Article

Theoretical implications of research on bilingual subject production: The Vulnerability Hypothesis

Ana de Prada Pérez
University of Florida, USA

Abstract

In this paper we propose a new hypothesis for the formal analysis of cross-linguistic influence, the Vulnerability Hypothesis (VH), with the support of data from subject personal pronoun use in Spanish and Catalan in Minorca, and contrast it to the Interface Hypothesis (IH). The VH establishes a categorical–variable continuum of permeability, that is, structures that show variable distributions are permeable while those that exhibit categorical distributions are not.

To test the predictions of the VH, Spanish language samples were collected from 12 monolingual Spanish speakers, 11 Spanish-dominant bilinguals, and 12 Catalan-dominant bilinguals, and Catalan language samples from 12 Catalan-dominant speakers. Following a variationist comparative analysis, 4,466 first person singular (1sg) and 1,291 third person singular (3sg) tokens were coded for speech connectivity, verb form ambiguity, and semantic verb type. The language-external variable included in the analysis was language group (Spanish monolinguals, Spanish-dominant bilinguals, Catalan-dominant bilinguals, and Catalan controls).

Results indicated that speech connectivity is the highest ranked variable in the Spanish control group (most categorical variable), while ambiguity and verb type are ranked lower, with only ambiguity reaching significance. The VH would, therefore, predict bilinguals would be similar to monolinguals in the most categorical variables, in this case, speech connectivity. This is in contrast to the IH, which would predict bilinguals would exhibit difficulty with the pragmatically driven distributions (e.g. speech connectivity), while they would show no contact effects or lesser effects with distributions at the lexico-semantic interface with syntax (e.g. verb form ambiguity and verb type). The prediction of the VH bears out in our data. Bilinguals do not differ with respect to speech connectivity. Ambiguity, on the other hand, is no longer significant in the bilingual groups and verb type reaches significance with 1sg (and not with 3sg) subjects. These results are discussed, redefining the concepts of convergence and simplification from language contact research to adapt to the variationist analysis used. Simplification is specified as the reduction of lower ranked predicting variables, while convergence is defined as an increase in parallels across languages with respect to the variables that are significant, their effect size (variable ranking), and the direction of effects (constraint ranking). Regarding language group, it was not returned as significant in 1sg

Corresponding author:
Ana de Prada Pérez, University of Florida, 170 Dauer Hall, PO Box 117405, Gainesville, FL 32611, USA.
Email: aprada@ufl.edu

data. Thus, the groups did not differ in their rates of overt pronominal expression. Differences, however, emerged across groups in the 3sg data, where bilinguals used significantly more overt pronominal subjects than monolinguals do.

This paper contributes to current discussions in the fields of language contact, second language acquisition, and bilingualism, introducing a new hypothesis and contrasting it with the IH. In addition, it contributes to variationist approaches by examining a novel community of bilingual speakers.

Keywords
Subject pronoun expression, Spanish-Catalan bilingualism, language contact outcomes, cross-linguistic influence, convergence, Interface Hypothesis, Vulnerability Hypothesis

Introduction

The present paper proposes a language contact outcome or cross-linguistic influence hypothesis, the Vulnerability Hypothesis (VH), illustrating its testability through the discussion of data from Spanish-Catalan bilingual subject personal pronoun expression (SPE), and comparing it to the productive Interface Hypothesis (IH; Sorace, 2011, 2012).

This paper is envisioned within the formal analysis of bilingual speech in order to determine what linguistic restrictions exist on cross-linguistic influence or language contact induced change. The data presented benefits from the comparative approach commonly used in variationist linguistics (Tagliamonte, 2002), whereby varieties (monolingual and bilingual) are compared based not only on rates of use of a specific form over another but also on the variables or factor groups that are involved, their magnitude of effect (the ranking of these variables or factor groups), and the direction of effect (constraint ranking). In particular, this paper discusses SPE data in light of the hypothesis proposed, and in contrast with the IH.

In the past decade an abundance of studies within the generative tradition have tested the IH, whereby structures that lie at the core syntax are said to be more impervious to cross-linguistic influence than those at the syntax interfaces with other modules, particularly those at the external interfaces, such as the well-researched syntax–discourse interface. It is noteworthy, however, that these studies tend to study phenomena that are not only at the syntax interface with other modules but also variable.

The paramount Spanish phenomenon both in the IH and variationist approaches is the distribution of null and overt pronominal subjects in Spanish. Within the generativist acquisition studies, the alternation between null and overt pronominal subjects has been attributed to the discourse context, where null subjects take place in topic continuation and overt pronominal subjects in topic shift contexts (however, see Holmberg, 2005; Liceras & Fernández Fuertes, 2017; Martínez, 2011; Sheehan, 2006). Although assumed to be a categorical distribution in monolingual Spanish in most generativist acquisition research (however, see Liceras, 2014, Liceras & Fernández Fuertes, 2017), variationist analyses converge on the multitude of variables that affect this distribution (including discourse context). Among these variables, the highest ranked across studies tend to be discourse context (also dubbed co-referentiality) and verb person and number. Variationist studies, thus, tend to examine the variable verb person and number, largely concluding a number effect (singular subjects favor overt pronominal subjects) or a person effect (1sg, in comparison with 3sg, favors overt pronominal subjects). Alternatively, they control for this variable by including only 1sg tokens in the analysis.

Generative acquisitionists either did not control for the variable (Rothman, 2009) or only included 3sg items (e.g. Montrul, 2004). To the best of our knowledge, the only exception is Liceras and Fernández Fuertes (2017), where they separate all six person and number combinations.

The paper is organized as follows. The second section introduces the hypothesis, placing it in the context of previous work dealing with the classic problem of cross-linguistic influence selectivity. The third section reviews previous research regarding subject expression in Spanish, both in monolingual and bilingual speakers. In the fourth section, the present project is introduced. Lastly, the fifth section offers a discussion and the sixth section the conclusions.

Linguistic constraints on language contact

Grounded in the notion of linguistic modules and interfaces between them (e.g. Jackendoff, 2002; Reinhart, 2006), the IH (Sorace, 2011) has evolved to a current account where the notion of interfaces is downplayed (Sorace, 2012, in response to commentaries to Sorace, 2011). Responses to Sorace (2011) raised a number of concerns with the IH. Crucially, several authors highlighted the role of complexity as being a better factor than interface (Hopp, 2011; Pires & Rothman, 2011). In her response to these commentaries, Sorace (2012) argued "again, the problem evaporates once we remove a rigid distinction between core syntax and interfaces and instead allow for a range of interface conditions, graded according to their computational complexity and their dependence on extra-linguistic factors." The details, however, remain largely unexplained. Thus, this paper aims to offer a detailed proposal of what this complexity might be as well as to contribute further evidence that being a linguistic phenomenon at the external interface is not what explains cross-linguistic influence selectivity, that is, the fact that grammar is more susceptible to cross-linguistic influence in certain domains than others (as suggested by Holmberg, 2005; Liceras & Fernández Fuertes, 2017; Sheehan, 2006). In fact, I argue here that complexity resides in variability, which is understood as the availability of more than one form for a specific paradigmatic gap such that the alternation between the two forms is not categorical. As will be further explained below, this alternation is not necessarily random either, as the probabilistic use of a form is determined by linguistic and extra-linguistic factors (Labov, 1972). Cross-linguistic influence selectivity is, thus, due to variability of the distribution in the receiving language, that is, the language being influenced.

A proposal along these lines does not rely on the structure of grammars and our changing understanding of linguistic modules and their interfaces. Instead it is based on an understanding of the grammar based on the factors that affect a specific distribution in a specific language, thus requiring a detailed description of the facts in the receiving language. Once the probability of producing one of the forms in the alternation in a specific language is known, predictions can be made about the effects of bilingualism on the distribution (relative frequencies) under study. The VH establishes a cross-linguistic permeability hierarchy along the variability continuum, which spans from categorical distributions, where a form occurs (near) 0% or 100% of the time in a specific context (e.g. overt pronominal subjects occur 0% of the time with impersonal predicates), to highly variable contexts, where the production of a specific form is (near) 50% of the time. Thus, this analysis is based on the relative frequencies of forms in a specific context. Under the VH, those distributions in the variable end of the continuum will be subject to cross-linguistic influence, whereas those that are on the categorical end of the continuum will not (1). As represented by the arrow, the more variable a distribution is the more susceptible to cross-linguistic influence it will be.

