt et al., 2015; Dominicis, 2014). Instruction on the use of *se* also provides a useful entry point for introducing R-impersonals, as these forms could be presented in conjunction with *se* as alternative language choices based on degree of impersonality. In this way, instruction of R-impersonals fits comfortably into language curricula and assists in pragmatic competence while also developing grammatical knowledge.

For Spanish-language learners whose first language is English, parallels between Spanish and English R-impersonals could also facilitate pragmatic acquisition of these forms. For instance, the Spanish second-person singular R-impersonal usage parallels that of the English *you*. Some studies even suggest that developing pragmatic awareness in a second or foreign language assists with intercultural awareness of one's native culture (Liddicoat, 2014), hence this sort of instruction with R-impersonals would support language learners in honing intercultural awareness and language identity.

Applications of R-impersonal forms should be incorporated in second and heritage language Spanish pedagogy. Given the availability of authentic language usage available through the internet and other free resources, instructors have the ability to present natural use of these forms in a variety of media (video, audio, and writing). At the same time, employing R-impersonals in target language use is a natural steppingstone for language learners as they progress through language acquisition and are able to move into more

complex communicative tasks, such as expressing opinions, making speculations, or generalizing.

While consideration of these forms will enhance pragmatic competence among second or foreign-language learners, their inclusion will also reinforce communicative competence among heritage populations. Prior approaches to teaching Spanish as a heritage language in the United States has taken an approach of correcting so-called 'defective' speech in favor of acquiring a standard variety (Beaudrie, 2015), but contemporary considerations recognize the need to validate students' home varieties and inherent language variation (Carreira, 2007). Illustrating how second, first, and third-person forms may be used both personally and impersonally in Spanish fits within a variationist approach favored for heritage language education. It is possible that heritage language learners employ these R-impersonal forms in their own speech habits, and showing how these forms, which were previously considered non-standard, actually fulfill pragmatic functions, will further demonstrate the validity of all language varieties. Many heritage language classrooms seek to assist learners in developing awareness of formal and informal registers (Beaudrie, 2015; Leeman, 2005), and instruction of R-impersonals demonstrates one avenue for this distinction. For instances, students may learn about how some impersonality mechanisms may be more common in formal speech, while others appear more frequently in informal contexts. These brief examples contextualize how instruction of R-impersonals fit within existing curricula for second and heritage language learners. The instruction of R-impersonals assists not only with acquisition of formal structures but also with communicative competence, pragmatic acquisition, and recognition of language varieties.

5.4 Study Limitations

The current investigation makes a significant contribution to the understanding of reference impersonal usage in U.S. Spanish. However, there are a number of limitations to the study that should be addressed. Firstly, the data coding of R-impersonal forms was largely

manual, and this limited the number of samples that could be included in the final analysis. Manual data coding was largely used due to two factors: 1) the highly-context dependent usage of R-impersonals (making these forms difficult to automatically code) and 2) the unavailability of additional resources (electronic and human) to assist with R-impersonal classification. Given additional resources to assist in data coding, such as access to machine learning technology or assistance from other individuals, the use of these forms could be considered in a greater number of language samples.

Restrictions of sample size due to data complexity coincides with a further limit to this study, which is that of a restricted modality and genre. Oral data from sociolinguistic corpora was chosen as the sole text type for the present analysis. Restricting modality and genre allowed for analysis of similar conversation topics among different speakers; this homogeneity of conversation topic is also a negative restricting factor, however, in that there were simply certain discourse topics that did not surface throughout the interviews. Including additional types of oral discourse, such as speeches, standard interviews, and spontaneous conversation, would allow for an examination of genre on R-impersonal usage, and opening consideration to written text as well would further explore R-impersonal usage in all language dimensions.

Finally, the restricted sample size from SIT also resulted in an unequal distribution of speakers by generation; in particular, third and fourth generation U.S. Spanish speakers were greatly underrepresented in the current analysis. Demographic information from the US. Census indicates that over half of the total U.S. population growth between 2010 and 2020 came from an increase in the population of U.S. Hispanics (Manuel Krogstad & Noe-Bustamente, 2021), and U.S. born-Spanish speakers aged 18-29 constitute the second-largest group of U.S. Latinos (Noe-Bustamente & Flores, 2019). Considering that many of these younger bilinguals are third and fourth generation Spanish speakers, including data from this

demographic will further elucidate the language habits of U.S. Spanish speakers with respect

to impersonality.

5.5 Future Research Possibilities

The current study serves as a significant source for quantitative and qualitative

analysis of R-impersonals in U.S. Spanish. At the same time, it is not all-inclusive and

provides further avenues for important research into how and why speakers engage with

impersonality. There are several existing studies of impersonality mechanisms in different

written text genres (Fernández, 2013; Lamas, 2015); future studies should also consider

further expansion of *oral* genres to examine differences in R-impersonal usage. By

comparing how R-impersonal usage differs across distinct genres in the oral modality, we

may better understand just how genre contributes to the types of forms used and purposes that

they employ. It would be particularly interesting to expand analysis of R-impersonal usage to

electronic media resources, such as videos and streamed content posted and disseminated on

social media.

