



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Alder Luis Pérez Córdoba

PRESENÇA/AUSÊNCIA DO PRONOME PESSOAL SUJEITO NO
ESPAÑHOL FALADO NO CARIBE COLOMBIANO

São José do Rio Preto

2019

Alder Luis Pérez Córdoba

PRESENÇA/ AUSÊNCIA DO PRONOME PESSOAL SUJEITO NO
ESPAANHOL FALADO NO CARIBE COLOMBIANO

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, como parte dos requisitos para obtenção do título de Doutor em Estudos Linguísticos.

Agência financiadora: AUIP/UNESP

Orientador: Prof. Dr. Roberto Gomes Camacho

São José do Rio Preto

2019

P438p

Pérez-Córdoba, Alder Luis

Presença/ausência do pronome pessoal sujeito no espanhol falado no caribe colombiano / Alder Luis Pérez-Córdoba. -- São José do Rio Preto, 2019

165 f. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Roberto Gomes Camacho

1. Linguística. 2. Sociolinguística variacionista. 3. Espanhol colombiano. 4. Sujeito. 5. Pronome pessoal. I. Título.

Alder Luis Pérez Córdoba

**PRESENÇA/ AUSÊNCIA DO PRONOME PESSOAL SUJEITO NO
ESPANHOL FALADO NO CARIBE COLOMBIANO**

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, como parte dos requisitos para obtenção do título de Doutor em Estudos Linguísticos.

Agência Financiadora: AUIP/UNESP

Comissão Examinadora

Prof. Dr. Roberto Gomes Camacho
UNESP – Câmpus de São José do Rio Preto
Orientador

Profa. Dra. Angélica Terezinha Carmo Rodrigues
UNESP – Câmpus de Araraquara

Profa. Dra. Sílvia Figueiredo Brandão
Universidade Federal do Rio de Janeiro

Profa. Dra. Sanderleia Roberta Longhin
UNESP – Câmpus de São José do Rio Preto

Profa. Dra. Talita Storti Garcia
UNESP – Câmpus de São José do Rio Preto

São José do Rio Preto

7 de agosto de 2019

*A mi esposa Marcia y a mis hijos Luis José y Daniel José;
a mis padres Filomena y Luis y a mi familia,
con amor y gratitud*

AGRADECIMENTOS

A Deus, primeiramente, porque Ele tudo dispõe no momento certo, mesmo sendo fraca minha fé. Seus propósitos estão acima de nosso entendimento.

A meu orientador, Prof. Dr. Roberto Gomes Camacho, por aceitar desde o início ser meu orientador, mesmo sem antes me conhecer; pela paciência, pela confiança em meu potencial, pelas orientações seguras e contribuições a este trabalho e pelas inúmeras correções de meu dialeto “portunhol”. Mas, sobretudo, por sua generosidade como pessoa e como acadêmico. Muito obrigado professor.

Ao PPGEL (Ibilce/Unesp), por aceitar-me no curso de doutorado, e aos professores pelos caros ensinamentos nas disciplinas e em outros contextos acadêmicos.

Às Profas. Dras. Sanderleia Roberta Longhin e Talita Storti Garcia (Ibilce/Unesp), pelas valiosas contribuições durante participação como avaliadoras no Exame de Qualificação. E a elas e as Profas. Dras. Angélica Terezinha Carmo Rodrigues (UNESP/ Araraquara) e Sílvia Figueiredo Brandão (UFRJ), pelas contribuições durante a defesa da tese.

Ao Prof. Rafael Orozco e à Profa. Paola Bentivoglio (*in memoriam*) pela disposição de compartilhar conhecimentos e pelas orientações seguras nas situações em que recorri a seu auxílio.

A minha esposa, Marcia, minha força, minha alegria, meu apoio, minha companheira, *gracias por estar siempre presente, aun cuando yo estuviera ausente. Gracias por todo.*

A meus pais Luis e Filomena, a meus irmãos Kelly, Katty, Luis e Ketty, e a minhas sobrinhas María L. e Gabriela, pelo apoio constante e pela ajuda generosa. Nunca me senti sozinho, mesmo estando distante de meu país, porque minha família está sempre presente em todos os momentos. Também aos meus tios Pérez López pela presença constante na distância.

A toda família Córdoba Mercado, especialmente minha sogra Luz Maria, pela ajuda alegre e desinteressada; e aos irmãos Mercado Hoyos e família, pela força que me deram nos momentos de dificuldades.

A minha família no Brasil: Maria do Perpétuo Socorro Teixeira, Márcia Ferrari, Deo da Silveira, Mislaine Silva e Sirlei Custódio, e aos amigos do grupo São Padre Pio, da comunidade da paróquia Nossa Senhora de Fátima e do bairro Jardim Nazareth. Muito obrigado pela acolhida e por fazer-nos sentir em casa.

A meus amigos colombianos Amparo Olmos, Alex Silgado e Alberto Espitia e as minhas colegas e amigas brasileiras Aline Garcia e Simone Cordeiro, pela amizade e pela disposição em dar ajuda quando mais necessitei. A Liliam Cuartas, por facilitar-me diligentemente o acesso ao *corpus* de Cartagena.

À AUIP/UNESP pelo apoio financeiro com a bolsa de doutorado para qualificar-me.

A todos, *mil gracias!*

RESUMO

Apesar da ampla bibliografia sobre a expressão do sujeito pronominal, como resultado das pesquisas desenvolvidas em diferentes dialetos do espanhol na América hispânica, nos Estados Unidos e na Espanha, encontram-se escassos trabalhos sobre duas variedades linguísticas colombianas. Tomando como suporte teórico a Teoria da Variação e Mudança (LABOV, [1972] 2008), o objetivo deste trabalho foi analisar a expressão ou não do pronome sujeito no espanhol falado no Caribe colombiano para determinar que fatores linguísticos e sociais motivam esse uso variável. A amostra esteve constituída por gravações de 18 informantes para cada *corpus* sociolinguístico (Barranquilla, Cartagena e Valledupar), que compõem a variedade falada no caribe colombiano, coletados com base na metodologia do PRESEEA (MORENO, 1996). Uma vez definido o envelope da variação, codificaram-se 7595 ocorrências para testar 15 variáveis independentes, onze das quais linguísticas (pessoa, número e especificidade do sujeito; paralelismo formal; modo; tempo verbal; progressividade; morfologia verbal; ambiguidade; tipo de oração; correferencialidade; turno da fala; classe semântica do verbo) e quatro sociais (faixa etária; sexo/gênero do falante; grau de escolaridade; subvariedade dialetal). Além disso, testou-se a variável independente posição pronominal, para saber quais fatores linguísticos e sociais determinam também a posposição. A análise quantitativa foi realizada mediante o uso do programa Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005) e Rbrul (JOHNSON, 2009, 2019). Os resultados de expressão pronominal de 41,9%, a baixa posposição pronominal e os altos índices percentuais das duas primeiras pessoas do singular revelaram que a variedade do Caribe colombiano integra, no que tange a esse fenômeno, o grande dialeto conhecido por espanhol caribenho. A análise multivariada mostrou que diversas variáveis linguísticas (por exemplo, pessoa, número e especificidade, correferencialidade, paralelismo formal, modo e tempo verbal, entre outras) e três variáveis sociais (sexo/gênero, faixa etária e grau de escolaridade) desempenharam um papel significativo na variação pronominal, confirmando a força dos grupos de fatores linguísticos mais do que os sociais. Análises bem detalhadas confirmaram que grupos aparentemente não muito significativos (por exemplo, ambiguidade morfológica) ganham relevância quando se fazem análises deles interagindo com outros fatores. Com efeito, os resultados obtidos apontaram não somente para a existência de semelhanças com o espanhol caribenho em relação à expressão e posição do sujeito pronominal, mas também em relação aos grupos de fatores mais significativos e específicos que parecem ser mais recorrentes nos dialetos caribenhos. Ressalte-se ainda que os resultados também apontaram para uma considerável convergência com o comportamento dos diferentes dialetos do espanhol, o que parece um indício significativo de superação de limites dialetais.

Palavras-chave: Pronome sujeito; espanhol falado; Caribe colombiano; variacionismo; Preseea.

RESUMEN

Pese a la amplia bibliografía sobre la expresión del sujeto pronominal, como resultado de las diferentes investigaciones sobre diferentes dialectos en Hispanoamérica, EEUU e España, se encuentran escasos trabajos sobre dos variedades lingüísticas colombianas. Tomando como fundamento teórico la Teoría de la variación y cambio lingüístico (LABOV, [1972] 2008), el objetivo de este trabajo fue analizar la expresión del pronombre personal sujeto en el español hablado en el Caribe colombiano, para determinar qué factores lingüísticos e extralingüísticos favorecen su uso variable. La muestra estuvo constituida por grabaciones de 18 informantes por cada corpus sociolingüístico (Barranquilla, Cartagena e Valledupar), que conforman la variedad hablada en el Caribe colombiano, recolectados con base en la metodología del PRESEEA (MORENO, 1996). Una vez definido el contexto variable, se codificaron 7595 casos para verificar la influencia de 15 variables independientes, once lingüísticas (persona, número y especificidad del sujeto, perseverancia, modo verbal, tiempo verbal, progresividad, morfología verbal, ambigüedad morfológica, tipo de oración, correfencialidad, turno de habla, clase semántica del verbo) y cuatro sociales (edad, sexo/género, nivel de instrucción y dialecto). Además, se analizó la variable independiente posición pronominal, para saber cuáles factores lingüísticos y sociales determinan también la posposición. El análisis cuantitativo se realizó con el programa Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005) y Rbrul (JOHNSON, 2009, 2019). La tasa de expresión pronominal de 41,9%, la baja posposición pronominal y los elevados porcentajes de las dos primeras personas del singular revelaron que la variedad del Caribe colombiano hace parte la variedad conocida como español caribeño. El análisis multivariado demostró que diversas variables lingüísticas (por ejemplo, persona, número y especificidad del sujeto, correfencialidad, perseverancia, modo verbal e tiempo verbal, entre otras) y tres sociales (sexo/género, nivel de instrucción, edad) juegan un papel significativo en la variación pronominal, confirmando la fuerza de los grupos de factores lingüísticos mucho más que los sociales. Análisis más detallados confirmaron que los grupos aparentemente poco significativos (ambigüedad morfológica, por ejemplo) adquieren bastante relevancia cuando se cruzan con otros factores. En efecto, los resultados obtenidos apuntalan no solo para la existencia de semejanzas con el español caribeño en cuanto a tasas y posición del sujeto pronominal, sino también en relación a los grupos de factores más significativos y específicos que parecen ser más constantes en los dialectos caribeños. Se subraya además que los resultados también apuntalan para una considerable convergencia con el comportamiento de diferentes variedades del español, lo que parece un indicio significativo de superación de límites dialectales.

Palabras clave: Pronombre personal sujeto; español hablado; Caribe colombiano; variacionismo; Preseea.

ABSTRACT

Despite of the broad literature on expression on pronominal subjects that has come out of research developed in different Spanish dialects from the Hispanic America, the United states, and Spain, there are few works on two Colombian linguistic varieties. Taking the theory of variation and change as theoretical support (LABOV, [1972] 2008), the aim of this study was to analyze the expression of subject pronoun in Spanish language spoken in the Colombian Caribbean to determine which linguistic and social factors motivate this variable usage. The sample was made up of 18 recordings for each sociolinguistic corpus (Barranquilla, Cartagena, and Valledupar), which make up the variety spoken in the Colombian Caribbean, and the elicitation was based on the PRESEEA methodology (MORENO, 1996). Once the variation envelope was defined, 7595 occurrences were coded to test 15 independent variables, eleven of which were linguistic (person, number, and specificity of the subject; priming; mode; verbal tense; progressivity; verbal morphology; ambiguity; sentence type; co-referentiality; speech shifts; verbal class), and four were social (age; gender of the speaker; educational level; dialectal subvariety). In addition, the independent pronominal position variable was tested to identify which linguistic and social factors also determine the postposition. The quantitative analysis was performed using the Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005) and RBrul (JOHNSON, 2009, 2019) programs. For the broad dialect known as Caribbean Spanish, the results revealed that the variety of the Colombian Caribbean integrates concerning this phenomenon, confirmed by the percentage of the pronominal expression (41.9%), the low pronominal postposition, and the high percentage indexes of the first two people of the singular. The multivariate analysis showed that several linguistic variables (*i.e.* person, number and specificity, co-referentiality, priming, mode, and verbal tense; etc.) and three social variables (gender, age, and educational level) played a significant role in the pronominal variation, confirming a highest strength of the groups of linguistic factors than the social ones. In fact, the results obtained pointed not only to the existence of similarities with Caribbean Spanish in relation to the expression and position of the pronominal subject but also in relation to the groups of the most significant and specific factors that seem to be more recurrent in the Caribbean dialects. It should be highlighted that the results also indicated a considerable convergence with the behavior of the different dialects of Spanish, which seems to be a significant indication of overcoming dialectal boundaries.

KEYWORDS: subject pronoun; spoken Spanish; Colombian Caribbean; variationism; Preseea.

LISTA DE ILUSTRAÇÕES

GRÁFICOS

Gráfico 1 – Línguas do Caribe colombiano multilíngue.....	27
Gráfico 2 – Gradualidade da prominência cognitiva e da informatividade textual.....	48
Gráfico 3 – Árvore de inferência condicional para correferencialidade e paralelismo formal... ..	122
Gráfico 4 – Árvore de inferência condicional para correferencialidade e ambiguidade.	123
Gráfico 5 – Árvore de inferência condicional para pessoa e tempo.....	125
Gráfico 6 – Árvore de inferência condicional para correferencialidade e pessoa.....	127
Gráfico 7 – Árvore de inferência condicional para sexo/gênero e faixa etária.....	135

MAPAS

Mapa 1 – Prováveis áreas dialetais do espanhol colombiano.....	22
Mapa 2 – Panorama multilíngue da região do Caribe.....	28
Mapa 3 – Zona e possíveis subzonas dialetais do espanhol caribenho colombiano.....	31

QUADROS

Quadro 1 – Classificação dialetal do espanhol da Colômbia.....	23
Quadro 2 – Características fonético-fonológicas presentes no espanhol colombiano.....	25
Quadro 3 – Principais características morfossintáticas do espanhol colombiano.....	25
Quadro 4 – Características fônicas e morfossintáticas do espanhol caribenho colombiano.....	30
Quadro 5 – Subdialetos do espanhol caribenho colombiano e áreas de dispersão.....	32
Quadro 6 – Distribuição dos informantes na amostra.....	64

LISTA DE TABELAS

Tabela 1 – Distribuição de presença e ausência de PPS.....	82
Tabela 2 – Frequência de pronomes sujeitos expressos em diferentes variedades.....	83
Tabela 3 – Hierarquia dos grupos de fatores linguísticos e sociais no Caribe colombiano comparados com os de Barranquilla.....	85
Tabela 4 – Grupo de fatores internos significativos para a presença do PPS no Caribe colombiano, ordenados pela amplitude, a partir de 7595 dados.....	89
Tabela 5 – Pessoa, número e especificidade do PPS.....	92
Tabela 6 – Correferencialidade.....	99
Tabela 7 – Paralelismo formal.....	102
Tabela 8 – Modo verbal.....	104
Tabela 9 – Tempo verbal.....	106
Tabela 10 – Tempo, modo e aspecto verbal.....	107
Tabela 11 – Morfologia verbal.....	109
Tabela 12 – Progressividade.....	110
Tabela 13 – Classe semântica do verbo.....	111
Tabela-13a – Classe semântica do verbo ampliada.....	114
Tabela 14 – Ambiguidade morfológica.....	115
Tabela 15 – Hierarquia de fatores do Caribe colombiano e Barranquilla.....	117
Tabela 16 – Cruzamento de correferencialidade e paralelismo formal.....	121
Tabela 17 – Cruzamento de correferencialidade e ambiguidade.....	123
Tabela 18 – Cruzamento de pessoa e tempo.....	125
Tabela 19 – Cruzamento de correferencialidade e pessoa.....	126
Tabela 20 – Grupo de fatores externos significativos para a presença do PPS no Caribe colombiano, ordenados pela amplitude, a partir de 7595 dados.....	128
Tabela 21 – Sexo/gênero.....	129
Tabela 22 – Grau de escolaridade.....	131
Tabela 23 – Faixa etária.....	132
Tabela 24 – Cruzamento de sexo/gênero e faixa etária.....	135
Tabela 25 – Percentual de pronomes antepostos e pospostos ao verbo.....	138
Tabela 26 – Índices percentuais de pronomes e sua posição nas variedades do espanhol.....	138

Tabela 27 – Grupos de fatores significativos para a Posposição do PPS no Caribe colombiano.....	140
Tabela 28 – Comportamento da expressão do PPS nas variedades caribenhas com foco na segunda do singular.....	146
Tabela 29 – Frequência léxica de verbos utilizados 10 ou mais vezes.....	161
Tabela 30 – Percentual de pronome por informante.....	164

LISTA DE ABREVIATURAS E SIGLAS

ADESSE	<i>Alternancias de Diátesis y Esquemas Sintácticos-Semánticos del Español</i>
ALEC	<i>Atlas Lingüístico y Etnográfico de Colombia</i>
BA	Barranquilla
CA	Cartagena
PPS	Pronome Pessoal Sujeito
PR	Peso Relativo
PRESEEA	<i>Proyecto para el estudio sociolingüístico del español de España y América</i>
VA	Valledupar

SUMÁRIO

INTRODUÇÃO	14
1 O ESPANHOL FALADO NO CARIBE COLOMBIANO	20
1.0 Introdução	20
1.1 As línguas e o espanhol da Colômbia	20
1.2 As línguas do Caribe colombiano	26
1.2.1 O espanhol do Caribe colombiano.....	29
1.2.2 Estudos sociolinguísticos sobre o espanhol caribenho colombiano.....	32
2 PRESSUPOSTOS TEÓRICOS	35
2.0 Introdução	35
2.1 A sociolinguística variacionista e a variação sintática	35
2.2 A variável <i>expressão do pronome pessoal sujeito</i>	41
2.2.1 A gramática tradicional.....	41
2.2.2 A Teoria de Princípios e parâmetros.....	42
2.2.3 Os estudos funcionalistas.....	43
2.2.4 O quadro variacionista.....	45
2.2.5 A posposição do sujeito pronominal.....	50
3 PROCEDIMENTOS METODOLÓGICOS	55
3.0 Introdução	55
3.1 Hipóteses de trabalho	55
3.2 Universo da investigação	63
3.3 O envelope da variação	65
3.4 Variáveis linguísticas e sociais	69
3.4.1 Variáveis linguísticas.....	69
3.4.2 Variáveis sociais.....	76
3.4.3 A variável dependente: posição do pronome sujeito.....	77
3.5 Análise quantitativa	77

4	A EXPRESSÃO DO PRONOME PESSOAL SUJEITO NO CARIBE COLOMBIANO.....	81
4.0	Introdução.....	81
4.1	Distribuição do PPS no espanhol caribenho colombiano e em outras variedades de espanhol.....	81
4.2	O efeito das variáveis linguísticas no uso do pronome sujeito no Caribe colombiano.....	83
4.2.1	Pessoa, número e especificidade do sujeito.....	92
4.2.2	Correferencialidade.....	98
4.2.3	Paralelismo formal.....	101
4.2.4	Tempo, modo e aspecto verbal.....	103
4.2.5	Classe semântica do verbo.....	111
4.2.6	Ambiguidade morfológica.....	115
4.2.7	Comparação da hierarquia de fatores significativos com Barranquilla.....	116
4.2.8	Interação entre variáveis independentes linguísticas.....	120
4.3	O efeito das variáveis sociais no uso do pronome sujeito no Caribe colombiano.....	127
4.3.1	Sexo/gênero.....	129
4.3.2	Grau de escolaridade.....	131
4.3.3	Faixa etária.....	132
4.3.4	Interação entre variáveis independentes sexo/gênero e faixa etária.....	134
4.4	A posposição do PPS no Caribe colombiano: fatores que condicionam seu uso.....	137
4.4.1	Distribuição da posição PPS no espanhol caribenho colombiano e em outras variedades.....	137
4.4.2	Variáveis linguísticas e extralinguísticas que favorecem a posposição do sujeito pronominal.....	139
5	CONSIDERAÇÕES FINAIS.....	143
	REFERÊNCIAS.....	149
	APÊNDICE A - Frequência e percentual léxico.	161
	APÊNDICE B - Frequência e percentual por informante.....	165

INTRODUÇÃO

Como a expressão ou não do pronome pessoal sujeito (doravante PPS) no espanhol é altamente variável (OTHEGUY; ZENTELLA; LIVERT, 2007), *yo nací/ Ø nací*, para muitos pesquisadores, essa variação se correlaciona com o paradigma verbal do espanhol: o fato de conter a marca de pessoa permite que o PPS possa ser explícito ou nulo na maioria dos contextos. O modo de explicar esse fenômeno variável difere conforme a perspectiva adotada: gramática tradicional, sintaxe gerativa ou sociolinguística variacionista.

A tradição gramatical basicamente afirma que o espanhol é uma língua de variação livre e explica a presença do PPS como um recurso para indicar ênfase, contraste e desambiguar o referente nos tempos e modos em que pode apresentar-se ambiguidade (RAE, 1973; FERNÁNDEZ SORIANO, 1999). Já para a Teoria de Princípios e Parâmetros (CHOMSKY, 1986), a explicação reside no fato de ser o espanhol uma língua de sujeito nulo, ou seja, que admite sujeitos vazios nas orações finitas que se refletem na flexão verbal. Nessa teoria, as flexões verbais têm um componente de concordância (AGR) com o qual é possível identificar o sujeito quando ele é foneticamente nulo. Diferentes estudos, contudo, especificamente os referentes à variedade caribenha antilhana, evidenciam uma perda clara das características das línguas de sujeito nulo (*pro-drop languages*) e, em compensação, a conversão para uma variedade contendo, às vezes, sujeito obrigatório (*non pro-drop languages*) e, às vezes, sujeito misto (ORTIZ LÓPEZ, 2016).

Estudos variacionistas vêm demonstrando que, em certos contextos, é obrigatória a presença do pronome sujeito (1); em outros, é obrigatória a ausência (2); e, em alguns outros, é variável a inserção ou não do pronome (3), e cada uma das variantes é condicionada por grupos de fatores gramaticais, pragmático-discursivos, semânticos e, com menor frequência, sociais (BENTIVOGLIO, 1987; CARVALHO; OROZCO; SHIN, 2015; OROZCO; GUY, 2008; SILVA-CORVALÁN, 2001, entre outros).

(01) [...] esa ovejita extreviao (extraviada)/ de pronto **soy yo** <silencio> (2 s) [VA-45-23H]¹

(02) [...] «**vamos arreglar** para pagarle a su hija / 559 y yo le pago a su hija». [BA-71-13H]

(03) bueno/ **nací** en Valledupar <silencio> (1'8 s.) en el barrio el Pupo// [VA-11-21M]

¹ O código pode ler-se assim: as duas primeiras letras significam a cidade (BA, Barranquilla; CA, Cartagena; VA, Valledupar). O número seguinte é o número da entrevista. A última parte do código significa o seguinte: o primeiro número faz referência ao grau de escolaridade: "1", baixo; "2", médio; "3", alto. O segundo é a faixa etária: "1", 20-34 anos; "2", 35-54; "3", 55 ou mais. Por fim, "H" é homem e "M" é mulher. Assim o código (VA-45-23H) deverá ser lido como: Valledupar, entrevista 45, informante com grau médio de escolaridade (12 anos aprox.), do terceiro grupo da faixa etária (55+ anos), homem. Se houver alguma dúvida, o capítulo 3 traz informações mais detalhadas sobre as características dos informantes e compreensão mais explícita da codificação.

Partindo do pressuposto sociolinguístico de que a língua é inerentemente variável e que a variação na língua é sistematicamente governada por fatores tanto internos quanto externos a ela (LABOV, [1972] 2008), o objetivo deste trabalho é analisar a variável expressão do pronome pessoal sujeito no espanhol falado no Caribe colombiano para determinar que fatores linguísticos e extralinguísticos a motivam, com base em três variedades, a de Cartagena, a de Barranquilla e a de Valledupar.

Para alcançar esse objetivo geral, propõem-se os seguintes objetivos específicos: i) descrever a distribuição dos pronomes sujeitos no espanhol falado no Caribe colombiano; ii) explicar a influência de fatores linguísticos e sociais na expressão do pronome sujeito na variedade objeto de estudo; iii) determinar a coerência intradialetal em relação ao fenômeno linguístico enfocado e contrastar seu comportamento com outras variedades do espanhol já estudadas.

Não é recente o interesse pelo estudo da variação do sujeito no espanhol e, particularmente, do PPS. Os trabalhos pioneiros remontam aos últimos anos da década de 1970 e início dos anos 1980 e versam sobre o espanhol de Buenos Aires (BARRENECHEA; ALONSO, 1977), o espanhol do oeste de Los Angeles (falado por mexicanos e mexicano-americanos) (SILVA-CORVALÁN, 1977), de Caracas (BENTIVOGLIO, 1980)², de Santiago de Chile (CIFUENTES, 1980-81) e de Porto Rico (MORALES, 1982), entre os mais conhecidos. Esses trabalhos despertaram o interesse pelo uso variável do sujeito em outras variedades do espanhol com comunidades monolíngues ou em situação de contato, fundamentalmente com o inglês.

Entre os trabalhos em comunidades monolíngues, além dos já mencionados de Buenos Aires, Caracas, Santiago de Chile e Porto Rico, estão os do espanhol de Santiago de los Caballeros (MONTENEGRO, 1984), Madri (ENRÍQUEZ, 1984), Porto Rico (MORALES, 1986), Porto Rico e Madri (CAMERON, 1992, 1993), Alcalá de Henares (BLANCO CANALES; 1999), e um grande volume de trabalhos desenvolvido nos últimos quinze anos em que se reativa o interesse pelo tema sobre o espanhol de Caracas (QUIJADA, 2006; PÉREZ, 2010, MARTÍNEZ, 2012), Mérida-Venezuela (BALASCH, 2008), espanhol peninsular (POSIO, 2008), Porto Rico (CLAES, 2011), Cuba (SÁNCHEZ ARROBA, 2013, ALFARAZ, 2018), Lima (CERRÓN-PALOMINO, 2014), Yucatán (MICHNOWICZ, 2015), Cidade do México (LASTRA; MARTÍN, 2015), Santo Domingo (ALFARAZ, 2015),

² Os trabalhos de Silva-Corvalán (1977) e Bentivoglio (1980) se referem a suas dissertações de mestrado, parcialmente publicadas em Silva-Corvalán (1982) e totalmente em Bentivoglio (1987), que são as versões mais conhecidas desses trabalhos.

Granada-Espanha (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), Xalapa (OROZCO, 2016). Alguns desses trabalhos analisam todos os pronomes pessoais e outros analisam só a primeira pessoa ou a terceira pessoa, por essas serem as mais frequentes nas entrevistas sociolinguísticas.

A situação de contato do espanhol com outras línguas é a que gera o maior número de pesquisas e publicações sobre a variável presença ou ausência do PPS. Centrados no contato com o inglês, além do estudo do espanhol de mexicanos e méxico-americanos em Los Angeles, encontramos também os desenvolvidos por colombianos em Miami (HURTADO, 2001, 2005 a e b), de diferentes comunidades hispânicas (Cuba, República Dominicana, Porto Rico, Colômbia, Equador e México) da variedade hispânica de Nova York (FLORES, 2004a; OTHEGUY; ZENTELLA; LIVERT, 2007, SHIN; OTHEGUY, 2009, OTHEGUY; ZENTELLA 2012; ERKER; GUY 2012;), do espanhol de Novo México (TRAVIS, 2007, TORRES CACOULOS; TRAVIS, 2015), de San Antonio, Texas (BAYLEY, CÁRDENAS, TREVIÑO, VÉLEZ, 2012), de porto-riquenhos em Miami (ABREU, 2012), entre outras variedades investigadas.

A respeito do contato do espanhol com outras línguas, há estudos de variedades do espanhol em contato com o português (CARVALHO; CHILD, 2011), com o Catalão (PRADA, 2009) e com línguas indígenas do México (MICHNOWICZ, 2015). Cabe esclarecer que esses trabalhos estão interessados em revisar a influência do contato linguístico em falantes, cuja língua materna é o espanhol, em falantes bilíngues com o espanhol como segunda língua ou como língua estrangeira.

Apesar da ampla bibliografia sobre o tema, como resultado das pesquisas desenvolvidas em diferentes dialetos do espanhol na América hispânica, na Espanha e nos Estados Unidos, há poucos trabalhos sobre as variedades linguísticas colombianas. Até hoje, registram-se os trabalhos de Orozco e Guy (2008), Orozco (2015), com os mesmos dados, Travis (2007) e Travis e Torres Cacoulios (2012), com os mesmos dados. Desses trabalhos, só os de Orozco e Guy (2008) e Orozco (2015) se concentram em todos os pronomes sujeitos do espanhol de Barranquilla, ao passo que as outras autoras pesquisam a expressão do pronome de primeira pessoa do singular (*yo*) no espanhol falado em Cali.

Todos esses trabalhos fazem constatações relevantes de interesse geral, como as seguintes: i) o uso do pronome expresso é motivado por razões de ênfase ou ambiguidade; ii) a mudança da referência e a mudança do turno conversacional contribuem para a expressão pronominal; iii) a semântica do verbo favorece a presença dos pronomes sujeitos; iv) os pronomes expressos na posição de sujeito são mais frequentes nas variedades caribenhas e na

chilena do que na da América continental e na península ibérica; v) a razão pela qual esse fenômeno ocorre com essa frequência elevada nas variedades caribenhas é a supressão do –s e do –n no final de verbos³ e a posição quase obrigatória da ordem SVO; vi) o contato de algumas variedades do espanhol com outras línguas, especialmente com o inglês, provavelmente contribui para uma alta frequência de expressão.

Este trabalho testa, nas variedades aqui investigadas, quase todas essas constatações e outras menos frequentes e recentemente analisadas visando com isso a fornecer um quadro mais amplo do comportamento do uso dos pronomes sujeitos no litoral do Caribe colombiano, além de Barranquilla. A expressão do pronome tem semelhanças e/ou diferenças nos subdialetos do caribe, que, por seu lado, mantêm similaridades e desigualdades com outros dialetos de língua espanhola na América e na Espanha, embora vários fatores pareçam transcender as diferenças dialetais, o que poderia ser um comportamento geral da língua espanhola.

A análise contrastiva é hoje possível uma vez que já há *corpora* disponíveis de três cidades do litoral colombiano, a saber: Barranquilla (RODRÍGUEZ CADENA, 2008b-2010), Cartagena (VÁSQUEZ; CUARTAS, 2017) e Valledupar (CALDERÓN NOGUERA, 2005), coletados e transcritos com base na metodologia do *Proyecto para el estudio sociolingüístico del español de España y América - PRESEEA* (MORENO FERNÁNDEZ, 1996).

Além disso, a homogeneidade dos *corpora* permite realizar comparações das variantes estudadas com critérios de grau mais elevado de confiabilidade, uma das considerações feitas por Silva-Corvalán (1992, 2001), que menciona a possibilidade com *corpora* comparáveis de pesquisa. Nesses trabalhos, ao discutir as debilidades dos estudos realizados sobre a variação dos pronomes sujeitos no espanhol até aquela data, a autora questiona a confiabilidade das comparações, precisamente em razão das diferenças salientes entre os *corpora* e dos critérios de análise. Hoje, essa realidade é possível graças aos *corpora* do PRESEEA e ao subprojeto sobre o pronome sujeito proposto em 2011 no interior desse macroprojeto (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011), no qual esta pesquisa também se enquadra.

Com base no exposto, este trabalho se debruça sobre as seguintes perguntas de pesquisa:

³ Mesmo sendo um aspecto interessante a ser pesquisado em variedades que se consideram caribenhas, esse traço morfológico não será testado neste trabalho pelas limitações do corpus, limitado ao transcrito e sem acesso ao oral, e principalmente porque outros estudos sociolingüísticos já constataram que o –s em posição de coda silábica tende a se conservar quando é marca morfológica em algumas variedades caribenhas colombianas (LAFFORD, 1980, OLMOS; GÓMEZ; PÉREZ, 2010; RODRÍGUEZ CADENA, 2009a). De qualquer modo, a atuação desse fator merece uma revisão em futuras análises.

- (1) Qual é a distribuição da presença/ausência dos pronomes sujeitos no espanhol do Caribe colombiano?
- (2) Dos grupos de fatores linguísticos e extralinguísticos investigados, quais são os que influenciam significativamente a expressão dos pronomes sujeitos no espanhol do Caribe colombiano?
- (3) A variedade caribenha colombiana mostra similaridades ou diferenças com o espanhol caribenho, em relação a percentuais de pronomes expressos, posição do pronome, fatores significativos?
- (4) A variedade estudada, comparada com outras do espanhol, mostra semelhanças ou diferenças nos índices percentuais de expressão dos pronomes sujeitos e nos grupos de fatores significativos?

O trabalho está organizado da seguinte maneira: inicialmente, no capítulo 1, apresenta-se uma breve descrição da situação linguística atual na Colômbia, com destaque para as características mais relevantes do que se poderia chamar do “espanhol colombiano”, ao mesmo tempo em que se discutem os trabalhos dialetológicos sobre essa variedade. Em seguida, descreve-se, brevemente, a situação linguística atual da região do Caribe colombiano e as características mais relevantes do espanhol caribenho da Colômbia, contextualizando-o sempre na variedade do espanhol caribenho. Por fim, discutem-se os trabalhos sociolinguísticos do espanhol da região.

O capítulo 2 começa com uma visão geral da teoria variacionista, introduzindo os conceitos de variação linguística e as definições relacionadas a conceitos-chave na perspectiva da sociolinguística laboviana, como os de variável e variante. Em seguida, define-se a variação sintática, que seguiu seu próprio caminho no quadro teórico variacionista no mundo todo, mas especialmente também na língua espanhola, reajustando-se também os conceitos de variável e variante. Por fim, discute-se por que se pode falar da variável expressão do sujeito, e quais são as variáveis linguísticas e extralinguísticas que mais contribuem para sua expressão, seja anteposta ou posposta, sem deixar de passar antes por uma síntese muito geral do que a gramática tradicional, a sintaxe gerativa e os quadros discursivos e pragmáticos têm dito sobre a expressão do sujeito pronominal no espanhol.

No Capítulo 3, apresentam-se detalhadamente os aspectos metodológicos da presente pesquisa. Primeiramente, levantam-se as hipóteses e se retomam as perguntas de investigação. Na sequência, discutem-se a natureza dos dados empregados e as informações sociais e demográficas dos participantes. A seguir, descreve-se a forma de análise, com base no exame qualitativo das entrevistas e da circunscrição do contexto da variável para a expressão do

sujeito, para, em seguida, apresentarem-se as variáveis independentes incluídas na análise e sua operacionalização. Finalmente, discute-se brevemente a natureza das análises estatísticas realizadas.

O capítulo 4 dedica-se à apresentação e à discussão dos resultados quantitativos obtidos mediante a aplicação do programa estatístico. Primeiramente, apresentam-se os resultados gerais, em termos das frequências e percentuais de pronomes pessoais sujeitos para cada variedade analisada isoladamente e para o espanhol do Caribe colombiano em geral; em seguida, comparam-se os resultados aos de outras variedades previamente estudadas; discutem-se, depois, as variáveis linguísticas e sociais que se mostram mais significativas nos diferentes modelos de regressão logística e também os dados estatísticos em relação a outras pesquisas prévias; por fim, procura-se mostrar para o que apontam os índices percentuais de pronomes sujeitos pospostos em relação aos fatores significativos que contribuem para a expressão nessa posição.