(1)   Cross-linguistic Permeability Hierarchy: Vulnerability Hypothesis
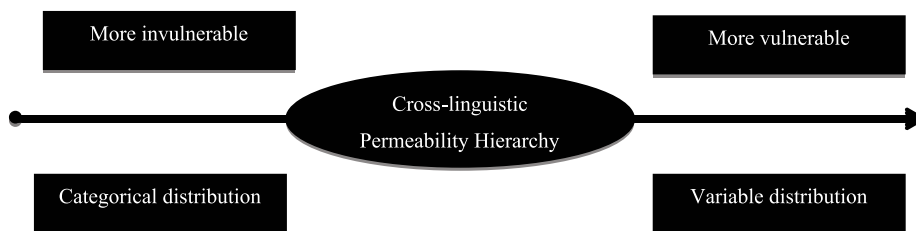
How far to the categorical end a speaker will exhibit cross-linguistic influence depends on individual and social factors, as a measure of language contact intensity. Thus, highly proficient bilinguals are expected to only exhibit influence in highly variable phenomena, while low proficiency bilinguals are expected to show evidence of cross-linguistic influence in less variable phenomena. However, this is an assumption that is left for further research in this paper as it would require testing of several bilingual groups.

Subject expression in Spanish and Catalan

The topic of subject expression in Spanish has received extant attention in both acquisitionist and variationist studies. Spanish and Catalan are both null subject languages (Camacho, 2013; Holmberg, 2005; Perlmutter, 1971; Sheehan, 2006), with similar null versus overt subject distributions (Prada Pérez, 2009, 2010a, 2006). Variationist analyses of Spanish subject expression indicate that the distribution of null and overt pronominal subjects in Spanish is regulated by a combination of variables, some related to discourse structure (co-referentiality, tense, aspect, mood, or TAM, continuity, connect or speech connectivity, and clause type, among others), and some to lexico-semantics and morphology (grammatical person, semantic verb type, verb form ambiguity, etc.).

Regarding those variables related to discourse structure, both co-referentiality (or switch reference) or whether the subject is the same subject as that of the previous clause or not, and TAM continuity, or whether the TAM is the same as in the previous clause, have been thoroughly examined in the literature. A combination of these two variables has been productively used, under the term Connect or Speech connectivity. Although more constraints have been proposed in earlier studies (Bayley & Pease-Álvarez, 1997; Paredes Silva, 1993), recently researchers have used a three-way distinction from the maximum level of connectedness (same referent and same TAM), as in (2a), to the lowest level of connectedness (different referents), as in (2b), through an intermediate level (same referent but different TAM), as in (2c), a coding practice that we follow here.[1]

(2)   Coding of language-internal variables: CONNECT or SPEECH CONNECTIVITY
   a.   *Y yo los bañaba, y los **vestía**, les daba de comer, los ponía a dormir.*
        "And I would bathe them, [I] would dress them, [I] would feed them, [I] would put them to sleep." [NMCOSS, 117–1A3: 248.-2512] (Travis, 2007)
   b.   *Entonces, había necesidad de trabajar porque se había muerto mi papá y **teníamos** una familia de seis personas.* (Participant #19)
        Then, there was the need to work because my father had died and [we] had a family of six people. (Holmquist, 2012)
   c.   *... y ha recibido hijos de las personas que ella **recibió**, o sea nietoh, cómo quien dice.*

[BF07116-117]
" … and [she] has delivered children of the people that she delivered, that is grandchildren, we could say." (Orozco, 2015)

In general terms, null subjects tend to indicate continuity in speech (Cameron, 1994). Clause type has also been explained in terms of discourse structure such that subjects of *embedded clauses* tend to be null (see Morales, 1997; Otheguy, Zentella, & Livert, 2007; Silva-Corvalán, 1994). Of these variables, this paper only includes the variable connect (speech connectivity) in the analysis. Co-referentiality and TAM continuity were, thus, combined instead of examined separately, as in recent previous studies **(Bayley & Pease-Álvarez, 1997; Otheguy et al., 2007; Paredes Silva, 1993; Prada Pérez, Ana de, 2015).** Clause type was excluded from the analysis, since only those subjects in main clauses were included. In embedded contexts in Spanish, overt subjects are rarely used in clauses whose subjects are co-referential with the matrix clause (Montalbetti's 1984 principle). Thus, we excluded embedded clauses to avoid the large effect that this variable could have.

In addition to cases where continuity in discourse is reduced, the use of the overt pronominal subject has also been identified with "speaker egocentrism" and verb form ambiguity. It has been widely attested in the literature that overt pronominal subjects are more frequent in 1sg, as in (3), than in 3sg, or any of the plural verb forms (Enríquez, 1984; Morales, 1997; Otheguy et al., 2007; Silva-Corvalán, 1982, 1994; Travis, 2007). Thus, the variable person (i.e. grammatical person and number) affects subject expression in Spanish.

(3) Coding of language-internal variables: Person
    Y luego ya no pude ir más, porque yo iba muy lejos pa' agarrar el bos,
    "And then (I) couldn't go anymore, because I had to go really far to catch the bus,"
        [NMCOSS, 76–1A1: 228–229] (Travis, 2007)

Some of the verb forms in Spanish, namely 1sg and 3sg forms, have the same form in several tenses (imperfect, conditional, present subjunctive, and related compound forms). These ambiguous forms have sometimes been attested with more overt pronominal subjects, with disambiguating purposes, as in (4).

(4) Coding of language-internal variables: Verb form ambiguity
    En la noche ella iba a mi lado y yo estaba temblando
    "At night she used to go by my side and I wasn't shaking." (Silva-Corvalán, 1994)

While overt subjects are favored with verbal forms that are *ambiguous* and not with *unambiguous* ones (Bayley & Pease-Álvarez, 1996, 1997; Silva-Corvalán, 1994; Travis, 2005), this result is not always attested (Casanova Seuma, 1999; Morales, 1997; Ranson, 1991). The distribution of overt subjects is also relevant to the establishment of the speaker's position on an idea. As a result, person in combination with semantic verb type affects the distribution. For instance, 1sg and verbs that express opinion or estimative verbs favor overt subjects (Enríquez, 1984; Morales, 1997; Otheguy et al., 2007; Silva-Corvalán, 1982, 1994; Travis, 2007). For example, Morales (1997) shows that the subjects of verbs like *pensar* "to think" may be produced even in topic continuation contexts, as in (5).

(5) Coding of language-internal variables: Verb type
    Parece que ellos piensan que es signo de cultura.
    "It seems as if they think that is it a sign of culture." (Morales, 1996)

More recently, the classification of verb type has been reduced to external actions, mental processes, and stative verbs (see Orozco, 2015, for a fuller description). In general, mental and stative predicates favor overt subjects, while external actions favor null pronominal subjects (however, see Orozco, 2015; Posio, 2015). In this paper, we included the variables verb form ambiguity and semantic verb type in the analysis. As for person, only 1sg and 3sg subjects were included. Since this variable was returned as having a large effect, separate analyses were run for 1sg and 3sg.

Numerous studies have examined the distribution of null and overt pronominal (and rarely lexical) animate subjects in Spanish across different varieties. In general, differences are reported for overt pronominal rates, particularly between the Caribbean and Mainland varieties (Alfaraz, 2015; Camacho, 2013; Cameron, 1995; Lastra & Butragueño, 2015; Orozco, 2015; Orozco & Guy, 2008; Otheguy & Zentella, 2012; Posio, 2015, among many others). Crucially, Spanish and Catalan exhibit the same variables that are significant, with the same ranking of variables (obtained through the range or effect size), and the same direction of effect (Prada Pérez 2009, 2010a, 2015). Prada Pérez (2009) examined SPE in four groups of speakers: Spanish monolinguals, Spanish-dominant bilinguals, Catalan-dominant bilinguals speaking Spanish, and Catalan-dominant bilinguals speaking Catalan, and across grammatical persons. The analysis included close to 14,000 tokens, including lexical, overt, and null pronominal subjects and the effect of 12 linguistic variables. Prada Pérez (2010) compared Spanish and Catalan based on those 12 linguistic variables. In a more restrictive analysis, Prada Pérez (2015) offered a comparison of SPE in the four groups of speakers only in 1sg subjects in main clauses in topic continuation contexts and broad focus, that is, in contexts where variation has been previously attested. The analysis examined the role of connect (speech connectivity), verb form ambiguity, and semantic verb type in addition to the extralinguistic variables: language group, gender, and age in SPE. The results indicated few differences in terms of the variables that were returned as significant across the groups (connect was significant for all groups, verb form ambiguity for the two control groups, and verb type for the two bilingual Spanish groups) as well as their ranking (connect was the highest ranked variable across groups) or the direction of effects (in all groups more connected speech, ambiguous verb forms and external activities favored more null subjects). In terms of rates, the two languages are also very similar in that respect: Spanish exhibits 19.8% overt pronominal subjects and Catalan 20.7% overt pronominal subjects.