Additionally, further research into R-impersonal usage specifically among U.S.

Spanish speakers is warranted. While there is documented analysis of R-impersonal usage for

Peninsular Spanish (De Cock, 2021; Posio, 2012, 2013, 2016, 2017), Colombian Spanish

(Dieck, 2016; Hurtado, 2001, 2015), and Chilean Spanish (Pulido Astorga & Rivadeneira

Valenzuela, 2017), there are less extant studies on U.S. Spanish (Gervasi, 2007; Lapidus &

Otheguy, 2005; Morales, 1995). The current study analyzes Spanish spoken by speakers in

Texas, but an exploration of R-impersonality usage in other regions of the country, such as

the West, Midwest, Upper Midwest, or Northeast, could further illuminate potential dialectal

variation. Further development and exploration of Spanish language corpora collected within

the United States will also strengthen the tradition of corpus linguistics and amplify the

prominence of U.S. Spanish at a global level.

A third avenue for further exploration of R-impersonals includes looking at the presence and function of impersonal object clitics, such as *te* (referring to second-person singular) or *nos* (referring to first-person plural). These forms may be present to represent impersonality when their own accompany R-impersonal verbal forms are not and should be considered as well.

Finally, it would be useful to consider the phenomenon of switch-reference or deictic change in reference to impersonality. Throughout the course of an enunciation, which may span sentences, speakers may employ different personal and impersonal forms, signaling a deictic shift in addition to impersonality. Analyzing which impersonal form(s) are most included in instances of switch-reference would contextualize this language tendency within a sociopragmatic communicative framework.

The pragmatic use of R-impersonals in Spanish is a vibrant, pertinent topic that is ripe for future investigation, especially in terms of their usage in U.S. Spanish. Those interested in sociolinguistics, pragmatics, and language pedagogy will benefit from continued research into this phenomenon by understanding how these forms function in both written and oral discourse.

REFERENCES

Abu-Bader, S. H. (2016). *Using statistical methods in social science research with a complete SPSS guide*. Oxford University Press.

Aijon Oliva, M., & Serrano, M. J. (2013). *Style in syntax: Investigating Variation in Spanish Pronoun Subjects (Linguistic Insights Studies in Language & Communication)*. Peter Lang Publishing.

Ajión Oliva, M. (2018). Not just you: The construction of radio audiences through second-person choice in Peninsular Spanish. *Language & Communication*, *60*, 80–93. https://doi.org/10.1016/j.langcom.2018.02.002

Ajión Oliva, M. (2020). It can be us or you. The desubjectification of viewpoint through person choice in Spanish oral and written media discourse. *Journal of Pragmatics*, *163*, 4–17. https://doi.org/10.1016/j.pragma.2020.04.009

Akinrẹmi, I. I. (2013). Impersonal constructions in Igbo. *Theory and Practice in Language Studies*, *3*(7), 1129–1135. https://doi.org/10.4304/tpls.3.7.1129-1135

Alcón, E., & Martínez-Flor, A. (2008). Introduction. In E. Alcón & A. Martínez-Flor (Eds.), *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing* (pp. 3–21). Multilingual Matters Ltd.

Alonso Alonso, M. (2011). Un estudio comparativo de la impersonalidad en el lenguaje académico a través de la UAM Corpus Tool. *Hesperia: Anuario de Filología Hispánica*, *14*(2), 23–38.

Alonzo, F. (2020). *Hispanic Heritage Month 2020* (No. CB20-FF.07; Profile America Facts for Features, p. 3). United States Census Bureau. https://www.census.gov/content/dam/Census/newsroom/facts-for-features/2020/CB20-FF.07.pdf

Anderson, G. (2010). How to use corpus linguistics in sociolinguistics. In A. O'Keeffe & M.

    McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 547–562).

    Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W.

    Spooren (Eds.), *Text Representation: Linguistic and psycholinguistic aspects* (pp. 29–

    87). John Benjamins Publishing Company.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press;

    ProQuest Ebook Central.

Barberà, G., & Cabredo Hofherr, P. (2017). Backgrounded agents in Catalan Sign Language

    (LSC): Passives, middles, or impersonals? *Language*, *93*(4), 767–798.

    https://doi.org/10.1353/lan.2017.0057

Bauman, K. (2017). *School enrollment of the Hispanic population: Two decades of growth*.

    United States Census Bureau. https://www.census.gov/newsroom/blogs/random-

    samplings/2017/08/school_enrollmentof.html

Beaudrie, S. (2015). Approaches to language variation: Goals and objectives of the Spanish

    heritage language syllabus. *Heritage Language Journal*, *12*(1), 1–21.