No Capítulo 5, responde-se às perguntas de pesquisa e se discutem as principais generalizações e implicações teóricas advindas da análise, além de acrescentar a contribuição que foi possível dar ao tema com os resultados desta pesquisa. Por fim, aparecem as referências e os apêndices do trabalho.

1 O ESPANHOL FALADO NO CARIBE COLOMBIANO

1.0 Introdução

Neste capítulo, sintetiza-se o contexto geográfico dialetal da pesquisa e se faz uma breve caracterização dos estudos sociolinguísticos na região. Inicialmente, apresenta-se uma sucinta descrição da situação linguística atual no país, destacando-se as características mais relevantes do que se poderia chamar “espanhol colombiano”, além de se repassar a literatura variacionista sobre o espanhol colombiano. Descrevem-se, em seguida, brevemente, a situação linguística atual da região do Caribe colombiano e as características mais relevantes do espanhol caribenho colombiano, inserindo-o sempre no contexto do dialeto espanhol caribenho. Por fim, apresentam-se os trabalhos sociolinguísticos do espanhol da região.

1.1 As línguas e o espanhol da Colômbia

A Colômbia, localizada na região noroeste da América do Sul, é o único país dessa parte do continente que tem acesso ao oceano Atlântico (através do mar Caribe) e ao oceano Pacífico. Tem, além do território continental, um território insular, sendo que as ilhas maiores e mais povoadas são as que constituem o arquipélago de San Andrés, Providencia e Santa Catalina. Segundo o Departamento Administrativo de Estatística da Colômbia (DANE⁴, 2018), a população colombiana é de 45.5 milhões de habitantes e é considerada multiétnica e multicultural de acordo com a Constituição Política (1991).

Essa característica ao mesmo tempo multiétnica e multicultural se reflete linguisticamente na pluralidade de línguas e dialetos espalhados pelo território nacional, até ser considerado o segundo país com mais línguas na América, depois do Brasil. Assim, na Colômbia residem 87 povos indígenas que falam 64 línguas indo-americanas⁵ e uma diversidade de dialetos dessas línguas que podem ser agrupadas em 13 famílias linguísticas. O número de indígenas é superior a um milhão, correspondentes a 2,2% da população, estando a

⁴ Segundo o site do DANE, os dados de 2 de novembro de 2018 são preliminares, mas abrangem geograficamente 99,8% do território. <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/cuantos-somos>

⁵ Sobre o número real de línguas ameríndias não parece haver consenso; provavelmente é por essa razão que se faz menção a um número não definitivo, defendido como 70 por Rodríguez Cadena (2008) e como 65 pelo Ministério da Cultura da Colômbia.

maior parte localizada nas zonas rurais e territórios indígenas reconhecidos oficialmente como *resguardos*⁶ (DANE, 2005).

Adicionalmente, destaca-se a presença de duas línguas crioulas faladas por afrodescendentes: o crioulo *sanandresano*, de base inglesa e o crioulo *palenquero*, de base espanhola. Além da língua *rom*, falada por ciganos situados em diferentes partes do país, pode-se mencionar a língua de sinais colombiana, reconhecida oficialmente em 1996 como meio legítimo de expressão da comunidade surda do país e também a língua espanhola falada pela maior parte da população, equivalente, segundo Moreno Fernández e Otero Roth (2016), a 99,2% da população. Por isso, a nação colombiana é considerada o segundo país com maior número de falantes hispânicos, depois do México.

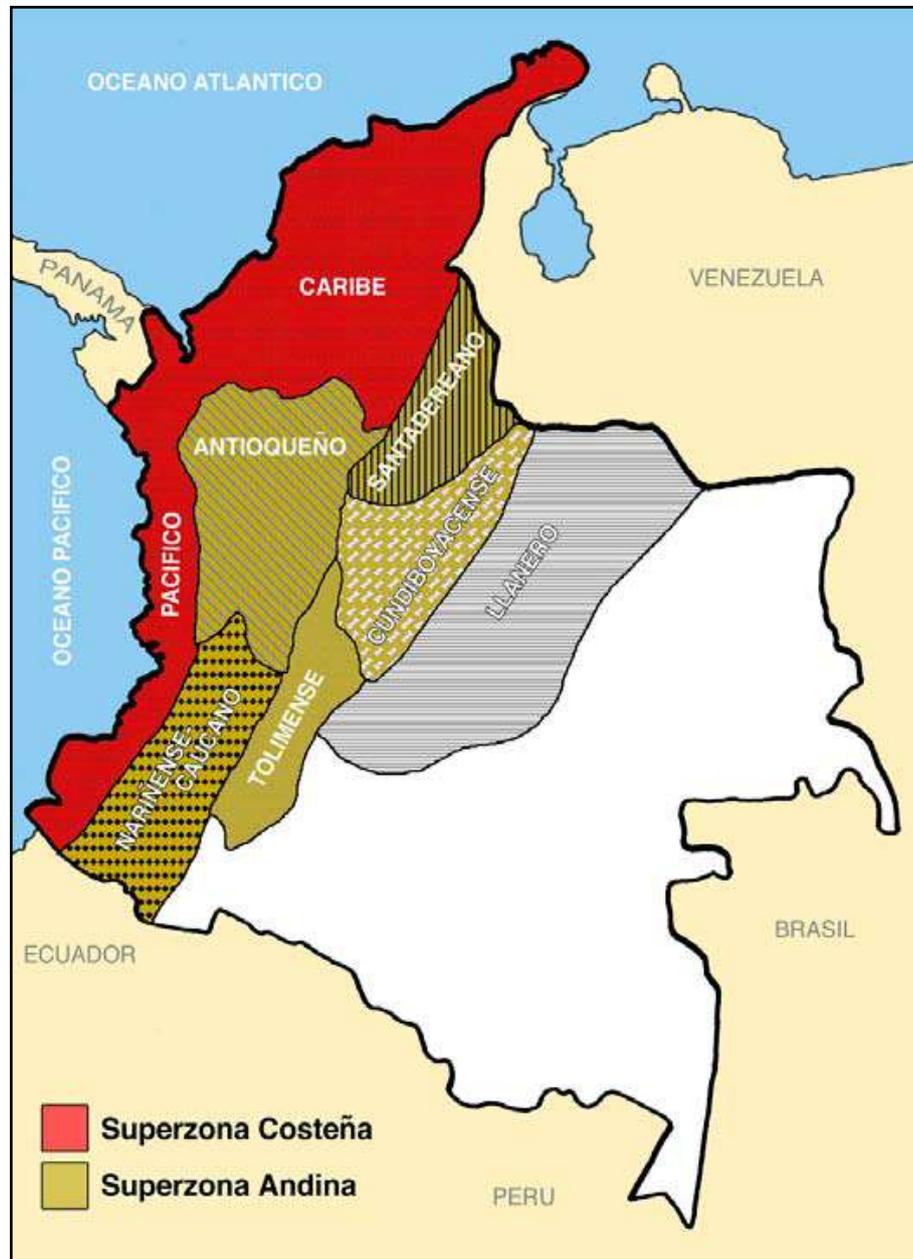
No que tange ao “espanhol colombiano”, cumpre destacar que este não é homogêneo; tanto do ponto de vista geográfico, quanto do ponto de vista social. O fenômeno da variação nessa região é o resultado de diversos fatores sociohistóricos, dentre os quais se destaca a influência das línguas indígenas em toda a extensão do território, e das línguas africanas nas zonas costeiras, de maneira especialmente notável em San Basilio de Palenque e no departamento do Chocó (OROZCO; DÍAZ CAMPO, 2016).

Desde as primeiras propostas de uma classificação dialetal do espanhol da América (HENRÍQUEZ URUEÑA, 1921), o espanhol da Colômbia apresentou pelo menos duas grandes áreas dialetais, uma que coincidia com o espanhol caribenho e outra que compartilha similaridades com o espanhol andino. Essa classificação, *grosso modo*, foi corroborada nas propostas dos dialetólogos colombianos elaboradas de uma perspectiva geopolítica e cultural com base nos dados coletados para o *Atlas Lingüístico y Etnográfico de Colombia (ALEC)*. A primeira classificação é a de Flórez (1961), que propôs as seguintes áreas dialetais: *costeño* (Atlântico e Pacífico), *antioqueño*, *nariñense-caucano*, *tolimense*, *cundiboyacense*, *santandereano* e *llanero*, conforme apresentamos no Mapa 1. De acordo com proposta de Flórez (1961) e a binária do dialetólogo espanhol Fernández Sevilla (1980, *apud* MONTES GIRALDO, 1982), Montes Giraldo (1982) propôs uma classificação do espanhol colombiano em dois macrodialetos: *andino* e *costeño* (ver o Mapa 1); esses, por sua vez, estão divididos

⁶ Os *resguardos* (reservas) *indígenas* são propriedade coletiva das comunidades indígenas a favor das quais se constituem, e conforme os artigos 63 e 329 da Constituição Política de Colômbia têm o caráter de inalienável, imprescritível e não sujeito a qualquer tipo de embargo. Os *resguardos indígenas* são uma instituição legal e sociopolítica de caráter especial, conformada por uma ou mais comunidades indígenas que, com um título de propriedade coletiva que goza das garantias da propriedade privada, possuem um território e se organizam para reger a administração dele e sua vida interna por uma organização autônoma amparada pelo foro indígena e seu sistema normativo próprio (decreto 2164 de 1995). <https://www.mininterior.gov.co/content/resguardo-indigena>

em dois outros dialetos e esses, em subdialetos diversos, como está representado no Quadro 1, elaborado com base em fenômenos fonéticos, morfossintáticos⁷ e lexicais.

Mapa 1: Prováveis áreas dialetais do espanhol colombiano.



Fonte: Flórez, 1961, p.104 e Montes Giraldo, 1982, p.71.

⁷ As amplas limitações dos aspectos gramaticais no questionário do ALEC e a dificuldade de pesquisar tais aspectos em um primeiro atlas geral faz com que sejam muito reduzidos os dados do questionário o que, portanto, não permite a possibilidade de delimitar isoglossas gramaticais (MONTES GIRALDO, 1982, p.42).

Quadro 1 – Classificação dialetal do espanhol da Colômbia

Superdialeto	Dialeto	Subdialeto
Andino/Central	Andino-Oriental	Tolimense-Huilense
		Cundiboyacense
		Santandereano
	Andino-Occidental	Antioqueño
		Nariñense-Caucano
Costeño	Costeño-Pacífico	Septentrional? ⁸
		Meridional?
	Costeño-Caribe	Cartagenero
		Samario
		Guajiro
		Caribe interior

Fonte: Montes Giraldo, 1982, p.42.

É o macrodialeto andino, correspondente à variedade falada no interior do país, conforme apresentado no Mapa 1, que é falado pela maior parte da população colombiana. Divide-se em seis variedades principais: *antioqueño*, situado entre o sul do litoral atlântico e o leste do Pacífico; *nariñense-caucano*, sudoeste da Colômbia até a fronteira com o Equador; *santandereano*, sudeste do litoral Atlântico até a divisa com a Venezuela; *cundiboyacense*, correspondente à *meseta* central andina; *tolimense*, no sudoeste da região *cundiboyacense* e, finalmente, as variedades *llaneras*⁹ nas planícies orientais até a fronteira com o oeste venezuelano.

Segundo Orozco e Díaz Campo (2016), essas últimas constituem as variedades menos estudadas do espanhol colombiano, o que se pode observar no ALEC, que contém pouca informação sobre essa vasta extensão territorial que compreende duas regiões naturais do país (Orinoquia e Amazônia)¹⁰, equivalentes a 657.735 km² (58% do território) e com uma população de aproximadamente 2,7 milhões de falantes (6% da nacional), a maioria multilíngue (RODRÍGUEZ CADENA, 2008).

⁸ Esses signos de interrogação são utilizados pelo autor, que, todavia, não fornece explicação sobre o que significam.

⁹ As variedades *llaneras* foram incluídas na proposta de Flórez (1961) como variedades independentes, mas em Montes Giraldo (1982) não fica muito claro a que macrodialeto pertencem; elas compartilham traços fônicos com o espanhol caribenho, mas também com o espanhol andino. Embora as análises dialetais recentes do léxico com base no ALEC (MORA *et al.*, 2004) pareçam confirmar a hipótese dos superdialeto de Montes Giraldo (1982), incluindo-se, nesse novo estudo, um dialeto mais no *andino*, o dialeto *llanero*, subdividido pela sua vez em *llanero norte* e *llanero sur*, confirmando-o como um dialeto como propôs Flórez (1961).

¹⁰ Pela topografia do território nacional, o país está dividido em cinco regiões naturais: Caribe, Andina, Pacífica, Orinoquia e Amazônia. As maiores regiões são a Orinoquia e Amazônia.

O macrodialeto *costeño*, subdividido em Caribe e Pacífico (MONTES, 1982), apresenta as características fônicas do espanhol das terras baixas latino-americanas e peninsulares (HENRÍQUEZ UREÑA, 1921; ZAMORA; GUITART 1982; LIPSKI, 1994, entre outros). Segundo Montes (1982), o critério mais importante que possibilita distinguir os dois dialetos é o *tuteo* do litoral caribenho frente ao *voseo* do Pacífico (o espanhol caribenho será descrito no item 1.2.1).

Segundo Lipski (1994), a Colômbia é uma das nações latino-americanas mais estudadas em relação à diversidade dialetal de que dispõe; a maior parte das descrições dialetológicas têm por base os materiais coletados pelo ALEC, que de fato consegue uma caracterização bastante detalhada do espanhol de grande parte do território nacional. Esse material foi coletado com uma metodologia própria da geografia dialetal e da dialetologia da metade do século XX, tornando-se um importante instrumento de estudo e análise dos dialetos colombiano e espanhol, em geral. Ainda assim, tem limitações quanto aos aspectos gramaticais das variedades colombianas e à carência de informação linguística para pesquisas mais atualizadas da fala, com foco em questões sociolinguísticas, por exemplo.

As caracterizações dialetais presentes no ALEC, e o próprio ALEC em si, foram constituídos há mais de 50 anos, havendo, portanto, a necessidade de incrementar informações mais atualizadas, enriquecidas com novas caracterizações e sustentadas em trabalhos com *corpora* mais recentes.

A fala mais bem caracterizada é a *bogotana*, não apenas por ser usada na capital, mas também por ser dotada de *corpora* mais recentes se comparado aos demais dialetos. Dentre os textos que compõem o *corpora*, destacamos *El estudio de la norma lingüística culta de Bogotá* (OTÁLORA, 1997), anexado ao *Proyecto de estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica*, e *El español hablado en Bogotá: análisis previo de su estratificación social* (MONTES GIRALDO *et al.*, 1998). A ressalva é a de que os estudos não são nem tão numerosos nem tão variados como seria de esperar.

As caracterizações mais recentes das variedades do espanhol da Colômbia resultam de pesquisas individuais ou de resultados de trabalhos sobre os *corpora* coletados no *Proyecto para el estudio sociolingüístico del español de España y América*, onde há registros de várias cidades de diferentes regiões colombianas.

A seguir, apresentamos os Quadros 2 e 3, que sintetizam as principais características do espanhol colombiano, decorrentes, principalmente, de estudos e pesquisas dialetais recentes, embora seja forçoso reconhecer a necessidade de atualização e de implementação

mais decisiva do material com base em estudos mais representativos do espanhol da atualidade.

Quadro 2 – Características fonético-fonológicas presentes no espanhol colombiano

Traço	Exemplo (s)	Variedade(s)
Aspiração e elisão do /-s/ na posição de coda silábica	[sej̥h. 'pe.soh] [sej̥. 'pe.so]	Regiões costeiras
Aspiração do /-s/ na posição anterior ao núcleo e intervocálica	[se. 'no. ra]>[xe. 'no. ra] [no. 'so.tros]> [no. 'xo.tros]	Todas
Elisão do /-r/ ao final da palavra	[sa. 'lir]>[sa. 'li]	Regiões costeiras
Velarização, glotalização e geminação dos sons líquidos /-r/ e /-l/	<i>Alberto el turco</i> > [ag. 'beg.to. eg. 'tug.ko] [aʔ. 'beʔ.to.eʔ. 'tuʔ.ko] [ab. 'bet.to.et. 'tuk.ko]	Sul da região costeira do Caribe
<i>Yeísmo</i>	cayo e callo = ['ka.jo]	Todas
Elisão do /-d-/ intervocálico	[a.ma. 'ra.ðo]>[a.ma. 'ra.o]	Todas
Elisão do /-d-/ em coda	[pa. 'reð]>[pa. 're]	Todas
Velarização do /w/	[ma.ri. 'ya.na]>[ma.ri. 'ɣya.na]	Todas
Ditongação dos hiatos	[pe. 'tro.le.o] >[pe. 'tro.ljo]	Todas

Fonte: adaptado de Orozco; Díaz Campo, 2016, p.344.

Quadro 3 – Principais características morfossintáticas do espanhol colombiano

Traço	Exemplo (s)	Variedade(s)
Predomínio do futuro perifrástico	<i>Voy a cantar esta semana</i>	Todas
Alternância de adjetivos possessivos com perífrases possessivas	<i>Mis amigos~los amigos míos</i>	Todas
<i>Ustedeo</i>	<i>(Usted) está contento</i>	Andina
<i>Tuteo</i>	<i>(Tú) estás contento</i>	Predomina na variedade <i>costeña</i> e é usado com menor frequência na variedade andina.
<i>Voseo</i>	<i>(Vos) estás contento</i>	<i>Costeña</i> pacífica, Antioquia
<i>Su merced</i> como pronome de segunda pessoa do singular	<i>Su merced está contento</i>	Cundiboyacense
<i>Ser</i> enfático ou focalizador	<i>(Nosotros) estamos es sorprendidos</i>	Todas
Reduplicação redundante dos pro- nomes átonos de objeto indireto	<i>Juan dijo que él me lo iba a traérmelo</i>	Fala vernácula <i>costeña</i>
Terminações em <i>-nos</i> para a primeira pessoa do plural	<i>(Nosotros) íbanos pa la casa</i>	Fala vernácula <i>costeña</i>
Pluralização do verbo <i>haber</i>	<i>Hubieron muchos favorecidos</i>	Todas

Fonte: adaptado de Orozco; Díaz Campo, 2016, p. 345.

Os aspectos fônicos e morfossintáticos, apresentados nos Quadros 2 e 3, bem como a variedade lexical influenciada por diversas línguas (indígenas, africanas e estrangeiras) definem os dialetos colombianos, que têm o espanhol culto *bogotano* a variedade de prestígio central, embora cada dialeto tenha sua própria variedade de prestígio. Segundo Moreno Fernández (2014), o espanhol colombiano conta com um elevado prestígio, em razão de seu caráter conservador em termos de maior grau de proximidade com a língua escrita, assumindo-se, portanto, como “espanhol colombiano” a variedade culta da capital.

Por outro lado, destacamos que o espanhol empregado por falantes bilíngues, em diferentes registros de fala, é um dos fatores que contribuem para o fato de serem as variedades do espanhol colombiano as menos exploradas no país. Em estudo sobre as características do espanhol falado por indígenas colombianos e, particularmente, por falantes das línguas nativas do Caribe, Rodríguez Cadena (2008a) destaca que a interferência das línguas ameríndias nas variedades do espanhol depende do grau de bilinguismo do falante e do grau de escolaridade. Assim, quanto maiores os graus de escolaridade e de bilinguismo, menor o número de traços de interferência de que dispõem os falantes; em contrapartida, quanto menores os graus de escolaridade e de bilinguismo, maior o grau de interferência.

Dentre os principais traços detectados do contato de línguas indígenas com o espanhol, pode-se destacar os fônicos, morfossintáticos (em sua maioria) e os semânticos (alguns verbos). Segundo Trillos Amaya (2018), esses aspectos interferem na comunicação que devem estabelecer os indígenas fora do *resguardo*.

1.2 As línguas do Caribe colombiano

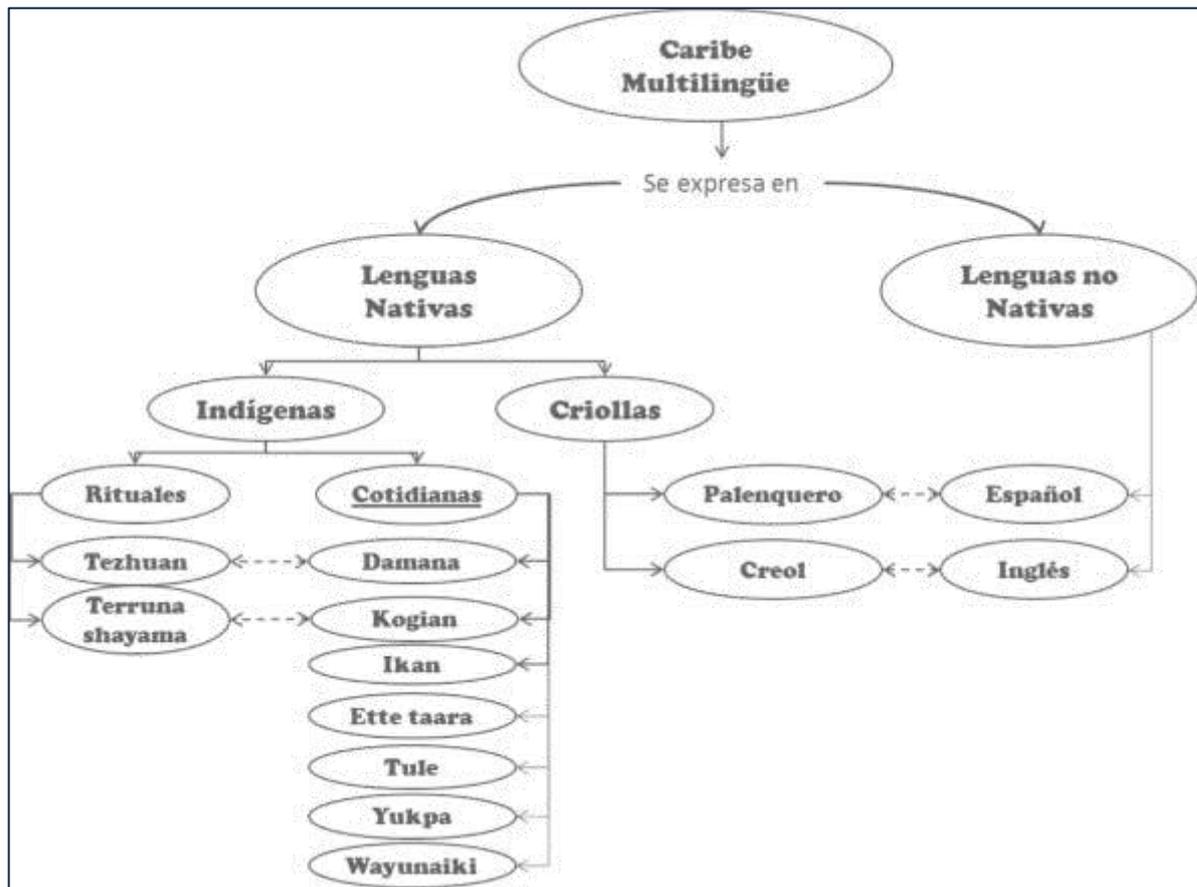
A região caribenha colombiana, uma das cinco que constituem o território nacional, está localizada no norte do país, em frente ao mar Caribe, identificada como a mais setentrional de todas, além de abranger áreas continentais e marítimas. Contudo, a região não se destaca apenas pelos aspectos naturais e geográficos, mas também por sua história, pelos aspectos étnicos, culturais e, sobretudo, linguísticos, que permitem identificar os habitantes regionais (*costeños*) e distingui-los dos demais. Essa identidade específica parece ser uma das mais bem definidas no país e, como tal, reconhecida pelo restante da população colombiana. Dados do DANE (2007) revelam que a população caribenha é 9.506.220 habitantes, equivalente a 22,6% da população nacional.

Trillos Amaya (2001), partindo de fontes documentais, caracteriza o Caribe colombiano como um grupo social multilíngue e pluricultural. Para a autora, a região é um

complexo cultural, geográfico e linguístico, que requer estudos sociolinguísticos aprofundados para determinar os tipos de bilinguismo que apresentam os falantes nas suas interações. Pode-se considerar o multilinguismo no Caribe colombiano como uma herança cultural legada pelos primeiros habitantes da região e por aqueles que, por diferentes motivos, foram chegando a ela.

Nessa região, 14 línguas¹¹ são faladas por um número variado de usuários. Os idiomas caribenhos estão ligados a povos ameríndios, africanos e mestiços. Em termos quantitativos, esse multilinguismo pode ser representado por 10 línguas indígenas, 2 crioulas, uma variedade de inglês caribenho, além do espanhol caribenho (TRILLOS AMAYA, 2009, 2018). No Gráfico 1 podemos observar o multilinguismo da região e no Mapa 2 visualizamos geograficamente a preponderância de uso das línguas.

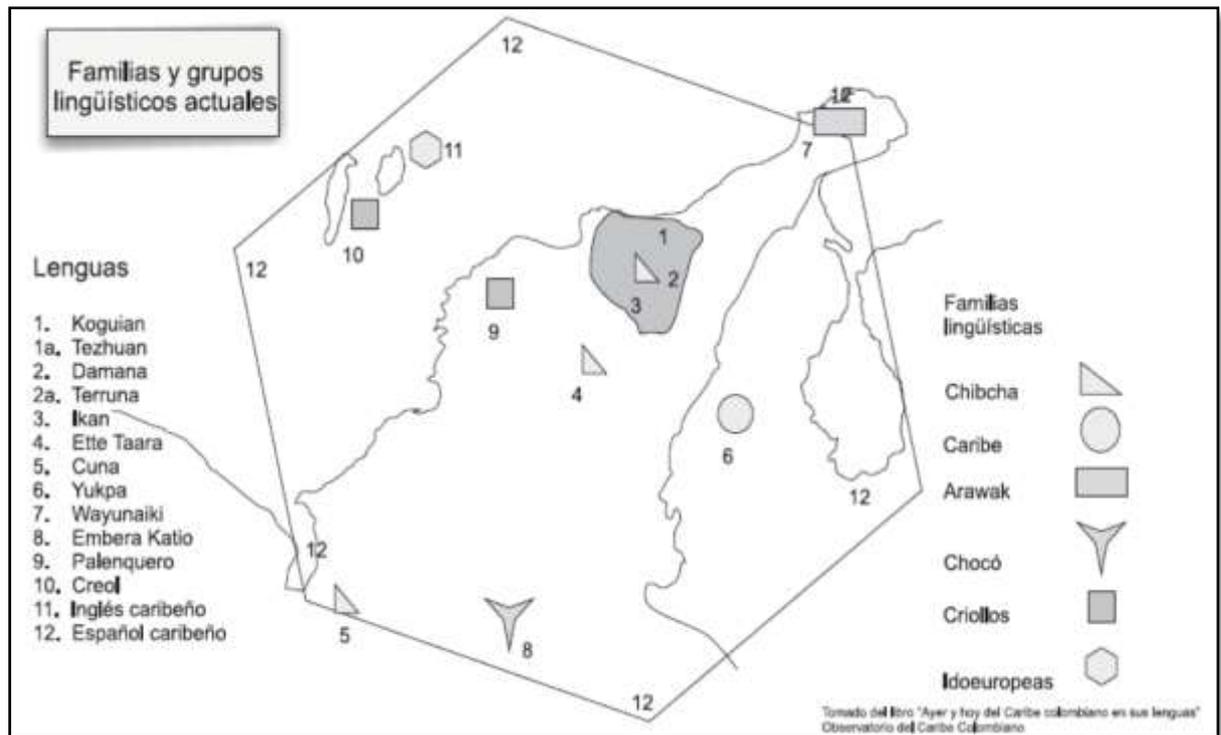
Gráfico 1 – Línguas do Caribe colombiano multilíngue.



Fonte: Trillos Amaya, 2018, p.74.

¹¹ Trillos Amaya (2009, 2018) sempre menciona 14 línguas, mas sempre inclui em suas descrições a língua *romani*, falada pelo grupo étnico cigano, com a qual haveria um total de 15 línguas.

Mapa 2 – Panorama multilíngue da região do Caribe.



Fonte: Trillos Amaya, 2009, p.191.

Segundo Trillos Amaya (2001, 2018), o multilinguismo apresentado no Gráfico 1 e no Mapa 2 tem, como característica, a presença de línguas nativas e não nativas. Conforme mostra o Gráfico 1, o primeiro grupo é constituído por línguas indígenas e crioulas. As línguas indígenas, por sua vez, estão divididas em duas outras destinadas aos usos rituais (*chibchas*) e sete línguas ao uso cotidiano, cinco das quais pertencem à família *chibcha*, uma caribe e uma *arawak*. As duas línguas crioulas são o *palenquero*, de base espanhola e o *creol*, de base inglesa – das Ilhas de San Andrés e Providencia; ambas dispõem de traços gramaticais de línguas africanas, embora provindos de famílias diferentes, como o bantu, a primeira e *kwa*, a segunda. Além disso, em Sabanalarga Atlântico, encontra-se uma comunidade linguística representante do povo rom, falantes de romani, língua de origem indo-europeia como o inglês falado no arquipélago de San Andrés, Providencia e Santa Catalina e o espanhol falado em toda a região.

Cumprir destacar, contudo, que as condições pelas quais ocorrem contatos interlinguísticos revelam que as línguas nativas, o romani e o inglês são línguas

*minoritárias*¹², com o uso do espanhol avançando para espaços tradicionalmente ocupados por elas, originando diversos fenômenos relacionados aos falantes, aos sistemas linguísticos e à sociedade (TRILLOS AMAYA, 2018). Em relação aos falantes, é possível identificar traços de diglossia como manifestações de mudança, alternância e interferência nos códigos em ambas as línguas; em relação aos sistemas linguísticos, dispõe-se de empréstimos e transferência de palavras de uma língua a outra; em relação à sociedade, bilinguismo, plurilinguismo e fenômenos de empréstimos, transferência, mudança linguística e, portanto, de variação linguística. Por outro lado, apresenta-se a imposição linguística, o deslocamento linguístico e a extinção de línguas. Essas situações necessitam de urgente elaboração de políticas e processos de planejamento linguístico, para tornar efetiva a Lei de Línguas 1.381 de 2010.

1.2.1 O espanhol do Caribe colombiano

Como é possível inferir das seções anteriores, o espanhol colombiano está em uma clara situação de vantagem em relação às demais línguas faladas no país. Essa situação se reflete no Caribe colombiano, uma região onde se falam quinze línguas, sendo o espanhol caribenho a variedade mais frequente. Essa variedade apresenta traços fonológicos, lexicais e gramaticais específicos que permitem distingui-la da variedade andina, no léxico, na fonética e na gramática (MONTES GIRALDO, 1982; TRILLOS AMAYA, 2001). O Quadro 4, apresentado a seguir, dispõe, em síntese, das principais características do espanhol caribenho colombiano.

¹² Utilizamos a palavra *minoritária* (*minorizada* em espanhol) para não esvaziar o sentido a ela atribuído, o de que essas línguas cumprem cada vez mais um papel de minoritárias, por pressão da língua dominante falada pela maioria da população e por falta de melhores políticas e planejamentos linguísticos que defendam e promovam o uso das línguas nativas por seus falantes como estabelece a Constituição Política (1991).

Quadro 4 – Características fônicas e morfossintáticas do espanhol caribenho colombiano.

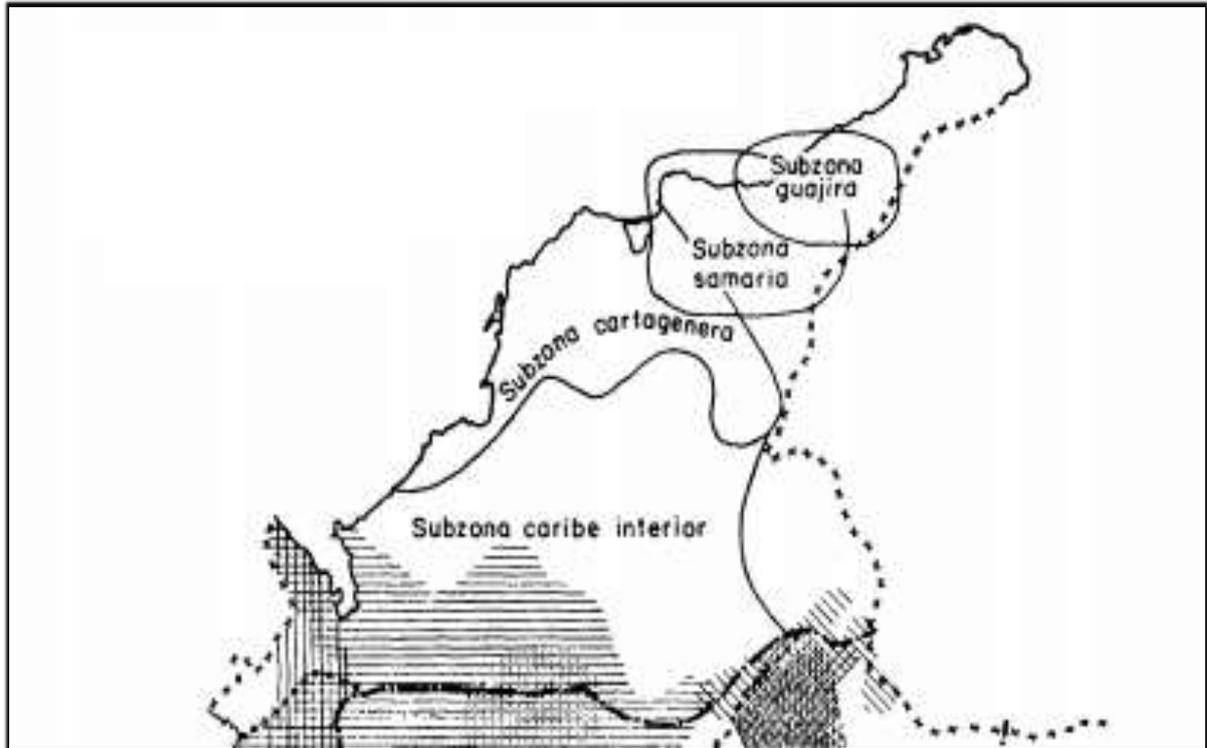
Traços	Espanhol caribenho colombiano
/-s/ implosivo	[-s], [-h] ou [-Ø]
/-r/ e /-l/ implosivo	Neutralização do /-L/ e perda com e sem geminação da consoante seguinte
/-n/ implosivo	Tendência à velarização
/tʃ/	Tendência à realização [tʰ]
/-d-/ intervocálico	Tendência à elisão
/y/ e /ɲ/	<i>Yeísmo</i>
<i>Voseo/ tuteo</i>	<i>Tuteo</i>
<i>Hacer</i> e <i>haber</i> impessoais	Tendência à pessoalização

Fonte: Adaptado de Montes Giraldo, 1996.

Rodríguez Cadena (2004) agrupa as características do dialeto espanhol caribenho colombiano com base num conjunto linguisticamente diversificado de traços. Inicialmente, de acordo com as tendências fonético-fonológicas, aponta os seguintes traços: a) aspiração e ensurdecimento de fonemas; b) duplicação de fonemas; c) velarização; d) elisão de fonemas em posição implosiva e final; e) entonação relacionada à duração das vogais; f) articulação aderente da africada palatal surda. Já de acordo com as tendências morfossintáticas, aponta os seguintes traços: a) o *voseo* verbal; b) alteração dos esquemas do plural; c) abreviação morfossintática. De acordo com as tendências semântico-lexicais, aponta os seguintes traços: a) a afixação hiperbólica; b) uso de marcadores discursivos, formas vocativas e fáticas peculiares.

Embora o Caribe seja considerado um dialeto, também é possível identificar a presença de subdialeto no seu interior, conforme mostra o Quadro 1, já apresentado, e visualizados graficamente no Mapa 3 abaixo, onde se podem observar as possíveis fronteiras desses subdialeto caribenhos.