The role of age and gender, however, remains largely variable across varieties (cf. Flores-Ferrán, 2007, for a review). In monolingual communities, age and gender are considered as variables indicating linguistic change (e.g. Bailey's [2004] "apparent time" for age[2]). Age has variably been found to condition subject expression in Spanish. For instance, Cameron (1992) did not report an age effect while Ávila-Jiménez (1995) did, both examining the same variety and region. Recently, younger generations have exhibited lower rates of overt pronominal expression for monolingual speakers from Barranquilla, Colombia (Orozco & Guy, 2008), Colombian Costeño Spanish (Orozco, 2015), and Mexico City Spanish (Lastra & Butragueño, 2015), while the opposite trend has been reported for Puerto Rico (e.g. Ávila-Jiménez, 1995; Flores-Ferrán, 2002; Lizardi, 1993) and the Dominican Republic (Alfaraz, 2015).

For gender, different results have been attested as well. Some studies report no gender effect (Holmquist, 2012; Orozco & Guy, 2008; Otheguy et al., 2007) while others find a *women effect*, in which female participants are leading the change toward expressed pronouns in monolingual and bilingual varieties (Bayley & Pease-Álvarez, 1996; Carvalho & Child, 2011; Otheguy & Zentella, 2012; Shin, 2013; Shin & Otheguy, 2013).

A comparison in subject pronoun expression that has been rather productive in the literature has been between monolingual and bilingual varieties of Spanish. Spanish in contact with English has reported conflicting results: while some do not report differences (Bayley & Pease-Álvarez, 1997; Flores & Toro, 2000; Flores-Ferrán, 2004; Liceras & Fernández Fuertes, 2017; Silva-Corvalán, 1994;

Torres-Cacoullos & Travis, 2010; Travis, 2007), others report an increase in overt pronominal subjects and changes in the distribution reflected in variables that are significant, their ranking, and/or their constraint ranking (Erker & Guy, 2013; Erker & Otheguy, 2016; Lipski, 1994, 1996; Otheguy & Zentella, 2012; Orozco, 2015; Shin, 2012; Shin & Otheguy, 2013; Toribio, 2004, among others).

In summary, the extensive and expanding literature on SPE in Spanish across varieties agrees on the patterned nature of the distribution of null versus overt pronominal subjects, which is delimited by well-studied linguistic factors. Differences in rates have been reported in the previous literature, with varieties of Spanish where overt pronominal subject rates are as low as 20% and those where they are close to 40%. Nonetheless, the effect of variables and direction of effects remain largely the same across varieties (Carvalho, Orozco, & Shin, 2015). The role of the language-external factors age and gender seem to depend largely on the variety, as they seem to be community-specific.

The present study

The main aim of this paper is to introduce the VH and test it comparing it to the IH. The IH considers that the main area of cross-linguistic influence is the syntax interface with pragmatics, while the VH considers that distributions that are more variable are more susceptible to cross-linguistic influence. In particular, the IH predicts difficulty with SPE as it relates to information structure. The VH, in contrast, predicts more similarities with monolingual Spanish speakers with respect to the variable related to information structure, connect, as it is a variable that has a strong effect (or a high ranked variable). Bearing in mind these differences, the current project seeks to test the predictions of these hypotheses against Catalan-Spanish bilingual data in contrast with monolingual Spanish and Catalan spoken by Catalan-dominant Catalan-Spanish bilinguals. In particular, the linguistic feature analyzed in these data is SPE.

Research questions and predictions

In particular, this study aims to answer the following questions.

(i) Subject expression in monolingual Spanish: What variables from those examined (speech connectivity, ambiguity, and verb type) predict the distribution of Spanish null and overt pronominal 1sg and 3sg subjects?[3] From the variables included, the distribution of Spanish null and overt pronominal subjects can be interpreted as lying at the syntax–pragmatics interface (speech connectivity), at the syntax–morphology interface (verb form ambiguity), and at the syntax interface with the lexico-semantics (verb type). In previous studies, all these variables have been found to be significant. Thus, we predict they will also play a role in the Spanish of the monolingual participants included in this study. The analysis will also examine if any of these variables is more predictive than other (i.e. if the magnitude of effect is larger, as per the range in constraints). If, as previously claimed, distributions at the syntax interface with pragmatics are more variable (as per the IH), it is expected that speech connectivity will be ranked lower than verb type and verb form ambiguity, a result not reported in the previous literature and not expected in the present data either. Crucially, these results will set the baseline for the comparison between the two hypotheses. While the IH predicts difficulty with SPE, particularly in the weakening of the variable connect, the VH predicts difficulty with the lower ranked variables (i.e. the variables with a smaller magnitude of effect). From previous studies it is likely that connect is a highly ranked variable both in monolingual Spanish and in Catalan, the two control groups. If that were the

case, the predictions of the two hypotheses would be different with respect to the variable connect: the IH would predict a weakening of the variable while the VH would predict no weakening. Lastly, comparisons will be made between 1sg and 3sg subjects, as some of the contradictions found in the previous literature on bilingual SPE in US Spanish may be largely due to studies examining 1sg versus 3sg. In particular, some studies examining different grammatical persons report a language contact effect manifested in a higher rate of overt pronominal subjects (Otheguy & Zentella, 2012) while others, which focus on 1sg, do not (Torres-Cacoullos & Travis, 2010).

(ii) Subject expression in bilingual Spanish: What variables from those examined (connect, ambiguity, and verb type) predict the distribution of Spanish null and overt pronominal 1sg and 3sg subjects in the Spanish spoken by bilinguals as compared to monolinguals? Previous studies in general show sensitivity to these variables in bilingual speakers, although the differences may be expected in the magnitude of effect of certain variables, as sometimes results are interpreted as a laxity in the variables. The IH predicts a weakening of the variable connect in bilingual Spanish, as it lies at the syntax–discourse interface. This is predicted to be so even if Spanish and Catalan are the same with respect to this variable (i.e. even if it significant and with a similar effect size in both languages). Sorace (2011) explains that the difficulty resides in the processing of null subjects with bilinguals' strained resources, which are the same irrespective of language pairing. The VH, in contrast, anticipates changes to target mostly lower ranked variables, which are likely not to be the variable connect. Given the differences in cross-linguistic effects reported in the previous literature between 1sg and 3sg, it is expected that more evidence of cross-linguistic influence will be found with subjects in 3sg than with 1sg subjects.

(iii) Cross-linguistic influence theories: If cross-linguistic influence is attested, can it be attributed to variability or to interface between linguistic and extra-linguistic modules? The IH predicts bilingual speakers will exhibit more cross-linguistic influence (or larger differences from the monolingual group) in the distribution of null and overt pronominal subjects that is restricted by syntax–pragmatics interface variables (connect or speech connectivity) and less so by syntax–semantics (in our study, verb type) or syntax–morphology (in this case, verb form ambiguity) interface variables, depending on proficiency. At higher levels of proficiency the IH still anticipates non-monolingual-like behavior with respect to the syntax–pragmatics interface. This is predicted to be the case even in a contact situation between two null subject languages, as the bilingual effect constitutes a simplification in processing to alleviate the burden of having two active languages. Sorace (2011) refers to the psycholinguistic literature, which concludes that processing null subjects in Spanish is more effortful than processing overt forms (Alonso-Ovalle, Clifton, Frazier and Férnandez-Solera, 2005). This is in contrast with the syntactic theoretical accounts, where the null is the unmarked form (Holmberg, 2005; Liceras & Fernández Fuertes, 2017; Martínez, 2011; Sheehan, 2006). The VH, in contrast, predicts that those variables that are lower ranked (i.e. where the distribution is more variable) in the monolingual grammar will be the target of cross-linguistic influence. In order to determine which variables are lower ranked, the monolingual data will be examined first. Bilinguals are expected to exhibit more differences in higher ranked variables as proficiency decreases. Thus, Catalan-dominant bilinguals are expected to exhibit more differences from Spanish monolingual speakers in Spanish than Spanish-dominant bilinguals.

In order to respond to these questions, data in Spanish from Spanish monolinguals and Catalan-Spanish bilinguals, of different dominance (Spanish-dominant and Catalan-dominant), as well as

data in Catalan from Catalan-dominant bilinguals, were collected and analyzed. The following section describes the participants included in the study.