Berry, R. (2009). You could say that: The generic second-person pronoun in modern English.

    English Today, *25*(3), 29–34. https://doi.org/10.1017/S0266078409990368

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis.

    Discourse Processes, *15*(2), 133–163. https://doi.org/10.1080/01638539209544806

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language*

    Structure and Use. Cambridge University Press.

Biber, D., Davies, M., Jones, J. K., & Tracy-Ventura, N. (2006). Spoken and written register

variation in Spanish: A Multi-Dimensional analysis. *Corpora : Corpus-Based

Language Learning, Language Processing and Linguistics.*, *1*(1), 1–37.

Blanco, J. (2016). *Portales 1*. Vista Higher Learning.

Blanco, J. (Ed.). (2018). *Portales 2*. Vista Higher Learning.

Blevins, J. (2003). Passives and impersonals. *Linguistics*, *39*(3), 473–520.

https://doi.org/10.1017/S0022226703002081

Blitt, M. A., Casas, M., & Copple, M. T. (Eds.). (2015). *Exploraciones: Curso intermedio*.

Cengage.

Borthen, K. (2010). On how we interpret plural pronouns. *Journal of Pragmatics*, *42*(7),

1799–1815. https://doi.org/10.1016/j.pragma.2009.02.008

Cabañas Maya, R. (2006). La animacidad como factor de variación entre la pasiva

perifrástica y la pasiva refleja. *Anuario de Letras. Lingüística y Filología*, *44*, 125–

145.

Cabredo Hofherr, P. (2003). Arbitrary readings of 3PL pronominals. *Proceedings of the

Conference "Sub7-Sinn Und Bedeutung,"* 81–94.

Cabredo Hofherr, P. (2006). "Arbitrary" pro and the theory of pro-drop. In P. Ackema, P.

Brandt, M. Schoorlemmer, & F. Weerman (Eds.), *Arguments and Agreement* (pp.

230–257). Oxford University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second

language teaching and testing. *Applied Linguistics*, *1*(1), 1–47.

https://doi.org/10.1093/applin/l.1.1

Cardona, M. D. (2015). La enseñanza de las construcciones pasivas e impersonales con se en

E/LE. ¿Cuántas distinciones son necesarias? *Didáctica: Lengua y Literatura*, *27*, 73–

96. https://doi.org/10.5209/rev_DIDA.2015.v27.51042

Carreira, M. (2007). Teaching Spanish in the U.S.: Beyond the one-size-fits-all paradigm. In K. Potowski & R. Cameron (Eds.), *Spanish in Contact: Policy, Social and Linguistic Inquiries* (pp. 61–79). John Benjamins Publishing Company.

Chandler, P. (1996). Idea: VOCES: A Mnemonic Device to Cue Mood Selection after Impersonal Expressions. *Hispania*, *79*(1), 126–128. https://doi.org/10.2307/345627

Chaves, M. P. (2006). Los pronombres reflexivos clíticos como operadores de destransitivación en español. *Revista de Filología y Linguística de La Universidad de Costa Rica*, *32*(2), 185–201. https://doi.org/10.15517/rfl.v32i2.4296

Cogen, C., & Herrmann, L. (1975). Interactions of the Expression "Let's Just Say" with the Gricean Maxims of Conversation. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, *1*, 60–68. https://doi.org/10.3765/bls.v1i0.2330

Corpas Pastor, G. (2021). Constructional idioms of 'insanity' in English and Spanish: A corpus-based study. *Lingua*, *254*, 1–20. https://doi.org/10.1016/j.lingua.2020.103013

Dahl, K. (2017). *Uno nunca sabe lo que se puede encontrar: Las construcciones impersonales con referencia humana en la traducción española de la novela negra Det som inte dödar oss de David Lagercrantz* [Bachelor's Thesis, Stockholms universitet]. DiVA Digitala Vetenskapliga Arkivet.

De Cock, B. (2011). Why we can be you: The use of 1st person plural forms with hearer reference in English and Spanish. *Journal of Pragmatics*, *43*, 2762–2775. https://doi.org/10.1016/j.pragma.2011.04.009

De Cock, B. (2021). Intersubjective impersonals in context: A multivariate analysis of Spanish uno and se in spoken language. *Catalan Journal of Linguistics*, *20*, 93–114. https://doi.org/10.5565/rev/catjl.310

den Dikken, M. (2011). Phi-feature inflection and agreement: An introduction. *Natural Language & Linguistic Theory*, *29*, 857–874. https://doi.org/10.1007/s11049-011-9156-y

Deutschmann, M. (2006). Social Variation in the use of apology formulae in the British National Corpus. In A. Renouf & A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics* (pp. 205–222). Rodopi.

Díaz Blanca, L. A. (2005). Oraciones impersonales en el habla de Mérida (Venezuela): Varias configuraciones, un fenómeno. *Boletín de Lingüística*, *24*, 33–54.