Mapa 3 – Zona e possíveis subzonas dialetais do espanhol caribenho colombiano.



Fontes: Adaptado de Montes Giraldo, 1982, pág. 92.

No Mapa 3, Montes Giraldo (1982) subdivide o dialeto caribenho em: *cartagenero*, *samario*, *guajiro* e *caribe interior*; sendo o primeiro o que mais concentra os traços caribenhos, enquanto o *samario* dispõe da conservação do /-s/ em posição de coda, traço que o distingue do *guajiro*.

Estudos mais recentes permitem sintetizar a diáspora das falas espanholas caribenhas colombianas. Reconhecendo a necessidade de estudos mais sistemáticos e aprofundados prévios para estabelecer uma divisão dialetal, Rodríguez Cadena (2004) propôs a seguinte classificação diatópica intradialetal:

- a) Córdoba, Sucre e Bolívar;
- b) La Guajira e Cesar;
- c) Magdalena e Atlântico¹³.

Para essa proposta de três sub-regiões dialetais, a linguista leva em consideração os fenômenos presentes com maior intensidade em todas essas variedades. Trillos Amaya (2001,

¹³ Ela não inclui o estado de San Andrés y Providencia que historicamente é considerado parte do Caribe colombiano, talvez pelas características próprias do espanhol das ilhas, que o difere do espanhol da região continental, além do fato de ser falado como primeira língua o crioulo e como segunda(s) língua(s), inglês e/ou espanhol.

2009), por sua vez, propôs uma subdivisão em seis subdialetos e suas áreas de influência com base em critérios fônicos e gramaticais (especificamente o *voseo*), como se observa no Quadro 5.

Quadro 5 – Subdialetos do espanhol caribenho colombiano e áreas de dispersão.

Fala	Dispersão	Processo linguístico
<i>Vallenata</i>	Para o sul de La Guajira	Presença de <i>voseo</i>
<i>Samaria</i>	Para o Cesar e o Atlântico	Escasso enfraquecimento do /-s/
<i>Cartagenera</i>	Para as savanas de Sucre e Córdoba	Perda de /s/ e geminação de /p,t,k/
<i>Mompoxina</i>	Zonas específicas de La Mojana	<i>Voseo</i> , perda de /s/ e geminação de /p,t,k/
<i>Barranquillera</i>	O município todo	Confluência de todos os subdialetos
<i>Sanandresana</i>	O arquipélago	Confluência de todos os subdialetos

Fonte: Adaptado de Trillos Amaya, 2009, p.192.

Segundo Trillos Amaya (2001), Barranquilla e San Andrés concentram todos os subdialetos caribenhos, além de outros do país, certamente em virtude do caráter cosmopolita dessas cidades, muito embora seja necessário empreender uma investigação acurada da variedade do espanhol das ilhas.

1.2.2 Estudos sociolinguísticos sobre o espanhol caribenho colombiano

Vem-se ressaltando, desde o final do séc. XX e início do XXI, a necessidade de realizar estudos linguísticos no Caribe colombiano, que superem a etapa dialetal do ALEC, segundo considerações de Rodríguez Cadena (2004) e Trillos Amaya (2001). Em 2004, Rodríguez Cadena assinalava a necessidade inadiável de novos estudos e a importância de se enfocar o campo dialetal, mas também o sociolinguístico, conforme suas próprias palavras:

A investigação sobre a fala da Região Caribe na Colômbia é uma necessidade inadiável, especialmente das perspectivas sociolinguística e discursiva, antes enunciadas, que produzam resultados confiáveis sobre as práticas comunicativas e suas complexas implicações socioculturais e ideológicas. Essa tarefa só será possível depois de definidos os subprojetos específicos com delimitações regidas por variáveis dependentes e independentes, como por exemplo: sub-regiões, comunidades de fala, traços socioculturais (educacional, econômico, sexo, idade,

profissão, práticas culturais) e fenômenos linguístico-discursivos, como os expostos nesse projeto (RODRÍGUEZ CADENA, 2004, tradução nossa).¹⁴

Sobre o Caribe colombiano, já se realizaram algumas pesquisas sociolinguísticas, como as sugeridas por Rodríguez Cadena (2004). No nível fonético-fonológico, destacam-se os estudos de Becerra (1985), File-Muriel (2012), Lafford (1980), Olmos, Gómez e Pérez (2010), Rodríguez Cadena (1997, 2005, 2008b, 2009a, 2009b, 2010), Rueda e Navas (2011). No nível morfossintático, Martínez (2007), Orozco (2005, 2007b, 2009b, 2012, 2015) e Orozco e Guy (2008). No nível discursivo, por sua vez, é possível destacar Cuartas (2011), Ortega, Primo e Acosta (2010), Rodríguez Cadena (1999), Sarabia, Bueno e Mejía (2010) e Vásquez Cantillo (2006)¹⁵. A importância desses estudos para este trabalho é que vários deles demonstram, por um lado, tendências similares entre os fenômenos estudados neles e os do Caribe antilhano, reforçando os argumentos dialetais de que a variedade local faz parte desse conjunto maior conhecido como espanhol caribenho. Por outro lado, a maioria desses estudos realizados com base nos *corpora* aqui utilizados, e que serão mais bem caracterizados no capítulo 3, evidencia que os fatores sociais contribuem para a variação das variedades faladas nessa região.

Além desses trabalhos, Rodríguez Cadena (2008a) apresenta resultados preliminares sobre algumas variedades do espanhol falado por indígenas com algum nível de bilinguismo nas línguas nativas faladas na região. Contudo, esse conjunto de trabalhos mostra que são muito reduzidos os estudos sociolinguísticos sobre o espanhol do Caribe colombiano, conforme destacam Orozco (2009a) e Rodríguez Cadena (2004; 2009a). Além disso, a maioria das pesquisas focaliza os níveis fonético-fonológico e discursivo, deixando uma grande lacuna em relação ao morfossintático. Essa situação se aplica também aos trabalhos dialetológicos e lexicográficos que têm sido desenvolvidos sobre essa variedade colombiana¹⁶.

¹⁴ La investigación sobre el habla de la Región Caribe en Colombia es una necesidad inaplazable, especialmente desde las perspectivas sociolingüísticas y discursivas arriba enunciadas, que arrojen resultados confiables sobre las prácticas comunicativas y sus complejas implicaciones socioculturales e ideológicas. Esta tarea sólo será posible una vez que hayamos definido subproyectos específicos con delimitaciones según variables dependientes e independientes, entre otras: subregiones, comunidades de habla, rasgos socioculturales (nivel educativo, económico; sexo, edad, ocupación, prácticas subculturales) y fenómenos lingüístico - discursivos, como los aquí planteados. (RODRÍGUEZ CADENA, 2004, [online](#)).

¹⁵ Vários das pesquisas realizadas com o *corpus* de Barranquilla aparecem resumidos nos capítulos iniciais dos volumes publicados do *corpus* (RODRÍGUEZ CADENA, 2008b, 2009b, 2010), por isso, nem todos esses nomes integram as referências bibliográficas deste trabalho.

¹⁶ A maior parte dos trabalhos dialetológicos foi o resultado do material coletado para o ALEC, que se centrou na coleta das variantes lexicais e fonéticas do país; é por isso que são limitadas as conclusões morfossintáticas extraídas dos materiais (MONTES GIRALDO, 1982). A tese de doutorado de Rodríguez Cadena (2009a) apresenta até esse ano uma boa síntese dos estudos linguísticos sobre o Caribe colombiano: 1) das línguas

Até aqui, foram apresentados vários pressupostos relevantes sobre a organização dialetal do caribe colombiano: 1) o espanhol da Colômbia apresenta diferenças notáveis, o que permite uma classificação em dois superdialetos, o *Costeño* e o *Andino*, que compreendem subdivisões internas; 2) o *Costeño* do Caribe (atlântico) colombiano apresenta também variação interna, podendo se dividir em três ou seis subdialetos, segundo diferentes propostas, e apresentando traços que o aproximam mais do espanhol caribenho: litoral oriental do México, Antilhas, litoral do Panamá e da Venezuela; 3) a maior parte dos trabalhos linguísticos sobre o espanhol do Caribe colombiano é de cunho dialetológico e lexicográfico, e alguns trabalhos sociolinguísticos estão centrados nos níveis fônico e discursivo; 4) os trabalhos no nível morfosintático são poucos, de qualquer perspectiva de investigação; focalizam-se no espanhol falado em Barranquilla e possíveis temas de pesquisa nesse nível são apenas hipóteses que precisam ser estudadas em profundidade.

Todos esses pressupostos compõem um conjunto considerável de razões que nos levam a empreender estudos na Colômbia, um país com uma grande diversidade dialetal, que compartilha traços com outras variedades do espanhol americano e que necessita de estudos sociolinguísticos recentes, especialmente aplicados às variedades do litoral a fim de possibilitar um melhor conhecimento da realidade linguística local e nacional.

2 PRESSUPOSTOS TEÓRICOS

2.0 Introdução

Esta pesquisa pretende analisar a expressão do pronome sujeito no espanhol falado no Caribe colombiano para determinar que variáveis linguísticas e extralinguísticas (sociais) motivam sua variação nesta comunidade. Antes de iniciar a análise propriamente dita da expressão do pronome sujeito, faz-se necessário esclarecer o conceito de variação linguística aqui adotado, que compreende, necessariamente, uma identificação precisa das noções dele derivadas de variável e variante, conceitos-chave na perspectiva variacionista laboviana. Em seguida, discute-se especificamente o modo como se problematizou o estudo da variação sintática, que acabou seguindo seu próprio caminho no quadro teórico variacionista em língua espanhola, com o reajuste também dos conceitos de variável e variante. Introduce-se, por fim, o próprio fenômeno em análise, a variável expressão pronominal do sujeito, e se apresentam as variáveis independentes de natureza linguística e extralinguística que mais contribuem para a expressão do sujeito, seja em posição pré-verbal seja em posição pós-verbal, sem deixar de fornecer, antes, uma síntese muito geral do que têm a dizer sobre a expressão do sujeito pronominal no espanhol a gramática tradicional, a sintaxe gerativa e os quadros funcionalistas de base discursivo-pragmática.

2.1 A sociolinguística variacionista e a variação sintática

Define-se a sociolinguística como a disciplina que se ocupa do estudo da covariação sistemática entre a língua e a sociedade (LABOV, [1972], 2008). Nesse quadro teórico, define-se a variável linguística, entendida como uma classe de variantes, “como o uso alternativo de formas diferentes de dizer a mesma coisa, que se pode encontrar praticamente em todos os níveis da língua, desde o mais concreto (fonético-fonológico) até o mais amplo (discurso, por exemplo), passando pelo gramatical e pelo lexical” (MORENO FERNÁNDEZ, 1998, p. 19, tradução nossa).¹⁷ Assim, o ponto de partida da análise variacionista é o conceito de variável linguística como unidade de estudo, com suas respectivas variantes. “Variantes linguísticas são, portanto, diversas maneiras de se dizer a mesma coisa em um mesmo contexto, e com o mesmo valor de verdade. A um conjunto de variantes dá-se o nome de

¹⁷ [...] definida como el uso alterno de formas diferentes de decir lo mismo, se puede encontrar prácticamente en todos los niveles de la lengua, desde el más concreto (fonético-fonológico) al más amplio (discurso, por ejemplo), pasando por la gramática y el léxico.

‘variável linguística’” (TARALLO, 1985). Então, nessa concepção clássica do paradigma laboviano, duas formas são consideradas alternantes ou variantes se não veiculam diferenças semânticas (LABOV, [1972], 2008), entendidas como diferenças referenciais que caracterizam o mesmo valor de verdade.

Embora essa definição de variável e variante tenha tido um sucesso notável nos estudos variacionistas centrados no nível fônico¹⁸ e ainda o continua tendo, não se pode dizer que teve o mesmo resultado quando os estudos começaram a analisar fenômenos que vão além desse nível, particularmente o sintático. O conhecido debate de que diferentes linguistas participaram (LAVANDERA, 1978; GARCÍA, 1985; LABOV, 1978, ROMAINÉ, 1979, entre outros)¹⁹ questionava se seria possível encontrar, no nível sintático, ‘variantes’ de uma variável que não envolvessem diferenças de significado, o que impediria a aplicação da teoria para além do nível fônico.

No centro desse debate, surgiu a proposta de Lavandera, que consiste em relativizar a exigência de que as variantes sejam iguais semanticamente, critério que seria substituído pelo de comparabilidade ou equivalência funcional (LAVANDERA, 1978) ou pragmática (SILVA-CORVALÁN, 2001; SERRANO, 2007, 2009). Este conceito de variável linguística no sentido mais amplo, que se desenvolverá posteriormente, recebe, hoje, amplo reconhecimento dos estudiosos da variação sintática no contexto hispânico e em outros lugares do mundo. Mas pode-se resumir-lo, com base em Torres-Cacoullós (2011, p. 151), que o conceito de variável linguística – duas ou mais expressões com a mesma função, neste caso, gramatical ou discursiva – “requer a definição do envelope de variação, ou contexto variável, que é o ambiente mais amplo no qual os falantes têm uma escolha entre diferentes formas”. As tensões sobre esse conceito se mantiveram ao longo do percurso do desenvolvimento da disciplina, como se resume na sequência.

Segundo Serrano (2009), os estudos sociolinguísticos da variação sintática vão passar por três momentos e, ao longo desse percurso, encontra-se um conjunto significativo de trabalhos variacionistas sobre sintaxe, centrados especialmente na língua espanhola; tanto nesses trabalhos realizados no âmbito hispânico, quanto nos estudos sintáticos em geral de

¹⁸ O estudo da variação começou pela fonologia, início justificável pela falta de necessidade de demonstrar que os segmentos fonológicos carecem de conteúdo semântico.

¹⁹ Para uma revisão desta discussão, ver Bentivoglio (1987b, 2001) e Serrano (2007, 2009), no espanhol, e Camacho (2003, 2013), no português. Vários dos autores de outras línguas que participam nas primeiras discussões são referenciados nos trabalhos de Bentivoglio e Serrano, pelo que não aparecem referenciados na bibliografia desta tese.

base variacionista, há algumas dificuldades de natureza teórica e metodológica que precisam ser resolvidas para dar continuidade às investigações.

Na primeira fase, encontram-se os estudos sociolinguísticos pioneiros da variação em sintaxe sobre o francês do Canadá de Sankoff (1972), que, pela primeira vez tenta demonstrar que a variação acontece também em outros níveis da língua, além do fonológico. Além desse trabalho estão também os do francês canadense de Sankoff e Tribault (1977), Laberge (1977) e Laberge e Sankoff (1979) e do inglês de Labov e Weiner (1977). Nesta série de estudos, a noção de variável fonológica se estende ao nível sintático, no qual se mantém inalterado o requerimento absoluto de equivalência semântica para as variantes de uma variável, o que resultou pouco revelador, e mesmo inadequado, para vários pesquisadores, a começar por Lavandera (SERRANO, 2009). No plano sintático, a única forma de obter essa restrição seria circunscrever a noção de significado ao conteúdo referencial, como está disposto na proposta de Labov (1978) e que seria efetivamente posto em prática por essa série de trabalhos sociolinguísticos pioneiros.

Para Bentivoglio (1987b, 2001), o auge dos estudos variacionistas da sintaxe, na década de 1970-80, nas diferentes línguas (inglês, francês canadense, português do Brasil e espanhol), vai-se reduzir drasticamente na década seguinte, como consequência dos fortes questionamentos de Lavandera (1978) a respeito da extensão *ipsis litteris* ao nível sintático dos métodos variacionistas aplicados ao nível fonológico. Outras pesquisadoras, como Romaine (1979) e García (1985), unem-se às críticas de Lavandera sobre a aplicação dos estudos variacionistas ao nível sintático, desenvolvidos segundo a abordagem clássica (*apud* BENTIVOGLIO, 1987b, 2001).

Os primeiros estudos revelaram que a variação sintática não é análoga à fonológica, pelas seguintes razões que Silva-Corvalán resume: 1) a variação sintática tem frequência menor que a fonológica; 2) a escassa frequência com a que se pode contar com um contexto de ocorrência e a dificuldade para a obtenção de exemplos de uso de uma ou outra variante tornam a variável sintática mais difícil de medir ou quantificar; 3) na variação sintática, os contextos de ocorrência são mais difíceis de identificar ou definir; 4) as possíveis diferenças de significado entre variantes são um problema na variação sintática; 5) a variação sintática normalmente não está estratificada estilística e socialmente; ela vem determinada por fatores meramente linguísticos (SILVA-CORVALÁN, 1988).

Por isso, para a autora, “as dificuldades implícitas no tratamento dos elementos portadores de significado têm constituído um desafio e têm relacionado mais estreitamente a sociolinguística com a semântica e pragmática do discurso” (SILVA-CORVALÁN, 1991,

p.122, tradução nossa).²⁰ Assim, os obstáculos e as duras críticas a esse tipo de estudo, em lugar de frear as pesquisas variacionistas sobre sintaxe, têm servido para reforçá-las (BENTIVOGLIO²¹, 2001).

Serrano (2009) resume assim essa fase:

as investigações estão centradas na natureza da variação sintática como uma simples extensão do plano fonético-fonológico, considerando temas centrais da discussão o aspecto da sinonímia das variantes, o da quantificação da variação e a sua natureza e – em conformidade com o anterior – a possibilidade ou não de que esta fosse socialmente significativa (SERRANO, 2009, p.158, tradução nossa).²²

Na segunda fase dos estudos variacionistas, que se dá de 1990 até hoje, observa-se a implantação relevante da pesquisa variacionista na sintaxe da língua espanhola, que, paralelamente a outros estudos realizados na América, contribui para desenvolver ainda mais decisivamente a metodologia da variação sintática e sua aplicação nessa língua. Com exceção de alguns trabalhos de variação sintática na década anterior (BENTIVOGLIO, 1987b; GARCÍA, 1986; MORALES, 1986; SILVA-CORVALÁN, 1982), registra-se depois de 1990 o grande volume de pesquisas sobre a variação sintática no espanhol. Nesse período, consolida-se a metodologia do estudo da variação sintática em uma perspectiva discursiva e pragmática (SERRANO, 2009).

A partir de 1995, aproximadamente, e depois de haver estudado já numerosas variáveis sintáticas, vai-se tornando habitual aplicar os mecanismos e ferramentas metodológicas da análise do discurso e da pragmática consideradas como consequência natural e óbvia da análise no plano morfossintático (SERRANO, 2009, p.160). Nos finais dos anos 1990, Serrano propõe relativizar a metodologia quantitativa, uma vez que se incorporam valores discursivos, pragmáticos e interacionais não somente mais difíceis de quantificar, mas também menos efetivos e operacionais, quando se trata de estudar a variação nesse nível de análise. Propõe também a seguinte definição de variável sintática: “esquemas de alternância contextual comparáveis entre si, que constituem funções que os elementos gramaticais

²⁰ [...] las dificultades implícitas en el tratamiento de los elementos portadores de significado han constituido un reto y han relacionado más estrechamente la sociolingüística con la semántica y pragmática del discurso.

²¹ Em um artigo de 1987a, Bentivoglio, observando a realidade dos estudos variacionistas da sintaxe no âmbito hispânico e aparentemente em diferentes línguas, nessa década, assinalava a estagnação dos estudos dessa abordagem nesse nível linguístico, porém na sua conferência de 2001 parece que a realidade é bem diferente no contexto hispânico, como também vai assinalar em seguida Serrano (2009).

²² [...] las investigaciones se centran sobre la naturaleza de la variación sintáctica como una simple extensión del plano fonético-fonológico, considerando temas centrales de discusión el aspecto de la sinonimia de las variantes, el de la cuantificación de la variación y su naturaleza y —a tenor de lo anterior— la posibilidad o no de que esta fuera socialmente significativa.

adquirem, e que se explicam com referência a uma forma” (1999b, p. 21, tradução nossa), acrescentando ainda que, na variação sintática, o importante não é verificar se as variantes são ou não sinônimas, iguais ou equivalentes (SERRANO, 2009).

Por sua vez, Silva-Corvalán (2001) toma o discurso em seu contexto sociolinguístico amplo como base da análise para propor, como uma resposta conciliadora aos problemas surgidos nos estudos variacionistas da sintaxe, a existência de um significado invariável, baseado na referência do estado de coisas envolvido, e a de outro significado, esse variável e relacionado a questões discursivas e pragmáticas.

Segundo Serrano (2009), nos finais dos 1990 e durante os anos 2000 se consolida o interesse por incorporar os fatores discursivos e pragmáticos como a melhor opção para estudar a sintaxe na perspectiva sociolinguística, que se percebia já a partir dos anos 1995 e 1996.

Na terceira fase, já na atualidade, considera-se a variação sintática metodologicamente dependente do fator comunicativo, pela inter-relação e inseparabilidade entre sintaxe, discurso e pragmática. Nesta fase, os estudos variacionistas no nível sintático seriam identificados com os postulados de Dubois e Sankoff (2003), segundo os quais 1) as variáveis discursivas envolvem um número finito de variantes independentes e autônomas que não formam parte de um mesmo conjunto de equivalências, mas que estão relacionadas entre si; 2) não é possível estabelecer um contraste entre presença/ausência de um mesmo fenômeno do mesmo modo como se faz nos estudos variacionistas em fonologia, porque as substituições entre as distintas manifestações discursivas participam de vários níveis ao mesmo tempo (pragmático, interacional, etc.); 3) o propósito da análise variacionista é iluminar o potencial significativo do discurso, isto é, os diferentes níveis do significado que intervêm na criação desse discurso (DUBOIS; SANKOFF, 2003, p.283, *apud* SERRANO, 2009).

O significado discursivo não é diferente de outros significados nem se trata de acepções, variantes de um significado maior e invariante; trata-se justamente do contrário: se for possível chegar a uma definição geral do significado de uma expressão ou de uma palavra é porque, precisamente, foram selecionados todos os traços possíveis para chegar a essa definição e, na seleção desses traços, o falante em si e suas características sociais têm um papel fundamental para a configuração desse significado.

O uso sistemático e regular de determinadas construções no discurso falado com um determinado valor acaba por incorporá-lo ao significado geral dessas construções. Além disso, argumenta-se favoravelmente à ideia da inexistência de um significado formal único e invariante, afastado da funcionalidade do uso no discurso falado e acredita-se ser mais

aceitável uma concepção do significado situado no contexto de uso, que transcende o puramente linguístico, e que está formado por vários elementos que procedem da intenção sistemática e regular fornecida em cada situação comunicativa. Assim, o significado na variação sintática deixa para trás a limitada e rançosa definição de variantes sintáticas como *duas formas de dizer o mesmo* (SERRANO, 2009).

A autora conclui que já não é possível denominar por variação sintática tudo o que até agora tem sido estudado e ao que se segue estudando atualmente. A variação na sintaxe é raramente sintática em sentido exclusivo, já que está vinculada obrigatoriamente a aspectos semânticos, discursivos, pragmáticos e comunicativos. É por isso que considera necessária uma reclassificação da variação sintática da seguinte forma: a) aquela que poderia ser, em raras ocasiões, uma variável exclusivamente sintática (variação do adjetivo demonstrativo, subida de clíticos, etc.); b) variação sintático-discursiva (expressão da forma pronominal de sujeito, alternância infinitivo/subjuntivo em orações completivas, etc.); c) variação sintático-discursivo-pragmática (alternância de formas verbais em orações condicionais) e d) variação discursivo-pragmática (marcadores discursivos, etc.). Sobre cada um deles, pode incidir de forma mais ou menos saliente o aspecto sociocomunicativo que deve ser analisado de acordo com os tipos de variação (SERRANO, 2009).

Como é possível observar nesta síntese teórica, a variação sintática é um campo de estudos hoje consolidado no âmbito hispânico, que recorre aos estudos do discurso e à pragmática para explicar melhor a variação na sintaxe. Não deve surpreender que se recorra a (ou “se apoie em”) explicações dessas disciplinas para descrever certos fenômenos gramaticais variáveis na língua, assim como sua implicação social. A metodologia também se consolidou com o passar dos anos. Assim, para Serrano (2009), não é rigoroso, nem científico, nem verdadeiro que a variação sintática seja um tema cuja problemática perdura na atualidade, já que isso parece constituir um sério despropósito que consiste em ignorar seu considerável e vertiginoso desenvolvimento. No entanto, é obvio que os estudos variacionistas da sintaxe no espanhol continuam a desenvolver seus constructos teórico-metodológicos. É com base nesse cenário que este trabalho procura explicar a variação do pronome pessoal sujeito nesse quadro teórico, particularmente nessa última fase de desenvolvimento, considerando que, no espanhol caribenho colombiano, também são diferentes fatores linguísticos e extralinguísticos que permitem dar uma explicação mais adequada de como varia a expressão pronominal. Além disso, o quadro teórico-metodológico permite fazer comparações mais confiáveis com outros estudos do mesmo fenômeno em diferentes variedades do espanhol.

2.2 A variável *expressão do pronome pessoal sujeito*

A expressão do pronome pessoal sujeito (PPS) no espanhol é altamente variável (OTHEGUY; ZENTELLA; LIVERT, 2007) como se vê em (4) e (5).

- (4) Sí/ **yo trabajé** en (battanteh) bastante (patte:) parte/ cocinando. [CA-10-12M]
 (5a) No/ (rettaurante) restaurante no/ **trabajé** en casa (e) de familia. [CA-10-12M]
 (5b) eso</expresivo> de ahí//bueno/ de ahí cuando llegamo/ **seguí yo trabajando** ahí [VA-05-11M]

Para muitos pesquisadores, essa variação é o resultado do paradigma verbal do espanhol que, por conter a marca de pessoa, permite que, em certos contextos, o pronome sujeito possa ser explícito ou nulo. Esta variação, percebida por quase todos os linguistas, tem sido explicada de maneira diferente pelas distintas perspectivas (gramática tradicional, sintaxe gerativa, os estudos funcionalistas, sociolinguística variacionista, etc). Na sequência, apresentam-se brevemente as principais ideias de alguns desses quadros teóricos sobre a expressão do sujeito pronominal.

2.2.1 A gramática tradicional

No tocante à expressão do pronome pessoal sujeito, a tradição gramatical considera a ideia de que o espanhol dispensa a necessidade da realização fonética do pronome por razões de redundância devido à riqueza flexional do verbo que indica a pessoa gramatical, embora, admita a presença dele por questões da ênfase, expressividade, contraste e ambiguidade (GILI GAYA, 1961; ALARCOS LLORACH, 1994; RAE, 1973).

Hernández Alonso (1996) resume com a seguinte citação as ideias sobre a expressão do pronome sujeito na tradição gramatical:

Dado que a flexão verbal no espanhol conservou de forma muito clara as diferenças desinenciais de pessoa, o pronome sujeito poderá omitir-se normalmente. No entanto, será necessária a sua presença ante o verbo em alguns casos, preferencialmente como signo diacrítico ou enfático. No primeiro caso, o pronome aparece para evitar a ambiguidade, para diferenciar o gênero no plural e para contrapor diferentes pessoas (exemplo: *Yo ya te he dado mi opinión, ahora tú sabrás lo que haces*). Com o valor enfático é o uso frequente desses pronomes, tanto para ressaltar o interesse de uma pessoa como a elipse verbal por zeugma (exemplo: *Tú has tenido toda la culpa. ¿Yo?*). Aparece também na frase atributiva que traz o

predicado antes do verbo e o sujeito posposto (Exemplo: *El estado soy yo*) (HERNÁNDEZ ALONSO, [1984] 1996, p.580, tradução nossa).²³

As principais fragilidades das explicações baseadas na tradição gramatical têm a ver com a superposição ou a carência de univocidade de alguns conceitos (MEYER-HERMANN, 1996). Provavelmente essas fragilidades sejam o resultado dos recursos teóricos e metodológicos, com baixo grau de detalhamento estrutural, de que dispunha a gramática tradicional, ou do pouco interesse em fornecer explicações do uso mesmo da língua, por se focar no normativo e não no descritivo, além do caráter geral que se atribui às propostas da tradição.

Também as gramáticas descritivas, *grosso modo*, mantêm o mesmo postulado (FERNÁNDEZ SORIANO, 1999; BOSQUE; GUTIÉRREZ, 2009), embora admitam que a expressão da função de foco contribui para a expressão do sujeito pronominal.

2.2.2 A Teoria de Princípios e parâmetros

Para a Teoria de Princípios e Parâmetros (CHOMSKY, 1986) as línguas seriam constituídas de princípios universais e parâmetros; os princípios constituem características gerais, comuns a todas as línguas e, portanto, propriedades inatas, e os parâmetros, características particulares específicas a cada língua, fixadas no período de aquisição, como um modo de detalhamentos da atuação dos Princípios.

Se, por um lado, for considerado o parâmetro do sujeito não nulo (*non pro-drop*), o sistema linguístico prevê a obrigatoriedade de preenchimento com material fonético de todas as posições de sujeito pronominal independentemente do contexto de uso. Se, por outro lado, for considerado o parâmetro do sujeito nulo, entende-se que, dentro de determinadas condições de licenciamento, o sujeito pronominal pode ser deixado nulo. Um exemplo de língua do primeiro tipo é o Inglês, em que somente são gramaticalmente aceitáveis sentenças com sujeito (*It rains; I have a green car*); um exemplo de línguas do segundo tipo é o

²³ Dado que la flexión verbal en español ha conservado de manera muy clara las diferencias desinenciales – personales, el pronombre sujeto podrá omitirse normalmente. Sin embargo, será necesaria su presencia ante el verbo en algunos casos, preferentemente como signo diacrítico o como enfático. En el primer caso aparece el pronombre para evitar el equívoco, para diferenciar el género, sobre todo en plural, y para contraponer diferentes personas (ejemplo: *Yo ya te he dado mi opinión, ahora tú sabrás lo que haces*). Con valor enfático es el uso frecuente de estos pronombres, tanto para resaltar el interés de una persona como la elipsis verbal por zeugma (ejemplo: *Tú has tenido toda la culpa. ¿Yo?*). Aparece también en la frase atributiva que lleva el atributo ante el verbo y el sujeto pospuesto (ejemplo: *El estado soy yo*).

espanhol, cujas sentenças não precisam necessariamente apresentar sujeito explícito (*Llueve; Tengo un coche verde*).

Segundo essa perspectiva teórica, portanto, o espanhol segue os princípios do parâmetro “língua de sujeito nulo”, que, como já se viu anteriormente, são línguas que admitem sujeitos vazios nas orações finitas em razão da robustez das distinções morfológicas na flexão verbal. Nessa teoria, é a categoria funcional concordância (AGR) a responsável pela identificação do sujeito foneticamente nulo.

No Brasil, investigando os procedimentos de preenchimento da posição de sujeito em sete peças de teatro populares escritas entre 1845 e 1992, Duarte (1993) detectou uma tendência inequívoca de redução progressiva de sujeitos vazios. A autora atribuiu o aumento expressivo de sujeitos preenchidos à perda da possibilidade de identificação mediante desinências verbais em razão de uma drástica redução do paradigma flexional do português brasileiro.

Em relação ao espanhol, diferentes estudos, mais especificamente os que se debruçam sobre a variedade caribenha antilhana, fornecem evidências de uma perda progressiva, similar ao que ocorre ao português, das características das línguas de sujeito nulo (pro-drop) e, em compensação, a conversão para uma variedade contendo, às vezes, sujeito obrigatório (non pro-drop) e, às vezes, sujeito misto (ORTIZ LÓPEZ, 2016). O caso da redução do paradigma flexional do português brasileiro, identificado como contendo apenas duas pessoas (*eu canto/tu, você, ele, ela, nós, vocês, eles canta*) não se aplica da mesma forma, como veremos mais adiante, às variedades caribenhas do espanhol colombiano.

2.2.3 Os estudos funcionalistas

Os estudos no quadro funcionalista, sustentados em várias perspectivas teóricas, que nem sempre compartilham os mesmos conceitos e categorias de análise, também fornecem explicações semânticas e pragmáticas para a expressão do PPS no espanhol. Morales (2007), por exemplo, com base numa série de estudos precedentes, faz relevantes deduções acerca da expressão do PPS.

Segundo Morales (2007), nas línguas que permitem a elisão do sujeito, como o espanhol atual, a simples realização do pronome sujeito pode representar um recurso de tematização contrastiva ou focalizada. Assim, tem-se comprovado nos contextos de possível confusão entre participantes, chamados sujeitos contrastivos. Nesses contextos, o falante de todos os dialetos (Buenos Aires, Madri e San Juan) impõe sua própria interpretação, ao

recorrer à realização anteposta do sujeito e, igualmente, nas orações relativas de objeto diante de um contexto com dois temas ativos, sujeito e objeto. Além disso, é frequente que, nos verbos de cognição e dicendi, os sujeitos pronominais estejam representados majoritariamente pela primeira pessoa, que aparece predominantemente em posição pré-verbal. O espanhol tem, assim, estratégias especiais de reforço e, como indicam os dados, elas privilegiam as duas pessoas do discurso.

Antes de mais nada, para Morales (2007), há que levar-se em conta que a realização pronominal anteposta ao verbo, que se percebe como um recurso marcado, pode prescindir de significado comunicativo especial. Se, numa comunidade linguística, as estratégias de tematização são recorrentes e se realizam especialmente com pronomes sujeitos tônicos, eles podem se converter, em longo prazo, em elementos clíticos obrigatórios. Esse poderia ser o caso do espanhol de Porto Rico, em que a tematização repetida atual não parece ser um recurso de expressão significativo, e a manifestação do actante é percebida como normal pelos porto-riquenhos.

Segundo Morales (2007), se partirmos do fato de que se trata unicamente de diferenças quantitativas e de que os processos de reforço do sujeito pronominal se movem para a mesma direção em todos os dialetos (estudados), são perceptíveis esses princípios pragmáticos universais compartilhados. O uso abundante do recurso na variedade porto-riquenha poderia obedecer a tendências dialetais estáveis incorporadas nos primeiros estágios de sua configuração dialetal e que seguem um rígido padrão linguístico SVO. Nesse sentido, o uso do sujeito pronominal, considerado o tempo todo como recurso “marcado” poderia ser universalmente “não marcado”, tanto em termos tipológicos como em termos de aquisição, se consideramos a linguagem infantil e as línguas crioulas.

Para Fanjul (2014), um pronome não comporta essencialmente informação nova, mas acessível (CHAFE, 1994 *apud* FANJUL, 2014), na perspectiva funcional que ele utiliza para analisar a expressão do PPS. Além disso, para o autor, os sujeitos pronominais no espanhol manifestam um valor de contraste não necessariamente limitado aos exemplos (6-7) do mesmo tipo, que foram mencionados por Morales (2007), mas podem exercer uma função de contraste fornecida pelo contexto discursivo (8).

(6) [...] en el cual **él** sacó una nota D y **yo** saqué una nota mejor (SJ MII) (Morales, 2007).

(7) porque **él** se fijó en una cosa que todo el mundo veía y **nadie** [...] todo el mundo miraba y **nadie** veía ¿no? (BA MII) (Morales, 2007).

(8) «Yo [y no tú u otra persona] creo que en eso estuvo mal» (Giardinelli *Oficio* [Arg. 1991] 162) (Diccionario Panhispánico de Dudas, online).

Ao contrário da tradição gramatical, que justifica a omissão do sujeito na ideia da riqueza da flexão verbal e a expressão, na ideia do estilo, Fanjul (2014)²⁴ não considera suficiente a primeira razão e não acredita que o problema seja resolvido em termos estilísticos. Para o autor, no quadro da perspectiva funcional adotada, a expressão do sujeito pronominal não é uma variável livre que dependa de decisões “estilísticas”, mas um processo orientado por fatores relacionados com a progressão temática, com consequências na interpretação, especificamente na identificação da referência. Assim, a presença do pronome sujeito no espanhol tem um efeito de contraste, que implica a necessidade de selecionar uma entre várias possibilidades, e pelos exemplos que fornece o autor, obtém-se esse efeito basicamente mediante a inserção do pronome na posição pré-verbal.