Participants

Data from bilingual and monolingual speakers were collected and divided into four groups: two control groups and two bilingual Spanish groups. The Spanish control group consisted of data from 12 Spanish monolingual speakers from Valladolid, a monolingual province in Spain, while the Catalan control group consisted of data from 12 Catalan-dominant speakers from villages in the center of Minorca, where Spanish is rarely spoken. The bilingual data comprised Spanish speech samples from 12 Catalan-dominant Catalan-Spanish bilinguals and 11 Spanish-dominant Catalan-Spanish bilinguals residing in Minorca. Each group comprised (roughly) the same number of male and female participants who were equally distributed into three age groups, as per their access to education in Catalan: age group 1 (ages 13–35), age group 2 (ages 36–64), and age group 3 (65 and over). Thus, there were two males and two females in each of the age groups, except in the Spanish-dominant Catalan-Spanish bilingual group where there was only one male, instead of two, in age group 3. Gender and age groups are not examined here (refer to Prada Pérez 2015, for an analysis including these variables) as neither the VH nor the IH make predictions about these. In any case, there was a balance across groups with representation of speakers of different genders and ages.

Materials and coding

The relevant data for this study were extracted from two sections of an oral interview conducted in Spanish by a native speaker from Valladolid and in Catalan by a native of Alaior, Minorca: the language background questionnaire, for participant profiling, and a sociolinguistic interview.

The sociolinguistic interview was transcribed and every subject form was coded for a number of linguistic and extralinguistic variables. For the purpose of the analysis, the only tokens included were those with null or overt pronominal subjects in 1sg or 3sg, in topic continuation contexts (i.e. exclusion of newly introduced referents), in broad focus (i.e. exclusion of tokens in narrow focus), and in main clauses.[4]

The final set of tokens were coded for subject form (null versus overt pronominal subjects), person and number (first versus third person singular subjects), connect or speech connectivity (same referent, same TAM; same referent, different TAM; different referent, cf. Otheguy et al., 2007), verb form ambiguity (ambiguous versus non ambiguous verb forms), and semantic verb type (stative, mental, or external actions, cf. Enríquez, 1984). The predictions were largely guided by antecedent research, where (i) 1sg, (ii) more connected speech, (iii) ambiguous verb forms, and (iv) stative and mental verbs favored overt pronominal subjects more than 3sg subjects, less connected speech, unambiguous forms, and external action verbs. Thus, the same distribution was expected in this study.

The data were also coded by individual characteristics of the speaker. The external variable included in the final analysis was language group (Spanish controls, Catalan controls, Spanish-dominant bilinguals, Catalan-dominant bilinguals).

All the data were submitted to statistical analysis using Goldvarb Yosemite software. This regression analysis provides the combination of variables that better explains the data. If a variable is categorical, a regression analysis is not viable. This was not the case in any of the analyses presented here. Only contexts where the alternation between null and overt pronominal subjects was variable were included in the analysis. For the variables selected as significant, Goldvarb returns factor weights for each of the constraints, which are said to favor overt subjects (the application value selected) if the weights are over .50 and to disfavor them if the weight is lower than .50,

Table 1. First and third person singular subject expression in all the groups. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first or third person singular overt pronominal subject versus a null subject; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.66**	27.4	2562
Same referent, different TAM	0.41	12.7	1275
Same referent, same TAM	0.34	8.5	1920
Range	32		
PERSON			
First singular	**0.56**	20.5	4466
Third singular	0.31	8.6	1291
Range	25		
AMBIGUITY			
Ambiguous	**0.59**	20.7	1365
Unambiguous	0.47	16.9	4392
Range	12		
VERB TYPE			
Mental	**0.54**	23.5	899
Stative	**0.53**	19.7	1069
External	0.48	15.9	3789
Range	6		
LANGUAGE GROUP			
Catalan L1 bilinguals	**0.52**	18.9	1580
Catalan control	**0.52**	18.6	1422
Spanish L1 bilinguals	**0.52**	19.1	1499
Spanish control	0.43	14.1	1256
Range	2		
Total N			5757
Corrected mean			.152
Log likelihood			−2477.539
Significance			.027

TAM: tense, aspect, mood; L1: first language.

indicating the direction of effect. The range in the factor weight for a variable determines the magnitude of effect, which allows the researcher to rank variables from largest effect to smallest effect. This analysis, thus, indicates what variables are significant, the size of their effect, and the direction of their effect. Comparing these three pieces of information across different groups allows us to test and contrast the predictions of the IH and the VH.

Results

Several analyses were run to gain a better understanding of the data. All the data (1sg and 3sg) were included in an initial analysis in order to determine if there was an effect for person.

As Table 1 shows, person was returned as significant. In particular, 1sg subjects favor overt pronominal subjects more than 3sg subjects.[5]

Table 2. First person singular subject expression in all the groups. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first person singular overt pronominal subject versus a null subject; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.65**	20.6	616
Same referent, different TAM	0.42	14.6	152
Same referent, same TAM	0.34	10.4	147
Range	31		
AMBIGUITY			
Ambiguous	**0.59**	25.6	237
Unambiguous	0.48	19.1	678
Range	11		
VERB TYPE			
Mental	**0.55**	24.5	208
Stative	**0.53**	24.7	178
External	0.48	18.3	529
Range	7		
LANGUAGE GROUP			
Catalan L1 bilinguals	**0.51**	21.3	274
Catalan control	**0.50**	20.7	234
Spanish L1 bilinguals	0.49	19.9	253
Spanish control	0.49	19.8	154
Range	2		
Total N			4466
Corrected mean			.186
Log likelihood			−2127.748
Significance			.012

TAM: tense, aspect, mood; L1: first language.

Since this result has been widely attested in the literature and the variable person occurs in collinearity with several other variables, the remainder of the paper presents 1sg data separately from 3sg data. The separate analysis can be particularly telling as it pertains to language group differences. As noted above, previous research is inconsistent with respect to the effects that language contact has: some reported a contact effect and others did not. We hypothesized that this could be due to different studies examining subject expression across linguistic persons, in 1sg or 3sg. Table 2 presents the results for the 1sg data, where differences did not emerge across language groups.

The data above indicate that the use of 1sg overt pronominal subjects hovers around 20% across speaker groups (range: 19.8–21.3%). As will be discussed below, when the data analysis is performed separately on each of the language groups some minor differences emerge between the bilingual and the control groups. In contrast, in Table 3, the data for 3sg subjects reveal a language group effect.

The rate of overt 3sg pronominal subjects varies from 4.8% in the Spanish control to 14.5% in the Spanish-dominant bilingual group. The data further indicate that all groups in Minorca favor the use of overt pronominal subjects (Spanish-dominant bilinguals [.65], Catalan-dominant bilinguals [.51], and Catalan controls [.59]), while the Spanish controls disfavor them [.37].

Table 3. Third person singular subject expression in all the groups. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a third person singular overt pronominal subject versus a null subject; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.71**	15.4	504
Same referent, different TAM	0.38	4.2	236
Same referent, same TAM	0.32	3.2	504
Range	39		
LANGUAGE GROUP			
Spanish L1 bilinguals	**0.65**	14.5	227
Catalan control	**0.59**	10.6	292
Catalan L1 bilinguals	**0.51**	8.2	294
Spanish control	0.37	4.8	478
Range	28		
VERB TYPE			
Stative	**0.53**	9.5	348
External	**0.50**	8.4	894
Mental	0.41	6.1	49
Range	12		
AMBIGUITY			
Ambiguous	**0.56**	10.4	441
Unambiguous	0.47	7.6	850
Range	9		
Total N			1291
Corrected mean			.062
Log likelihood			−340.105
Significance			.000

TAM: tense, aspect, mood; L1: first language.

These data so far reveal an interesting interaction between language group and person. While no statistical difference emerges in the overall rates of overt 1sg pronominal subjects across groups, differences emerge with respect to 3sg. In the remainder of this section, thus, we examine the variables involved in subject expression in 1sg and 3sg subjects in each of the language groups.

Spanish controls. The data presented above revealed great differences in rates of overt pronominal expression in 1sg versus 3sg subjects in the Spanish controls: 19.8% overt pronominal subjects in 1sg versus 4.8% overt pronominal subjects in 3sg. First and third person subjects differ in significant ways, for example, first person is deictic while third person is referential. As can be seen in Tables 4 and 5, they do not differ in the linguistic variables that are significant in the distribution of overt versus null pronominal subjects in Spanish.

In 1sg, two variables were selected as significant, namely connect and ambiguity. As expected, the more connected the speech is, the lower the odds of producing an overt pronominal subject, and the more ambiguous a verb form is, the higher the odds of producing an overt pronominal subject. As can be seen in Table 5, similar results are returned for 3sg.