Dieck, M. (2016). ¿Y qué más hace uno, pues? La expresión de la impersonalidad en el español de Medellín. *Lingüística y Literatura*, *69*, 145–175. https://doi.org/10.17533/udea.lyl.n69a06

Dominicis, M. C. (Ed.). (2014). *Repase y escriba: Curso avanzado de gramática y composición* (7th ed.). John Wiley & Sons, Inc.

Enrique-Arias, A. (2016). Sobre la noción de perspectiva en lingüística de corpus: Algunas ventajas de los corpus paralelos. In J. Kabatek (Ed.), *Lingüística de corpus y lingüística histórica iberorrománica* (pp. 21–39). Walter de Gruyter.

Espinoza, A. M. (1997). Contrastive analysis of the Spanish and English passive voice in scientific prose. *English for Specific Purposes*, *16*(3), 229–243. https://doi.org/10.1016/S0889-4906(96)00034-8

Feeny, T. (1978). Using classroom objects in teaching the subjunctive with impersonal expressions. *Hispania*, *61*(2), 328–329. https://doi.org/10.2307/340880

Fernández, S. (2008). Generalizaciones y evidencialidad en español. *Revue Romane*, *43*(2), 217–234. https://doi.org/10.1075/rro.43.2.04fer

Fernández, S. (2013). Impersonality in Spanish personal pronouns. In K. Jeppersen Kragh & J. Lindschouw (Eds.), *Deixis and pronouns in Romance languages* (pp. 87–107). John Benjamins Publishing Company.

Fishman, J. (1964). Language maintenance and language shift as a field of inquiry: A definition of the field and suggestions for its further development. *Linguistics*, *2*(9), 32–70. https://doi.org/10.1515/ling.1964.2.9.32

Fratini, V., Acha, J., & Laka, I. (2014). Frequency and Morphological Irregularity Are Independent Variables. Evidence from a Corpus Study of Spanish Verbs. *Corpus Linguistics and Linguistic Theory*, *10*(2), 289–314. https://doi.org/10.1515/cllt-2013-0028

Fried, M. (2006). Agent back-grounding as a functional domain: Reflexivization and passivization in Czech and Russian. In B. Lyngfelt & T. Solstad (Eds.), *Demoting the Agent: Passive, middle, and other voice phenomena* (pp. 83–109). John Benjamins Publishing Company.

García, P. (2007). Pragmatics in academic contexts: A spoken corpus study. In M. C. Campoy & M. J. Luzón (Eds.), *Spoken Corpora in Applied Linguistics* (pp. 97–126). Peter Lang Publishing.

Gelabert-Desnoyer, J. (2008). Not so impersonal: Intentionality in the use of pronoun uno in contemporary Spanish political discourse. *Pragmatics*, *18*(3), 407–424. https://doi.org/10.1075/prag.18.3.03gel

Gervasi, K. (2007). The use of Spanish imersonal forms in monolingual and bilingual speech. *Hispania*, *90*(2), 342–353.

Gil-Vallejo, L., Castellón, I., & Coll-Florit, M. (2018). Verb similarity: Comparative analysis between theoretical linguistics and corpus data. *Revista Signos*, *51*(98), 310–332. https://doi.org/10.4067/S0718-09342018000300310

Giordano, R. (2015). Las oraciones impersonales: Un análisis comparativo español-italiano. *Estudios Románticos*, *24*, 145–156.

Gómez Torrego, L. (2013). *La impersonalidad gramatical: Descripción y norma* (4th ed.). Arco Libros S. L.

Guirado, K. (2011). Uso impersonal de tú y uno en el habla de Caracas y otras ciudades. *Círculo de Lingüística Aplicada a La Comunicación*, *47*, 3–27. https://doi.org/10.5209/rev_CLAC.2011.v47.39017

Hermon, G. (2001). Non-canonically marked A/S in Imbabura Quechua. In A. Aikhenvald, R. M. W. Dixon, & M. Onishi (Eds.), *Non-canonical marking of subjects and objects* (Vol. 46, pp. 149–176). John Benjamins Publishing Company.

Hernández, R. M. (2012). Teaching subject position in Spanish: Gustar verbs, unaccusatives and the impersonal passive. *Journal of Language Teaching and Research*, *3*(2), 321–330. https://doi.org/10.4304/jltr.3.2.231-330

Hidalgo Navarro, A. (1996). Sobre los mecanismos de impersonalización en la conversación coloquial: El tú impersonal. *Estudios de Lingüística: Universidad de Alicante*, *11*, 163–176. https://doi.org/10.14198/ELUA1996-1997.11.08

Hunston, S. (2002). Introduction to a corpus in use. In *Corpora in Applied Linguistics* (pp. 3–24). Cambridge University Press.

Hurtado, L. M. (2001). *La variable expresión del sujeto en el español de los Colombianos y Colombo -Americanos residentes en el condado de Miami -Dade* [Doctoral Dissertation]. University of Florida.