Por fim, García Salido (2011)²⁵ realiza um trabalho com dados conversacionais, que tem por objetivo avaliar (i) se é adequado atribuir caráter focal a qualquer pronome pessoal, cuja expressão seja opcional por contar com afixo verbal que o reproduz; e (ii) se o fato de estar focalizado o pronome pessoal não forçaria uma interpretação de contraste. O autor conclui o seguinte:

Comprovou-se, pois, que não é possível considerar focal e contrastivo qualquer instância de um pronome pessoal que disponha de uma marca sufixal no predicado e, no caso de existirem indícios que apontem para o caráter focal de determinado pronome, não se pode atribuir-lhe caráter de foco marcado ou estreito e interpretação contrastiva ou exaustiva (GARCÍA SALIDO, 2011, p.43-44, tradução nossa)²⁶

2.2.4 O quadro variacionista

A sociolinguística variacionista, possivelmente o quadro teórico-metodológico que mais tem desenvolvido estudos para explicar a presença ou ausência do pronome sujeito, têm sua própria interpretação do fenômeno. Embora as gramáticas tradicionais deem algumas explicações válidas sobre a expressão do sujeito, os estudos variacionistas demonstram, à luz das descobertas empíricas mais recentes, que merece ser revista a maioria dessas explicações.

²⁴ Fanjul adverte que, apesar de não desconhecer o comportamento dos dialetos caribenhos, sua explicação incide basicamente sobre o espanhol peninsular.

²⁵ Este autor faz uma análise de *corpora* de diferentes dialetos do espanhol peninsular.

²⁶ Se ha comprobado, pues, que no es posible considerar focal y contrastiva cualquier instancia de un pronombre personal que cuente con una marca afijal en el predicado y, en caso de que existan indicios que apunten al carácter focal de un determinado pronombre, no se le puede atribuir inmediatamente carácter de foco marcado o estrecho e interpretación contrastiva o exhaustiva.

Muitos trabalhos de cunho variacionista mostram que, nos contextos em que a gramática tradicional recomenda a não expressão da forma pronominal do sujeito, os falantes fazem uso explícito dele e, naqueles em que se considera necessária ou preferível a presença do pronome sujeito, os falantes não sempre o exprimem. Silva-Corvalán (2004), uma das primeiras pesquisadoras a abordar esse fenômeno em língua espanhola, afirma:

O estudo desta variável tem mostrado que sua “opcionalidade” é válida só em um número limitado de contextos. Não se trata de duas formas iguais para dizer a mesma coisa. Existem fatores pragmático-discursivos que condicionam a possibilidade de expressar o sujeito (SILVA-CORVALÁN, 2004, p.63, tradução nossa).²⁷

Os estudos variacionistas que tratam desse fenômeno têm demonstrado que, em certos contextos, é obrigatória ou categórica a manifestação do pronome sujeito: a) uso de expressões idiomáticas; b) usos enfáticos com *mismo/misma* [mesmo/mesma]; c) sujeitos focais. Há também, por outro lado, contextos em que é categórica a ausência do pronome pessoal, tais como, o uso dos verbos *hacer* [fazer] e *haber* [haver] e o emprego das orações relativas com sujeito relativizado (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011).

Nos contextos em que não é obrigatória ou categórica a expressão ou a ausência do sujeito, os fatores que condicionam a variação são de natureza morfossintática, como ambiguidade morfológica, pessoa e número gramatical, TMA verbal; são de natureza semântica, como o tipo semântico do verbo ou a frequência lexical, de tipo pragmático-discursivo, vista como a mudança de referência e a paralelismo formas ou *priming effect*, dentre outros (SILVA-CORVALÁN, 1982; BENTIVOGLIO, 1987; CAMERON, 1992, 1993; FLORES-FERRÁN, 2004ab; HURTADO, 2001; TRAVIS, 2007; OTHEGUY; ZENTELLA; LIVERT, 2007; OROZCO; GUY, 2008; MARTINEZ-SANZ, 2011, entre outros).

Blas Arroyo (2005) sintetiza na lista abaixo os fatores associados à expressão dos sujeitos pronominais no espanhol:

- 1) A mudança da referência com respeito ao sujeito precedente;
- 2) Os contextos em que se faz necessário distinguir o sujeito de outras funções sintáticas;
- 3) O significado de certos verbos;
- 4) O tempo verbal;

²⁷ El estudio de esta variable ha mostrado que su “opcionalidad” es válida sólo en un número limitado de contextos. No se trata de dos formas iguales de decir lo mismo. Existen factores pragmático-discursivos que condicionan la posibilidad de expresar el sujeto.

- 5) A necessidade de desfazer possíveis ambiguidades na morfologia verbal;
- 6) A expressão da informação nova ou contrastiva;
- 7) A pessoa e o número gramatical;
- 8) A compensação funcional pela supressão do /-s/ final da segunda pessoa nas variedades do espanhol em que se produz este fenômeno;
- 9) A mudança do turno conversacional;
- 10) A ênfase expressiva;
- 11) A tendência a fixar a ordem das palavras não marcada no espanhol: sujeito-verbo-objeto;
- 12) A existência de diferenças paramétricas relevantes entre diversos dialetos do espanhol;
- 13) A influência de uma língua de sujeito obrigatório como o inglês.

Embora possam parecer significativas nas variedades do espanhol em situação de contato com o inglês (OTHEGUY; ZENTELLA, 2012), as variáveis sociais têm demonstrado ser pouco relevantes nos trabalhos das variedades monolíngues (LASTRA; MARTÍN, 2015). Nessas variedades, a variável diatópica resulta mais significativa do que as sociais, embora essas ainda não tenham sido examinadas sistematicamente (SILVA-CORVALÁN, 2004). Ainda que os trabalhos recentes venham demonstrando a influência de certas variáveis sociais, elas não atuam de modo constante nos diferentes estudos e não têm tanto peso estatístico quanto as variáveis internas ou linguísticas. Nesses trabalhos, mostraram-se significativas: sexo/gênero, idade, nível socioeconômico (OROZCO; GUY, 2008; OROZCO, 2015, LASTRA; MARTÍN, 2015).

Segundo Silva-Corvalán (2001, 2004), um estudo dos pronomes deve começar por estabelecer as correlações internas (sintáticas, semânticas e pragmáticas) e, uma vez estabelecidas essas correlações, é possível estabelecer correlações sociais. Contudo, é importante esclarecer que o fato de não serem identificadas correlações sociais claras não enfraquece a importância do tratamento, porque estudos de variação sintática não têm necessariamente que testar e/ou estabelecer possíveis correlações sociolinguísticas para justificar sua importância (SERRANO, 2007, 2009; SILVA-CORVALÁN, 2001, 2004).

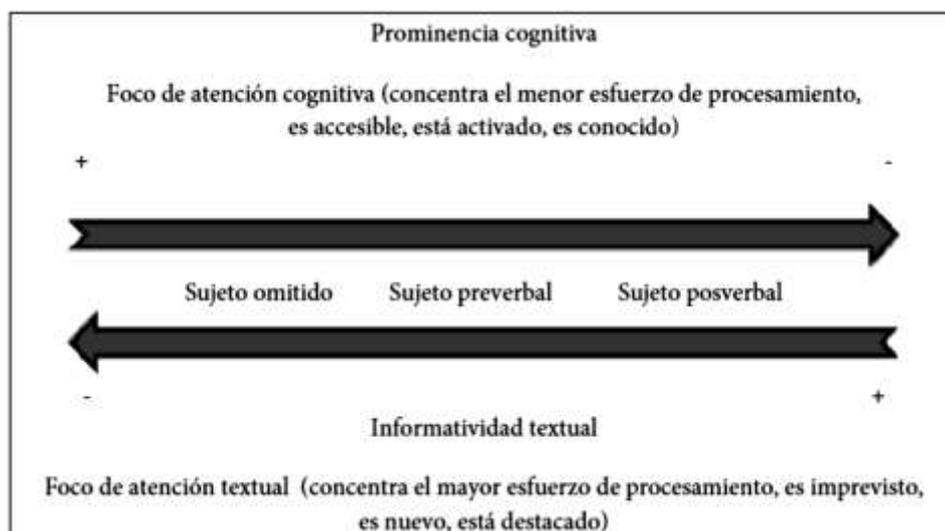
Em suma, os resultados encontrados até agora fornecem um panorama do uso dos pronomes em espanhol em distintas áreas dialetais, cujos resultados parecem indicar que a expressão do sujeito é regida por fatores semânticos e pragmáticos, fato que poderia ser próprio do espanhol em geral (MARTÍN BUTRAGUEÑO, 1994). Contudo, os resultados não podem ser comparados linearmente, porque, de uma perspectiva metodológica, não

respondem do mesmo modo às mesmas perguntas de pesquisa, nem seguem os mesmos critérios de análise.

A chamada que Silva-Corvalán fez em 1992 sobre a necessidade de se disporem de *corpora* coletados sob os mesmos critérios metodológicos e se abordarem temas de pesquisa que compartilhem alguns requisitos básicos comuns começa a tornar-se realidade no interior do PRESEEA. O grupo de análise que enfoca a expressão do sujeito pronominal, liderado por Bentivoglio, Ortiz López e Silva-Corvalán, fornece um roteiro com um mínimo de variáveis linguísticas e sociais que devem ser submetidas à análise sistemática em toda pesquisa sobre o tema que se realize com base nos *corpora* mencionados, para facilitar resultados contrastivos confiáveis (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011). Esta pesquisa segue esse roteiro e outras precisões metodológicas apresentadas em vários dos trabalhos referenciados.

Por outro lado, os estudos de Serrano (2013a, 2013b), no quadro variacionista diferente do laboviano, analisam a variação do PPS da perspectiva cognitiva e textual, que relaciona cada variante com a possibilidade de contribuir para a criação de estilos comunicativos no discurso. Dessa perspectiva recorrem às noções de *prominência cognitiva* e de *informatividade textual* para propor uma escala que adiciona valores opostos a cada uma das variantes do PPS, de forma que o sujeito omitido é a forma mais saliente, o que é expresso pós-verbalmente, o mais informativo e o que é expresso pré-verbalmente ocupa a posição intermediária entre os polos da *prominência cognitiva* e de *informatividade textual*. O Gráfico 2 faz uma representação dessa escala.

Gráfico 2 – Gradualidade da *prominência cognitiva* e da *informatividade textual*.



Fonte: Serrano, 2013a, p.289.

Por conseguinte, segundo o Gráfico 2, a expressão do sujeito tem um grau maior de informatividade que a não expressão; a expressão pré-verbal é, por seu lado, menos informativa que a pós-verbal, mas adquire maior saliência que essa última. A colocação posposta do sujeito corresponde com o extremo de grau máximo de informatividade textual e com a mínima prominência cognitiva.

Os resultados das pesquisas de Serrano (2013a; 2013b) confirmam que a posição do sujeito e seu significado co-ocorrem com certos elementos linguísticos variáveis, mas também com outros fatores de natureza social e situacional. Ambas as dimensões acabam por perfilar o significado da variação no discurso, constituída a partir da inter-relação das categorias mencionadas. Desse modo, o significado da posição pós-verbal é basicamente relacionado com o caráter epistêmico e com o caráter focal do referente graças ao aumento de informatividade textual, quando comparado com a variante pré-verbal, e adquire certo predomínio nas situações comunicativas em que o falante emprega a primeira pessoa (yo) em interações de tipo transacional. Nessas situações, o falante pode se ocultar ou “se tornar objeto”, em comparação ao destaque de sua presença e da agentividade que implicam a posição pré-verbal. O fato de serem elevadas as frequências percentuais na função comunicativa interpessoal para esse pronome se justifica no fato de ser ele mais frequente nas conversações espontâneas.

Além disso, os resultados desse trabalho confirmam a existência de pautas comunicativas diferentes (e em ocasiões opostas) entre homens e mulheres. Esse contraste pode indicar certo interesse por parte de ambos os grupos sociais de querer diferenciar-se linguisticamente um do outro e de construir estilos comunicativos paralelos, mas diversos, em consonância com as descobertas mais recentes do campo, que tendem a dar primazia à criatividade do falante na construção das identidades feminina e masculina (SERRANO, 2013b), nos termos da autora.

Em concordância com diferentes autores mencionados em seus trabalhos, Serrano (2013b) considera que os comportamentos femininos e masculinos se definem respectivamente em função de suas próprias necessidades comunicativas a cada situação, razão pela qual, em diferentes ocasiões, os comportamentos de homens e mulheres são muito similares. E por fim, que a relação entre linguagem e sexo/gênero é complexa e multidimensional.

A seção seguinte focaliza a expressão do sujeito pronominal em posição pós-verbal, porque nem sempre as razões para sua expressão nessa posição são as mesmas para a posição pré-verbal.

2.2.5 A posposição do sujeito pronominal

São muito limitadas as explicações que as gramáticas e os estudos teóricos e empíricos fornecem sobre a posposição do sujeito pronominal no espanhol. Embora, inicialmente fosse possível entender a explicação acerca da expressão do sujeito pronominal em geral, na posposição, parecem atuar fatores discursivos e pragmáticos nem sempre similares aos que motivam a posição pré-verbal. Na sequência, discutimos alguns enfoques que tratam desse tema.

As mesmas razões que a tradição gramatical fornece para a expressão do sujeito pronominal podem ser inferidas para qualquer posição que ele ocupe, seja a pré-verbal seja a pós-verbal. Em outros termos, a posposição se justificaria pelas motivações assinaladas acima de ênfase, contraste, expressividade e ambiguidade na flexão verbal e seria obrigatória na “frase predicativa que leva o predicado antes do verbo [...] (exemplo: *El estado soy yo*)”²⁸ (HERNÁNDEZ ALONSO, [1984] 1996, p.580, tradução nossa). As gramáticas descritivas, que defendem os mesmos argumentos, acrescentam como explicação que a posposição tem motivação focal. Para Fernández Soriano (1993), os fatores que incidem sobre a ordem dos constituintes principais da oração são lexicais (a natureza do verbo), sintáticos (as diversas configurações), informacionais (distinção do tema e rema) e de processamento (extensão do sujeito).

Procurando dar uma explicação das funções pragmáticas da posposição do PPS, os estudos funcionalistas assinalam que, embora o pronome posposto possa cumprir a função de foco, não é todo PPS posposto que exerce a função de foco (contrastivo) (MEYER-HERMAN, 1996; GARCÍA SALIDO, 2011)²⁹. Para Bentivoglio (1997, *apud* MORALES, 2007), a posposição do sujeito pronominal pode marcar continuidade tópica, nos termos da autora. A falta de consenso pode ser resultado também das distintas perspectivas e conceitos divergentes que assumem os pesquisadores, mesmo utilizando a mesma terminologia.

São os estudos variacionistas que têm presença mais constante na análise dos fatores que contribuem para a expressão do sujeito, e são eles também que fornecem um número maior de razões sobre a posposição. Embora, inicialmente não faça distinção entre os pronomes expressos na posição pré-verbal e na posição pós-verbal, que entendem como variantes da mesma variável, alguns trabalhos mostram que os grupos de fatores que

²⁸ en la frase atributiva que lleva el atributo ante el verbo

²⁹ O trabalho de Morales (2007) voltado para três variedades do espanhol (Buenos Aires, Madri e San Juan) demonstra que os falantes recorrem à manifestação anteposta de sujeitos contrastivos.

contribuem indistintamente para a expressão são os que também atuam na posposição, com a ressalva de que são reduzidos os casos de posposição pronominal em vários dialetos³⁰. Para López-Ortíz (2016), a expressão do sujeito pronominal na posição pré-verbal tem contribuído quase categoricamente para a ordem SV(O) no espanhol do Caribe antilhano (ECA), indistintamente na língua falada e na escrita.

Uma das principais hipóteses fornecidas pela gramática tradicional de que a expressão do sujeito pronominal é um recurso para desambiguar o referente tem-se confirmado em alguns estudos variacionistas que analisam particularmente a posposição. Silva-Corvalán (1982) salienta o fato de que a forma verbal ambígua favorece a expressão posposta do sujeito. Essa hipótese, *grosso modo*, foi constatada em pesquisas posteriores. A investigação de Hurtado (2001, p.121-122) mostra que, de 309 casos de sujeitos pospostos³¹, apenas 56 não apresentam algum tipo de ambiguidade. Os casos de ambiguidade ficaram assim distribuídos: 137 casos de ambiguidade morfológica de *usted, uno* e a terceira pessoa do singular, embora 113 casos se esclarecessem no contexto; 49 casos de ambiguidade morfológica entre *ustedes* e a terceira pessoa do plural, estando 47 deles esclarecidos no contexto, e finalmente 24 casos de ambiguidade morfológica entre *yo* e a terceira pessoa do singular, também esclarecidos na sua totalidade no contexto. Dito em outros termos, o contexto discursivo deixa clara a maioria dos casos de ambiguidade morfológica. Para Hurtado (2001), as frequências de uso mencionadas acima confirmam as deduções a que chegou Silva-Corvalán (1982), revelando que a posposição responde principalmente pela ambiguidade morfológica entre *usted, uno* e a terceira pessoa do singular.

Lastra e Martín (2015) confirmam também que a ambiguidade morfológica é um dos fatores mais significativos da posposição do sujeito pronominal, o terceiro com a amplitude mais robusta fornecida por resultados estatísticos do Goldvarb³². O estudo deles revela que as formas verbais ambíguas favorecem muito significativamente a expressão do PPS (.73), que se distancia bastante das formas não ambíguas de expressão, que claramente a desfavorecem, como indica o peso relativo de (.39).

Segundo os autores, dois aspectos parecem convergir para chegar a esse resultado. Em primeiro lugar, a presença de ocorrências no tempo imperfeito motiva a inserção de PPS

³⁰ Vários estudos sobre a variedade de Porto Rico confirmam a alta frequência de sujeitos pronominais antepostos e o caráter categórico em outros contextos (MORALES, 2007); Orozco e Guy (2008) encontraram um baixo percentual de sujeitos pospostos em Barranquilla (3,9%).

³¹ Tanto Hurtado como Silva-Corvalán analisam sujeitos pronominais e nominais (para a terceira pessoa).

³² A amplitude (*range*) mede a diferença entre o peso relativo mais baixo e o peso relativo mais alto: quanto mais elevada a amplitude maior a força de uma dada variável independente (TAGLIAMONTE, 2006).

estritamente para evitar a ambiguidade (*vivía, comía, ganaba*), além das formas do condicional e das formas com *usted, ustedes*. Em segundo lugar, a elevada frequência de *uno* e *yo* pospostos implica naturalmente a presença de formas morfológicas de primeira e terceira pessoa do singular, precisamente aquelas que estão mais relacionadas com a ambiguidade morfológica.

Em Lastra e Martín (2015), outro grupo de fatores selecionado como significativo foi pessoa e número gramatical, que consiste na variável com maior amplitude. Nesse grupo dois dos fatores tiveram percentuais notáveis: *uno* (77,8%, 21/27), cuja posição como objeto poderia estar associada com o seu significado genérico, e *yo* (12,5%, 29/232). Depois da análise individual de todos os casos, os autores descobriram que muitos deles estavam formulados no tempo imperfeito e no condicional, como um fragmento descritivo numa narração mais ampla (frequentemente na abertura): *tenía yo, estaba yo, tendría yo, etc.* Outro grupo de casos aparece no pretérito num ponto importante da narrativa: *nací yo, me casé yo*, entre outros.

As variáveis sociais também produzem resultados significativos em alguns trabalhos variacionistas. Lastra e Martín (2015) mostram que o grau de escolaridade é uma variável significativa na posposição do PPS, mantendo a segunda amplitude mais significativa. Um cruzamento dessa variável com a variável pessoa e número revelou um padrão significativo. Todos os casos de *uno* posposto (24) foram realizados pelas pessoas de grau mais baixo de escolaridade. Além disso, os 29 casos de *yo* pospostos mostram também uma distribuição bem estratificada: 19 no grau baixo, 7 no grau médio e unicamente 3 no grau alto; grau de escolaridade é, então, uma força social ativa na posposição.

Outra variável social selecionada como significativa foi sexo/gênero; homens favorecem a posposição pronominal, mas são eles que mantêm menos número de casos de expressão de PPS. Como um contraponto, em geral, as mulheres favorecem a expressão de PPS. Quando se cruza essa variável com idade, os resultados mostram que 40/47 casos são realizados por homens de 55+ anos. A conclusão a que chegam os autores é a de que as pessoas que realizam PPS pospostos com mais frequência são falantes masculinos, adultos, e com baixo grau de escolaridade. No entanto, fazem uma ressalva, ao afirmar que a posição pré-verbal ou pós-verbal não é uma questão determinística, havendo, portanto, um alto grau de liberdade para tomar decisões.

Uma hipótese sociodialetal testada em alguns trabalhos é a proposta por Cameron (1993): Se um dialeto A tem um percentual menos elevado de expressão de PPS do que o dialeto B, então o dialeto A teria um percentual mais frequente de sujeitos pospostos do que o

dialeto B. Se o dialeto B tem um percentual mais elevado de expressão de PPS do que o dialeto A, então o dialeto B teria um percentual menos frequente de sujeitos pospostos do que o dialeto A.

Segundo Lastra e Martín (2015), a principal deficiência da hipótese de Cameron (1993) tem a ver com a base de dados em que se assenta sua proposta. Cameron (1993) emprega os dados coletados por Silva-Corvalán (1977) sobre o espanhol mexicano falado no Oeste de Los Angeles (Califórnia), cujos resultados indicam 65% de sujeitos pré-verbais e 35% de sujeitos pós-verbais, com um índice geral de expressão do sujeito de 39%; utiliza também os dados de Morales (1982) sobre o espanhol de Porto Rico, cujos resultados chegaram a 81% de sujeitos antepostos e 19% de pospostos, com um índice geral de expressão do sujeito de 57%. A fragilidade da hipótese tem a ver com fato de que Cameron (1993) faz uma predição da expressão de PPS e não de sujeitos em geral, e os percentuais de Silva-Corvalán (1977) e Morales (1982) são precisamente sobre sujeito e não exclusivamente sobre PPS.

Para Lastra e Martín (2015), os resultados de sua própria pesquisa sobre o espanhol da Cidade de México refutariam a hipótese de Cameron (1993). Suas descobertas sobre a variedade da Cidade do México mostram que há 85% de PPS antepostos e 14,4% de PPS pospostos. Esses dados são bastante similares aos do espanhol de Yucatán no México, cujos percentuais são 85,6% para a posição pré-verbal e 12% para a pós-verbal (MICHNOWICZ, 2015). Esses resultados aproximariam mais as duas variedades mexicanas da variedade de Porto Rico do que da variedade de mexicanos de Los Angeles³³, sendo que seus índices de expressão de PPS na variedade da Cidade do México e de Yucatán pertenceriam aos dialetos tipo A.

No entanto, as comparações que Lastra e Martín (2015) realizaram também apresentam os mesmos equívocos que eles próprios destacaram em relação à hipótese de Cameron (1993), ou seja, comparar dados de posição do PPS com dados de posição do sujeito em geral³⁴. De fato, trabalhos mais recentes sobre a variedade de Porto Rico, focalizados na expressão do pronome sujeito (MORALES, 2007; ORTIZ LÓPEZ, 2016), mesmo sem fornecer dados quantitativos, atestam que a posição pré-verbal do sujeito pronominal é quase categórica em todos os contextos, o que permite inferir um percentual de pronomes sujeitos

³³ Eles fazem agora a comparação de seus dados com os dados de Morales (1982) e Silva-Corvalán (1977) que, como se disse, refletem dados do sujeito em geral e não simplesmente do sujeito pronominal, fenômeno sobre o qual incide o trabalho sobre a variedade da Cidade do México.

³⁴ Equívocos dos quais os autores estão cientes especialmente por mencionarem a importância de não comparar literalmente os resultados da expressão do sujeito com resultados de expressão do PPS.

pospostos muito mais baixo que o fornecido antes pelo mesmo autor (MORALES, 1982). Assim, com certeza o índice de pronomes sujeitos pospostos em Porto Rico é inferior ao do mesmo fenômeno nas variedades da Cidade do México e Yucatán, o que parece sustentar, em vez de anular, a hipótese de Cameron (1993).

Outros trabalhos parecem confirmar a hipótese. Num estudo descritivo sobre a fala culta de Havana³⁵, González e Pérez (2010) confirmam uma tendência geral à anteposição do sujeito pronominal. A posposição ocorre quase unicamente nas orações do tipo *¿qué sé yo?*, que consistem em frases feitas com significado unitário, inclusive, quase esvaziadas de significado. Orozco e Guy (2008) registram percentuais muito baixos de pronomes sujeitos pospostos na variedade de Barranquilla (3,9%), em que há percentuais elevados de expressão pronominal (35,7%). Numa pesquisa sobre uma variedade chilena, Van Esbroeck (2014), por outro lado, concluiu que o percentual de sujeito pronominal não é muito elevado (25,93%), se comparado com o das variedades caribenhas, além de dispor de um considerável percentual de sujeitos pronominais pospostos (26,14%). Assim, essa variedade que parece estar numa posição intermediária entre os dialetos caribenhos com mais sujeitos expressos e os continentais que menos expressam o sujeito (OTHEGUY; ZENTELLA; LIVERT, 2007; OTHEGUY; ZENTELLA, 2012), inserindo-se, portanto, na classificação de dialeto A de Cameron (1993), deveria ter um percentual também intermediário em relação a sujeitos pronominais pospostos.

É possível concluir esta seção com as palavras seguintes de Meyer-Hermann (1990):

Se se quiser resumir os resultados de todas as pesquisas que temos realizado em torno do mesmo tema, tropeça-se em grandes dificuldades. Em suma, os conhecimentos sobre esse assunto não estão tão avançados como seria de se esperar. Isso é devido, sobretudo, ao fato de os estudos mencionados não poderem ser comparados entre si. Sobre o tema, não há dois estudos de dois autores diferentes baseados na mesma concepção teórica ou nos mesmos fatos empíricos (1990, p.74, tradução nossa)³⁶.

³⁵ A variedade da Havana, pelos resultados dos trabalhos sobre o espanhol cubano (OTHEGUY; ZENTELLA; LIVERT, 2007; OTHEGUY; ZENTELLA, 2012), pode ser considerada uma variedade do tipo B na classificação de Cameron (1993).

³⁶ Si se quieren resumir los resultados de todas estas investigaciones realizadas alrededor de un mismo tema, se tropieza con grandes dificultades. En suma, los conocimientos en este asunto no están tan adelantados como se pudiera esperar. Esto es debido sobre todo a que los estudios mencionados no pueden ser comparados entre ellos. Sobre este tema, no hay dos estudios de dos autores diferentes que se basen en la misma concepción teórica o en los mismos hechos empíricos.

3 PROCEDIMENTOS METODOLÓGICOS E UNIVERSO DE INVESTIGAÇÃO

3.0 Introdução

Neste capítulo, em que se apresentam os aspectos metodológicos, primeiramente, levantam-se as hipóteses e se discute o universo de investigação em termos da natureza dos dados empregados e das informações sociodemográficas dos participantes. A seguir, descreve-se o modo como os dados foram analisados, a partir da análise qualitativa das entrevistas e da circunscrição do contexto da variável para expressão dos sujeitos; na sequência, apresentam-se as variáveis independentes incluídas na análise e, simultaneamente, sua operacionalização. Finalmente, discute-se brevemente a natureza das análises estatísticas realizadas.

Tratemos, inicialmente, do método. Os sociolinguistas concordam que o desenvolvimento da teoria variacionista no transcurso desses cinquenta anos tem contribuído para estabelecer seu status epistemológico e também para aperfeiçoar os procedimentos metodológicos (LÓPEZ MORALES, 1994; MORENO FERNÁNDEZ, 1998; SILVA-CORVALÁN, 2001; HERNÁNDEZ CAMPOY; ALMEIDA, 2005). É justamente esse desenvolvimento que permitiu a consolidação de metodologias pertinentes à realidade do objeto a ser investigado.

Silva-Corvalán (2001), por exemplo, enuncia os procedimentos da pesquisa sociolinguística que coincidem, em grande medida, com os parâmetros próprios da pesquisa nas ciências sociais: a) formula-se uma hipótese de trabalho; b) selecionam-se os falantes; c) coletam-se os dados; d) analisam-se os dados qualitativamente (identificação de variáveis linguísticas, identificação de contextos linguísticos, identificação de variáveis extralinguísticas e codificação das variáveis) e quantitativamente (quantificação e aplicação de procedimentos estatísticos); e) interpretam-se os resultados das análises. Nas seções seguintes explicitamos justamente esses parâmetros, iniciando com as hipóteses.

3.1 Hipóteses de trabalho

Uma hipótese geral, que pretendemos confirmar é a de que, por ser considerada parte do espanhol caribenho segundo dialetólogos (HENRÍQUEZ UREÑA, 1921; ZAMORA; GUITART, 1982; LIPSKI, 1994; QUESADA, 2000, entre outros), a variedade caribenha

colombiana apresenta um comportamento bastante próximo das variedades caribenhas em relação ao uso dos pronomes pessoais sujeitos.

A variedade caribenha do espanhol colombiano dispõe de um percentual superior a 30%³⁷ de presença de pronomes sujeitos por fazer parte do espanhol caribenho, que se caracteriza justamente por contar com índices mais elevados de explicitação que as variedades continentais do espanhol da América e as variedades da Espanha (CAMERON, 1992, 1993, 1995; FLORES-FERRÁN, 2004a; OTHEGUY; ZENTELLA; LIVERT 2007; OROZCO; GUY, 2008, entre outros).

Estudos anteriores do espanhol de Barranquilla no Caribe colombiano apurou que os índices de frequência da presença do pronome sujeito confirmam, para esse fenômeno morfossintático, a hipótese de que o espanhol dessa região faz parte do espanhol do Caribe (OROZCO; GUY, 2008; OROZCO, 2015)³⁸.

Cameron (1993, p. 321) fez uma predição que ainda não recebeu uma análise definitiva nos estudos variacionistas sobre a variação do sujeito pronominal. Segundo essa predição, se uma variedade A tiver um percentual relativamente menos elevado de expressão de sujeito pronominal do que a variedade B, então a variedade A teria um percentual mais elevado de sujeitos pospostos do que a variedade B. Pelo contrário, se a variedade B tiver um percentual mais elevado de expressão de sujeito pronominal do que a variedade A, então a variedade B teria um percentual menor de sujeitos pospostos do que a variedade A.

De conformidade com a hipótese de Cameron, consideramos que, por comportar-se como um dialeto B, a variedade caribenha colombiana, deve apresentar um percentual mais elevado de PPS e, conseqüentemente, uma frequência mais baixa de pronomes pospostos do que os dialetos continentais ou peninsulares que se comportam como o dialeto A (CAMERON, 1993; VAN ESBROECK, 2014; LASTRA; MARTÍN, 2015)³⁹. Essa hipótese

³⁷ Com base uma revisão de diferentes trabalhos sobre a expressão pronominal, Orozco (2015) assinala que, nos dialetos caribenhos, é um lugar comum estarem os índices percentuais acima de 30%, o que pode ser corroborado com os dados de uma revisão da literatura fornecida pelo mesmo Orozco (2015), como também com os dados de outros trabalhos não referenciados por ele que fazem análises de PPS em dialetos caribenhos a não caribenhos.

³⁸ Tentaremos fazer uma revisão do ponto de vista da sociolinguística *de contraste* para detectar se os grupos de fatores e os fatores mais significativos nesta pesquisa são bastante parecidos ou não aos de variedades caribenhas, ainda que outros trabalhos afirmem o contrário (CLAES, 2011), e as diferenças metodológicas dificultem o contraste. Em trabalho anterior (PÉREZ CÓRDOBA; CAMACHO, 2019), estabelece-se um contraste entre os grupos de fatores e fatores significativos selecionados em cada subdialeto, cujos resultados apontam para grande similaridade entre os três.

³⁹ Sobre quais grupos de fatores linguísticos e sociais favorecem percentualmente a posposição do sujeito, consideram-se os já selecionados em pesquisas anteriores, ou seja, ambigüidade, pessoa e número gramatical, mudança na referência e no grau de escolaridade; no entanto, as diferenças entre os dialetos estudados e os

de Cameron (1993), de acordo com Morales (1982), está atrelada ao argumento de que o dialeto porto-riquenho tende a fixar a ordem SVO, refletida nos índices percentuais mais elevados de PPS na posição pré-verbal. Esse aspecto levaria a considerar as variedades antilhanas e as caribenhas colombianas como mais inovadoras em relação às outras variedades do espanhol americano e peninsular que seriam consideradas mais conservadoras em relação à expressão e posição do sujeito pronominal.

Por outro lado, as subvariedades aqui investigadas, Barranquilla, Cartagena e Valledupar não devem mostrar diferenças significativas entre si. Por fazerem parte do dialeto espanhol caribenho colombiano, formam uma variedade bastante coesa em relação ao fenômeno estudado como acontece com alguns outros fenômenos variáveis, seja fonológicos, seja morfossintáticos, que permitem agrupá-los numa mesma variedade dialetal (MONTES GIRALDO, 1982).

Uma segunda hipótese que se quer examinar é a de que os grupos de fatores internos ou linguísticos são mais determinantes do que os grupos de fatores sociais para a expressão dos pronomes sujeitos na variedade caribenha do espanhol colombiano por conta do que afirmam estudos prévios sobre outras variedades do espanhol (BENTIVOGLIO, 1987a; OROZCO; GUY, 2008; OROZCO, 2015; LASTRA; MARTÍN, 2015). A principal razão para esse comportamento provável dos dados é a de que a expressão ou não de PPS não é uma variável com valores de prestígio e de estigmatização associados às variantes.

Nesse contexto de grupos de fatores linguísticos, uma hipótese a confirmar ou rejeitar é a influência de pessoa, número e especificidade do sujeito. As formas verbais singulares, especialmente as formas de primeira e de segunda pessoa do singular, favorecem PPS, enquanto as formas plurais a desfavorecem (MORALES, 1986, 2007; SILVA-CORVALÁN, 2001; OROZCO; GUY, 2008; OROZCO, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

O grupo de fatores [+/-] especificidade não tem apresentado resultados significativos em vários trabalhos que o analisam (LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). No entanto, a análise da influência de segunda pessoa do singular tem demonstrado diferenças significativas na expressão do pronome pessoal sujeito (CAMERON, 1993; HURTADO, 2001). Por isso, considera-se, com base nesses estudos

prévios, que os pronomes *tú* [-esp] e *usted* [-esp] favorecem a expressão do pronome sujeito em comparação com as formas [+esp], que não a favorecem⁴⁰.

Outro grupo de fatores é o de correferencialidade: os contextos de referência variável favorecem a presença de PPS, enquanto os contextos de referência idêntica favorecem a ausência (BENTIVOGLIO, 1987a; SILVA-CORVALÁN, 2001; OROZCO; GUY, 2008; OROZCO, 2015; LASTRA; MARTÍN, 2015). Adicionalmente, correferência parcial, isto é, os casos em que o sujeito tem como referente o complemento da oração anterior, mostra efeito pouco significativo, que se traduz quantitativamente nem por favorecimento nem por desfavorecimento de PPS (BENTIVOGLIO, 1987a; OROZCO; GUY, 2008; OROZCO, 2015; LASTRA; MARTÍN, 2015).