The variables connect and ambiguity continue to be significant and exhibit the same patterns, where the more connected the speech, the lower the use of overt pronominal subjects, and the more ambiguous the verb form is, the higher the use of overt pronominal subjects.

Table 4. First person singular subject expression in the Spanish control group. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first person singular overt pronominal subject versus a null subject in the Spanish control group; factor groups selected as significant in gray background.

Spanish control

	Factor weight	%	N
CONNECT			
Different referent	**0.65**	29.3	389
Same referent, different TAM	0.40	12.6	175
Same referent, same TAM	0.30	8.4	214
Range	35		
AMBIGUITY			
Ambiguous	**0.63**	28.4	637
Unambiguous	0.47	17.9	141
Range	16		
VERB TYPE			
Mental	**0.59**	25.9	135
Stative	**0.55**	23.1	130
External	0.46	17.3	513
Range	13		
Total N			778
Corrected mean			.175
Log likelihood			−359.123
Significance			.007

TAM: tense, aspect, mood.

Overall, Spanish controls indicate differences across persons in terms of rates of overt pronominal subjects. Nonetheless, two linguistic variables affect the distribution of null and overt pronominal subjects irrespective of person: connect and verb form ambiguity. Thus, although differences in rates are significant across persons, the distribution is subject to the same variables.

Catalan controls. In the Catalan control group differences between 1sg and 3sg emerged with respect to overt pronominal subject rates, which were 20.7% in 1sg versus 10.6% in 3sg. As can be seen in Tables 6 and 7, the difference between 1sg and 3sg is also noticeable in their distribution of null versus overt pronominal subjects, which differ in the linguistic variables that are significant.

As was the case in the Spanish controls, the linguistic variables that are significant in Catalan are connect and ambiguity, with connect presenting the largest effect size (range: 33). Also similarly, the direction of effect is the same: more connected speech favors null subjects and ambiguous verb forms favor overt subjects. Although the trends are the same, as can be seen in Table 7, only connect is significant in 3sg.

As can be seen above, the same direction of effect is observed in 3sg with a large effect size (range: 40).

Unlike in Spanish, in Catalan the difference in rate of use of overt pronominal subjects between 1sg and 3sg is accompanied by a difference in significant variables. Only the highest ranked variable, connect, remains significant across persons.

Table 5. Third person singular subject expression in the Spanish control group. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a third person singular overt pronominal subject versus a null subject for Spanish controls; factor groups selected as significant in gray background.

Spanish controls			
	Factor weight	%	*N*
CONNECT			
Different referent	**0.74**	10.1	199
Same referent, same TAM	0.26	1.5	196
Same referent, different TAM	N/A	0	83
Range	48		
AMBIGUITY			
Ambiguous	**0.68**	7.8	141
Unambiguous	0.42	3.6	337
Range	26		
VERB TYPE			
External	**0.52**	5.4	351
Stative	0.43	3.7	107
Mental	N/A	0	20
Range	9		
Total *N*			478
Corrected mean			.038
Log likelihood			−76.899
Significance			.019

TAM: tense, aspect, mood.

Spanish-dominant bilinguals. As seen in Tables 2 and 3, the Spanish-dominant bilingual group does not differ noticeably in the percentage of overt pronominal subjects in 1sg (19.9%) versus 3sg (14.5%), exhibiting the highest rate of overt pronominal subjects in 3sg of all groups while using similar rates of overt pronominal subjects as the other groups. In 1sg (Table 8), some differences from both control groups are attested.

In particular, unlike in both control groups, ambiguity is not returned as a significant variable, while verb type is. The highest ranked constraint, connect, however, remains significant. Also relevant is the fact that the trends are similar across groups, even if they do not reach significance in all groups: ambiguous forms favor overt pronominal subjects more than unambiguous forms; mental and stative verbs favor overt subjects more than external actions; and less connected speech favors more overt subjects than more connected speech. As was the case with the Catalan control group, the data from the Spanish-dominant bilinguals show some differences between 1sg and 3sg, not so much in terms of percentage of use of overt pronominal subjects as in terms of significant variables.

Although the difference in rates between 1sg and 3sg was not as noticeable in this speaker group, only the variable connect is significant in 3sg (Table 9).

In conclusion, Spanish-dominant bilinguals exhibit similar overt pronominal subject rates in 1sg and 3sg. Nonetheless, differences exist in the variables that are returned as significant, where fewer variables are significant in the 3sg.

Catalan-dominant bilinguals. As in the Spanish control group, Catalan-dominant bilinguals use a significantly higher rate of overt pronominal subjects in 1sg (13.1%) than in 3sg (8.2%).

Table 6. First person singular subject expression in the Catalan control group. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first person singular overt pronominal subject versus a null subject in the Catalan control group; factor groups selected as significant in gray background.

Catalan controls

	Factor weight	%	N
CONNECT			
Different referent	**0.62**	27.9	545
Same referent, different TAM	0.46	18.3	289
Same referent, same TAM	0.31	9.8	296
Range	33		
AMBIGUITY			
Ambiguous	**0.63**	29.4	255
Unambiguous	0.46	18.2	875
Range	17		
VERB TYPE			
Stative	**0.56**	24.9	185
External	**0.51**	21.3	658
Mental	0.44	16.7	287
Range	12		
Total N			1130
Corrected mean			.191
Log likelihood			−547.111
Significance			.021

TAM: tense, aspect, mood.

Examining significant variables and their rankings, Catalan-dominant bilinguals exhibit a contrast between 1sg and 3sg similar to that of the Spanish-dominant bilingual group.

Table 10 indicates that in 1sg, Catalan-dominant bilinguals, like Spanish-dominant bilinguals and unlike both control groups, return two linguistic variables as significant: connect and verb type, while ambiguity is not significant. The patterns across all groups are the same for all these variables, even if they do not reach significance across the groups. These trends are also present in 3sg but the only variable that reaches significance is connect.

In 3sg (Table 11), both bilingual groups and the Catalan control group only return connect as a significant variable.

Language group comparisons. In terms of rates of overt pronominal subjects, there is a larger difference between 1sg and 3sg in the Spanish controls than in the Catalan controls, where the percentage of use of overt pronominal subjects is higher in 3sg than in the Spanish controls. The difference is also smaller in the two bilingual groups, particularly in the Spanish-dominant bilingual group. Within persons, no difference was found across groups in 1sg, while differences were attested across groups in 3sg. This result may explain some of the discrepancies previously found in the literature, an issue that we expand on in the next sections.

With respect to the patterns of use observed, the two control groups are identical in the significant variables in 1sg but not in 3sg, where the Catalan control group is not sensitive to verb form ambiguity. For Spanish-dominant bilinguals, differences in variables that are significant are found in 1sg between this bilingual speaker group and the two control groups. Regarding 3sg, however,

Table 7. Third person singular subject expression in the Catalan control group. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a third person singular overt pronominal subject versus a null subject for Catalan controls; factor groups selected as significant in gray background.

Third person singular		Catalan controls	
	Factor weight	%	N
CONNECT			
Different referent	**0.70**	18.2	121
Same referent, different TAM	0.46	7.5	67
Same referent, same TAM	0.30	3.8	104
Range	40		
VERB TYPE			
Mental	**0.72**	23.1	13
Stative	**0.50**	10.6	94
External	0.48	9.7	185
Range	24		
AMBIGUITY			
Ambiguous	**0.54**	12.2	115
Unambiguous	0.47	9.6	177
Range	7		
Total N			292
Corrected mean			.087
Log likelihood			−92.110
Significance			.002

TAM: tense, aspect, mood.

this bilingual control group patterns similarly to the Catalan control group. Recall, however, that this is the person where differences are attested between Spanish and Catalan.

Discussion

The results indicate interesting contrasts across speaker groups both in terms of rates of use of overt pronominal subjects and patterns of use, as attested in the variables returned as significant, the effect size of the variables, and the direction of effects. Crucially, contrasts between 1sg and 3sg were rather revealing.