Hurtado, L. M. (2015). El predominio de uno en Bogotá, Colombia: ¿Estrategia para desfocalizar o para subjetivizar? *Anuario de Letras. Lingüística y Filología*, *3*(2), 105–151.

Ibáñez, C. V. (2018). Primera persona del plural en los juicios orales. Valor representativo y estrategia atenuante. *RILCE: Revista de Filología Hispánica*, *34*(3), 1056–1080. https://doi.org/10.15581/008.34.3.1056-80

Jaeggli, O. (1986). Arbitrary Plural Pronominals. *Natural Language & Linguistic Theory*, *4*(1), 43–76. https://doi.org/10.1007/BF00136264

Jørgensen, A. M. (2013). Spanish teenage language and the COLAM-corpus. *Bergen Language and Linguistics Studies*, *3*(1), 151–166. https://doi.org/10.15845/bells.v3i1.368

Keenan, E. (1985). Passive in the world's languages. In T. Shopen (Ed.), *Clause Structure* (Vol. 1, pp. 325–361). CUP.

Kerstens, J. (1993). Agreement. In *The syntax of number, person and gender: A theory of phi-features* (pp. 43–69). Mouton de Gruyter.

Kitagawa, C., & Lehrer, A. (1990). Impersonal uses of personal pronouns. *Journal of Pragmatics*, *14*(5), 739–759. https://doi.org/10.1016/0378-2166(90)90004-W

Kluge, B. (2016). Generic uses of the second person singular—How speakers deal with referential ambiguity and misunderstandings. *Pragmatics*, *26*(3), 501–522. https://doi.org/10.1075/prag.26.3.07klu

Lamas, A. (2015). Análisis de la impersonalidad en textos jurídicos. Estudio aproximativo de Carteles y Edictos venezolanos. *Revista Miradas*, *1*(13), 100–113. https://doi.org/10.22517/25393812.12261

Lapidus, N., & Otheguy, R. (2005). Overt nonspecific Ellos in Spanish in New York. *Spanish in Context*, *2*(2), 157–174. https://doi.org/10.1075/sic.2.2.03lap

Leeman, J. (2005). Engaging critical pedagogy: Spanish for native speakers. *Foreign Language Annals*, *38*(1), 35–45. https://doi.org/10.1111/j.1944-9720.2005.tb02451.x

Liddicoat, A. (2014). Pragmatics and intercultural mediation in intercultural language learning. *Intercultural Pragmatics*, *11*(2), 259–277. https://doi.org/10.1515/ip-2014-0011

Light, C., & Wallenberg, J. (2015). The expression of impersonals in Middle English. *English Language and Linguistics*, *19*(2), 227–245. https://doi.org/10.1017/S1360674315000076

Lope Blanch, J. (1990). *El español hablado en el suroeste de los Estados Unidos: Materiales para su estudio*. Universidad Autónoma de México.

Malamud, S. A. (2012). Impersonal indexicals: One, you, man, and du. *Journal of Comparative Germanic Linguistics*, *15*(1), 1–48. https://doi.org/10.1007/s10828-012-9047-6

Malchukov, A., & Ogawa, A. (2011). Towards a typology of impersonal constructions: A semantic map approach. In A. Malchukov & A. Siewierksa (Eds.), *Impersonal constructions: A cross-linguistic perspective* (pp. 19–56). John Benjamins Publishing Company.

Malchukov, A., & Siewierksa, A. (Eds.). (2011). *Impersonal Constructions: A cross-linguistic perspective*. John Benjamins Publishing Company; Proquest Ebook Central.

Manuel Krogstad, J., & Noe-Bustamente, L. (2021). *Key facts about U.S. Latinos for National Hispanic Heritage Month*. Pew Research Center. https://www.pewresearch.org/fact-tank/2021/09/09/key-facts-about-u-s-latinos-for-national-hispanic-heritage-month/

Mautner, G. (2012). Corpora and Critical Discourse Analysis. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 32–46). Continuum.

McCarthy, M. (2002). Good Listenership Made Plain: British and American Non-minimal Response Tokens in Everyday Conversation. In R. Reppen, S. Fitzmaurice, & D.

Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 73–90). John

Benjamins Publishing Company.

Mendikoetxea, A. (2008). Clitic impersonal constructions in Romance: Syntactic features and

semantic interpretation. *Transactions of the Philological Society*, *106*(2), 290–336.

https://doi.org/10.1111/j.1467-968X.2008.00210.x

Morales, A. (1995). The loss of the Spanish impersonal particle se among bilinguals: A

descriptive profile. In C. Silva-Corvalán (Ed.), *Spanish in the four continents: Studies

in language contact and bilingualism* (pp. 148–162). Georgetown University Press.

Moreno Castrillón, F. (2013). *Gramática y ortografía al día: Guía con las principales

novedades de la Real Academia Española*. Universidad del Norte.