Quanto ao grupo de fatores tempo verbal, diferentemente de outros trabalhos que codificam tempo, modo, aspecto verbais juntos, optamos por seguir a proposta de Bentivoglio, Ortiz López e Silva-Corvalán (2011) de analisá-los separadamente. As formas imperfeitas de indicativo e do condicional favorecem a expressão do PPS, enquanto as formas de presente, pretérito perfeito simples de indicativo e as formas compostas a desfavorecem, (SILVA-CORVALÁN, 1997 *apud* SILVA-CORVALÁN, 2001; OROZCO; GUY, 2008; OROZCO, 2015; LASTRA; MARTÍN, 2015).

Como as flexões do imperfeito, do condicional e do subjuntivo são morfologicamente indiferenciadas na primeira e na terceira pessoa do singular, é provável que os falantes optem pela expressão dos pronomes para evitar ambiguidade na identificação do referente, diferentemente do que ocorre com outros tempos verbais (HOCHBERG, 1986)⁴¹. Portanto, formas morfologicamente ambíguas são mais prováveis de ocorrer com PPS explícitos do que formas não ambíguas (HOCHBERG 1986; HURTADO, 2001; TRAVIS, 2007; ERKER; GUY, 2012; MARTÍNEZ, 2012; LASTRA; MARTÍN, 2015; MICHNOWICZ, 2015).

Similarmente ao grupo de fatores tempo, o grupo flexões verbais não diferenciadas para diferentes pessoas (ausência de marcação explícita da pessoa) favorece mais

⁴⁰ Segundo Limerick (2018) é a especificidade que favorece a expressão do sujeito nos pronomes de segunda pessoa do singular em duas variedades do espanhol mexicano.

⁴¹ Silva-Corvalán (2001) postula que é o tempo, mais especificamente a função do tempo no discurso, que se correlaciona com a expressão do sujeito. Em razão disso, a autora propõe que, em decorrência da função pragmática do tempo no discurso, seria possível esperar uma incidência menor de sujeito expresso com verbos no pretérito, que vai aumentando progressivamente com verbos do pretérito para o presente e do presente para o imperfeito. Essa predição está baseada no pressuposto de que pronomes sujeitos expressos atraem a atenção para o sujeito, reduzindo, conseqüentemente, a atenção prestada ao evento (SILVA-CORVALÁN, 2001).

provavelmente a presença do PPS para desambiguar o referente do sujeito (HOCHBERG, 1986)⁴².

Quanto ao grupo de fatores modo verbal, os usos de formas subjuntivas desfavorecem a presença de PPS e o de formas indicativas exercem efeito reduzido sobre a variável (LASTRA; MARTÍN, 2015). O desfavorecimento da variante ‘pronomes expressos’ pode justificar-se na função subordinada do subjuntivo.

Quanto à morfologia verbal, e tendo presente os resultados de estudos que agrupam tempo, modo e aspecto, considera-se que as formas simples favoreçam a expressão pronominal e as formas compostas a desfavoreçam.

O grupo de fatores turno de fala se justifica no fato de que a retomada do turno na entrevista sociolinguística força o falante a tornar explícito o sujeito pronominal e o fato da manutenção do turno favorece a ausência do PPS (BENTIVOGLIO, 1987a, MARTÍNEZ, 2012; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

Outro grupo de fatores é o de *paralelismo formal*⁴³. Por paralelismo formal, entende-se que a presença de PPS é favorecida quando a oração precedente já dispõe de um PPS e, *mutatis mutandis*, favorece a ausência do PPS quando a oração precedente também é marcada pela ausência de pronome sujeito. Outros tipos de sujeito na oração precedente não têm um efeito claro na expressão do PPS analisado (CAMERON; FLORES-FERRÁN, 2004; FLORES-FERRÁN, 2004b, TRAVIS, 2007; PÉREZ, 2010; OROZCO, 2015).

O roteiro para codificação do PPS do PRESEEA propõe um grupo de fatores denominado “tipo de frase”, que busca verificar se a natureza da oração, afirmativa, negativa ou interrogativa, favorece ou não a expressão pronominal. Estudos prévios (LASTRA;

⁴² O enfraquecimento do –s e, em certos contextos de –n em posição final de palavra nas variedades caribenhas, pode contribuir para a simplificação do paradigma flexional do verbo e, conseqüentemente, para uma maior ambigüidade do referente que deveria ser explicitado pelo pronome expresso ou pelo contexto. Essa ideia é defendida por alguns pesquisadores que analisam as variedades antilhanas (HOCHBERG, 1986) e refutada por outros (CAMERON, 1993). Essas diferenças de resultados deveriam ser um argumento relevante para incluir um grupo de fatores diferente que controlasse o comportamento da expressão do pronome pessoal sujeito quando se suprime o –s ou incluir mais fatores que analisassem essa complexidade no grupo de fatores ambigüidade morfológica; na nota 3, contudo, explicam-se as razões pelas quais não se aborda esse aspecto nesta pesquisa.

⁴³ Em português, esse grupo de fatores é bem conhecido pelo nome de “paralelismo formal” (CAMACHO, 2013), e, no espanhol, pode ser conhecido pelos conceitos de perseverança, persistência, efeito repetição ou *priming effect*. Camacho (2013, p. 219) define o paralelismo como “a repetição de formas, considerada tanto na sucessão de construções no discurso quanto na preservação da primeira posição estrutural para abrigar o mesmo referente numa sucessão de sentenças”. Para o autor esse fenômeno “se identifica com uma tendência mecânica de preservação de estruturas sintaticamente paralelas”. O paralelismo formal é uma variável relativamente pouco estudada nos trabalhos variacionistas do PPS na língua espanhola, quando comparada com os estudos do efeito dela em outros fenômenos linguísticos. Além disso, quase todos analisam o efeito desse grupo de fatores na primeira pessoa do singular (yo), e tomam em consideração até dez orações prévias para fazer a análise do efeito. O trabalho conhecido nesse domínio, que analisa o comportamento de todos os pronomes, mas tomando em consideração unicamente a oração precedente, é o de Orozco (2015).

MARTÍN, 2015; LIMERICK, 2018⁴⁴) mostram que orações com ilocução interrogativa e afirmativa são contextos de favorecimento da presença de PPS, enquanto sentenças negativas são contextos de desfavorecimento. Conforme postulam Lastra e Martín (2015), o efeito desfavorável do PPS com construções negativas pode estar relacionado ao agrupamento de orações negativas, que estão simultaneamente em contextos essenciais para a progressão do discurso, um parâmetro que promove a ausência de PPS. De qualquer modo, efeitos de polaridade podem ser mitigados no caso de haver ou não uma mudança no referente (mesma referência vs. mudança de referência). Além disso, a presença de operadores pré-verbais de negação pode representar um fator desfavorecedor da presença do PPS.

Também o roteiro propõe o grupo de fatores progressividade, contendo duas variantes, progressiva e não progressiva. Postula-se que formas verbais progressivas não favorecem a presença de PPS, enquanto as formas não progressivas parecem não exercer um efeito muito nítido sobre a presença ou ausência do PPS. Esse grupo de fatores não teve um resultado significativo em alguns estudos que o analisam (LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

Finalmente, dentro das hipóteses linguísticas considera-se que a classe semântica do verbo tem um efeito diferenciador para a presença ou ausência do PPS. Como a classificação assumida neste estudo não é a tradicionalmente utilizada nos estudos sobre a expressão pronominal, faz-se uma descrição breve dela para, em seguida, passar a formular uma hipótese.

A classificação assumida neste estudo é a da Base de Dados ADESSE (*Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español*) que propõe uma classificação de seis macroprocessos: 1) mental: compreende os verbos em que “uma entidade dotada de vida psíquica (A1) mantém ou experiencia algum tipo de estado, mudança de estado ou atividade interior perceptiva, sensitiva e/ou cognitiva (A2)”: *gustar, querer, ver, recordar, creer, elegir*, entre outros; 2) relacional: compreende os verbos de atribuição e posse: *ser, estar, parecer, nombrar*, entre outros; 3) material: compreende todos os verbos de processos físicos (não mentais): *ir, señalar, mecer, fabricar, romper, destruir*, entre outros; 4) dicendi: compreende os verbos de comunicação e os que manifestam valorização ou opinião; exemplos: *decir, hablar, criticar, pedir*, entre outros; 5) existencial: inclui verbos de existência, tempo-fase e vida: *ocurrir, empezar, durar, nacer, morir*, entre outros; 6) de modulação: compreende os

⁴⁴ Esses autores reagrupam as formas negativas com as interrogativas em um mesmo grupo de fatores que denominam por não afirmativas; no entanto, assume-se aqui o que se diz a respeito sobre não afirmativas exclusivamente para as negativas, e se diferenciam das interrogativas, que podem favorecer a presença de PPS.

verbos que funcionam como auxiliares ou semiauxiliares, ou como suportes: *ayudar, imponer, tratar, atreverse*, entre outros (GARCÍA-MIGUEL; ALBERTUZ, 2005; GARCÍA-MIGUEL; COSTAS; MARTÍNEZ, 2005)⁴⁵.

De acordo com essa classificação de ADESSE e os estudos prévios que utilizam essas categorias (PÉREZ, 2010; MARTÍNEZ, 2012), é provável que os mentais, os *dicendi*, e os existenciais favoreçam a presença de PPS enquanto os outros tipos de verbos a desfavoreçam (OROZCO; GUY, 2008; OROZCO, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Devido à sua função contrastiva implícita e à função semântica de expressar um ponto de vista, os verbos de processo mental promovem PPS explícitos, que são usados para afirmar o papel do falante (TRAVIS, 2007; LIMERICK, 2018).

Como já mencionado anteriormente, a terceira hipótese deste trabalho é a de que os grupos de fatores sociais ou externos não tenham relevância sobre a variável ou, se tiverem, disponham de menor grau de pertinência que os grupos de fatores linguísticos para a expressão dos pronomes sujeitos na variedade caribenha do espanhol colombiano, levando em consideração os estudos prévios que enfocaram outras variedades do espanhol (BENTIVOGLIO, 1987a; OROZCO; GUY, 2008; PÉREZ, 2010; MARTÍNEZ, 2012; OROZCO, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). De qualquer modo, os grupos de fatores sociais aqui enfocados são sexo/gênero, grau de escolaridade e faixa etária.

Começando por discutir sexo/gênero, é possível afirmar que nem sempre é clara a relação entre esse grupo de fatores sexo/gênero e a expressão do sujeito gramatical; no entanto, vários estudos comprovaram que as mulheres são mais propensas que os homens ao emprego de pronomes (PÉREZ, 2010; CARVALHO; CHILD, 2011; OTHEGUY; ZENTELLA, 2012; ALFARAZ, 2015; OROZCO, 2015).

De acordo com esses trabalhos, especialmente as descobertas sobre a variedade de Barranquilla (OROZCO, 2015), prevemos que esse grupo de fatores poderá ter efeito significativo para a expressão do PPS. Embora seja o efeito de sexo/gênero interpretado de diferentes maneiras, uma das interpretações possíveis é sua ligação com o valor de prestígio (ALFARAZ, 2015) e mesmo com a sugestão de possível mudança em progresso ou inovação (OTHEGUY; ZENTELLA, 2012).

⁴⁵ Todo sobre ADESSE pode ser consultado em <http://adesse.uvigo.es/> (GARCÍA-MIGUEL; ALBERTUZ, 2005; GARCÍA-MIGUEL; COSTAS; MARTÍNEZ, 2005). A classificação de ADESSE não coincide totalmente com as seguidas nos trabalhos variacionistas do PPS, mas também não difere substancialmente delas. Os únicos trabalhos que até agora preservaram a mencionada classificação são os que incidem sobre a expressão do PPS na fala de Caracas (PÉREZ, 2010; MARTÍNEZ, 2012).

Embora o grau de escolaridade não seja comumente considerado significativo, seus efeitos também são pouco examinados (LIMERICK, 2018). Sobre a variedade da Cidade de México (LASTRA; MARTÍN, 2015) e sobre a de Granada (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), constatou-se que esse grupo de fatores, mesmo não sendo selecionado como significativo pelo programa estatístico, interage com outras variáveis sociais, especialmente faixa etária.

Em ambas as variedades, os dados percentuais mostram que os informantes com menor grau de escolaridade favorecem a presença do PPS mais do que aqueles com grau superior de escolaridade, efeito que se justifica no fato de serem as regras gramaticais prescritivas mais fixadas nos níveis superiores de escolaridade. As conclusões de Ávila-Jiménez (1996) e Otheguy e Zentella (2012) são consistentes com essa explicação, pois encontraram um uso desfavorável de PPS com informantes desse grau de escolaridade. Se neste trabalho for significativa essa variável, prevê-se que o grau superior de escolaridade é um contexto desfavorecedor da presença do PPS.

Quanto ao grupo de fatores faixa etária⁴⁶, apenas muito recentemente a literatura sobre o assunto vem apontando para alguma relação substancial entre idade e expressão do pronome sujeito (OROZCO; GUY, 2008; CARVALHO; CHILD, 2011; MARTÍNEZ, 2012; ALFARAZ, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

A tendência desses estudos é a de que o grupo de informantes idosos favorece a presença do PPS e o grupo de jovens a desfavorece, mas pode haver outros trabalhos com uma posição oposta (FLORES- FERRÁN, 2004⁴⁷). Em alguns trabalhos, esse grupo de fatores interage de maneira significativa com outros fatores, como demonstrado no trabalho sobre a variedade de Barranquilla (OROZCO, 2015), contexto em que a diferença etária teve efeitos significativos; todavia, em interação com sexo/gênero, constata-se algum grau de significância no comportamento de mulheres adultas, que favorecem mais claramente a presença de pronome sujeito e dos homens mais jovens, que favorecem a ausência. Se, nos dados deste trabalho, esse grupo de fatores tiver alguma significância, prevê-se que, de acordo

⁴⁶ Mesmo havendo muitas propostas de agrupamento por faixa etária, neste trabalho se utilizam as estabelecidas pelo PRESEEA, que já foram testadas em vários trabalhos com esses *corpora* e com outros que seguem mais ou menos os mesmos critérios de classificação etária, e que não se distanciam das propostas de diferentes trabalhos sociolinguísticos em diferentes línguas. A proposta do PRESEEA é a seguinte: 20-34 anos (1); 35-54 anos (2) e 55+ anos (3).

⁴⁷ Mas esse trabalho é com falantes bilíngues porto-riquenhos que nasceram nos Estado Unidos, o que pode ser efeito do contato de língua nos mais jovens.

com a maioria desses trabalhos, os falantes adultos favoreçam a expressão do PPS e os mais jovens a desfavoreçam.

Essas hipóteses podem ser traduzidas nas seguintes perguntas de pesquisa:

- (1) Qual é a distribuição da presença/ausência dos pronomes sujeitos no espanhol do Caribe colombiano?
- (2) Dos grupos de fatores linguísticos e extralinguísticos investigados, quais são os que influenciam significativamente a expressão dos pronomes sujeitos no espanhol do Caribe colombiano?
- (3) A variedade caribenha colombiana mostra similaridades ou diferenças com o espanhol caribenho, em relação a percentuais de pronomes expressos, posição do pronome, fatores significativos?
- (4) A variedade estudada, comparada com outras do espanhol, mostra semelhanças ou diferenças nos índices percentuais de expressão dos pronomes sujeitos e nos grupos de fatores significativos?

Conhecidas as hipóteses e perguntas, passa-se a delimitar o universo de investigação em que se assenta a pesquisa.

3.2 Universo de investigação

A variedade do espanhol do Caribe colombiano enfocada aqui é constituída pelos *corpora* coletados em três cidades da costa caribenha: Barranquilla (RODRÍGUEZ CADENA, 2008b, 2009b, 2010), Cartagena (VÁSQUEZ; CUARTAS, 2017) e Valledupar (CALDERÓN NOGUERA, 2005). Esses *corpora* foram coletados e transcritos sob os critérios do PRESEEA; que é um projeto de criação de um *corpus* de língua espanhola falada em todo o mundo hispânico em suas variedades geográficas e sociais. Os materiais coletados, que estão reunidos de acordo com a diversidade sociolinguística das comunidades de língua espanhola, resultam do trabalho coordenado de pesquisadores comprometidos com uma metodologia comum com a finalidade de constituir um banco coerente de dados que favoreça uma aplicação educacional e tecnológica (MORENO FERNÁNDEZ, 1996)⁴⁸.

Em conformidade com o PRESEEA, o *corpus* de Barranquilla é constituído por 72 entrevistas, cada qual com um tempo de gravação de 45 a 60 minutos. Os *corpora* de Cartagena e Valledupar dispõem de 54 informantes cada, com o mesmo tempo de duração. As

⁴⁸ Para mais informações, consulte: <http://presea.linguas.net/>

entrevistas foram presenciais e, na maioria delas, encontravam-se apenas entrevistador e informante. Em casos excepcionais, havia mais de um entrevistador, mas sempre com a participação mais saliente de um deles. Em certas ocasiões, havia também mais de um informante acompanhando o principal, mas é este que se destaca por sua participação. Nesse contexto, a pessoa que mais sobressaiu na entrevista foi eleita como informante. Além disso, o levantamento procurou dar representatividade aos informantes de diferentes níveis socioeconômicos e de bairros tradicionais e modernos.

As entrevistas seguem os módulos temáticos sugeridos pelo PRESEEA, tendo como centro a vida cotidiana do informante, ou seja, sua família, seu trabalho, lembranças da cidade. Assim, foi possível documentar a fala das pessoas de diferentes tipos de origem social, nos seus próprios contextos (casa ou trabalho).

A amostra para este estudo, seguindo o roteiro para a pesquisa da expressão do sujeito no PRESEEA (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011), é constituída por um conjunto de 18 gravações de cada *corpus*⁴⁹, o que representou um total de 54 indivíduos-falantes do Caribe colombiano. Esses informantes estão distribuídos segundo as variáveis sociais pré-estratificadas, ou seja, por grau de escolaridade (baixo: seis anos de escolaridade ou menos; médio: de sete a doze anos de escolaridade; alto: mais de treze anos); por faixa etária (1: 20-34 anos, 2: 35-54 anos, 3: 55 ou mais); por sexo/gênero (feminino e masculino). A representatividade, de três informantes por célula, acha-se demonstrada no Quadro 6.

Quadro 6 – Distribuição dos informantes na amostra.

Faixa etária \ Grau de escolaridade	20 a 34 (1)		35 a 54(2)		55 ou mais (3)	
	H	M	H	M	H	M
Baixo (5 anos aprox.) (1)	3	3	3	3	3	3
Médio (12 anos aprox.) (2)	3	3	3	3	3	3
Alto (15 anos aprox.) (3)	3	3	3	3	3	3

Fonte: elaborado pelo autor

⁴⁹ O PRESEEA propõe 18 como um número mínimo de falantes por cada variedade, para ter representatividade por cada uma das variáveis sociais pré-estratificadas do projeto. Não entanto, o número de informantes por variedade não é homogêneo em outros estudos com outros *corpora*.

Uma vez descritas as características dos *corpora* utilizados para este trabalho, e os traços sociais mais relevantes dos falantes selecionados, passa-se a definir o envelope da variação, determinante em todo trabalho realizado no quadro variacionista laboviano (TAGLIAMONTE, 2006).

3.3 O envelope da variação

Nos procedimentos metodológicos foram levados em consideração os seguintes parâmetros: i) as observações do pesquisador sobre a variável em estudo; ii) o roteiro de codificação de Bentivoglio, Ortiz López e Silva-Corvalán (2011) e o Manual de Codificação de Otheguy, Zentella e Heidrick (2012); iii) as metodologias das pesquisas recentes que utilizam *corpora* do PRESEEA de San Juan (CLAES, 2011) e Cidade de México (LASTRA; MARTÍN, 2015)⁵⁰.

Definiu-se a variável dependente (expressão do pronome pessoal sujeito) em relação a duas variantes identificadas com a alternância e a ausência ou a presença do PPS em orações com verbos finitos, como se pode observar, respectivamente, em (9a) e (9b):

(9)a **Tengo** como tre (tres) dosaño (dos años)/ dosaños (dos años) por ahí// si dos años/ [VA-02-11H]

(9)b **nosotros nos tiramos** de todas partes de ahí del puente/ /[VA-02-11H]

Ainda que essa variável seja denominada ‘expressão do pronome pessoal sujeito’, é necessário fazer um esclarecimento adicional em relação ao tipo de distinção presença/ausência, que ficou clara já no título, mas que é importante reiterar: esta pesquisa não analisa sujeitos lexicais, que podem recorrer no discurso falado e que se comportam, por isso, como se fossem variantes em competição, em razão dos seguintes postulados: em primeiro lugar, a maior parte de trabalhos que aborda a variável sujeito analisa somente a alternância entre presença/ausência do pronome; em segundo lugar, os trabalhos que analisam a variável, incluindo sujeitos lexicais, têm que definir qual é a distância da oração em foco como outra variante do mesmo tipo, ou, pelo contrário, ser uma primeira menção⁵¹.

⁵⁰ Discutem-se também critérios metodológicos aplicados a outras pesquisas e pertinentes para a compreensão dos dados deste trabalho, como se pode observar no percurso da descrição metodológica.

⁵¹ Dumont (2006, *apud* Limerick, 2018) descobriu que a existência de um maior número de orações entre duas com o mesmo referente e o mesmo sujeito resultou em elevado número de sujeitos lexicais repetidos e baixo número de sujeitos pronominais, especialmente quando os sujeitos estão a uma distância de 10 ou mais orações;

Então, o procedimento analítico para a coleta dos dados levou em conta a análise apenas das orações em que aparece um verbo em forma finita, o que totalizou 150 casos por falante, incluindo todos os pronomes (*yo, tú, usted, él, ella, nosotros, nosotras, ustedes, ellos, ellas, uno*), para um total de 2700⁵² ocorrências por cidade e um total de 8200 para a variedade caribenha colombiana. Excluíram-se dessa seleção inicial 605 casos contendo o pronome indefinido *uno*, por não ser considerado um pronome pessoal, restando, portanto, um total de 7595 ocorrências.

O número de casos por falantes pode ser discutível, não há consenso na literatura sobre esse aspecto, nem do número ideal de falantes de uma variedade para dar conta do fenômeno em consideração, no entanto há tendências nos estudos. Nos estudos sobre os corpora do PRESEEA, há uma tendência para manter o número de 150 casos por falante do roteiro do mesmo projeto no estudo de México (LASTRA; MARTÍN, 2015), de Porto Rico (CLAES, 2011) e de Caracas⁵³ (PÉREZ, 2010; MARTINEZ, 2012). Em relação ao número de informantes por corpus, eles podem conter de 10 (CAMERON, 1993) até 140 (OTHEGUY; ZENTELLA, 2012) com características de pessoas e dialetos muito diferentes.

Excluíram-se da codificação os contextos em que são categóricas tanto a presença quanto a ausência do pronome. Nos seguintes contextos a presença do pronome é obrigatória:

a) usos enfáticos com *mismo/misma* [mesmo/mesma]:

(10) no pasó más/ ya / debido de pronto a la crianza a la obediencia que (**elloh mimmom**) **ellos mismos me (incuccaron) inculcaron** a mi/ ya/ [CA-20-21M]

b) sujeitos focais:

i) O sujeito é uma informação nova que responde a uma pergunta:

(11) yo recuerdo que cuando me fui a matricular me dijeron “y dónde está la muchacha que (uhté) usted va a (matriculá) matricular” y digo “no si **el muchacho soy yo mismo**”. [CA-39-32H]

Limerick (2018), pela sua vez, constatou que depois de 5 orações ou mais inseridas entre os dois referentes, o sujeito lexical é quase categórico. Aqui está um exemplo de sujeito léxico competindo com pronome ou com a ausência de pronome: *o sea mi papá// era un hombre <(e::)>/ yo digo que puede también ser por su historia// mi papá fue un hombre que él no <(estudio::)>* [VA-18-31M]

⁵² Orozco (2015) compara a taxa de incidência pronominal de seu trabalho em Barranquilla (34,2%), acrescentando 2000 ocorrências, em relação ao trabalho piloto de Orozco e Guy (2008) sobre a mesma cidade, cujo percentual é de 35,7%. Esse enfoque comprova que os resultados não diferem significativamente, comparação que leva o autor a sugerir, como uma consequência metodológica, que um conjunto de 1000 ocorrências aproximadamente já seria suficiente para estabelecer uma taxa pronominal confiável.

⁵³ Como analisam unicamente uma pessoa, as pesquisas de Caracas selecionam 50 casos por falantes; ou seja, uma média de 50 casos para cada uma das três pessoas.

ii) O sujeito é indiscutivelmente necessário por ser foco de contraste:

(12) **Mi (hemman) hermana se tiraba** en un rincón y **yo me tiraba** en otro/ [CA-06-13M]

c) O sujeito é um pronome indefinido (alguns, todos, etc), mesmo seguido do pronome *nosotros*:

(13) [...] y se cayó: se cayeron (unah) unas papayas/ y (noh) nos (asuttamo) asustamos y **todos (toos) salimos (salimok corriendo) corriendo**/ y así son cosas (cosah) así ((ruido)) que:/ nos (noh) han pasado (pasao).[CA-02-11 M]

Também foram excluídos os contextos em que não é possível a presença de um pronome pessoal, como os seguintes:

d) Verbos impessoais como *hacer* (14) [fazer], *haber* (15) [haver]:

(14) [...] uno no tira pedal para nada/ ahí se nos cayó un compañero **hace** poco/ **hace** como dos meses se nos cayó/ joda se nos cayó en una bajada/ VA-02-11H]

(15) primero que todo a (veceh) veces **hay** (pedsonah) personas que andan con (loh amigoh) los amigos y le brindan (drogah) drogas ¿ya:?:/ [CA-01-11H]

e) Orações com sujeito relativizado⁵⁴:

(16) en la puerta de la iglesia dicen (muchoh) muchos que han (vihto) visto entrar a un matrimonio/ lo entran/ lo:/ **un primo mío que vivió** frente a la iglesia él dice él él (vacilación) ha dicho que/ ha (vitto)[..] [CA-02-11 M]

f) Orações com *se* impessoal (17) e com terceira pessoa do plural com sentido impessoal (18)⁵⁵:

(17) (simultáneo) por eso se le (simultáneo) debía sí: esa deuda (**je se le debía** a Electrocosta/ y: eso (caa) cada rato venían y (nok cottaban) nos cortaban la (lu) luz [CA-05-13H]

(18) el caballo sin cabeza también **dicen** que sale/ en la puerta de la iglesia dicen (muchoh) muchos que han (vihto) visto entrar a un matrimonio/ CA-02-11 M]

g) Expressões idiomáticas ou marcadores do discurso⁵⁶ (*digamos, no sé qué*, etc) (19-20):

⁵⁴ Os *corpora* dispõem de um número muito reduzido de casos com sujeito pronominal resumptivo como: *mi papá fue un hombre que él no <(estudio::)> (estudió)/ que tuvo le <palabra cortada> que tuvo que levantarse solo/ desde niño/ [VA-18-31M]*

⁵⁵ Vários autores optaram já por excluir esses casos por acharem que não constituem casos de variação (MORALES, 1986, MARTÍNEZ, 2012; LASTRA; MARTÍN, 2015).

⁵⁶ Vásquez (2006) e Cuartas (2011) enumeram entre os diferentes marcadores do discurso em Barranquilla e Cartagena, respetivamente, a presença em ambos os dialetos de marcadores discursivos como: *sabes, tú sabes, no sé qué, qué se yo, que yo sepa, te repito, digamos, digamoslo así, qué te digo, no sabría decirte, me entiendes*. Para Morales (1986, p. 84), a segunda pessoa do singular fornece itens de índole muito diversa, porque juntam-se

- (19) [...] voy voy(vacilación) a ayudar a (lop pobre) los pobres voy ayudar **no sé qué**/ (pueh) pues a mí me (guttaría) gustaría primero/ [CA-02-11 M]
- (20) yo no (vacilación) nunca (jui) fui rebelde siempre (jui) fui como sumisa/ **me entiendes**. [CA-20-21M]

h) Orações imperativas:

- (21) [...] nos juimos (fuimos) entonce (entonces) él dijo/ bueno no **se vaya a demorá (demorar)** pa (para) que **venga y sirva**/ <expresivo> no <vacilación> no yo no me voy a demorar [VA-05-11M]

i) Verbos com referentes inanimados⁵⁷:

- (22) [...] **la visión cósmica del mundo/ y religiosa y religiosa del hombre** también es universal entonces **trasciende** más allá de una creencia y de un dogma/ **trasciende** hacia la identidad del ser en (buhqueda) búsqueda de una satisfacción/ [CA-39-32H]

j) Estruturas tipo *vamos a + infinitivo*⁵⁸, com sentido inclusivo ou exortativo:

- (23) bueno **vamo (vamos) a ve (ver)** pa ve (para ver) si no doy/ tonce (entonces) tiene uno que mostrá (demostrar) que si llego uno al fondo/ [VA-02-11H]

Também é importante mencionar que, em orações repetidas idênticas ou quase idênticas, que não representam estados de coisas diferentes, só se elege uma das orações para ser codificada, sendo que a outra é excluída (24); no caso de haver orações truncadas, seleciona-se a que tem sentido completo (25), sendo excluídos os truncamentos não retomados em orações.

- (24) 908 A: sí sí (repetición de palabra) / **ellos son así / ellos son así** / (simultáneo) la 909 mayoría son así (simultáneo). (BA-01-31M)
- (25) porque él era un señor que con esa / **como él tiene es** (palabra cortada) / bueno / **como él / si él lo tiene / esa cara así** toda / como agría / cuando él sonrío / se cambia / enseguida / (BA-01-31M)

aos usos pessoais regulares as frases feitas que muitos falantes repetem na segunda pessoa do singular e que não sempre se distinguem claramente dos primeiros.

⁵⁷ Embora verbos com sujeitos inanimados possam aparecer com PPS expreso sua ocorrência é extremamente rara (OTHEGUY; ZENTELLA 2012: 240-241). Em nossos dados se registram dois casos com dois falantes *el amor de (Dioh) Dios (eh) es eterno él nunca se acaba entonces si tú dices que tú (amah) amas a tu esposa* [CA-36-23 M]; *Barranquilla es una ciudad que ella es ambiente:sa / que ella es rumbera ¿no? / al fin costeña / ella es ambien:tosa / el clima me gusta* (BA-01-31M)

⁵⁸ Essas orações foram excluídas no trabalho de Caracas (PÉREZ, 2010) e de Cidade do México (LASTRA; MARTÍN, 2015). Eles não dão muita explicação, o trabalho de Pérez (2010) só assinala que no seu *corpus* essas estruturas jamais tiveram um PPS e Lastra e Martín parecem denominá-las inclusivas, infere-se isso pelo exemplo que eles colocam, ou seja, o falante inclui nessa forma plural *ele* e o interlocutor *dele*. Neste trabalho são chamadas de orações exortativas, que são as que não permitem variação, e que no *corpus* efetivamente jamais levam pronome.

Definida a amostra com a inclusão e a exclusão de casos envolvendo a expressão ou não de sujeito, procedemos à identificação dos grupos de fatores ou variáveis independentes da análise.

3.4 Variáveis linguísticas e sociais

Um dos principais objetivos desta tese é investigar e contrastar os resultados desta pesquisa com os de outras variedades já pesquisadas com base em *corpora* dentro da metodologia do PRESEEA e com o de outras variedades monolíngues também já estudadas. Em razão disso, discutimos os grupos de fatores que foram selecionados, que foram todos propostos pelo roteiro metodológico de Bentivoglio, Ortiz López e Silva-Corvalán (2011) e de outros estudos sobre o espanhol caribenho (CAMERON; 1996; CLAES, 2011; OROZCO, 2015). A seleção das variáveis independentes teve por critério, portanto, facilitar a comparabilidade entre os resultados desta pesquisa e os do estudo de variedades monolíngues do espanhol caribenho e os de outras variedades da América e da Espanha.

Esta pesquisa testou 15 variáveis independentes, onze das quais linguísticas e quatro sociais, conforme as hipóteses levantadas na seção anterior. As linguísticas ou internas são as seguintes: 1) pessoa, número e especificidade do sujeito; 2) paralelismo formal; 3) modo; 4) tempo verbal; 5) progressividade; 6) morfologia verbal⁵⁹; 7) ambiguidade; 8) tipo de frase; 9) correferencialidade; 10) turno da fala; 11) classe semântica do verbo. Já as variáveis externas ou sociais são as seguintes: 1) faixa etária; 2) sexo/gênero do falante; 3) grau de escolaridade; 4) subvariedade dialetal.

A descrição com uma breve discussão das variáveis independentes, acompanhada de exemplos está na seção seguinte.

3.4.1 Variáveis linguísticas

Pessoa, número e especificidade do sujeito

A análise prevista para a atuação desse grupo inclui nove fatores, que correspondem às três pessoas do singular e as três do plural: primeira pessoa do singular (*yo*), primeira pessoa do plural (*nosotros /nosotras*), segunda pessoa do singular (*tú [+esp]*, *tú [-esp]*, *usted[+esp]*)

⁵⁹ Segue-se a Lastra e Martín (2015) que preferem utilizar esse nome em vez de “perfectividade” que é utilizado no roteiro.

e *usted*[-esp]), segunda do plural (*ustedes*), terceira pessoa do singular (*él/ ella*) e terceira pessoa do plural (*ellos/ellas*), como exemplificado abaixo. A codificação da especificidade do sujeito de segunda pessoa tem por base os procedimentos de Cameron (1992, 1993; 1996) e Hurtado (2001). Usa-se *tú* [-esp] não específico, e consequentemente *usted* [-esp], quando um falante relata uma experiência pessoal e a generaliza de tal forma que se torna aplicável a qualquer pessoa que tenha um conjunto de circunstâncias semelhantes (CAMERON, 1996)⁶⁰.

Primeira pessoa

- (26) No:/ a (veceh) veces **yo la llevo y me baño** con ella/ (**toy**) **estoy pendiente** a ella ahí (mihmo) mismo al (lao) lado/ [CA-01-11H]
- (27) no tiene: con quien jugar y yo soy la que/ hago (ep papel) el papel de su amiguita/ (toce) entonces ese (eh) es mi tiempo la paso con ella o a veces (salimoh) salimos en familia (todoh) todos a/ a pasear/ (**vamoh**) **vamos** a playa: (**vamoh**) **vamos** a a (vacilación) cine: o (**vamoh**) **vamos** a comer afuera/ y así. [CA-02-11 M]

Segunda pessoa

- (28) *Tú* [+esp]
(x) [...] digo «niña pero tú que estás / caram:ba / **tú tienes** bastante familia en Cartagena / ¿y **tú no has ido** a Cartagena ni un día? / **estás fregada** ¿oíste?» // (BA-01-31M)
- (29) *Tú* [-esp]
639 B: ¿allá no?
640 A: allá no / (simultáneo) allá / allá **no consigues** un trabajo / uno va allá 641 (simultáneo).(BA-62-12H)
- (30) *Usted* [+ esp]
(uish) / **usted viera** a Palenque / si u 950 (palabra cortada) / **si usted fuera a** Palenque / (BA-01-31M)
- (31) *Usted* [- esp]
el diezmo se guarda en aquel tiempo hay que gastarlo comprar lo necesario/ lo que (**utté**) **usted necesite/ tiene que darle** a comer al: al (vacilación) hambriento o sea/ [CA-27-22H]
- (32) *Ustedes*
¿**Ustedes tienen** cuantos años? unos me dijeron/ no/ tengo diecisiete/ tengo dieciocho/ bueno// Yo quiero hacerles una pregunta/ **ustedes dicen** mucho que “el señor me tocó”/ “el señor me tocó”/ [CA-27-22H]

Terceira pessoa

- (33) [...] **se quedó** así / y **entró** / y **vio** todo (principio ético) en (palabra 565 cortada) / y **dijo** // (BA-15-32H)
- (34) Mire (e): ((Ruido)) Bueno yo creo que diferencias ((ruido)) hay muchas// muchas/ de pronto: **se sienten** (máh) más creyentes que nosotros/ **se sienten** más (santoh) santos que nosotros ¿sí?// **Ellos tienen** muchas cosas/ **hacen** una algarabía/ una bulla/ [CA-27-22H]

⁶⁰ Não se decidiu codificar especificidade como um grupo de fatores independente porque, entendido dessa forma, a variável não fornece resultados significativos na maioria dos trabalhos que a levam em conta. Além disso, há muita variação de critério nesses trabalhos que codificam outras pessoas (*ellos*) como não específicas: alguns trabalhos incluem formas verbais em terceira pessoa do plural com sujeito (ausente) não específico que em outros trabalhos são considerados formas impessoais, que jamais veiculam sujeito expresso no espanhol.