In 1sg, the rate of overt pronominal subjects was similar across groups (around 20%) while in 3sg, differences were attested (the rate varied from 4.8% in the Spanish control to 14.5% in the Spanish-dominant bilingual group). These differences were further supported by the variable language group (i.e. speaker group), which was returned as significant only in 3sg. This result confirms our intuition that differences across studies could be due to differences in the grammatical persons included across studies. Overall, these results are consistent, thus, with the previous literature that finds no contact effects in 1sg (Bayley & Pease-Alvarez, 1997; Flores & Toro, 2000; Flores-Ferrán, 2004; Silva-Corvalán, 1994; Torres-Cacoullos & Travis, 2010; Travis, 2007), but also with the previous literature that reports contact effects in 3sg or in the inclusion of several persons in the analysis (Erker & Guy, 2013; Erker & Otheguy, 2016; Lipski, 1994, 1996; Otheguy & Zentella, 2012; Shin, 2012; Shin & Otheguy, 2013; Orozco, 2015; Toribio, 2004). This language effect differentially affecting grammatical persons is particularly noticeable in the Spanish-dominant bilinguals, whose

Table 8. First person singular subject expression in Spanish-dominant bilinguals. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first person singular overt pronominal subject versus a null subject in Spanish-dominant bilinguals; factor groups selected as significant in gray background.

Spanish L1 bilinguals

	Factor weight	%	N
CONNECT			
Different referent	**0.68**	32.2	513
Same referent, different TAM	0.42	13.5	462
Same referent, same TAM	0.35	10.4	462
Range	33		
VERB TYPE			
Mental	**0.58**	28.2	211
Stative	**0.55**	24.6	188
External	0.47	17.0	873
Range	11		
AMBIGUITY			
Ambiguous	**0.54**	22.9	280
Unambiguous	0.49	19.1	992
Range	5		
Total N			1272
Corrected mean			.178
Log likelihood			−590.402
Significance			.040

Note: The variable age shows evidence of collinearity as the factor weights and percentages do not correspond. We leave them as three groups in this analysis for comparisons across ages. In Author (2015) the two youngest groups, who seem to behave similarly and have had access to education in Catalan, were merged into a single group.
TAM: tense, aspect, mood; L1: first language.

rate of overt pronouns in 3sg (14.5%) is close to that of 1sg (19.9%). Sorace (2011) anticipates the overuse of overt pronominal subjects in bilinguals as an economic processing strategy to avoid holding the referent in memory. Since the third person is referential, while the first is deictic, it follows that the processing burden is different in the third than in the first person. Thus, Sorace's assumption can be applied to the person difference attested here. If the processing burden lies in holding the referent in memory, it is expected that the contact effect be more prevalent in the third person, as reported in this analysis. These data, however, are not consistent with the inclusion of interfaces as the defining factor in cross-linguistic influence. Although subject expression has been considered to lie at the syntax interface with discourse-pragmatics, it is also dependent on other internal interfaces, such as the lexico-semantic interface (semantic verb type) or the morphology–syntax interface (verb form ambiguity). In fact, bilinguals in this study did not differ from monolingual speakers in their rates of use of overt pronominal subjects in contexts with a different referent from the previous referent or in contexts with same referent both in 1sg and 3sg. Thus, no loss of pragmatic content of overt pronominal subjects was attested in this study. In contrast, there were some differences between bilingual and monolingual speakers with respect to the variables verb type (at the lexico-semantic interface) and verb form ambiguity (at the interface with morphology). Thus, the data from both 1sg and 3sg subject expression in these groups of Spanish-Catalan bilinguals are more consistent with the predictions of the VH. For these data, the IH predicted a loss of pragmatic content in the

Table 9. Third person singular subject expression in Spanish-dominant bilinguals. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a third person singular overt pronominal subject versus a null subject for Spanish-dominant bilinguals; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.66**	22.5	111
Same referent, different TAM	0.32	6.5	31
Same referent, same TAM	0.35	7.1	85
Range	61		
VERB TYPE			
Stative	**0.53**	16.9	77
External	0.48	14.1	142
Mental	N/A	0	8
Range	5		
AMBIGUITY			
Ambiguous	**0.51**	15.4	78
Unambiguous	0.50	14.1	149
Range	1		
Total N			227
Corrected mean			.132
Log likelihood			−87.171
Significance			.006

TAM: tense, aspect, mood.

use of overt pronominal subjects (i.e. bilinguals' extension of overt pronominal subjects to contexts where there is no switch in reference), while the VH predicted differences between monolingual and bilingual speakers in the lower ranking constraints (verb form ambiguity and verb type in our data). Our results were, thus, consistent with the predictions of the VH, since differences between monolingual and bilingual speakers were reported only with respect to the variables verb form ambiguity and verb type.

No differences in rates were attested across groups in 1sg. In fact, the direction of effect was the same in all the groups. Nonetheless, some differences emerged in the variables that were significant for each of the groups, and as predicted by the VH, only in the lower ranked variables. Connect was returned as the highest ranked variable, with range from 30 to 39. Being such a highly ranked variable in Spanish, the VH anticipated this variable not to be so permeable to language contact effects,[6] a result attested in the data. The other significant variable both in Catalan and Spanish was ambiguity. The two bilingual groups, however, returned this variable as not significant and verb type as significant. These differences can also be explained as simplification, as the lower ranked variables exhibit more variability and are, thus, predicted by the VH to be affected by language contact, an effect attested in our data. The IH, in contrast, predicts that bilinguals differ from monolinguals more in the syntax–pragmatics interface, instantiated in the variable connect, than in the internal interfaces with syntax, instantiated in the variables verb type and verb form ambiguity. The similarities and differences between monolingual and bilingual speakers in this study, thus, were not predicted by the IH.

In 3sg, on the other hand, contact effects are stronger. In fact, significant differences are reported in rates of overt pronominal subjects, the variable language group is returned as significant in the

Table 10. First person singular subject expression in Catalan-dominant bilinguals. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a first person singular overt pronominal subject versus a null subject in Catalan-dominant bilinguals; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.66**	32.8	564
Same referent, different TAM	0.39	13.3	278
Same referent, same TAM	0.36	11.7	444
Range	30		
VERB TYPE			
Mental	**0.58**	30.0	240
Stative	**0.57**	25.6	195
External	0.46	17.9	851
Range	12		
AMBIGUITY			
Ambiguous	**0.54**	23.4	248
Unambiguous	0.49	20.8	1038
Range	5		
Total N			1286
Corrected mean			.193
Log likelihood			−621.557
Significance			.010

TAM: tense, aspect, mood.

regression, and differences in significant variables and their rankings are attested across groups. In this case too, though, all the groups were sensitive to the pragmatic condition of the distribution, as connect was significant and the highest ranked constraint in all speaker groups. In 3sg, unlike in 1sg, differences existed between Catalan and Spanish, Spanish had lower rates of overt pronominal subjects than Catalan and was subject to more constraints (connect and verb form ambiguity), while for Catalan only connect was significant. In this case, since differences between the languages exist, the language contact effect predicted could be interpreted not only as simplification, as in 1sg, but also as convergence, where the rates of overt pronominal subjects and the rankings of the variables could be indicative of cross-linguistic influence. This is a common prediction in the IH and the VH. The difference between the VH and the IH is still in the areas of the grammar that would be affected by language contact (i.e. the external interface, instantiated in connect in SPE, in the case of the IH, and in the lower ranked variables, verb form ambiguity, and verb type, in the case of the VH). In particular, both hypotheses predict that the bilinguals exhibit higher rates of overt 3sg pronominal subjects in Spanish than the monolinguals do, albeit for different reasons. The IH predicts this result irrespective of differences between the languages, while the VH only predicts this result in cases where the languages in contact differ. Given that Catalan has a higher rate of overt pronominal subjects than Spanish does in 3sg, the VH also predicts a higher rate in bilinguals. In addition, it also predicts that bilinguals show a diminished sensitivity to the variable verb form ambiguity, since it is lower ranked and it is not significant in Catalan. Overall, the predictions of the VH are the same as in 1sg, that is, language contact effects take place only in the lower ranked constraints. The bilingual groups exhibited higher rates of overt pronominal subjects, even higher than the Catalan control group, and, as was the case with the Catalan control group, subject expression was only constrained by the variable

Table 11. Third person singular subject expression in Catalan-dominant bilinguals. Multivariate regression analysis of the contribution of internal and external factors to the probability of producing a third person singular overt pronominal subject versus a null subject for Catalan-dominant bilinguals; factor groups selected as significant in gray background.

	Factor weight	%	N
CONNECT			
Different referent	**0.72**	15.0	120
Same referent, different TAM	0.46	5.5	55
Same referent, same TAM	0.28	2.5	119
Range	46		
AMBIGUITY			
Ambiguous	**0.51**	8.8	107
Unambiguous	0.49	8.2	187
Range	2		
VERB TYPE			
Stative	**0.51**	8.6	70
External	0.50	8.3	216
Mental	N/A	0	8
Range	1		
Total N			294
Corrected mean			.064
Log likelihood			−75.856
Significance			.002

TAM: tense, aspect, mood.

connect. In this case, since differences existed between Catalan and Spanish, the differences attested between monolingual and bilingual Spanish could be due to convergence, simplification, or a combination of both. In 3sg, the predictions of the VH bear out in our data.