Munuera, K. (2018). *Mecanismos de impersonalidad en el español mexicano: Un análisis de

epístolas de los siglos XVI, XVII y XVIII* [MA Thesis]. University of Houston.

Nieuwenhuijsen, D. (2016). Notas sobre la aportación de análisis estadístico a la lingüística

de corpus. In J. Kabatek (Ed.), *Lingüística de corpus y lingüística histórica

iberorrománica* (pp. 215–237). Walter de Gruyter.

Noe-Bustamente, L., & Flores, A. (2019). *Facts on Latinos in the U.S.* Pew Research Center.

Oca, E. (1914). El pronombre se en nominativo. In *Boletín de la Real Academia Española* (p.

573'581). La Real Academia Española.

Palacios, S. P., & Olivares, Y. G. (2015). Teaching Spanish three passive constructions

versus English passive voice. *Asian Journal of Education and E-Learning*, *3*(6), 387–

391.

Parodi, G. (Ed.). (2007). *Working with Spanish Corpora*. Continuum; ProQuest Ebook

Central.

Posio, P. (2012). Who are "we" in spoken Peninsular Spanish and European Portuguese?

Expression and reference of first person plural subject pronouns. *Language Sciences*,

34, 339–360. https://doi.org/10.1016/j.langsci.2012.02.001

Posio, P. (2013). The expression of first-person-singular subjects in spoken Peninsular

Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia

Linguistica*, *47*(1), 253–292. https://doi.org/10.1515/flin.2013.010

Posio, P. (2016). You and we: Impersonal second person singular and other referential

devices in Spanish sociolinguistic interviews. *Journal of Pragmatics*, *99*, 1–16.

https://doi.org/10.1016/j.pragma.2016.04.014

Posio, P. (2017). Entre lo impersonal y lo individual: Estrategias de impersonalización

individualizadoras en el español y portugués europeos. *Spanish in Context*, *14*(2),

209–229. https://doi.org/10.1075/sic.14.2.03pos

Pulido Astorga, P., & Rivadeneira Valenzuela, M. (2017). "En la vida teníh que luchar para

salir adelante". Variación pragmático-discursiva y sociolingüística en los usos no

deícticos de la segunda persona del singular en el español de Chile. *Onomázein:

Revista de Lingüística, Filología y Traducción*, *37*(3), 16–40.

https://doi.org/10.7764/onomazein.37.03

Romero, J. R., & Ormazabal, J. (2019). The formal properties of non paradigmatic "se."

Borealis (Tromsø), *8*(1), 55–84. https://doi.org/10.7557/1.8.1.4704

Said-Mohand, A. (2007). Discurso tú sabes en el habla de jóvenes bilingües estadounidenses.

Southwest Journal of Linguistics, *26*(2), 67–93.

Serrano, M. J. (2013). El pronombre tú como recurso objetivador en español: Variación

textual y discursiva. *Borealis: An International Journal of Hispanic Linguistics*, *2*(1),

179–197. https://doi.org/10.7557/1.2.1.2519

Serrano, M. J. (2018). Deictic and desubjectivizing meaning in advertising discourse: The

    usage of Spanish te and se clitics. *Lingua*, *216*, 28–46.

    https://doi.org/10.1016/j.lingua.2018.10.005

Serrano, M. J., & Ajión Oliva, M. (2014). Discourse objectivization, social variation and

    style of Spanish second-person singular tú. *Folia Linguistica*, *48*(1), 225–254.

    https://doi.org/10.1515/flin.2014.007

Siewierska, A. (2008a). Introduction: Impersonalization from a subject-centred vs. Agent-

    centred perspective. *Transactions of the Philological Society*, *106*(2), 115–137.

    https://doi.org/10.1111/j.1467-968X.2008.00211.x

Siewierska, A. (2008b). Ways of impersonalizing: Pronominal vs. Verbal strategies. In M. de

    los Ángeles Gómez-González, L. Mackenzie, & E. González Alvárez (Eds.), *Current

    Trends in Contrastive Linguistics* (pp. 3–26). John Benjamins Publishing Company.

Siewierska, A. (2011). Overlap and complementarity in reference impersonals: Man-

    constructions vs. Third-person plural-impersonals in the languages of Europe. In A.

    Malchukov & A. Siewierksa (Eds.), *Impersonal Constructions: A cross-linguistic

    perspective* (pp. 57–90). John Benjamins Publishing Company.

Siewierska, A., & Papastahi, M. (2011). Towards a typology of third person plural

    impersonals. *Linguistics*, *49*(3), 575–610. https://doi.org/10.1515/ling.2011.018

Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Clarendon

    Press.

Silva-Corvalán, C., & Enrique-Arias, A. (2017). *Sociolingüística y pragmática del español*

    (2nd ed.). Georgetown University Press.