Correferencialidade

Os procedimentos aplicados ao estudo da correferência podem receber diferentes graus de detalhamento e especificidade, mediante o uso de medidas exaustivas de distância entre os elementos correferentes (PAREDES DA SILVA, 1988, HURTADO, 2001), ou mediante a limitação a considerações de continuidade da referência ou de mudança (BENTIVOGLIO, 1987a; SILVA-CORVALÁN, 1982). A proposta que se adota aqui assume uma posição intermediária com um grupo constituído por quatro fatores para medir a correferência: não correferência ou correferência distante: caso em que o sujeito ou seu correferente está localizado antes da oração imediatamente anterior; correferência parcial: o correferente mais próximo é um argumento não sujeito; correferência com o sujeito prévio, e não aplicação na situação em que houver mudança de turno.

Não correferência ou com correferência distante

- (35) No:/ a (veceh) veces yo la llevo y me baño con ella/ **(toy) estoy pendiente** a ella ahí (mihmo) mismo al (lao) lado/ (poe) porque **tú (sabe) sabes** que a (veceh) veces el mar con (ehtoh tiempoh) estos tiempos que hace brisa (tá) está (picao) picado [CA-01-11H]

Correferência parcial

- (36) “algún día que yo tenga (hijoh) hijos voy a sacar mis (hijoh) hijos de la duda que **no le pidan** nada al niño (Dioh) Dios/ (pokque) porque **él es** (injuhto) injusto” (ruido)/ [CA-06-13M]

Correferência com o sujeito prévio

- (37) (Da mierda: si **me (ganaría) ganara** la lotería ¿qué **(hacía) haría** con esa plata? **me (haría) hacía** una casa de dos plantas y **me traía** a mi papá y a mi mamá (pa) para acá/ ¿ya?// y **compraría** (otrah doh casah máh) otras dos casa (máh) más y **se (lah) la diera** a (mih hermanoh) mis hermanos/ a cada hermano **le compraría** una casa si **me ganara** una lotería así/ ¿ya?/ (da) mierda (ujualá) ojalá (risas). [CA-01-11H]

Mudança de turno

- (38) 639 B: ¿allá no?
640 A: allá no / (simultáneo) allá / allá **no consigues** un trabajo / uno va allá 641 (simultáneo).(BA-62-12H)

Tempo verbal

A seleção desta variável acompanhou o roteiro de codificação de Bentivoglio, Ortiz López e Silva-Corvalán (2011), que propõem seis fatores: presente, tempos do pretérito⁶¹, imperfeito, condicional, futuro, futuro perifrástico.

Presente

- (39) No:/ a (veceh) veces **yo la llevo y me baño** con ella/ (**toy**) **estoy pendiente** a ella ahí (mihmo) mismo al (lao) lado/ (poe) porque **tú (sabeh) sabes** que a (veceh) veces_el mar con (ehtoh tiempoh) estos tiempos que hace brisa (tá) está (picao) picado y uno tiene que (ehtá) estar pendiente/ a la chama/ y a (veceh) veces (pelaitos) peladitos que también (ehtán) (ehtán) están por ahí/ también (**toy**) **estoy pendiente** (pokque) porque a (veceh) veces hay (pedsonah) que no saben (nadá) nadar (ininteligible)/ ahí **se entretienen** (demasiado) demasiado/ (veceh) a veces la ola se lo lleva/ las maretas se lo lleva/ y **se ahoga// yo/ yo (vacilación) sé (nadá) nadar** aquí donde (**ehtoy**) **estoy/ yo tengo** (ek cudso abietto) el curso abierto en Bocagrande [CA-01-11H]

Pretérito

- (40) [...] el hombre: no aceptó / al día siguiente **fui** a buscarlo al 563 apartamento y cuando **vio** / diez / tres carros así con todo ese poco de 564 hombres / **se quedó** así / y **entró** / y **vio** todo (principio ético) en (palabra 565 cortada) / y **dijo** / «¿oye y tú qué? / ¿te **metiste** a narcotraficante o: / o 566 (vacilación) es que tienes culebras?» y (ininteligible) **dije** / «¿tú no sabes 567 que a mí me andan buscando para ma (palabra cortada) / para: / para 568 (vacilación) secuestrarme?» (BA-15-32H)

Imperfeito

- (41) [...] porque mi mamá no tenía fuerza/ en este sentido/ hermano/ que (noh) nos pedían (libroh) libros/ (cuadernoh) cuadernos/ y si **ella le compraba** a mí (hemman) hermana (treh cuadernoh) tres cuadernos y me **compraba** a mí (treh) tres/ ya **no me podía (comprá) comprar** lo otro (ruido) (pokque) porque **tenía que esperar** a que a ella le pagaran lo poquito/ para cuando/ **compraba** comida y cuando sobraba lo de la comida era que **ella** (noh) nos **iba a hacer** (lah cosah) las cosas (ruido)/ a (veceh) veces **me/ me (vacilación) quedaba yo/** (hemmano) hermano/ (noh) nos pedían (trabajoh manualeh) trabajos manuales y **yo no (loh) los hacía/** (pokque) porque mi mamá no tenía para (dannos) darnos la plata/ y cuando ya (noh) nos **daba** la plata ya era/ ya/ ya (vacilación) ese trabajo había pasado y **tenía que/ que (vacilación) hacer** otro. [CA-06-13M]

Condicional

- (42) [...] a cada hermano **le compraría** una casa si me ganara una lotería así/ [CA-01-11H]

Futuro

- (43) ojalá/ que un hijo mío se dedique a eso/ **yo lo ayudaré / haré** todo (loh effuezzo) los esfuerzos para para (vacilación) que él logre eso. [CA-15-12H]

⁶¹ Essa etiqueta agrupa os tempos pretérito perfeito simples e pretéritos compostos (perfeito e mais-que-perfeito)

Futuro perifrástico

- (44) [...] si (vacilación) me van a matar / **yo** ahora **voy a disfrutar** hasta: / hasta 469 (vacilación) cuando me maten» / pero eso era el: enemigo. (BA-15-32H)

Ambigüidade morfológica

A ambigüidade morfológica compreende duas variantes: formas ambíguas e não ambíguas. As formas ambíguas incluem verbos em primeira e terceira pessoa do singular nos tempos imperfeito e condicional de indicativo e em todos os tempos do subjuntivo, tempos e modos em que as formas verbais são iguais para *yo*, *usted*, *él* e *ella*. Por outro lado, a categoria não ambígua inclui todas as formas verbais restantes, em que há, de fato, a distinção morfológica.

Ambíguo

- (45) [...] dejó de hablar / (e:) / que teníamos que cargarla todos 584 los días / para poder: / para que **puñera** hacer sus necesidades / había que 585 bañarla / (BA-06-31H)

Não ambíguo

- (46) pero (em muchacho) el muchacho (corrección) **el señor se murió**/ y **lo vio** en la parte de (atrás) atrás de/ del: del (vacilación) patio/ con un balde/ [CA-02-11 M]

Classe semântica do verbo⁶²

Como já mencionado, a base de dados ADESSE propõe uma classificação dos verbos em seis macroprocessos que se consideraram os fatores desta variável: i) mentais (47); ii) relacionais (48); iii) materiais (49); iv) dicendi⁶³ (50); v) existenciais (51); e vi) de modulação (52).

- (47) 870 A: sí: / es así / y él está bastante **yo** (vacilación) tampoco **sabía** que él tenía 871 setenta y dos años. (BA-01-31M)
- (48) 870 A: sí: / es así / y él está bastante yo (vacilación) tampoco sabía que **él tenía** 871 setenta y dos años. (BA-01-31M)

⁶² Não seguimos para esta variável o roteiro de Bentivoglio, Ortiz López e Silva-Corvalán (2011), nem de outros trabalhos que a analisam e tem propostas de classificação parecidas (BENTIVOGLIO, 1987a; ENRÍQUEZ, 1984, entre outros).

⁶³ A palavra de ADESSE para esse macroprocesso é “*verbal*”, que foi modificado para evitar ambigüidades, e porque *dicendi* é muito mais conhecido na literatura.

- (49) si en ese tiempo si el hospital Rosario Pumarejo de López con el doctor Hermes Pumarejo tuve mis tres hijos// en el hospital/ **yo trabajaba** con el club Valledupar/ el club Valledupar nos pagaba los médicos/ [VA-53-33M]
- (50) 648 A: porque: / **yo dije** / «no más literatura». (BA-22-33H)
- (51) 775 A: (simultáneo) y llegaba (simultáneo) hasta el Paseo de / Bolívar / e incluso / 776 uno la / la (repetición de palabra) seguía a pie / nosotros / **yo vivía** en el 777 barrio: / Olaya // entre Olaya y Delicias / sesenta y nueve «c» con treinta 778 y cinco / por ahí por la Clínica Bautista /. 780 A: por ahí **vivía yo** / (BA-19-33M)
- (52) 779 A: en clase / y porque quiero que los (pelaos) pelados: / se entusiasmen también 780 porque / bueno **yo: / trato de cumplir** con mi trabajo / lo mejor posible / 781 de no engañar a nadie // pero no: / no (vacilación) es / no es no / no es 782 (vacilación) mi pasión / (BA-22-33H)

Modo verbal

O grupo modo verbal incluiu dois fatores: indicativo (53) e subjuntivo⁶⁴ (54).

- (53) [...] **dejó** de hablar / (e:) / que **teníamos que cargarla** todos 584 los días / para poder: / para que pudiera hacer sus necesidades / había que 585 bañarla / (BA-06-31H)
- (54) [...] dejó de hablar / (e:) / que **teníamos que cargarla** todos 584 los días / para poder: / para que **pudiera** hacer sus necesidades / había que 585 bañarla / (BA-06-31H)

Turno de fala

Este grupo tem dois fatores: mesmo turno (55) e mudança de turno⁶⁵ (56).

- (55) (Da:) mierda: si me (ganaría) ganara la lotería ¿qué (**hacía**) **haría** con esa plata? **me (haría)** **hacía** una casa de dos plantas y **me traía** a mi papá y a mi mamá (pa) para acá/ ¿ya?// [CA-01-11H]
- (56) 639 B: ¿allá no?
640 A: allá no / (simultáneo) allá / allá **no consigues** un trabajo / uno va allá 641 (simultáneo).(BA-62-12H)

⁶⁴ Verbos no modo imperativo, como mencionado acima, foram excluídos por serem categoricamente nulos.

⁶⁵ Segundo Bentivoglio (1987a), ao retomar o turno de fala, o falante tende a tornar expresso o sujeito pronominal; apesar disso, essa autora analisou somente a primeira pessoa, mas vale a pena examinar se esse fator tem efeito também com todos os pronomes.

Paralelismo formal

Orozco (2015) entende que um dado tipo de sujeito promove a ocorrência de outros do mesmo tipo; em função disso, propõem-se aqui quatro fatores para esse grupo: pronome⁶⁶ (57), nulo (58), frase nominal (59) e outro (60), incluindo o sujeito de uma oração que não é pronome pessoal, nem frase nominal.

- (57) [...] que llaman (risa) / de verdad / todo chévere / que le gus (palabra cortada) 949 o sea que pone a la gente a bailar: / (uish) / usted viera a Palenque / si u 950 (palabra cortada) / **si usted fuera a Palenque** / y **usted viera** / **usted se 951 quedaría** / tendría que llevar una filmadora / (BA-01-31M)
- (58) “yo creo que sí/ (demmen) denme (lah) las (herramienta) herramientas y yo veo si (ep posible) es posible”/ hablé con el ingeniero y **me dio** (to:da) todas (lah) las (herramienta) herramientas **me dio** (catapila) catapilas **me dio**: (retroexcavadora) retroexcavadora **me dio** todo/ y **le hice** un canal a (Policappa) Policarpa de tres metros de ancho (pod doh) por dos de (profundidá) profundidad/ [CA-23-23H]
- (59) así como la conoce mi mamá / que en / **mi mamá se ubica** en cualquier pal (palabra cortada) / **yo no** / todavía es la hora y **no ubico** todavía en cualquier (palte) parte / yo conozco setenta y dos / (BA-01-31M)
- (60) **hay** barrios que **no conozco**. (BA-01-31M)

Tipo de frase

De acordo com o roteiro de Bentivoglio, Ortiz López e Silva-Corvalán (2011), esse grupo inclui quatro fatores: afirmação (61), negação (62), pergunta com palavra interrogativa (63) e pergunta sem palavra interrogativa (64):

- (61) **tú tienes** bastante familia en Cartagena / ¿y tú no has ido a Cartagena ni un día? / **estás fregada** ¿oíste?» // (BA-01-31M)
- (62) en toda toda (repetición de palabra) Barranquilla **no la conozco** // hay barrios que **no conozco**. (BA-01-31M)
- (63) (Da:) mierda: si me (ganaría) ganara la lotería ¿qué (**hacía**) **haría** con esa plata? [CA-01-11H]
- (64) / ¿y **tú no has ido** a Cartagena ni un día? / estás fregada ¿oíste?» // (BA-01-31M)

⁶⁶ O pronome pessoal antecedente pode ser o mesmo da oração enfocada ou pode ser algum outro pronome pessoal expresso.

Progressividade

Essa variável, pouco testada na literatura variacionista do PPS, foi proposta aqui para ser analisada conforme o roteiro de Bentivoglio, Ortiz López e Silva-Corvalán (2011) com dois fatores: não progressivo (65) e progressivo (66).

- (65) **yo conozco** setenta y dos / **yo: conozco** el cen:tro / conozco cositas por allá arribota / (BA-01-31M)
- (66) [...] tonces (entonces) en el baño había una <vacilación> una basura/ entonce (entonces) **yo la estaba recogiendo/ yo estaba peleando** con el tío mío [VA-05-11M]

Morfologia verbal

Pouco testado na literatura como um grupo independente, o roteiro do PRESEEA (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011) propõe dois fatores: tempo simples (67) e tempo composto (68).

- (67) 466 A: y yo pues / ese era / mi visión / y **yo siempre hacía** así / pero al verme que 467 me iban a matar / **yo dije** «bueno **yo** para que **tengo** todo este poco de cosas 468 si: / (BA-15-32H)
- (68) «¿cómo que 535 te lo robaron? / ¿y por qué no **me habías dicho?**» porque / (BA-15-32H)

3.4.2 Variáveis sociais

As variáveis sociais incluem faixa etária, sexo/gênero e grau de escolaridade. A codificação do grupo de fatores identificado como faixa etária teve por critério três fatores, representando três diferentes gerações: 20 a 34 anos, 35 a 54 anos e 50+ anos. O grupo de fatores sexo/gênero foi codificado apenas como masculino e feminino⁶⁷, o que lhe dá um caráter mais de sexo biológico do que de gênero, ou características socialmente adquiridas. Por fim, o grupo de fatores grau de escolaridade foi codificado nos dados como baixa escolaridade (até 5 anos de escolaridade), média escolaridade (12 anos), e alta escolaridade (15 ou mais anos de escolaridade). Quanto às subvariedades dialetais, foram codificadas as ocorrências que identificam os *corpora* de Barranquilla, Cartagena e Valledupar.

⁶⁷ De conformidade com Serrano (2011, *apud* SERRANO, 2013b, p. 279), utilizamos a denominação *sexo/gênero* somando-nos à tendência geral na teoria sociolinguística, para a qual o uso desses termos isoladamente não seria recomendável dada a intrínseca relação teórica e hermenêutica existente entre ambos os conceitos.

3.4.3 A variável dependente: posição do pronome sujeito⁶⁸

A análise da posição do pronome sujeito em relação ao verbo exclui, em princípio, os casos de ausência de pronome. A expressão do pronome foi codificada com duas variantes: anteposição (69) e posposição (70).

- (69) 776 uno la / la (repetición de palabra) seguía a pie / nosotros / **yo vivía** en el 777 barrio: / Olaya // entre Olaya y Delicias / sesenta y nueve «c» con treinta 778 y cinco / por ahí por la Clínica Bautista /. 780 A: por ahí vivía yo / (BA-19-33M)
- (70) 776 uno la / la (repetición de palabra) seguía a pie / nosotros / yo vivía en el 777 barrio: / Olaya // entre Olaya y Delicias / sesenta y nueve «c» con treinta 778 y cinco / por ahí por la Clínica Bautista /. 780 A: por ahí **vivía yo** / (BA-19-33M)

A expressão pronominal com suas duas variantes foi identificada como a variável dependente; em função dessa decisão metodológica, a análise aqui proposta se debruça sobre os grupos de fatores linguísticos e sociais anteriormente descritos que determinam a posposição do sujeito.

3.5 Análise quantitativa

A análise quantitativa foi realizada com base no programa da regra variável Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005), Rbrul (JOHNSON, 2009) e Language Variation Suite (SCRIVNER; DÍAZ-CAMPOS, 2016).

O programa Goldvarb, por um lado, fornece evidência quantitativa em três níveis de análise: a relevância estatística do efeito, ou seja, as variáveis que são significativas com base no critério de se enquadrar no nível máximo aceitável de desvio, que é de 0,05 e, portanto, de elevada significância (*significance*); o grau de amplitude do efeito (*range*) (TAGLIAMONTE, 2006), entendido como a diferença entre o menor e o maior peso relativo, que determina a validade dos dados estatísticos, e a hierarquia da influência das variáveis, definida pela ordenação dos pesos relativos no interior de cada grupo de fatores fornecida pelo programa estatístico (TAGLIAMONTE, 2006, p. 235).

⁶⁸ Além disso, a posição do pronome, embora apresentada como uma variável independente em Lastra e Martín (2015), não é, de fato, uma variável independente e não funciona como uma. Em uma versão anterior de seu estudo, esses pesquisadores esclareceram que eles analisam como outras variáveis independentes restringem a posição pronominal em vez de empregar a posição pronominal como uma variável independente em si. Este fator, portanto, não foi incluído como variável independente no presente estudo.

Além disso, o efeito de cada grupo de fatores sobre as variantes da variável dependente é determinado pelo critério de relevância probabilística: quanto maior o grau de influência, mais próximo o peso relativo de .99. Do mesmo modo, por se tratar de uma variável binária, índices acima de .50 mostram efeito positivo do fator sobre a variante, enquanto índices abaixo desse ponto médio mostram desfavorecimento. Os resultados dos pesos relativos do Goldvarb X se apresentam com valores não centrados e com valores centrados⁶⁹. Se houver diferenças significativas entre peso centrado e não centrado em algum grupo de fatores, a análise retomará os pesos centrados.

O programa Rbrul, por outro lado, esboçado por Johnson (2009), pode ser rodado em um contexto *R* (<https://www.r-project.org/>). O Rbrul, diferentemente do Goldvarb X, tem a vantagem de poder executar principalmente uma análise de efeitos mistos. Um modelo de efeitos mistos combina fatores (grupo de fatores em Goldvarb) de efeitos fixos e de efeitos aleatórios. Dentre os primeiros, incluir-se-iam sexo/gênero ou posição do sujeito em relação ao verbo, enquanto os efeitos aleatórios são produzidos pelos itens lexicais específicos ou por falante. Assim, os fixos constituem um número pequeno de níveis (fatores) e são replicáveis, enquanto os aleatórios são extraídos de populações maiores, e, por isso, impossíveis de serem replicáveis, como falantes individuais ou palavras. É desejável um modelo misto porque, além de incluir os efeitos aleatórios, também pode detectar os efeitos fixos que emergem na análise (JOHNSON, 2009).

O Rbrul apresenta os resultados de regressão logística escalonada: as variáveis são os *fatores*, e cada uma das variantes é um *nível* (fator no Goldvarb). Se, por um lado, o Goldvarb fornece resultados dos efeitos mediante os pesos relativos numa escala de 0 a 1, o Rbrul, por outro, toma como primários os *log-odds*, ou seja, os ‘coeficientes ou razões logísticas’, que

⁶⁹ *Goldvarb* permite fazer os cálculos pesando (*weighting*) ou centrando (*centering*) os grupos de fatores. Segundo Martín (no prelo), centrar os fatores tem certas vantagens técnicas, "since weighted factors are the default in GoldVarb, they are widely used and reported" (PAOLILLO 2002, p. 168, *apud* MARTÍN, no prelo). A importância de centrar os fatores, segundo o manual de *Rbrul* (JOHNSON, 2010, *apud* MARTÍN, no prelo, p.4-5), é a seguinte: "In a group with two factors, if factor weights are not centered, the factor with more tokens comes out closer to 0.5, while the one with fewer tokens is shifted towards more extreme values. I have not found a convincing explanation for doing this uncentering, but recognize that many may be used to this behavior. Note that whether or not factor weights are centered, all the differences between factors in a group remain constant (on the log-odds scale). This reflects the fact that what any regression really estimates are the differences between variables' effects, not their absolute values. [...] All in all, it is recommended to use the centered factor weights, or to look at the log-odds coefficients directly. This is mandatory if continuous predictors are in the model, because no meaningful factor weights can be assigned to them. Note that the magnitude of log-odds coefficients can be compared more fruitfully than can factor weights. The premise of logistic regression is that effects are additive on the log-odds scale, so looking at that scale directly, rather than a translation of it into factor weight probabilities, is generally to be preferred".

podem ser positivos ou negativos, podem ser contrastados entre si, ao mesmo tempo em que também fornecem os pesos relativos tradicionais do Goldvarb (JOHNSON, 2009).

Além disso, o Rbrul⁷⁰ permite operar sem abreviar os nomes dos níveis (ou fatores), permite analisar preditores numéricos contínuos ou discretos, facilita o trabalho com as interações entre fatores e é possível ajustar o umbral de significação. E o mais importante, como já foi dito, permite trabalhar com fatores de efeitos fixos e de efeitos mistos. (JOHNSON, 2009)

É importante também assinalar que, enquanto o Goldvarb tende a aumentar os erros de tipo I (a apresentar como significativos efeitos que não o são na realidade), o Rbrul pode produzir mais erros do tipo II, ou seja, não identificar alguns efeitos significativos (JOHNSON, 2009)⁷¹.

Praticamente três razões motivaram a utilização do programa Rbrul neste trabalho: a primeira, corroborar a ausência de erros do tipo I com o Goldvarb, ou seja confirmar a identidade dos grupos de fatores significativos; a segunda, facilitar o amalgamento de fatores (grupos de fatores no Goldvarb), obtendo percentuais e pesos relativos, de modo a predizer, portanto, que fatores interagem; a terceira, se deve ao fato de o Rbrul permitir testar dois preditores de efeitos aleatórios, o indivíduo e o verbo. Mas como o programa não mostrou convergência desses preditores, eles acabaram finalmente não fazendo parte da análise; por isso, só são apresentados no apêndice e retomados em alguns trechos da análise com o fim de projetar uma possível revisão em análises futuras.

Além disso, a fim de explorar a relação entre as variáveis independentes e corroborar as potenciais interações entre eles conforme propõe o Rbrul, procedeu-se a tabulações cruzadas e análise de árvore de inferência condicional. Esses tipos específicos de análise revelam que variáveis interagem para restringir a expressão do sujeito. As tabulações cruzadas e as árvores de inferência condicional ajudam a revelar essas relações complexas no conjunto de dados.

As tabulações cruzadas foram rodadas também com o programa Goldvarb X e as árvores de inferência condicional com o programa *Language Variation Suite* (SCRIVNER; DÍAZ-CAMPOS 2016). Essas árvores geram modelos não paramétricos de regressão, que têm a vantagem de expor os resultados de forma tal a permitir a interpretação visual (SCRIVNER; DÍAZ-CAMPOS 2016, pp. 4 e 10), auxiliando o pesquisador a localizar e visualizar

⁷⁰ A versão do Rbrul utilizada na análise foi a 3.1.3 de janeiro 5 de 2019 (Johnson, 2019).

⁷¹ Um resumo e uma aplicação do Rbrul na análise dos dados da expressão do PPS na Cidade de México pode ser vista em Martín Butragueño (no prelo).

interações de forma mais eficiente em seus dados (TAGLIAMONTE; BAAYEN, 2012). O programa *Language Variation Suite* foi criado juntamente com pacote estatístico R e projetado com uma interface amigável (SCRIVNER; DÍAZ-CAMPOS 2016). Uma melhor explicação de como se visualiza e dever ser interpretada a informação estatística com esse programa aparece no item 4.2.8.

Este capítulo tratou dos procedimentos metodológicos, incluindo as hipóteses de pesquisa, e o universo de investigação, uma preparação para a análise quantitativa e qualitativa a ser discutida no capítulo seguinte.

4 A EXPRESSÃO DO PRONOME PESSOAL SUJEITO NO CARIBE COLOMBIANO

4.0 Introdução

Este capítulo, que se dedica à apresentação e à discussão dos resultados quantitativos obtidos mediante a aplicação dos programas estatísticos Goldvarb X, Rbrul e *Language Variation Suite*, está organizado da seguinte maneira: a seção 4.1 apresenta os resultados gerais, em termos das frequências e percentuais do uso de pronomes pessoais sujeitos, para cada variedade analisada isoladamente e para o espanhol do Caribe colombiano em geral, para compará-los em seguida com a análise de outras variedades previamente estudadas; a seção 4.2 discute as variáveis linguísticas que se mostram significativas nos diferentes modelos de regressão logística, como é o comportamento do dialeto caribenho colombiano quando é comparado com Barranquilla e como interagem as variáveis linguísticas entre si; a subseção 4.3 discute as variáveis sociais selecionadas que também se mostram significativas e como interagem entre si; a seção 4.4 discute, por fim, os percentuais de posposição de pronomes e os grupos de fatores favorecedores da expressão de sujeito nessa posição.

4.1 Distribuição do PPS no espanhol caribenho colombiano e em outras variedades de espanhol

A Tabela 1 apresenta a frequência e as médias percentuais de ausência e de presença do PPS nas variedades analisadas e a do dialeto caribenho colombiano em geral. A retomada das frequências e percentuais tem a função de mostrar as similaridades nesses subdialeto, como foi já justificado em um trabalho parcial prévio sobre as mesmas variedades (PÉREZ; CAMACHO, 2019). A reduzida diferença percentual parece indicar a coerência interna do dialeto, em relação ao fenômeno analisado, resultado que a análise de regressão logística corrobora, uma vez que não selecionou a variável ‘dialeto’ como significativa.

Tabela 1. Distribuição de presença e ausência de PPS

PPS	Barranquilla-BA		Cartagena-CA		Valledupar-VA		Caribe colombiano	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Presença	1100	43.3	1060	42.4	1024	40.1	3184	41.9
Ausência	1439	56.7	1441	57.6	1531	59.9	4411	58.1
Total	2539		2501		2555		7595	

Fonte: elaborada pelo autor

A distribuição das variantes apresentada na Tabela 1 reflete a predominância de sujeitos nulos nas distintas variedades do espanhol Caribenho colombiano⁷², confirmando-se, nesse caso, a preservação do parâmetro *pro drop* do dialeto (58,1%), uma característica paramétrica da língua espanhola em geral.

A percentagem de pronomes sujeito expressos (41,9%) agrupa tanto os da posição pré-verbal como os da pós-verbal; embora a maioria ocorra na posição anteposta, com um baixo percentual de casos pospostos (5,3%) (na seção 4.4 serão discutidas as várias hipóteses sobre a posposição do PPS). Esse índice expresso também permite identificar linguisticamente essa variedade diretamente com o dialeto caribenho (HENRÍQUEZ UREÑA, 1921; LIPSKI, 1994; MONTES, 1995-1996), no qual são comuns os percentuais de presença do PPS acima de 30% (CAMERON, 1993; OTHEGUY; ZENTELLA; LIVERT, 2007; OROZCO; GUY, 2008; CLAES, 2011).

Passemos, agora, à comparação dos índices de presença do PPS em distintas variedades recentemente pesquisadas do espanhol americano e peninsular, tal como se apresenta na Tabela 2. É recomendável alguma reserva na comparação dos resultados, pois pequenas diferenças na seleção dos dados nas diferentes pesquisas podem contribuir para pequenas diferenças de frequência percentual. Por exemplo, orações imperativas foram excluídas deste trabalho e do de Claes (2011), mas não no de Orozco e Guy (2008), Orozco (2015) e Lastra e Martín (2015); as orações na terceira pessoa do plural com sentido impessoal foram excluídas deste trabalho e no de Lastra e Martín (2015), mas não fica claro se também o foi do de Manjón-Cabeza, Pose e Sánchez (2016) e outros trabalhos. Além de disporem de frequências baixas, esses casos podem gerar diferenças nos índices percentuais, ainda que muito sutis e, portanto, pouco significativas.

⁷² A partir de agora só se retomarão os percentuais do dialeto caribenho colombiano para análise e comparação com outros dialetos do espanhol.

Tabela 2. Frequência de pronomes sujeitos expressos em diferentes variedades

Variedade	%
Santo Domingo, República Dominicana (Alfaraz, 2015)	42
Caribe colombiano (este estudo)	41.9
Dominicanos, recém-chegados a Nova Iorque (OTHEGUY; ZENTELLA; LIVERT, 2007)	41
San Juan, Porto Rico (CLAES, 2011)	38.8
Caribe, recém-chegados a Nova Iorque (OTHEGUY; ZENTELLA; LIVERT, 2007)	36.0
Barranquilla, Colômbia (OROZCO; GUY 2008, 72)	35.7
Porto-riquenhos, recém-chegados a Nova Iorque (OTHEGUY; ZENTELLA; LIVERT, 2007)	35
Barranquilla, Colômbia (OROZCO, 2015)	34.2
Cubanos, recém-chegados a Nova Iorque (OTHEGUY; ZENTELLA; LIVERT, 2007)	33
Xalapa, México (OROZCO, 2016)	25
Continente, recém-chegados a Nova Iorque (OTHEGUY; ZENTELLA; LIVERT, 2007)	24.0
Cidade do México (LASTRA; MARTÍN, 2015)	21.7
Granada (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016)	17.5

Fonte: elaborada pelo autor

Os resultados da Tabela 2 também confirmam que o dialeto colombiano estudado, pelo menos em termos percentuais, faz parte da área dialetal caribenha (San Juan, Cubanos, Dominicanos, Porto-riquenhos e caribenhos recém-chegados a Nova Iorque), que se caracteriza pela frequência mais elevada de expressão de pronomes sujeitos do que da dos dialetos das terras altas da América ou continentais (Cidade do México, Xalapa, continentais recém-chegados a Nova Iorque) e peninsulares (Granada). Por outro lado, pode-se observar um acréscimo do percentual do Caribe colombiano (41, 9%) quando comparado com os de Barranquilla (35,7% e 34,2%); as implicações dessas diferenças serão discutidas nas próximas seções. A seguir, a análise dos grupos de fatores linguísticos e sociais que restringem a variável dependente.

4.2 O efeito das variáveis linguísticas no uso do pronome sujeito no Caribe colombiano

Os dados foram processados mediante o uso do programa Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005) e seus efeitos devem ser compreendidos de acordo com o

se discutiu na seção 3.5. A análise quantitativa teve seus resultados corroborados com a análise multivariada no programa Rbrul (JOHNSON, 2009), não apresentando diferenças significativas nem em relação aos grupos de fatores selecionados nem em relação a valores da amplitude (*range*). Diferentemente de alguns trabalhos, que fizeram as rodadas dos grupos de fatores linguísticos e extralinguísticos separadamente (LASTRA; MARTÍN, 2015), em razão da amplitude de grupos de fatores em análise, esta investigação segue a regra geral de outros trabalhos de fazer rodadas com todas as variáveis.

Procedeu-se à execução de várias rodadas, sendo a última delas unicamente com as variáveis selecionadas pelo programa como significativas nas outras rodadas precedentes. As variáveis linguísticas e sociais submetidas às primeiras rodadas são as já discutidas no capítulo 4. Nas diferentes rodadas, tomaram-se decisões metodológicas que vale a pena destacar com antecedência, sem prejuízo de rediscussão na análise de seus respectivos grupos de fatores.

Em relação ao grupo *pessoa, número e especificidade do sujeito*, o fator *Usted [-esp]* foi reagrupado com o fator *Usted [+esp]* por não apresentar diferenças significativas nem em termos de percentual nem em termos de pesos relativos (53,1%; 54,6%); além disso, a frequência de uso de *Usted [-esp]* corresponde ao percentual mais baixo de todas as ocorrências (0,4%; N=32), o que é prejudicial à ortogonalidade dos dados⁷³.

Quanto à variável *tempo verbal*, reagruparam-se num só os fatores *futuro perifrástico, condicional e futuro*. Desses fatores, estar no futuro perifrástico não favorece o uso pronominal (39,6%), estar no condicional tem um efeito neutro (50%) e estar no futuro favorece a expressão do PPS (51,7%); além disso, esses dois últimos fatores têm um número extremamente baixo de ocorrências (0,5%; N=39; 0,4%; N=29). É justamente por essa razão que outros investigadores também já os reagruparam de uma maneira similar (LASTRA; MARTÍN, 2015; LIMERICK, 2018).

Por último, em relação à variável *grau de escolaridade*, reagruparam-se *grau médio e grau baixo* como um único fator, em razão do fato de que o detalhamento não apresenta diferenças quantitativamente significativas (43,1%; 42,3%). Todos esses reagrupamentos não geraram diferenças significativas nem em relação aos demais fatores do mesmo grupo, nem em relação à amplitude, nem em relação ao *Log likelihood*, e muito menos em relação à significância que ficou mais próxima do valor zero.

⁷³ Para Guy (1988 apud GUY; ZILLES, 2007) a ortogonalidade tem a ver com a co-ocorrência livre dos grupos de fatores, de modo que se tem que evitar que um fator de um grupo co-ocorra exclusivamente com apenas um fator de algum outro grupo.

Das quinze variáveis incluídas na análise quantitativa, os dois programas, o Goldvarb X (2005) e o Rbrul (2019) excluíram duas internas e uma externa como não significativas na análise de regressão logística escalonada: *tipo de frase, turno de fala e dialeto*.

Passemos, agora, à análise dos dados da Tabela 3, que mostra o efeito dos grupos de fatores linguísticos e sociais sobre a variável dependente.

Tabela- 3 Hierarquia dos grupos de fatores linguísticos e sociais no Caribe colombiano comparados com os de Barranquilla

Caribe colombiano (este estudo)	Amplitude	Barranquilla (Orozco, 2015)	Amplitude
Pessoa, número e especificidade do sujeito	37	Pessoa e número do sujeito	41
Correferencialidade	31	Correferencialidade	32
Paralelismo formal	19	Tempo, modo e aspecto verbal	27
Modo verbal	13	Classe de verbo	20
Tempo verbal	12	Paralelismo formal	17
Progressividade	10	Tipo de oração ⁷⁴	7
Classe semântica do verbo	10		
Morfologia verbal	7	Pessoa e número do sujeito anterior ⁷⁵	25
Ambiguidade Morfológica	6	Conteúdo léxico do verbo ⁷⁶	9
Sexo/gênero	8	Sexo e faixa etária	10
Faixa etária	4	Condições da conversação	8
Grau de escolaridade	4	NS	

Fonte: elaborada pelo autor

O modo de exposição da Tabela 3 parte da perspectiva da hierarquia dos grupos de fatores a partir da amplitude⁷⁷, índice que, como já vimos, representa a diferença entre o valor

⁷⁴ Não tem correspondência, com o tipo de fraes de nosso trabalho, esse grupo referente à variedade de Barranquilla, por incorporar a codificação de orações independentes, coordenadas e subordinadas, o que não se aplica a esta pesquisa.