Conclusion

This paper introduces the VH as a proposal for examining and understanding language contact. In particular, it proposes a continuum of language contact effects depending on the how variable or categorical, based on relative frequency, a linguistic distribution is, such that variable distributions, where more than one form can be used, are more susceptible to cross-linguistic influence than categorical distributions, where a specific form is used (near) exclusively. Using subject expression in Spanish-Catalan bilinguals, the predictions of the IH (Sorace, 2011) and those of the VH were contrasted. In particular, this paper examines the use of null and overt pronominal subjects as constrained by the variable connect, which lies at the syntax–pragmatics interface but is highly ranked in Spanish, verb form ambiguity, which lies at the morphology–syntax interface and is low ranked in Spanish, and verb type, which lies at the lexico-semantic interface with syntax and is not significant in Spanish. The IH would predict that bilinguals differ from monolinguals in their distribution of null and overt subjects more when it is constrained by pragmatic factors (connect) than when it is constrained by morphology (verb form ambiguity) or lexico-semantics (verb type). The VH, on the contrary, would predict that bilinguals differ from monolinguals in their distribution of null and

overt pronominal subjects more when it is constrained by lower ranked variables (verb type and verb form ambiguity) than by higher ranked variables (connect). The data from 1sg and 3sg subjects in Spanish are consistent with the predictions of the VH, as no difference was attested in bilingual and monolingual speakers' use of overt and null subjects in contexts of same or different referents. In contrast, differences between monolingual and bilingual speakers were attested with respect to their distribution across verb types and in the presence or absence of morphological ambiguity.

The VH is one more approach to examining and understanding differences between bilingual and monolingual speech. In line with Thomason and Kaufman (1988), it assumes differences in the outcomes depending on the intensity of language contact, an idea that requires further testing and that was not evaluated in the present paper. It, however, expands on previous attempts at explaining the selectivity of language contact effects, that is, the internal factors affecting language contact outcomes. It proposes that changes are not affecting specific structures because they lie at specific interfaces. Such a proposal poses problems with identifying what interface a structure belongs to. In addition, often times structures lie at more than one interface. Instead, the current proposal is grounded in the notion of variation, which is measurable. The VH predicts that categorical distributions are not affected by language contact, while variable distributions may be affected by language contact. In this paper, the focus has been on subject expression, with a comparison of contexts where the uses are more predictable (as in same versus different referent contexts) and those where the uses are less predictable (as with ambiguous versus with unambiguous verb forms).

There are data from other studies that seem consistent with this proposal. For instance, examining subject position in Spanish, several authors report more difficulty in the use of postverbal subjects in Spanish in narrow focus, which is regulated by the syntax–pragmatics interface but with a not near categorical distribution, than with unaccusative subjects, which is regulated by the syntax–semantics interface and in a near categorical distribution, in bilingual speakers of different language pairings (English-Spanish heritage speakers: Gómez Soler, 2013; Montrul, 2003; Prada Pérez & Pascual y Cabo, 2012; English L1 Spanish second language learners: Domínguez & Arche, 2008; Hertel, 2003; and Spanish-Catalan bilinguals: Prada Pérez, 2010b). Future research should examine other structures and language pairings in order to test this hypothesis further.

Lastly, this proposal is compatible with several theoretical approaches to linguistics. Variation is acknowledged as part of the grammar in generative approaches to syntax (Richards, 2008: 114). Variation is an integrative part of the grammar in functional approaches. From an acquisition point of view, it is reasonable to posit that variable distributions are inconsistent in the input and, thus, may pose more difficulty in acquisition (Papp, 2000; Sorace, 2000). Lastly, this approach is also compatible with sociolinguistic studies on language contact, as external sociolinguistic factors also play a role in the outcomes of contact between two or more languages.

Notes

1. In the examples in (2), the target verb is highlighted in bold.
2. In variationist sociolinguistics, "apparent time" is a cross-sectional study where a representative sample is collected from speakers of different ages under the assumption that differences across age groups are indicative of linguistic change.

3. Only two persons were included in this analysis for theoretical as well as practical reasons. Previous literature seems to report conflicting results as to the effect of English on Spanish SPE in the USA. Those studies not reporting an effect seem to focus on 1sg, while those reporting an effect either include all persons or focus on 3sg. The rates and uses of overt pronouns seem to be different in 1sg and 3sg in monolingual Spanish, given the deictic nature of the former and the referential nature of the latter. Lastly, for practical reasons 2sg was not included because in sociolinguistic interviews there are hardly any tokens in 2sg.

4. These exclusions were made to avoid contexts were variation is limited, as in narrow focus, where null subjects are rarely used (however, see Amaral & Schwenter, 2005) or newly introduced referents, where an overt lexical subject would generally be used.

5. This table includes data from all speaker groups, that is, data both in Catalan and in Spanish. Since the effect of person was so large, the results from the other variables are difficult to interpret with both grammatical persons included in the analysis. A comparison among the groups will be discussed below.

6. In this case, since both languages are similar with respect to this variable, cross-linguistic influence might not be expected to take place. However, other types of language contact effects, such as simplification, may still take place in these cases, as argued by Sorace (2011).

References

Alfaraz, G. G. (2015). Variation of overt and null subject pronouns in the Spanish of Santo Domingo. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 17–37). Washington, DC: Georgetown University Press.

Alonso-Ovalle, L., Clifton, C., Frazier, L., & Férnandez-Solera, S. (2005). Null vs. overt pronouns and the topic–focus articulation in Spanish. *Journal of Italian Linguistics*, *14*, 151–169.

Amaral, P. M., & Schwenter, S. A. (2005). Contrast and the (non-) occurrence of subject pronouns. In D. Eddington (Ed.), *Selected proceedings of the 7th Hispanic linguistics symposium* (pp. 116–127). Somerville, MA: Cascadilla Proceedings Project.

Ávila-Jiménez, B. I. (1995). A sociolinguistic analysis of a change in progress: Pronominal overtness in Puerto Rican Spanish. *Cornell Working Papers in Linguistics*, *13*, 25–47.

Bailey, G. (2004). Real and apparent time. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 312–332). Oxford, UK: Blackwell Publishing.

Bayley, R., & Pease-Álvarez, L. (1996). Null and expressed pronoun variation in Mexican-descent children's Spanish. In J. Arnold, R. Blake, & B. Davidson (Eds.), *Sociolinguistic variation: Data, theory, and analysis* (pp. 85–99). Stanford, CA: CSLI Publications.

Bayley, R., & Pease-Álvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, *9*, 349–371.

Camacho, J. (2013). *Null subjects*. Cambridge, UK: Cambridge University Press.

Cameron, R. (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. PhD dissertation, University of Pennsylvania, Philadelphia, PA.

Cameron, R. (1994). Switch reference, verb class and priming in a variable syntax. In K. Beals, Chicago Linguistic Society, & University of Chicago, *Papers from the 30th regional meeting of the Chicago Linguistic Society 1994: Volume 2* (pp. 27–45). Chicago, IL: Chicago Linguistic Society.

Cameron, R. (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics*, *6*(7), 1–27.

Carvalho, A. M., & Child, M. (2011). Subject pronoun expression in a variety of Spanish in contact with Portuguese. In J. Michnowicz & R. Dodsworth (Eds.), *Selected proceedings of the 5th workshop on Spanish sociolinguistics* (pp. 14–25). Somerville, MA: Cascadilla Proceedings Project.

Carvalho, A. M., Orozco, R., & Shin, N. L. (Eds.). (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.

Casanova Seuma, L. (1999). El sujeto en catalán coloquial. *Revista española de lingüística*, *29*(1), 105–131.

Prada Pérez, A. de. (2015). First person singular subject pronoun expression in Spanish in contact with Catalan. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 121–142). Washington, DC: Georgetown University Press.

Domínguez, L., & Arche, M. J. (2008). Optionality in L2 grammars: The acquisition of SV/VS contrast in Spanish. In H. Chan, H. Jacob, & E. Kapia (Eds.), *BUCLD 32 proceedings* (pp. 96–107). Sommerville, MA: Cascadilla Press.

Enríquez, E. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid, Spain: Consejo Superior de Investigaciones Científicas.

Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, *88*, 526–557.

Erker, D., & Otheguy, R. (2016). Contact and coherence: Dialectal leveling and structural convergence in NYC Spanish. *Lingua*, *172–173*, 131–146.