Solé, Y. R. (1990). Bilingualism: Stable or transitional? The case of Spanish in the United

    States. *International Journal of the Sociology of Language*, *84*, 35–80.

    https://doi.org/10.1515/ijsl.1990.84.35

Speshilova, Y., & Gulyás, N. F. (2014). Impersonals and passives in contemporary Udmurt. *Finnisch-Ugrische Mitteilungen*, *38*, 59–91.

Stewart, M. (1999). *The Spanish language today*. Routledge.

Suñer, M. (1976). Demythologizing the impersonal "se" in Spanish. *Hispania*, *59*(2), 268–275. https://doi.org/10.2307/339502

Szmrecsanyi, B. (2017). Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *The Canadian Journal of Linguistics*, *62*(4), 685–701. https://doi.org/10.1017/cnj.2017.34

Th. Gries, S. (2009). Useful statistics for corpus linguistics. *Linguistics*, 5–27.

Togini Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14–27). Routledge.

Tolchinsky, L., & Rosado, E. (2005). The effect of literacy, text type, and modality on the use of grammatical means for agency alternation in Spanish. *Journal of Pragmatics*, *37*(2), 209–237. https://doi.org/10.1016/j.pragma.2004.08.008

Tremblay, A. (2005). The L2 Acquisition of Spanish passive and impersonal se by French- and English-speaking adults. In L. Dekydtspotter et al. (Ed.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* (pp. 251–268). Cascadilla Proceedings Project.

Viana, V., Zyngier, S., & Barnbrook, G. (Eds.). (2011). Principles and applications of corpus linguistics. In *Perspectives on corpus linguistics* (pp. 155–170). John Benjamins Publishing Company.

Vila, M. R. (1987). La segunda persona gramatical en función no deíctica. *Revista Española de Lingüística*, *1*(1), 57–68.

Westphal, G. F. (1999). Sobre las oraciones impersonales y pasivas con 'se.' *Onomazein :*

    Revista de Linguística y Traducción Del Instituto de Letras de La Pontificia

    Universidad Católica de Chile, *4*, 155–167.

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre

    awareness, linguistic knowledge, and writing competence. *Journal of Second*

    Language Writing, *20*(2), 111–133. https://doi.org/10.1016/j.jslw.2011.03.001

APPENDIX A: DATA CODING KEY

Dependent Variable: Impersonal Form

Impersonal Form	Code Value
Impersonal *se*	1
Uno	2
2SG	3
1PL	4
3PL	5

Independent Variables

1. Pronoun option for 2SG, 1PL, and 3PL

Circumstance	Code Value
Overt	0
Non-overt	1

2. Verb Tense

Verb Tense	Code Value
Present	1
Past: Preterit	2
Past: Imperfect	3
Future	4
Preterit Perfect	5
Past Perfect	6
Conditional	7

3. Verb Mood

Verb Mood	Code Value
Indicative	1
Subjunctive	2

4. Adverbs of Time

Context	Code Value
Present (clause level)	0
Not present	1

5. If Clauses

Context	Code Value
Present (clause level)	0
Not present	1

6. Speaker Gender

Gender	Code Value
Male	1
Female	2

7. Speaker Generation

Generation	Code Value
1	1
2	2
3	3
4+	4

8. Speaker Age Range

Age Range	Code Value
18-35	1
36-54	2
55+	3

9. Speaker Educational Level

Educational Level	Code Value
Elementary	1
High School / GED	2
Associates / Technical	3
Bachelors	4
Postgraduate (MA, JD, MD, etc.)	5

APPENDIX B: PYTHON CODES FOR DATA CLEANING

Code Set B1:

```
#For finding the verbs associated with "se"#

import openpyxl
from nltk import word_tokenize
import re

path = "C:/Users/Ally/Desktop/U of H/D/Data/CESA/CESA_WSTags_files_se_concordance
list.xlsx"
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active

rownum = 7
for i in sheet_obj:
    sephraselist = []
    cell_obj = sheet_obj.cell(row = rownum, column = 2)
    cell_obj2 = word_tokenize(cell_obj.value)
    for i in range(len(cell_obj2)):
        cell_obj2[i] = cell_obj2[i].lower()
    seindex = cell_obj2.index('se')
    for j in cell_obj2[seindex:]:
        sephraselist.append(j)
    if 'vlfin' in sephraselist:
        Vindex = sephraselist.index('vlfin')
        print(rownum, sephraselist[Vindex+2])
        rownum = rownum + 1
    elif 'vmfin' in sephraselist:
        Vindex = sephraselist.index('vmfin')
        print(rownum, sephraselist[Vindex+2])
        rownum = rownum + 1
    elif 'vefin' in sephraselist:
        Vindex = sephraselist.index('vefin')
        print(rownum, sephraselist[Vindex+2])
        rownum = rownum + 1
    elif 'vhfin' in sephraselist:
        Vindex = sephraselist.index('vhfin')
        print(rownum, sephraselist[Vindex+2])
        rownum = rownum + 1
    else:
        print(rownum, "**PROBLEM**")
        rownum = rownum + 1
```