⁷⁵ Orozco (2015) realizou análises multivariadas com o Goldvarb X, utilizando duas configurações diferentes, para evitar a superposição e interação de fatores e preservar a ortogonalidade, como sugere Tagliamonte (2006). Portanto, fez uma rodada de todas as variáveis, excetuando *semântica verbal* e *pessoa e número do sujeito anterior*; na outra rodada incluiu estas variáveis e excluiu *conteúdo léxico do verbo* e *correferencialidade*.

⁷⁶ Orozco (2015, p. 25) faz também uma rodada mais detalhada da variável *conteúdo léxico do verbo*, especificando os verbos mais frequentes desses fatores e agrupando o resto em outro fator, obtendo uma amplitude maior (28).

do fator com o peso relativo mais elevado e mais baixo dentro do mesmo grupo, de modo que a amplitude de ponderação dos fatores é uma indicação da força relativa de cada variável (TAGLIAMONTE, 2006).

Tomou-se a decisão metodológica de organizar separadamente as variáveis linguísticas e extralinguísticas. O grupo de fatores internos que tem maior poder preditivo para o uso de PPS é *pessoa, número e especificidade do sujeito* e o de menor poder é *ambigüidade*. Em relação à influência de fatores externos no uso pronominal, eles estão entre os de amplitude mais baixa, particularmente faixa etária e grau de escolaridade, como parece ser uma regra geral do espanhol (BENTIVOGLIO, 1987a; SILVA-CORVALÁN, 2001; OTHEGUY; ZENTELA, 2007, 2012; OROZCO; GUY, 2008, entre outros).

Tagliamonte (2002), no que denomina “sociolinguística comparativa”, assinala quatro condições para a possibilidade de se detectarem contrastes entre duas ou mais variedades. A primeira é a de que o programa estatístico selecione as mesmas variáveis independentes; a segunda é a de que a amplitude seja a mesma para os grupos de fatores, ou seja, a hierarquia de restrições dos grupos seja a mesma; a terceira requer que os fatores tenham a mesma força, isto é, no interior de cada grupo as variedades comparadas disponham dos mesmos pesos relativos; a condição final requer que os fatores exerçam o mesmo efeito no interior do grupo favorecendo ou desfavorecendo a variante analisada da mesma forma e na mesma ordem. É apenas com base na contemplação desse conjunto de condições que é possível deduzir a identidade do sistema que subjaz. Na sequência, analisam-se apenas as duas primeiras condições nas variedades comparadas na Tabela 3, as outras servirão de contraste quando for analisada cada variável independentemente.

Não seria possível fazer uma comparação estreita entre os dois estudos, uma vez que cada qual não levou em conta as mesmas variáveis independentes, e as variáveis similares analisadas não foram codificadas da mesma maneira, o que poderia contribuir para a existência de diferenças na seleção e na hierarquia das amplitudes das variáveis significativas. Pode-se mencionar, como exemplo, que as outras duas variáveis linguísticas excluídas pelos programas não foram analisadas em Barranquilla. Pelo contrário, além das registradas na Tabela 3, analisaram-se em Barranquilla as seguintes variáveis linguísticas e sociais: *estilo do*

⁷⁷ A amplitude é um dos procedimentos (o mais utilizado nos trabalhos sociolinguísticos) para apresentar a hierarquia dos grupos de fatores significativos; é importante assinalar, todavia, que todos eles foram selecionados como significativos no modelo estatístico. Nesse sentido, a amplitude marca a força do efeito do grupo, que será tanto maior quanto mais elevada a amplitude e vice-versa; os valores intermediários entre esses extremos serão considerados como uma força intermediária.

discurso; regularidade do verbo; tempo, modo e aspecto do verbo anterior; pessoa e número do sujeito anterior e distância do sujeito correferencial prévio e nível socioeconômico.

Há, portanto, diferenças marcantes em relação à codificação dos grupos de fatores que interagiram no modelo estatístico, na reagrupação de fatores no interior de um grupo e na configuração de certos grupos; em Barranquilla, por exemplo, agruparam-se *tempo, modo e aspecto* em uma única variável, fatores que se acham separados nos dados desta pesquisa, mas ainda como quatro fatores: *tempo verbal, modo, morfología verbal e progressividade*⁷⁸.

Não obstante essas divergências, os dados da Tabela 3 mostram haver grupos quase estatisticamente equivalentes nos dois trabalhos, selecionados, inclusive, entre os seis de melhor amplitude em ambos os modelos estatísticos, e ocupando posições bastante próximas: *pessoa, número e especificidade do sujeito e correferencialidade* ocupam os dois primeiros lugares respectivamente em termos de amplitude, o que parece constituir uma regra do espanhol geral. Os outros quatro grupos de maior amplitude são *paralelismo formal* (3 vs 5); *modo, tempo e aspecto* (4 vs 3) e *classe semântica do verbo* (5 vs 4), cujas diferenças se justificam na separação dos acidentes verbais em nosso trabalho e a algumas diferenças em alguns fatores da variável *classe semântica do verbo*⁷⁹. Também a variável externa *sexo/gênero* foi selecionada em ambos os trabalhos, e ocupa praticamente a mesma posição em termos de amplitude; ou seja, as seis variáveis com maior capacidade de predição foram selecionadas nos dois trabalhos, indicando que, em relação à primeira condição postulada por Tagliamonte (2002), as duas variedades parecem constituir de fato um dialeto.

Além de comparar a seleção e a ordem das variáveis na hierarquia, outro parâmetro relevante de comparação, aplicado em pesquisas anteriores do PPS se assenta nos valores da

⁷⁸ Embora pareça uma decisão arbitrária a de “reagrupar” os grupos *tempo, modo e aspecto verbal* numa mesma “variável”, representada pela cor verde na Tabela 3, foi essa a única maneira de tornar possível alguma comparação com o trabalho de Orozco (2015), que foi levado em conta por ser o único realizado previamente e com outro *corpus* de uma das subvariedades que fazem parte da variedade aqui analisada. Outras rodadas dos dados recodificados com um novo grupo que reúne *Tempo, modo e aspecto* revelam uma amplitude de 14 e mantêm a quarta posição; mas um fato que pode favorecer um grau maior de amplitude no trabalho de Barranquilla e que leva para a terceira posição esse grupo de fatores é a inclusão de orações imperativas, que com certeza aumentam as formas nulas porque, como afirma o mesmo Orozco (2015), essas formas desfavorecem claramente a expressão pronominal. Além disso, as amalgamas de fatores tem influência nas amplitudes.

⁷⁹ De fato, a divergência dos resultados também não deveria surpreender. O trabalho de Orozco (2015) difere na seleção, ordem e valor das amplitudes de vários grupos de fatores do trabalho prévio de Orozco e Guy (2008), mesmo com iguais informantes, embora parte desses dados teve ajustes metodológicos. Também o trabalho de Limerick (2018), seguindo os mesmos critérios metodológicos e com informantes, ao menos parcialmente, da mesma cidade, só selecionou as duas primeiras variáveis e a última na mesma ordem, mas difere em relação a outras variáveis selecionadas, à ordem delas e à amplitude, quando comparado com o trabalho da Cidade do México (LASTRA; MARTÍN, 2015). Parece que as duas primeiras variáveis seguem uma tendência geral do espanhol enquanto as outras mostram certas divergências.

amplitude para cada variável (CLAES, 2011; OTHEGUY; ZENTELLA, 2012). Por exemplo, Otheguy e Zentella (2012, p.169) descobriram que a amplitude para certas variáveis (*correferencialidade, tipo de oração*) era mais baixas para bilíngues nascidos e/ou criados em Nova Iorque do que para imigrantes recém-chegados. Esses resultados indicam, para os autores, uma redução de força e um enfraquecimento de sensibilidade em relação a bilíngues.

Retomando a Tabela 3 para analisar a amplitude de cada variável, é possível deduzir a existência de semelhanças e diferenças nas duas variedades. Com efeito, os dois primeiros grupos de fatores, além de compartilharem as mesmas posições, também compartilham os valores aproximados de amplitude (37 vs 41; 31 vs 32). Os grupos de fatores *paralelismo formal* e *sexo/gênero*, mesmo não ocupando igual posição no ranqueamento, compartilham o mesmo valor aproximado da amplitude (19 vs 17; 8 vs 10).

Pelo contrário, os grupos *tempo, modo e aspecto verbal* e *classe semântica do verbo* são os que mantêm uma diferença em grau mais elevado em relação ao valor de suas amplitudes ('13' vs 27; 10 vs 20) resultado que deve estar relacionado à diferença na separação dos acidentes verbais em nosso trabalho e aos fatores definidos em cada pesquisa para o grupo *classe semântica do verbo*.

Apesar das diferenças metodológicas, o dialeto do Caribe colombiano e o de Barranquilla mostram mais proximidade do que diferença, como se esperava e como demonstraram as descobertas de um trabalho anterior sobre as principais variáveis linguísticas nas três variedades aqui agrupadas como espanhol caribenho colombiano (Barranquilla, Cartagena e Valledupar) (PÉREZ; CAMACHO, 2019). Também nessa segunda condição proposta por Tagliamonte (2002), verifica-se haver considerável consistência intradialetal como mostram dados próprios e dados de outros trabalhos.

Passemos, agora, à análise dos dados da Tabela 4, que demonstra com dados estatísticos como cada grupo de fatores linguísticos restringe a variável dependente, neste caso a expressão do pronome pessoal sujeito.

Tabela 4. Grupo de fatores internos significativos para a presença do PPS no Caribe colombiano, ordenados pela amplitude, a partir de 7595 dados

Grupo de fatores	%Total	Freq.	%	PR não centrado	PR centrado
Pessoa e especificidade do sujeito					
Tú -esp	1,8	78/135	57,8	.66	.66
Yo	52,2	1901/3962	48,0	.57	.58
Ud ⁸⁰	1,7	70/129	54,3	.56	.57
Tú +esp	3,8	140/287	48,8	.53	.54
Ustedes	0,5	21/40	52,5	.52	.53
Él, Ella	22,1	609/1676	36,3	.45	.46
Ellos, Ellas:	9,0	188/685	27,4	.36	.36
Nosotros	9,0	177/681	26,0	.29	.29
<i>Amplitude</i>				37	
Correferencialidade					
Não correferencialidade	36,2	1544/2751	56,1	.66	.64
Mudança de turno	8,5	331/647	51,2	.59	.57
Correferencialidade parcial	11,2	329/850	38,7	.48	.46
Correferencialidade com o sujeito anterior	44,1	980/3347	29,3	.35	.33
<i>Amplitude</i>				31	
Paralelismo formal					
Pronome	25,2	958/1917	50,0	.62	.62
Outro	21,8	817/1653	49,4	.49	.48
Frase nominal	16,6	541/1257	43,0	.48	.47
Nulo	36,4	868/2768	31,4	.43	.42
<i>Amplitude</i>				19	
Modo verbal					
Indicativo	95,3	3059/7236	42,3	.51	.57
Subjuntivo	4,7	125/359	34,8	.38	.43
<i>Amplitude</i>				13	

⁸⁰ Usted [+esp] (53/97; 54,6%; 1,3 total) e usted [-esp](17/32; 53,1%; 0,4 total)

Grupo de fatores	%Total	Freq.	%	PR não centrado	PR centrado
Tempo verbal					
Imperfeito	21,7	722/1648	43,8	.54	.55
Presente	44,4	1536/3371	45,6	.53	.54
Outros tempos ⁸¹	3,3	108/254	42,5	.46	.47
Pretérito	30,6	818/2322	35,2	.42	.43
<i>Amplitude</i>				12	
Progressividade					
Não progressivo	97,7	3121/7423	42,0	.50	.55
Progressivo	2,3	63/172	36,6	.40	.45
<i>Amplitude</i>				10	
Classe semântica do verbo					
Dicendi	11,9	458/907	50,5	.56	.55
Existencial	4,0	116/302	38,4	.53	.53
Mental	21,8	785/1655	47,4	.51	.51
Relacional	25,5	809/1939	41,7	.50	.50
Material	30,1	829/2289	36,2	.47	.46
Modulação	6,6	187/503	37,2	.46	.46
<i>Amplitude</i>				10	
Morfologia verbal					
Tempo composto	4,6	143/352	40,6	.57	.54
Tempo simples	95,4	3041/7243	42,0	.50	.46
<i>Amplitude</i>				7	
Ambiguidade Morfológica					
Ambígua	16,9	619/1280	48,4	.55	.53
Não ambígua	83,1	2565/6315	40,6	.49	.47
<i>Amplitude</i>				6	
		3184/7595	41,9		
<i>Input .405</i>				<i>Log likelihood = -4619.338</i>	<i>Significância = 0.031</i>
<i>Input .365</i>				<i>Log likelihood = -4619.347</i>	<i>Significância = 0.031</i>

Fonte: elaborada pelo autor

⁸¹ Futuro perifrástico (74/187; 39,6%; 2,5% total); condicional (19/38; 50%; 0,5% total) e futuro (15/29; 51,7%;0,4%)

A primeira coluna apresenta cada um dos fatores testados para saber o alcance da restrição deles em relação ao *valor aplicado* (TAGLIAMONTE, 2006), a presença do pronome sujeito, em contraste com os *valores não aplicados*, que são, neste caso, a ausência do sujeito pronominal. Em alguns grupos, esses fatores já aparecem reagrupados como resultado das diferentes rodadas, que foram feitas com a finalidade de evitar a dispersão dos dados e uma maior eficácia do modelo estatístico (MARTÍN, no prelo).

A segunda coluna mostra o percentual total do fator em relação a total de ocorrências analisadas, o que é importante apresentar porque algumas decisões metodológicas tomadas para o agrupamento de fatores se assentaram em baixos índices percentuais, que podem contribuir para a não ortogonalidade e para as consequências já discutidas acima⁸². Assim, por exemplo, em relação ao grupo *tempo verbal*, os baixos índices percentuais para condicional (0,5%) e futuro (0,4%) sugeriram reagrupá-los como futuro perifrástico, o que contornou um problema de dispersão.

A terceira coluna apresenta a frequência de casos do valor aplicado, expressão do PPS, dividido pelo número de ocorrências do fator em questão, ou seja, 177/681 para o fator *nosotros* no grupo *pessoa, número e especificidade do sujeito* significa que há 177 casos de sujeito expressos das 681 ocorrências de verbos que correspondem a esse tipo de pessoa.

A quarta coluna representa o percentual equivalente à frequência de pronomes expressos, ou seja, 26% de *nosotros* é o mesmo que 177 casos.

A quinta e a sexta colunas mostram os pesos relativos, não centrados e centrados, dos fatores de cada grupo selecionado pelo programa como significativo; estando o peso relativo mais próximo de 1 indica favorecimento de pronome expresso; estando mais próximo de 0 indica desfavorecimento do pronome expresso; estando próximo de 0,5, o peso relativo tem um efeito neutro, isto é, nem favorece nem desfavorece a variante em questão⁸³. Assim, por exemplo, no grupo *correferencialidade*, o fator *não correferencialidade* tem um peso relativo de .66 (não centrado), o que indica que favorece a expressão do sujeito pronominal; o fator *correferencialidade com o sujeito anterior* tem um peso de .35, indicando desfavorecimento da expressão pronominal, e *correferencialidade parcial* tem um peso de .48, indicando um efeito mais próximo do neutro.

⁸² Embora uma distinção mais detalhada possa ser relevante para uma análise qualitativa, e conseqüentemente possa ser discutida com uma precisão relevante, entende Martín (no prelo) que a construção de modelos quantitativos requer certa simplificação. A análise pode, às vezes, ser reducionista, mas acabar sendo mais eficaz para o tratamento quantitativo.

⁸³ Acima de .50 favorecem a variante em análise e abaixo desse valor a desfavorecem. Para este trabalho foi estabelecido a amplitude entre .52 a .59, para baixa significância; de .60 a .69, para média significância; e de .70 a .99, para elevada significância.

O valor da amplitude de um grupo é obtido a partir da diferença do peso relativo mais elevado e mais baixo do grupo, como já mencionado antes. Por exemplo, no grupo *correferencialidade*, o peso mais elevado do grupo é *não correferencialidade* com .66, e o mais baixo é *correferencialidade com o sujeito anterior*, com .35: da diferença deles resulta a amplitude 31, que, por seu valor, insere esse fator no segundo lugar da hierarquia de restrições.

Por fim, na parte baixa da tabela aparecem o *input*, o logaritmo de verossimilhança e a significância, lembrando que os valores de significância igual ou inferior a 0,05 são admitidos como significativos na maioria das pesquisas das ciências sociais.

Na sequência da exposição, discutem-se os resultados de cada grupo de fatores linguístico por vez.

4.2.1 Pessoa, número e especificidade do sujeito

A Tabela 5 mostra os resultados para a variável *pessoa, número e especificidade do PPS* para cada pronome.

Tabela 5. Pessoa, número e especificidade do PPS

Grupo de fatores	%Total	Freq.	%	PR
Pessoa e especificidade do sujeito				
Tú -esp	1,8	78/135	57,8	.66
Yo	52,2	1901/3962	48,0	.57
Ud ⁸⁴	1,7	70/129	54,3	.56
Tú +esp	3,8	140/287	48,8	.53
Ustedes	0,5	21/40	52,5	.52
Él, Ella	22,1	609/1676	36,3	.45
Ellos, Ellas:	9,0	188/685	27,4	.36
Nosotros	9,0	177/681	26,0	.29
Amplitude				37
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood</i> = -4619.338		<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

⁸⁴ Usted [+esp] (53/97; 54,6%; 1,3 total) e usted [-esp](17/32; 53,1%; 0,4 total)

Em relação à amplitude, o grupo de fatores envolvendo pessoa, número e especificidade do sujeito teve o índice respectivo mais elevado em nossos dados (.37). O mesmo se aplica aos resultados obtidos por Otheguy, Zentella e Livert (2007), Otheguy e Zentella (2012), Claes (2011), Orozco (2015), Lastra e Martín (2015), Manjón-Cabeza, Pose e Sánchez, 2016, entre outros. Esses autores agrupam pessoa e número do PPS tanto para as variedades americanas quanto para as peninsulares, e, inclusive, para a situação de contato do espanhol com outra língua.

Essa variável pode apresentar também uma amplitude elevada quando se analisam dois fatores do grupo: a primeira pessoa do singular e a primeira do plural (BENTIVOGLIO, 1987a); parece, todavia, não haver o mesmo efeito, quando se analisam unicamente as duas terceiras pessoas (MARTÍNEZ, 2012), o que pode indicar que são as duas primeiras pessoas que apontam para uma diferença mais elevada na expressão do PPS.

Em geral, os pronomes pessoais singulares são expressos com um índice percentual mais elevado que o dos plurais, resultado convergente com um grande número de trabalhos prévios (BENTIVOGLIO 1987a; FLORES-FERRÁN, 2004a; OTHEGUY; ZENTELLA; LIVERT, 2007; PRADA PÉREZ, 2009, CARVALHO; CHILD, 2011; CLAES 2011; ABREU 2012; ERKER; GUY 2012; OTHEGUY; ZENTELLA, 2012; OROZCO, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016, entre outros). Vejamos os seguintes exemplos da primeira do singular (71) e da primeira do plural (72):

- (71) con (cuakquiera) cualquiera carrera que (elloh) ellos quisieran (cogé) coger/ **yo se la apoyaría** con mucho (guhto) gusto/ mientras que **yo (ehté) esté** vivo/ (pokque) porque si yo/ (pokque) porque si **yo** (vacilación) **me muero** paila/ o mientras **yo (ehté) esté** vivo yo: en (cuakquiera) cualquier cosa que (**loh**) **los pueda (ayudá) ayudar yo/ yo** (vacilación) (**loh**) **los ayudo**. [CA-01-11H]
- (72) y yo soy la que/ hago (ep papel) el papel de su amiguita/ (toce) entonces ese (eh) es mi tiempo la paso con ella o a veces (**salimoh**) **salimos** en familia (todoh) todos a/ a pasear/ (**vamoh**) **vamos** a playa: (vamoh) **vamos** a a (vacilación) cine: o (vamoh) **vamos a comer** afuera/ y así. [CA-02-11 M]

Claes (2011, p. 200) afirma que os resultados de pesquisas anteriores “incita a pensar que o comportamento associado à pessoa e ao número gramatical não constitui uma característica da gramática dos dialetos em análise, mas do diassistema hispânico”⁸⁵. Orozco (2015, p. 29), por sua vez, assinala que, como o efeito condicionador de pessoa e número do

⁸⁵ Claes (2011, p. 200) “incita a pensar que el comportamiento asociado a la persona y el número gramatical no constituye una característica de la gramática de los dialectos bajo análisis, sino del diassistema hispánico”.

PPS afeta praticamente todas as variedades do espanhol, é bem possível estarmos lidando com uma tendência geral, que remete a uma explicação linguística de caráter universal.

Geralmente os trabalhos apresentam resultados de dois fatores do plural, primeira e terceira pessoa, por excluir-se da análise a segunda pessoa do plural *ustedes* ou se a reagrupa com a terceira com base na forma verbal, em razão do baixo número de ocorrências de formas de segunda pessoa do plural; o que não é uma decisão estranha em muitos trabalhos (MARTÍNEZ, 2011; LASTRA; MARTÍN, 2015).

Uma decisão como essa geralmente oculta os resultados dos índices percentuais e pesos relativos para a segunda pessoa do plural, que em toda América Latina é *ustedes*. Então, em nosso trabalho, a única pessoa do plural que favorece o uso pronominal é *ustedes*, ressaltando-se, todavia, que a frequência correspondente é de 0,5% dos dados, o que pode contribuir para a elevação estatística de seus valores.

Esses resultados coincidem com os da pesquisa sobre a variedade de Barranquilla, cujo percentual demonstra favorecimento de PPS com um índice de 38,9% (OROZCO, 2015)⁸⁶, embora na análise probabilística o autor reagrupe todos os fatores do plural, incluindo a segunda pessoa em razão do baixo número de ocorrências. Os elevados índices percentuais de uso do pronome *ustedes* nas diferentes variedades podem justificar-se pela necessidade de desambiguar as pessoas *ustedes* e *ellos/ellas* que dispõem de morfologia verbal idêntica em todos os tempos, o que está plenamente de acordo com o argumento de Orozco (2015) para os resultados da variedade de Barranquilla. O exemplo seguinte (73) ilustra o uso elevado do pronome:

(73) ¿Por qué dice eso?/ Porque es así// **Ustedes siguen** más a un pastor que a Dios/ **ustedes le creen** más a un pastor que a Dios ((ruido)) [CA-27-22H]

Pronomes de terceira e de primeira pessoa do plural são fatores que não favorecem a presença de PPS, inclusive a primeira aponta para um índice mais baixo de expressão, como é comum ocorrer em outras variedades (CAMERON, 1992, 1993; CLAES, 2011; MARTÍNEZ, 2011; LASTRA; MARTÍN, 2015); todavia, nos dados da variedade de Barranquilla, os

⁸⁶ Também na variedade da República Dominicana, o pronome *ustedes* favorece a expressão pronominal com 61,5% (MARTÍNEZ, 2011); todavia, Martínez (2011) exclui esse grupo de fatores da análise de regressão logística em razão de ele dispor de um número muito baixo de ocorrências. Assim, o fato de não coincidirem as decisões metodológicas acaba por gerar diferenças em relação aos pesos relativos.

percentuais estão invertidos, sendo os mais baixos os referentes à terceira do plural (OROZCO, 2015)⁸⁷.

As formas no singular, por sua vez, são as que mais favorecem o uso do pronome, embora mesmo para essas pessoas haja diferenças correspondentes às variedades dialetais. Neste trabalho, as formas da segunda pessoa do singular são as que favorecem, em relação aos índices percentuais, a presença do pronome; em relação ao peso relativo, no entanto, o fator mais influente é o uso de *tú* [-esp], seguido do uso de *yo* e de *usted*; o uso de *tú* [+esp] favorece, com baixa significância, a expressão do pronome, mas que está claramente distanciada da forma não específica.

Por fim, a terceira pessoa do singular desfavorece a expressão da forma pronominal (ver Tabela 5). É possível dar a esses resultados algumas explicações que podem justificar a diferença deles em comparação com os resultados de outros trabalhos, cuja tendência vai na direção da sequência primeira > terceira > segunda do singular, em relação à expressão pronominal (CLAES, 2011; OROZCO, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

A primeira razão tem a ver com o amalgamento dos pronomes da forma verbal da terceira do singular, ou seja, *ellos*, *ellas* e *uno* (OROZCO, 2015; LASTRA; MARTÍN, 2015⁸⁸; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), o que pode aumentar consideravelmente os índices percentuais e pesos relativos para os pronomes de terceira do singular, se levamos em consideração que *uno* tem, em quase todas as variedades, índices mais elevados de uso quando não categóricos como na variedade da Cidade de México (LASTRA; MARTÍN, 2015)⁸⁹. Um trabalho prévio de Pérez e Camacho (2019) em que se codifica o pronome *uno*, corresponde a 7,4% dos dados, e 80% deles são expressos, com pesos relativos acima de .80, o que mostra forte influência na expressão pronominal. Assim, ser amalgamado com a terceira do singular seguramente contribui para que os dados relativos a essa pessoa disponham de percentuais elevados e conseqüentemente pesos relativos significativos.

⁸⁷ Essa inversão das pessoas do plural também ocorre em Granada (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Acreditamos que a principal diferença entre os dados deste trabalho e os relativos às pesquisas das variedades de Barranquilla e de Granada tem a ver com a exclusão das orações impessoais de terceira pessoa do plural, que podem estar presentes nesses trabalhos, gerando assim essa diferença de ordem percentual.

⁸⁸ Estes autores também incluem as formas com *usted* na terceira pessoa do singular por compartilharem a mesma flexão verbal.

⁸⁹ Um exemplo de elevada frequência de *uno* em nossos *corpora* é o seguinte: *No / ya uno (deppué) después que hace una cosa/ que ya uno la tu (palabra cortada) la la: (vacilación) hizo o/ ya uno ya se arrepiente/ de lo que uno hizo/ y le pide (peddón) perdón a (Dio) Dios (pol:o) por lo que uno/ (aggún) algún día si uno tuvo un error/ ya uno no lo (vuebve) vuelve a hacer (má) más/ ya (lak cosas) las cosas son (diferente) diferentes. [CA-10-12M]*

A segunda razão tem a ver com não diferenciação em outros trabalhos dos fatores [+específico] e [-específico] em relação à segunda pessoa, que parece marcar diferenças entre as variedades latino-americanas e as peninsulares (CAMERON, 1996). Por exemplo, na variedade de Barranquilla (2015), há uma razão para que os pronomes de segunda pessoa não sejam percentualmente significativos e para que seu peso relativo seja de baixa significância, mais próximo de neutro, diferentemente de outras variedades caribenhas: é que, além da não separação entre formas específicas e não específicas, ocorre a codificação das formas imperativas⁹⁰, claramente desfavorecedoras da expressão pronominal (OROZCO; GUY, 2008; OROZCO, 2015), e que com certeza vão estar expressas em segunda pessoa do singular como é o mais frequente em espanhol.

Os trabalhos sobre as variedades caribenhas convergem em grande parte no favorecimento que as pessoas do singular exercem sobre a expressão do sujeito pronominal, especialmente as primeiras duas, seguidas da terceira (MORALES, 1984, 2007; CAMERON, 1993; FLOREZ-FERRAN, 2004a; MARTÍNEZ, 2011;). Nossos dados sobre o Caribe colombiano compartilham essa tendência, que, entre outras razões, é possível creditar ao reflexo da hierarquia temática que rege os referentes do discurso (GIVÓN, 1976, *apud* MORALES, 2007), sendo os pronomes pessoais e, particularmente as duas primeiras pessoas, que figuram no topo dessa escala; o fato de que são essas pessoas os interlocutores da comunicação (MORALES, 2007).

Por outro lado, como se sabe, a ambiguidade das formas verbais, particularmente representadas pela primeira e pela terceira pessoa do singular de vários tempos, é ainda mais frequente nas variedades caribenhas, em que o enfraquecimento da pronúncia de /s/ na morfologia verbal da segunda pessoa do singular aumenta potencialmente ainda mais o caráter ambíguo da oração, que seria compensado pela presença da forma pronominal (HOCHBERG, 1986)⁹¹. Outro fator que responde por uma complexidade ainda mais grave do paradigma verbal é o uso de *usted* e *uno*, cujas formas coincidem com as de terceira pessoa do singular.

⁹⁰ Na variedade da Cidade de México (LASTRA; MARTÍN, 2105), as formas imperativas aparecem categoricamente com sujeitos nulos, e a segunda pessoa do singular *tú* não favorece a expressão pronominal, o que parece mostrar uma interação entre o imperativo e a segunda pessoa do singular.

⁹¹ Como já foi mencionado em outros seções deste trabalho, o efeito do enfraquecimento do –s no foi testado neste trabalho, mas também poderia ser um argumento a favor dos elevados percentuais de *tú*.

Especificidade e não especificidade da segunda pessoa do singular

De conformidade com outros trabalhos, que consideram que os falantes das variedades latino-americanas empregam com maior frequência pronomes expressos de segunda pessoa do singular quando se comportam como não específicos (CAMERON, 1993, 1996; HURTADO, 2001), este trabalho assume também essa distinção de pronomes de segunda pessoa de singular [\pm específico].

Embora nossos resultados apresentem índices percentuais e pesos relativos relevantes para esses pronomes não específicos, o uso de *usted* [-esp], não representa um comportamento evidente nos dados, resultado provavelmente da frequência muito reduzida de ocorrências. O uso das formas com *usted* [-esp] favorece a expressão pronominal (.57), mas não apresenta diferenças significativas do uso de *usted* [+esp], que igualmente favorece a expressão da forma pronominal (.56); também não apresentam diferenças significativas em termos de índices percentuais (53,1% vs 54,6%), em razão do que foram amalgamados para constituírem um único fator.

Por sua vez, a presença do pronome *tú* não específico (.66) é mais favorecida que a de *tú* específico (.53), o que demonstra um grau de confiabilidade mais elevado da variável *especificidade* em razão de uma frequência robusta de ocorrências representando os dois fatores. Vejamos os exemplos dos usos não específicos (74) e específicos (75):

- (74) 901 A: por ejemplo / **tú alquilabas** un disfraz / por horas / un: disfraz de tigre / **tú 902 querías disfrazar** de tigre / y **tú ibas** a esa casa (suspiro) / y te / te (repetición 903 de palabra) alquilaban / un disfraz de tigre / y los disfraces debían tener: / 904 un número / una placa. (BA-19-33M)
- (75) alguien me dijo una vez / «nojoda / ¿**tú eres** 634 cristiano / (e:) / por convicción / o **eres** cristiano por tradición?». / (ñerda) 635 mierda ¿cómo así que por tradición?» / «por tradición / porque **vas** todos 636 los domingos / y **vas** todos los miércoles / y **vas** todos los viernes / ahora 637 por convicción es porque hayas / **has creído** (simultáneo) (ininteligible) (BA-06-31H)

Esses resultados confirmam a hipótese que formulamos em função das descobertas da variedade de Porto Rico, em que são as formas de *tú* não específicas (.72) que favorecem a expressão pronominal em comparação com as específicas (.51) (CAMERON, 1993). Outros trabalhos que analisam a especificidade ou não especificidade do pronome de segunda pessoa do singular na variedade de Porto Rico também confirmam essa hipótese corroborada pelos pesos relativos mais elevados para o uso de formas não específicas (.58) que específicas (.52) (FLORES-FERRÁN, 2004a). Também o estudo das variedades colombianas em Miami

mostra valores percentuais mais elevados para as formas de *tú* não específicas (.55) do que para as específicas (.38) (HURTADO, 2005a).

Por outro lado, um estudo na variedade peninsular de Granada (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), parece confirmar a hipótese contrária para as variedades peninsulares, em que as formas verbais de *tú* [-esp] não favorecem a expressão pronominal.

Esses resultados, com dados mais recentes e mais informantes, parecem confirmar a hipótese para as variedades latino-americanas e peninsulares⁹². Cameron (1996) analisou detalhadamente o comportamento do *tú* [+/-esp] em variedades americanas (de San Juan, Buenos Aires e Santiago) e peninsulares (Madri e Sevilha), postulando a existência de um padrão de comportamento inverso em ambas as variedades: as americanas favorecem a expressão do *tú* [-esp] em contextos de específico, enquanto as peninsulares favorecem *tú* [+esp] em contextos não específicos.

Com base nesses resultados, conclui o autor que esse comportamento indica uma reanálise operada pelos dialetos latino-americanos da natureza quantitativa do *tú* [-esp] por analogia com a de *uno*, não específico, ou de *usted* [-esp], que apontam para elevadas frequências de expressão pronominal pessoal em todos os dialetos. Finalmente, a revisão das frequências da expressão pronominal de segunda pessoa *tú* em vários dialetos permite prever como os dialetos ainda não analisados mostrarão a classificação de *tú* [+/- esp].

4.2.2 Correferencialidade

O grupo de fatores correferencialidade vem sendo abordado desde os primeiros trabalhos sobre a expressão do sujeito (SILVA-CORVALÁN, 1982 *apud* SILVA-CORVALÁN, 2001). A hipótese de que a mudança no referente contribui significativamente para a presença do sujeito foi testada com êxito nesses primeiros trabalhos e continua sendo uma das variáveis mais relevantes e influentes em comparação com os diferentes grupos de fatores aplicados ao fenômeno em quase todas as pesquisas.

Enquanto a maioria de trabalhos aplicam três fatores, correferencialidade total, parcial e não correferencialidade, este trabalho adiciona ao grupo de fatores a mudança do turno, por

⁹² O trabalho sobre a variedade da República Dominicana (MARTÍNEZ, 2011) mostra que o pronome com valor específico apresenta um índice percentual mais elevado (88,7) que o do valor não específico (78,1%); como ambos constituem os fatores mais elevados, eles acabaram sendo amalgamados. Esses índices percentuais parecem estar em contradição com os resultados de Cameron (1993, 1996) e com os deste trabalho, mas a autora não apresenta pesos relativos para cada tipo de *tú*.

acreditar-se que é um parâmetro relevante para veicular a expressão do sujeito, independentemente de correferência.

Vejamos na Tabela 6 os resultados desse grupo de fatores nos dados da variedade aqui analisada.

Tabela 6. Correferencialidade⁹³

Grupo de fatores	%Total	Freq.	%	PR
Correferencialidade				
Não correferencialidade	36,2	1544/2751	56,1	.66
Mudança de turno	8,5	331/647	51,2	.59
Correferencialidade parcial	11,2	329/850	38,7	.48
Correferencialidade com o sujeito anterior	44,1	980/3347	29,3	.35
Amplitude				31
		3184/7595	41,9	
<i>Input .405</i>		<i>Log likelihood = -4619.338</i>		<i>Significância = 0.031</i>

Fonte: elaborada pelo autor

Os resultados da Tabela 6 corroboram a relevância teórica desse grupo de fatores como um dos mais significativos estatisticamente, como indica a amplitude 31, de grande robustez estatística, o que dá a essa variável o segundo lugar na hierarquia de relevância, como se aplica a outros trabalhos que a testam (BENTIVOGLIO, 1987a; OROZCO, 2015; LASTRA; MARTÍN, 2015), apontando para o favorecimento do fator *mudança da referência* na expressão do sujeito pronominal. Nos dados aqui analisados, a mudança completa do referente, ou seja, a não correferencialidade, favorece a presença de PPS (.66) em contraste com a continuidade de referência ou *correferencialidade com o sujeito anterior*, que claramente a desfavorece (.35). Vejamos o efeito dos fatores não correferencialidade (76) e correferencialidade com o sujeito anterior (77) nos seguintes exemplos:

- (76) y ahora que/ <vacilación> que no quieren que le calienten la silla en la casa suya pues coja/ y **él me dijo** Mary **tú te hubieras parado** de ahí/ **yo no vuelvo** mas/ de verdad/ pues de ahí seguimos las relaciones/ [VA-30-22M]
- (77) vea/ **yo** un día **salí/ no pedí** permiso/ **salía** era de <(no:che)> (noche)/ y me **cogió mi papá** y **me pego// él** naa (nada) mah (más) **me pegó** doh (dos) vece (veces) a mi/[VA-19-12H]

⁹³ Embora esse grupo de fatores ainda não tenha sido testado na variedade de Barranquilla, temos a expectativa de que seja significativo.