Flores, N., & Toro, J. (2000). The persistence of dialect features under conditions of contact and leveling. *Southwest Journal of Linguistics*, *19*, 31–42.

Flores-Ferrán, N. (2002). *Subject personal pronouns in Spanish narratives of Puerto Rican in New York City: A sociolinguistic perspective*. Munich, Germany: Lincom Europa.

Flores-Ferrán, N. (2004). Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change*, *16*, 49–73.

Flores-Ferrán, N. (2017). A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistic Compass*, *1*(6), 624–652.

Gómez Soler, I. (2013). Non-native trends in the acquisition of split-intransitivity and its relation to variable input. Paper presented at the *17th Hispanic Linguistics Symposium*, Ottawa, Canada.

Hertel, T. J. (2003). Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research*, *19*(4), 273–304.

Holmberg, A. (2005). Is there a little pro? Evidence from Finnish. *Linguistic Inquiry*, *36*, 533–564.

Holmquist, J. (2012). Frequency rates and constraints on subject personal pronoun expression: Findings from the Puerto Rican highlands. *Language Variation and Change*, *24*, 203–220.

Hopp, H. (2011). Extended patterns and computational complexity. *Linguistic Approaches to Bilingualism*, *1*(1), 43–47.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Cambridge, MA: The MIT Press.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lastra, Y., & Butragueño, P. M. (2015). Subject pronoun expression in oral Mexican Spanish. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 41–60). Washington, DC: Georgetown University Press.

Liceras, J. (2014). Linguistic theory and the acquisition of Spanish as a non-primary language: Wavering between formal and empirical approaches, is there a happy medium? Paper presented at the *Hispanic Linguistics Symposium (HLS 2014)*, Purdue, IN.

Liceras, J., & Fernández Fuertes, R. (2017). Subject omission/production in child bilingual English and child bilingual Spanish: The view from linguistic theory. *Probus*. doi:10.1515/probus-2016–0012

Lipski, J. M. (1994). *Latin American Spanish*. London, UK: Longman.

Lipski, J. M. (1996). Patterns of pronominal evolution in Cuban-American bilinguals. In A. Roca & J. B. Jensen (Eds.), *Spanish in contact: Issues in bilingualism* (pp. 159–86). Somerville, MA: Cascadilla.

Lizardi, C. M. (1993). *Subject position in Puerto Rican wh-questions: Syntactic, sociolinguistic and discourse factors*. PhD Dissertation, Cornell University, Ithaca, NY.

Martínez, C. (2011). *Null and overt subjects in a variable system: The case of Dominican Spanish*. PhD dissertation, University of Ottawa, Canada.

Montalbetti, M. (1984). *After binding. On the interpretation of pronouns*. PhD dissertation, MIT, Cambridge, MA.

Montrul, S. (2003). Bilingual inaccusitivity. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the annual Boston conference on language development 27* (pp. 556–567). Somerville, MA: Cascadilla Press.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition*, *7*(2), 125–142.

Morales, A. (1997). La hipótesis funcional y la aparición del sujeto no nominal: El español de Puerto Rico. *Hispania*, *80*, 153–165.

Orozco, R. (2015). Pronominal variation in Colombian Costeño Spanish. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 17–37). Washington, DC: Georgetown University Press.

Orozco, R., & Guy, G. (2008). El uso variable de los pronombres sujetos: ¿Qué pasa en la costa Caribe colombiana? In M. Westmoreland & J. A. Thomas (Eds.), *Selected proceedings of the fourth workshop on Spanish sociolinguistics* (pp. 70–80). Somerville, MA: Cascadilla Proceedings Project.

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York. Language contact, dialectal leveling, and structural continuity*. Oxford, New York: Oxford University Press.

Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language*, *83*, 1–33.

Papp, S. (2000) Stable and developmental optionality in native and non-native Hungarian grammars. *Second Language Research*, *16*(2), 173–200.

Paredes Silva, V. L. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, *5*, 35–49.

Perlmutter, D. M. (1971). *Deep and surface structure constraints in syntax*. New York, NY: Holt, Rinehart, and Winston.

Pires, A., & Rothman, J. (2011). An integrated perspective on comparative bilingual differences: Beyond the Interface problem? *Linguistic Approaches to Bilingualism*, *1*(1), 74–78.

Posio, P. (2015). Subject pronoun usage in formulaic sequences: Evidence from Peninsular Spanish. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A c-dialectal perspective* (pp. 59–80). Washington, DC: Georgetown University Press.

Prada Pérez, Ana de. (2009). *Subject expression in Minorcan Spanish: The syntactic and pragmatic consequences of contact with Catalan*. PhD. Dissertation. The Pennsylvania State University, University Park, PA.

Prada Pérez, Ana de. (2010a). Variation in Subject Expression in Western Romance. In S. Colina, A. Olarrea & A. M. Carvalho (Eds.), *Romance Linguistics 2009: Selected papers from the 39th Linguistic Symposium on Romance Languages (LSRL)* (pp. 267–284). Amsterdam, Netherlands: John Benjamins.

Prada Pérez, A. de, & Pascual y Cabo, D. (2012). Interface heritage speech across proficiencies: Unaccusativity, focus, and subject position in Spanish. In K. Geeslin, & M. Díaz-Campos (Eds.), *Selected proceedings of the 14th Hispanic linguistics symposium* (pp. 308–318). Somerville, MA: Cascadilla Proceedings Project.

Prada Pérez, Ana de. (2015). First Person Singular Subject Pronoun Expression in Spanish in Contact with Catalan. In A. M. Carvalho, R. Orozco & N. L. Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 121–142). Washington, DC: Georgetown University Press.

Ranson, D. L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change*, *3*(2), 133–152.

Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. Cambridge, MA: The MIT Press.

Richards, M. (2008). Two kinds of variation in a minimalist system. In F. Heck, G. Müller, & J. Trommer (Eds.), *Varieties of competition* (pp. 133–162). Leipzig, Germany: University of Leipzig.

Rothman, J. (2009). Pragmatic deficits with syntactic consequences: L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics*, *41*, 951–973.

Sheehan, M. (2006). *The EPP and null subjects in Romance*. PhD dissertation, Newcastle University, England.

Shin, N. L. (2012). Variable use of Spanish subject pronouns by monolingual children in Mexico. In K. Geeslin & M. Díaz-Campos (Eds.), *Selected proceedings of the 2010 Hispanic linguistics symposium* (pp. 130–141). Somerville, MA: Cascadilla Proceedings Projects.

Shin, N. L. (2013). Women as leaders of language change: A qualification from the bilingual perspective. In A. M. Carvalho & S. Beaudrie (Eds.), *Selected proceedings of the 6th workshop on Spanish sociolinguistics* (pp. 135–147). Somerville, MA: Cascadilla Proceedings Projects.

Shin, N. L., & Otheguy, R. (2013). Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society*, *42*, 429–452.

Silva-Corvalán, C. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elías-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93–120). Cambridge, MA: Cambridge University Press.

Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford, UK: Clarendon.

Sorace, A. (2000). Syntactic optionality in non-native grammars. *Second Language Research*, *16*, 93–102.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sorace, A. (2012). Pinning down the concept of interface in bilingual development: A reply to peer commentaries. *Linguistic Approaches to Bilingualism*, *2*(2), 209–217.

Tagliamonte, S. (2002). Comparative sociolinguistics. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 729–763). Oxford, UK: Blackwell Publishers.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.

Toribio, A. J. (2004). Convergence as an optimization strategy in bilingual speech: Evidence from code-switching. *Bilingualism: Language and Cognition*, *7*, 165–173.

Torres-Cacoullos, R., & Travis, C. E. (2010). Variable Yo expression in New Mexico: English influence? In S. Rivera-Mills, & D. Villa (Eds.), *Spanish of the U.S. Southwest: A language in transition* (pp. 185–206). Madrid, Spain: Iberoamericana/Vervuert.

Travis, C. E. (2005). The yo-yo effect: Priming in subject expression in Colombian Spanish. In R. S. Gess, & E. J. Rubin (Eds.), *Theoretical and experimental approaches to romance linguistics: Selected papers from the 34th linguistic symposium on romance languages* (pp. 329–349). Amsterdam, Netherlands: John Benjamins.

Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, *19*, 101–135.

Author biography

Ana de Prada Pérez is an Assistant Professor at the University of Florida. She examines the effects of cross-linguistic influence and code-switching on the speech of bilinguals in two different communities with different language pairings: Spanish in contact with English in the U.S. and Spanish in contact with Catalan in Minorca, Spain. Her research has been published in several journals (Studies in Hispanic and Lusophone Linguistics, Spanish in Context, etc.) as well as several edited books.