Code Set B2:

```
#Finding uno verbs#

import openpyxl
from nltk import word_tokenize
import re
```

```
path = "C:/Users/Ally/Desktop/U of H/D/Data/SITCLEAN/Data/lva.xlsx"
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active

rownum = 5107
for i in sheet_obj:
    cell_obj = sheet_obj.cell(row = rownum, column = 3)
    cell_obj2 = word_tokenize(cell_obj.value)
    for i in range(len(cell_obj2)):
        cell_obj2[i] = cell_obj2[i].lower()
    if 'vlfin' in cell_obj2:
        Vindex = cell_obj2.index('vlfin')
        print(rownum, cell_obj2[Vindex+2])
        rownum = rownum + 1
    elif 'vmfin' in cell_obj2:
        Vindex = cell_obj2.index('vmfin')
        print(rownum, cell_obj2[Vindex+2])
        rownum = rownum + 1
    elif 'vefin' in cell_obj2:
        Vindex = cell_obj2.index('vefin')
        print(rownum, cell_obj2[Vindex+2])
        rownum = rownum + 1
    elif 'vhfin' in cell_obj2:
        Vindex = cell_obj2.index('vhfin')
        print(rownum, cell_obj2[Vindex+2])
        rownum = rownum + 1
    elif 'vsfin' in cell_obj2:
        Vindex = cell_obj2.index('vsfin')
        print(rownum, cell_obj2[Vindex+2])
        rownum = rownum + 1
    else:
        print(rownum, "**PROBLEM**")
        rownum = rownum + 1
```

Code Set B3:
```
#looking for the presence of explicit pronouns in clauses#

import openpyxl
from nltk import word_tokenize
import re

path = "C:/Users/Ally/Desktop/U of H/D/Data/SITCLEAN/Data/lva.xlsx"
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active

rownum = 2

for i in sheet_obj:
    cell_obj = sheet_obj.cell(row = rownum, column = 3)
    cell_obj2 = word_tokenize(cell_obj.value)
```

```
for i in range(len(cell_obj2)):
    cell_obj2[i] = cell_obj2[i].lower()
if 'tú' in cell_obj2:
    print(rownum, 'yes', 'tú')
    rownum = rownum + 1
elif 'nosotros' in cell_obj2:
    print(rownum, 'yes', 'nosotros')
    rownum = rownum + 1
elif 'ellos' in cell_obj2:
    print(rownum, 'yes', 'ellos')
    rownum = rownum + 1
else:
    print(rownum, 'no', "N/A")
    rownum = rownum + 1
```

Code Set B4:
```
#adverbs of time#
adverbsoftime = ['hoy', 'actualmente', 'ahora', 'anoche', 'anteriormente', 'antes', 'antiguamente',
'asiduamente', 'aún', 'ayer', 'constantemente', \
            'contemporáneamente', 'cuando', 'desde', 'después', 'enseguida', 'entretanto',
'eternamente', 'finalmente', 'frecuentemente', \
            'inicialmente', 'inmediatamente', 'instantáneamente', 'jamás', 'luego', 'mañana',
'mientras', 'momentáneamente', \
            'normalmente', 'nunca', 'ocasionalmente', 'posteriormente', 'primeramente', 'pronto',
'puntualmente', 'recién', 'recientemente', \
            'siempre', 'simultáneamente', 'tarde', 'temprano', 'ya']

import openpyxl
from nltk import word_tokenize
import re


path = "C:/Users/Ally/Desktop/U of H/D/Data/SITCLEAN/Data/4520B710.xlsx"
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active

rownum = 2
for i in sheet_obj:
    cell_obj = sheet_obj.cell(row = rownum, column = 3)
    cell_obj2 = word_tokenize(cell_obj.value)
    for j in range(len(cell_obj2)):
        cell_obj2[j] = cell_obj2[j].lower()
    if set(cell_obj2).intersection(adverbsoftime):
        for k in cell_obj2:
            if k in adverbsoftime:
                print(rownum, k)
    else:
        print(rownum, "*")
    rownum = rownum + 1
```

APPENDIX C: LIST OF ADVERBS OF TIME

- Actualmente
- Ahora
- Anoche
- Anteriormente
- Antes
- Antiguamente
- Asiduamente
- Aún
- Ayer
- Constantemente
- Contemporáneamente
- Cuando
- Desde
- Después
- Enseguida
- Entretanto
- Eternamente
- Finalmente
- Frecuentemente
- Hoy
- Inicialmente

- Inmediatamente
- Instantáneamente
- Jamás
- Luego
- Mañana
- Mientras
- Momentáneamente
- Normalmente
- Nunca
- Ocasionalmente
- Posteriormente
- Primeramente
- Pronto
- Puntualmente
- Recién
- Recientemente
- Siempre
- Simultáneamente
- Tarde
- Temprano
- Ya