A correferência parcial, ou seja, os casos que contêm um argumento não sujeito na oração precedente, tem um efeito neutro sobre o fenômeno variável investigado. Pode-se deduzir desses resultados que a correferência contribui para desambiguar o referente da oração enfocada.

Os valores dos PRs são muito próximos estatisticamente dos obtidos em pesquisas de outras variedades do espanhol, incluindo Barranquilla, Caracas, Cidade do México, Porto Rico, Oeste de Los Angeles, Nova Iorque, (SILVA-CORVALÁN, 1982, BENTIVOGLIO, 1987a; CAMERON, 1995; FLORES-FERRÁN, 2004a; OTHEGUY; ZENTELLA; LIVERT, 2007; OROZCO; GUY, 2008; CLAES, 2011; OTHEGUY; ZENTELLA, 2012; OROZCO, 2015; LASTRA; MARTÍN, 2015; entre outros), independentemente de testarem dois ou três fatores.

Diferentemente de outros trabalhos similares, um fator aqui incluído nesse grupo foi o de mudança de turno, com base na hipótese de que, ao iniciar um novo turno, principalmente por interrupção do interlocutor, o falante tende a tornar expresso especialmente o pronome de primeira pessoa do singular *yo*, independentemente de todos os demais fatores, porque a interferência induz o falante a restabelecer o referente deduzindo que ele já não esteja mais disponível no estoque mental do interlocutor (BENTIVOGLIO, 1987a). Em sua pesquisa sobre comportamento do pronome *yo* no espanhol de Madri, Davidson (1996) argumenta que o uso de pronomes se deve a finalidades pragmáticas e metalinguísticas, dentre as quais se incluem a negociação de turnos da fala e o envolvimento do falante.

Este trabalho testou esse fator com todos os pronomes, com um resultado significativo para o uso da forma pronominal (.59), como se observa no exemplo seguinte (78):

- (78) 847 B: pero **está joven: tu papá.**
848 A: **él es** joven. (BA-01-31M)

O cruzamento de pessoa com mudança de turno nesse grupo de fatores mostra que a primeira do singular (53%), a segunda pessoa (51%) e particularmente a terceira do singular (58%) favorecem a expressão do pronome nesse contexto; demonstra ainda que, de todas as pessoas, é a terceira do singular que mais favorece a expressão pronominal.

Esses resultados estão de acordo com a análise do uso de pronomes de terceira pessoa na variedade de Caracas, que também selecionou a mudança de turno como um fator significativo para a expressão da terceira pessoa (MARTÍNEZ, 2012); com toda a certeza, é a terceira do singular a que vai interagir com esse fator, porque, como demonstram os dados

desta variedade e os da variedade de Caracas, as formas plurais não favorecem a expressão pronominal.

Os resultados de Cameron (1993, p. 315), por outro lado, apontam para diferenças em relação aos índices percentuais por tratar de variedades dialetais distintas, mas distribuição coincidente em relação aos pesos relativos, ao comparar as variedades de San Juan (mudança na referência, 66%= **.65** vs mesma referência, 35%= **.35**) e de Madri (mudança na referência, 38%= **.66** vs mesma referência, 14%= **.35**). Entende o autor que esses resultados fornecem evidências adicionais de comprovação da influência, sobre o fenômeno, da mudança sistemática de referência, além de padrões regularmente semelhantes nas variedades do espanhol. Esses resultados comprovam ainda que a mudança de referência exerce um efeito consistente na frequência de ocorrência de PPS para pessoa e número, de modo tal a haver certa uniformidade gramatical atuando ao longo de diferentes restrições relacionadas ao fenômeno em análise (CAMERON, 1993).

Pesquisas recentes sobre variedades com índices percentuais diferentes de pronomes expressos permitem reafirmar essas descobertas de Cameron (1993). Por exemplo, na variedade da Cidade de México (mudança da referência 32,7%, **.70** vs mesma referência 16,3%, **.39**), com um índice de expressão pronominal de 21,7% (LASTRA; MARTÍN, 2015); na variedade de Lima (mudança da referência 24,1%, **.64** vs mesma referência 9,4%, **.36**), com um índice de expressão pronominal de 16,8% (CERRÓN-PALOMINO, 2014); na variedade de San Juan (mudança da referência 54,8%, **.68** vs mesma referência 25,9%, **.35**), com um índice de expressão pronominal de 38,8% (CLAES, 2011); e na variedade de Barranquilla (mudança da referência 45,9%, **.67** vs mesma referência 24,3%, **.35**), com um índice de expressão pronominal de 34,2% (OROZCO, 2015)⁹⁴. Dito em outros termos: a probabilidade de expressão do pronome sujeito varia em função de ele ter a mesma referência ou uma referência diferente e essa variação é sistemática e regular num grande número de variedades do espanhol independentemente dos índices específicos de expressão para cada variedade.

4.2.3 Paralelismo formal

A variável *paralelismo formal* é pouco testada nos trabalhos que investigam a expressão do pronome sujeito, se comparada com trabalhos que se dedicam ao levantamento

⁹⁴ Em Orozco e Guy (2008) os pesos relativos **.61** e **.35**, respectivamente, estão muito próximos dos que nossos dados registram para os mesmos fatores.

de aspectos fônicos, especialmente a expressão ou não de /s/ em posição de coda silábica. Mesmo assim já demonstrou ter significância para o fenômeno variável objeto deste trabalho quando testada em pesquisas do uso da primeira pessoa (FLORES-FERRÁN, 2004b; TRAVIS, 2007; PÉREZ, 2010). Dos poucos trabalhos que testam o efeito do fator *paralelismo formal* no uso ou não da expressão pronominal com todos os pronomes pessoais destaca-se o de Orozco (2015), que utiliza três fatores para o grupo, ao qual se adicionou um quarto neste trabalho, que é *outro*. Os resultados da análise da variável em relação ao grupo de fatores *paralelismo formal* estão dispostos na Tabela 7.

Tabela 7. Paralelismo formal

Grupo de fatores	%Total	Freq.	%	PR
Paralelismo formal				
Pronome	25,2	958/1917	50,0	.62
Outro	21,8	817/1653	49,4	.49
Frase nominal	16,6	541/1257	43,0	.48
Nulo	36,4	868/2768	31,4	.43
<i>Amplitude</i>				19
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood</i> = -4619.338		<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

Esse grupo de fatores ocupou a terceira posição na hierarquia de significância na expressão pronominal, com a amplitude 19, similarmente ao que ocorre em outras pesquisas que testam a variável, como a de Pérez (2010) e Orozco (2015), em que a posição desse grupo de fatores se abriga entre as cinco primeiras variáveis com amplitude mais significativa.

Os resultados da Tabela 7 mostram que um pronome sujeito expresso na oração anterior promove a ocorrência de pronome sujeito expresso na oração seguinte conforme indica o peso relativo de .62, de significância média. Orações impessoais e orações com outros tipos de sujeito e orações com sujeitos nominais precedentes têm um efeito neutro sobre a variável, como indicam os pesos relativos de .49 e .48, respectivamente, enquanto a ausência de uma forma pronominal na oração precedente desfavorece a presença pronominal, como mostra o PR de .43, ou seja, nulo favorece nulo. Vejamos o comportamento de uso do PPS quando a oração anterior leva um pronome expresso (79) e quando é nulo (80) nos seguintes exemplos:

- (79) **yo digo** que eso me hizo ser a mi una persona/ <sollozo> lo que **soy hoy**/ es más <(yo::)> (**yo**)/ **hoy me pongo a pensar**/ y **yo digo** bueno/ yo con respecto a esto// **yo que podría decir**/ porqué// <expresivo> **yo digo** que **yo** eso **se lo agradezco** a mi papá </expresivo>// como muchas vece (veces) **yo le digo** a mis hijos (mis hijos)/ y también la forma como se hace porque **yo les digo** a mis niños// <expresivo> estudien </expresivo>// <(e:)>// [VA-18-31M]
- (80) <expresivo> y yo he visto a Dios que me ha sacado adelante con mis embarazos cuando **me he sentido** mal/ cuando **he tenido** un miedo profundo a la muerte/ cuando **me// he sentido** que de pronto **me puedo quedar** <(en:)> (en) <vacilación> en medio de la cirugía// en medio de la cesaria </expresivo>/ [VA-18-31M]

Esses resultados são muito similares aos da variedade de Barranquilla: pronome (.62), outro (.49) e nulo (.45) (OROZCO, 2015); confirmando, na análise desse grupo de fatores, um comportamento de pertencimento a um mesmo dialeto. Em outros termos: um tipo específico de sujeito favorece a ocorrência de outro do mesmo tipo, de modo que sujeitos expressos favorecem sujeitos expressos e nulos favorecem nulos.

Em relação aos sujeitos pronominais expressos, os tipos que favorecem a manifestação positiva são os de sujeitos diferentes, o que apontaria para o fato de ser a mudança de referência que contribui para a expressão; sujeitos de mesma pessoa e número do singular também já começam a favorecer a expressão da forma pronominal (CAMERON; FLORES-FERRÁN, 2004; FLORES-FERRÁN, 2004b; TRAVIS, 2007; PÉREZ, 2010), mas apenas cruzamentos mais refinados serão capazes de revelar mais detalhadamente esse comportamento.

O espanhol do Caribe colombiano, como também Orozco (2015) destaca em relação à variedade de Barranquilla, está fortemente motivado pelo efeito de *paralelismo formal*, um fenômeno que se estende muito além do uso do PPS.

4.2.4 Tempo, modo e aspecto verbal⁹⁵

4.2.4.1 Modo

Em relação ao modo, o roteiro de análise aqui adotado exclui o modo imperativo, no que difere de outros trabalhos que o incluem (OROZCO; GUY, 2008; OROZCO, 2015;

⁹⁵ Como a maioria de trabalhos que analisam a expressão do pronome sujeito juntam os fatores tempo, modo e aspecto, tomou-se a decisão metodológica aqui de fazer o mesmo para facilitar a ordem, mas cada um desses fatores será analisado separadamente em razão de terem sido codificados como variáveis diferentes. Portanto, mesmo que *morfologia verbal* se identifique por dispor de uma amplitude mais baixa que a de outras variáveis, esse fator será analisado nesse grupo.

LASTRA; MARTÍN, 2015). Essa exclusão se justifica especialmente no fato de que os verbos no imperativo, via de regra, identificam-se justamente pela ausência de PPS. Vários estudos adotando a perspectiva variacionista não o levam em conta (CLAES, 2011) ou apenas confirmam a ausência categórica de PPS neles (OROZCO, 2015; LASTRA; MARTÍN, 2015). Os resultados da análise desse grupo de fatores estão contidos na Tabela 8.

Tabela 8. Modo verbal⁹⁶

Grupo de fatores	%Total	Freq.	%	PR
Modo verbal				
Indicativo	95,3	3059/7236	42,3	.51
Subjuntivo	4,7	125/359	34,8	.38
<i>Amplitude</i>				13
		3184/7595	41,9	
Input .405		Log likelihood = -4619.338		Significância = 0.031

Fonte: elaborada pelo autor

Na maioria de trabalhos que analisam tempo, modo e aspecto verbal (TMA) como fatores do mesmo grupo, os procedimentos estatísticos apontam para uma amplitude elevada, ocupando o segundo ou terceiro lugar na hierarquia de grupos de fatores. Diferentemente desses trabalhos, a análise separada desses fatores feita aqui não parece indicar a mesma força estatística, tendo em vista o valor baixo da amplitude, que é de 13, embora a pesquisa sobre a variedade da Cidade do México, que também separa esse grupo de outros acidentes verbais, apresente uma amplitude mais elevada, de 29 (LASTRA; MARTÍN, 2015).

Uma diferença nos procedimentos metodológicos, que poderia estar exercendo influência sobre esses valores, seria o fato que a pesquisa sobre a variedade mexicana incluiu o modo imperativo, cujo sujeito é categoricamente nulo, em razão do que o imperativo foi amalgamado com o subjuntivo. No entanto, as variáveis também não se comportam sempre da mesma maneira nas variedades. Além disso, em outras pesquisas, esse grupo de fatores não foi selecionado como significativo (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016; LIMERICK, 2018).

Embora os resultados obtidos mostrem que estar no modo indicativo tem um efeito neutro na presença de PPS (.51), a principal dedução que se pode extrair desses resultados é a

⁹⁶ Essa variável não foi selecionada pelo GoldVarb X no *corpus* de Barranquilla, o que explica porque nos referimos somente às variedades dessas duas cidades.

de que estar o verbo no subjuntivo é um contexto desfavorecedor da presença do PPS (.38); esse resultado sugere uma possível interação com outros grupos de fatores⁹⁷. Vejamos o efeito do fator subjuntivo (81) no exemplo seguinte:

- (81) Bueno al día que **tenga**:/ plata que **pueda (movilizamme) movilizarme** (pa) para (aggún) algún lugar me movilizo/ pero ahora no/ [CA-10-12M]

Esses resultados são similares aos de Lastra e Martín (2015) para o espanhol de Cidade do México (.52 e .23, respectivamente). Pela forma como reagrupa os fatores na variável TMA, também para a variedade de Barranquilla, nem o subjuntivo nem o imperativo favorecem o uso pronominal (OROZCO, 2015), em relação às formas indicativas, se fossem agrupadas, teriam um efeito neutro, muito próximo de nossos dados.

Concordamos com Lastra e Martín (2015) que os resultados são consistentes com o uso regular do subjuntivo em orações subordinadas e, em consequência, com uma provável correlação de curta distância entre o sujeito e os elementos anteriores. É preciso ressaltar, no entanto, que a grande maioria dos casos tem o verbo no modo indicativo, 95,3%, em comparação com apenas 4,7% de verbos no modo subjuntivo; por isso, é bem provável que as formas indicativas predominem em todas as estruturas.

4.2.4.2 *Tempo*

Como dito acima, numa primeira rodada dos dados, optamos por preservar todos os grupos de fatores previstos pelo roteiro de codificação postulado pelo PRESEEA (BENTIVOGLIO; ORTIZ LÓPEZ; SILVA-CORVALÁN, 2011), mas devido ao grau de dispersão nos dados de pouca frequência, toma-se aqui a decisão metodológica de basear-se no trabalho de Lastra e Martín (2015), com a finalidade de contrastar os resultados desta pesquisa com os desses autores. Os fatores ficaram organizados da seguinte maneira: Presente, Tempos do Pretérito, Imperfeito e outros (agrupados em “outros” os tempos futuro

⁹⁷ Quando se analisam os pesos relativos centrados, um fato se mostra revelador: o indicativo já não tem um efeito neutro sobre a expressão pronominal, mas, apesar de uma baixa significância (.57), é um fato que não pode passar despercebido, enquanto o subjuntivo continua mostrando-se desfavorecedor da expressão pronominal (.43). Esse elemento corrobora a ideia de Johnson (2009) de que se devem tomar em consideração os pesos relativos centrados como fornece o Rbrul (e também Goldvarb X) e não os não centrados fornecidos pelo Goldvarb X e que são os mais utilizados pelos pesquisadores. Para uma melhor compreensão disso, remete-se o leitor à nota 68.

morfológico e perifrástico, condicional⁹⁸). A Tabela 9 mostra os resultados relativos a esse grupo de fatores.

Tabela 9. Tempo verbal

Grupo de fatores	%Total	Freq.	%	PR
Tempo verbal				
Imperfeito	21,7	722/1648	43,8	.54
Presente	44,4	1536/3371	45,6	.53
Outros tempos ⁹⁹	3,3	108/254	42,5	.46
Tempos de Pretérito	30,6	818/2322	35,2	.42
<i>Amplitude</i>				12
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood</i> = -4619.338		<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

O tempo, assim como o modo verbal, também teve uma amplitude baixa (12) e muito próxima ao anterior.

Como parece ser regular e sistemático na maioria das variedades analisadas, os tempos com maior frequência nos *corpora* deste trabalho, presente, pretérito e imperfeito, mostram um comportamento similar ao detectado nessas variedades. Apesar das sutis diferenças de reagrupamento porque passam os dados nos diferentes enfoques, é normal que os tempos presente, pretérito e imperfeito, sejam quase sempre analisados separadamente.

Para facilitar a comparação entre esses resultados e os do trabalho de Barranquilla (OROZCO, 2015), que integra em um grupo de fatores tempo, modo e aspecto, foi feita uma recodificação que mostra os resultados¹⁰⁰ na Tabela 10. Nessa recodificação, também os tempos do pretérito, que aparecem agrupados na Tabela 9, são separados em pretéritos compostos (perfeito e mais-que-perfeito) e pretérito perfeito simples.

⁹⁸ Só as formas do subjuntivo ficaram com seus respectivos tempos, presente e imperfeito.

⁹⁹ Futuro perifrástico (74/187; **39,6%**; 2,5% total); condicional (19/38; **50%**; 0,5% total) e futuro (15/29; **51,7%**; 0,4%)

¹⁰⁰ Estes resultados foram rodados no programa Rbrul (2019).

Tabela-10. Tempo, modo e aspecto verbal

Tempo verbal	logodds	ocorrências	%¹⁰¹	PR	p.value
Imperfeito Indicativo	0.234	1538	0.443	.56	3.19e-12
Presente Indicativo	0.226	3116	0.463	.56	
Condicional y Futuro	0.155	66	0.5	.54	
Formas Perfeitas Indicativo	0.089	330	0.412	.52	
Futuro Perifrástico	-0.161	187	0.396	.46	
Pretérito simples Indicativo	-0.22	1968	0.343	.44	
Subjuntivo	-0.323	350	0.346	.42	
					grand.
model.basics	total.n	df	intercept	input.prob	proportion
	7555	30	-0.498	0.378	0.419
model.fit	deviance	AIC	AICc	Somers.Dxy	R2
	9.183.355	9.243.355	9.243.602	0.43	0.18

Fonte: elaborada pelo autor

Ainda que não estejam todos os tempos amalgamados exatamente como o foram para a análise da variedade de Barranquilla¹⁰², eles têm um comportamento probabilístico muito próximo: imperfeito de indicativo (.59), formas perfeitas de indicativo e condicional (.51), presente (.50), pretérito simples de indicativo (.46) e subjuntivo, futuros e imperativo (.32) (OROZCO, 2015). Também o comportamento dos tempos na variedade de Cidade de México segue o mesmo padrão: imperfeito (.60), presente (.49), pretérito (.44) e outros (.40) (LASTRA; MARTÍN, 2015).

A principal diferença entre os trabalhos comparados tem a ver com o fato de que, nos resultados deste trabalho, o presente de indicativo, mesmo com uma baixa significância, favorece a expressão pronominal (.56) e nos outros tem um efeito neutro. Todos os resultados das pesquisas mostram uma tendência detectável no imperfeito de favorecer a presença do sujeito, no pretérito, de desfavorecê-la e no presente, de manifestar um valor neutro. Esse

¹⁰¹ Os percentuais em Rbrul se expressam numa escala de 0 a 1 correspondente à de 0 a 100

¹⁰² Em quanto ao novo grupo *Tempo, modo e aspecto*, fizeram-se várias rodadas de dados, recodificados e reagrupados de uma forma bastante parecida com os dados de Barranquilla (2015), mas não idêntica, porque o objetivo deste trabalho não é simplesmente comparar seus próprios dados com os de aquele trabalho, e também porque não é possível tomar decisões metodológicas que não ficam claras no trabalho de Barranquilla, nem outras que não permitem a codificação dos dados deste trabalho, como por exemplo, a ausência do modo imperativo.

favorecimento das formas imperfeitas pode estar associado à ambiguidade desse tempo para as formas de primeira, segunda (*tú e usted*) e terceira do singular, mas também fica a pergunta: por que formas em tempos que não permitem essa ambiguidade, como o presente de indicativo, favorecem a presença pronominal? A discussão do item 4.2.6 ajudará a ampliar o efeito da ambiguidade. Vejamos o comportamento do PPS com imperfeito (82) e com pretérito (83) nos exemplos seguintes:

- (82) **yo: jugaba: yo (ettaba) estaba** e:n/ cuando:/ estaba en mi pueblo **practicaba** el: el (vacilación) <basketball>/ porque era u:n equipo del colegio/ [CA-02-11 M]
- (83) al día siguiente **fui** a buscarlo al 563 apartamento y cuando **vio** / diez / tres carros así con todo ese poco de 564 hombres / **se quedó** así / y **entró** / y **vio** todo (principio ético) en (palabra 565 cortada) / y **dijo** / «¿oye y tú qué? / ¿te **metiste** a narcotraficante o: / o 566 (vacilación) es que tienes culebras?» y (ininteligible) **dije** / (BA-15-32H)

Silva-Corvalán (2001) postula que é o tempo, mais especificamente ligado à função que desempenha no discurso, que se correlaciona com a expressão do sujeito. Em razão disso, defende a autora que, em decorrência da função pragmática do tempo no discurso, seria possível esperar uma incidência mais baixa de sujeito exposto com pretérito, mas com um aumento progressivo do pretérito para o presente e do presente para o imperfeito. Essa predição se baseia no pressuposto de que pronomes sujeitos expostos atraem a atenção para o sujeito, reduzindo, conseqüentemente, a atenção prestada ao evento (SILVA-CORVALÁN, 2001). Por outro lado, entende a autora que a comprovação dessa hipótese requer necessariamente levar em consideração os tipos de textos e fazer cruzamentos com as categorias de tempo, uma vez que é bem provável que interajam.

Dois aspectos, todavia, que são importantes ressaltar: poucos trabalhos corroboram até agora a frequência mais elevada de presença do sujeito em alguns tipos de textos (TRAVIS, 2007; LASTRA; MARTÍN, 2015), tendo como resultado adicional que o tipo de texto que se mostra significativo para a expressão do pronome na pesquisa de uma variedade não o é no mesmo grau que na de outra variedade. Além disso, nenhuma tenta examinar a interação possível de tempo e tipo de texto com o fenômeno variável em estudo. Finalmente, é importante ressaltar que, nestes *corpora*, predomina o texto narrativo, sendo nele que aparecem os tempos mais frequentes.

4.2.4.3 Morfologia verbal

Poucos trabalhos realizados sobre os pronomes sujeito analisaram o que Bentivoglio, Ortiz López e Silva-Corvalán (2011) denominam por morfologia verbal, mas nessas pesquisas esse fator não resultou significativo, pelo que não foi incluído na análise de regressão logística (LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Neste trabalho, esse grupo de fatores teve uma atuação com uma baixa amplitude, como se mostra na Tabela 11.

Tabela 11- Morfologia verbal

Grupo de fatores	%Total	Freq.	%	PR
Morfologia verbal				
Tempo composto	4,6	143/352	40,6	.57
Tempo simples	95,4	3041/7243	42,0	.50
<i>Amplitude</i>				7
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood</i> = -4619.338		<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

A amplitude desse fator (7) está entre as mais baixas das variáveis linguísticas analisadas e de todas as demais variáveis que interagem no modelo estatístico, o que indica que tem um baixo efeito sobre a expressão pronominal do sujeito.

Os resultados mostram que os tempos compostos têm um efeito significativo baixo para a expressão pronominal, enquanto os tempos simples mostram um efeito neutro. Vejamos o efeito dos tempos compostos (84) no seguinte exemplo:

- (84) pero **yo he sido** un pobre// que a mi Dios/ me ha favorecido mucho <lapso> (4 s.) todo **me lo ha dado él** <silencio> (2 s.) a pesar que yo soy <silencio> (2 s.) un mal cristiano <silencio> [VA-45-23H]

Para a variedade de Barranquilla (OROZCO, 2015), em que se amalgamaram as formas perfeitas e as condicionais por terem um índice percentual similar, esse grupo de fatores tem um comportamento neutro (.51); já em relação à variedade de San Juan (CLAES, 2011), os tempos perfeitos desfavorecem o uso pronominal (.34), ou seja, as formas compostas não mostram ter efeito similar nas diversas variedades em que se analisa esse fator.

4.2.4.4 Progressividade

Poucos trabalhos sobre o pronome sujeito analisaram a *progressividade* e, como esse grupo não se mostrou significativo, ele não foi incluído na análise de regressão logística (LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Neste trabalho esse grupo de fatores teve um comportamento pouco significativo para os cálculos do Goldvarb X e seus resultados se mostram na Tabela 12.

Tabela 12. Progressividade

Grupo de fatores	%Total	Freq.	%	PR centrado
Progressividade				
Não progressivo	97,7	3121/7423	42,0	.55
Progressivo	2,3	63/172	36,6	.45
Amplitude				10
		3184/7595	41,9	
<i>Input</i> .365		<i>Log likelihood</i> = -4619.347		<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

A amplitude 10, de baixa robustez estatística, demonstra também pouco efeito de *progressividade* na expressão do PPS. Uma das limitações desse grupo de fatores é o reduzido número de ocorrências de construções progressivas (2,3%) em comparação com o de não progressivas (97,7%), como também destacam Lastra e Martín (2015) em sua análise da variedade da Cidade de México. Apesar desses comportamentos, as formas não progressivas tendem a favorecer a expressão pronominal (.55) e as progressivas, a desfavorecê-la (.45).

O que os resultados indicam com maior grau de clareza é que as formas progressivas não favorecem a expressão do sujeito pronominal. Vejamos o uso do PPS com formas progressivas (85) no seguinte exemplo:

(85) pero si no se tienen en cuenta todas estas cosas que **te estoy diciendo** que **te estoy manifestando**/ imposible que la familia se pueda sostener hoy en día/ [VA-34-32M]

Não é possível fazermos comparações entre os dados deste estudo e os de outros sobre o mesmo fenômeno, já que nenhum deles se preocupa com a análise desses fatores. O que é possível ressaltar é um fato descoberto na análise da variável *perífrase* na variedade de Granada (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016): o de não favorecimento nem da

perífrase modal (.47) nem da aspectual (.37); se a esse dado adicionar-se o fato que, nestes dados, estar o verbo no futuro perifrástico não favorece a expressão do pronome sujeito¹⁰³, é possível inferir que as formas progressivas também não a favorecem.

4.2.5 Classe semântica do verbo

A variável *classe semântica do verbo* foi aplicada à análise de diferentes pesquisadores (ENRÍQUEZ, 1984; BENTIVOGLIO 1987a; HURTADO, 2001; 2005b; CERRÓN-PALOMINO, 2014; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), com classificações bastante parecidas. Neste trabalho, tomou-se a decisão de empregar como parâmetro a classificação de processos de ADESSE¹⁰⁴, que propõe seis macroprocessos como se pode observar na Tabela 13.

Tabela 13. Classe semântica

Grupo de fatores	%Total	Freq.	%	PR
Classe semântica do verbo				
Dicendi	11,9	458/907	50,5	.56
Existencial	4,0	116/302	38,4	.53
Mental	21,8	785/1655	47,4	.51
Relacional	25,5	809/1939	41,7	.50
Material	30,1	829/2289	36,2	.47
Modulação	6,6	187/503	37,2	.46
<i>Amplitude</i>				10
		3184/7595	41,9	
<i>Input .405</i>		<i>Log likelihood = -4619.338</i>		<i>Significância = 0.031</i>

Fonte: elaborada pelo autor

A amplitude 10 indica que o grupo de fatores tem um efeito pouco significativo na expressão pronominal, resultado similar ao do trabalho sobre a terceira pessoa na variedade de

¹⁰³ Também não favorece o uso pronominal em Barranquilla (OROZCO, 2015) nem Porto Rico (CLAES, 2011).

¹⁰⁴ Para mais informação, consultar o site do grupo de pesquisa ADESSE: <http://adesse.uvigo.es/>.

Caracas (MARTÍNEZ, 2012), mas diferente do resultado do trabalho sobre a primeira pessoa da variedade de Caracas (PÉREZ, 2010). Esses são os dois trabalhos que até agora também testaram a classificação de ADESSE.

Esse grupo de fatores foi selecionado entre as cinco amplitudes mais robustas em vários estudos que o analisam sob as outras classificações (BENTIVOGLIO 1987a; HURTADO, 2001; 2005b; CERRÓN-PALOMINO, 2014; OROZCO, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), mas também não chegou nem mesmo a ser selecionado como significativo no trabalho sobre a variedade da Cidade de México (LASTRA; MARTÍN, 2015).

Mesmo necessitando de uma análise mais profunda, os trabalhos que se debruçaram sobre a variedade de Caracas parecem indicar uma associação entre semântica verbal e uma pessoa específica para favorecer o uso da expressão de sujeito. O trabalho de Pérez (2010) sobre a primeira pessoa selecionou como mais significativo o fator *tipo dicendi* (.66), o segundo fator significativo para a presença do pronome é *tipo mental* (.52) e o fator que não contribui para a presença do pronome é claramente o *tipo modulação* (.30).

O trabalho de Martínez (2012) sobre a terceira pessoa selecionou como mais significativo o fator *tipo existencial* (.61), o segundo fator mais significativo para a presença do pronome é *tipo mental* (.54), e o fator que não contribui para a presença do pronome é o *tipo modulação* (.43). Em ambas as pesquisas os outros fatores têm um efeito neutro para a expressão pronominal.

Nos dados da variedade aqui analisada, os fatores que favorecem, com uma baixa significância, a presença pronominal são o tipo *dicendi* (.56) e o tipo *existencial* (.53). O fator que claramente não favorece a presença pronominal é o tipo *modulação* (.46). Esses resultados estão *grosso modo* de acordo com os resultados da variedade de Caracas, agrupando as primeiras e terceiras pessoas. Vejamos exemplos da presença do PPS nos tipo *dicendi* (86), tipo existencial (87) e tipo modulação (88):

(86) 468 A: o sea / ante todo / **yo nunca** / **hablo** mal de la competencia / ¿sí? / (BA-27-21M)

(87) 622 A: **ellos viven** ahí. (BA-59-12M)

(88) A: sí siempre / me **ha hecho** bastantes favores. (BA-62-12H)

Retomando então essa questão: como este trabalho inclui todas as pessoas do paradigma verbal, e os trabalhos sobre a variedade de Caracas, só a primeira e a terceira pessoa, os resultados, quando comparados com eles, parecem indicar que a semântica verbal

interage com a pessoa gramatical, favorecendo os verbos *tipo dicendi* a presença de pronomes de primeira pessoa e os verbos *tipo existencial*, a presença de pronomes de terceira pessoa. Os de *tipo modulação*, a ausência de pronome em todas as pessoas. A revisão da literatura não permitiu encontrar qualquer trabalho que tenha procedido ao cruzamento entre semântica verbal e as diferentes pessoas, o que pode auxiliar nas explicações mais aprofundadas da expressão pronominal. Também o fato de se empregarem os mesmos critérios de ADESSE pode propiciar a situação de poder contar com análises comparativas mais confiáveis sobre a semântica do verbo e sua relação com a presença e a ausência pronominal.

Considerando que uma análise mais detalhada dos subprocessos propostos em ADESSE poderia revelar diferenças no interior dos macroprocessos, fez-se uma rodada com esse grupo tendo os fatores discriminados¹⁰⁵ como aparece na Tabela 13a.

¹⁰⁵ Essa rodada dos dados foi feita no programa Rbrul (2019).

Tabela-13a. Classe semântica do verbo ampliada

Classe semântica do verbo							
Macroprocesso	Subprocesso	logodds	ocorrências	%	PR	p.value	
EXISTENCIAL	Vida	0.427	163	0.448	0.60	3.54e-06	
RELACIONAL	Atribuição	0.28	835	0.459	0.57		
DICENDI ¹⁰⁶	Comunicação	0.263	886	0.502	0.56		
MENTAL	Cognição	0.138	786	0.517	0.53		
MENTAL	Sensação	0.0282	383	0.441	0.51		
MENTAL	Percepção	0.0151	466	0.431	0.50		
MODULAÇÃO ¹⁰⁷	Causação	-0.00143	71	0.38	0.5		
MATERIAL	Comportamento	-0.038	209	0.392	0.49		
MATERIAL	Outros fatos	-0.0494	447	0.376	0.48		
MATERIAL	Espaço	-0.119	1317	0.358	0.47		
RELACIONAL	Posse	-0.138	1104	0.386	0.47		
MODULAÇÃO	Disposição	-0.159	61	0.344	0.46		
EXISTENCIAL ¹⁰⁸	Fase-tempo	-0.186	139	0.309	0.45		
MATERIAL	Mudança	-0.21	316	0.339	0.45		
MODULAÇÃO	Verbos de apoio	-0.25047	344	0.349	0.44		
		model.basics	total.n	df	intercept	input.prob	grand.prop ortion
			7527	39	-0.484	0.381	0.418
		model.fit	deviance	AIC	AICc	Somers.Dxy	R2
			9.114.971	9.192.971	9.193.387	0.436	0.185

Fonte: elaborada pelo autor

Os resultados da Tabela 13a mostram que certos macroprocessos são bastante consistentes na contribuição ou não da expressão pronominal. Por exemplo, todos os subprocessos do macroprocesso *dicendi* favorecem o PPS e todos os subprocessos do macroprocesso *material* e do macroprocesso *modulação* desfavorecem a expressão de PPS. Em outros macroprocessos, porém, entrevê-se uma contraposição no comportamento dos

¹⁰⁶ Os verbos *dicendi* todos favorecem a expressão pronominal, os de *emissão de som* (2.201; **4; 0.75**) e também os de *valoração* (0.026; **17; 0.59**)

¹⁰⁷ Os únicos verbos que parecem favorecer a expressão pronominal, mesmo que pouco frequentes, são os de *aceitação* (1.084; **27; 0.70**).

¹⁰⁸ Os verbos de *existência* não favorecem a expressão pronominal ficando no extremo da tabela ainda que pouco frequentes (-1.333; **8; 0.25**).

subprocessos. Por exemplo, o subprocesso *vida*, que faz parte do *existencial*, favorece com uma significância média a expressão pronominal (.60) e, nesse caso, opõe-se aos outros subprocessos dentro de seu macroprocesso. Igual resultado ocorre com os de *atribuição* (.57), que fazem parte do macroprocesso *relacional*, favorecendo a expressão pronominal e opondo-se aos de *posse*. Por fim, os verbos de *cognição* (.53), do macroprocesso *mental*, parecem tender a favorecer a expressão pronominal, mesmo com significância baixa.

São muito diversificadas as classificações semânticas dos verbos nos trabalhos sobre o PPS e sobre outros fenômenos linguísticos do espanhol. É provável que uma classificação mais homogênea permita fazer comparações mais consistentes e, nesse aspecto, os resultados podem contribuir para a classificação do ADESSE.

Ressalte-se ainda que uma análise mais detalhada dos subprocessos começa a revelar que também essas categorias podem disfarçar diferenças no interior delas. Esse procedimento permite postular que é mais apropriado fazer uma análise da frequência verbal e observar como se comportam os verbos mais frequentes, para determinar com maior certeza quais favorecem realmente a expressão pronominal e quais não (ERKER; GUY, 2012). No Apêndice A, a Tabela 29, que mostra os verbos com dez ou mais ocorrências com seus respectivos percentuais, permite observar, nos verbos mais frequentes, claras tendências de alguns favorecerem a expressão do sujeito e de outros a desfavorecerem, análise que ficará para investigações futuras.

4.2.6 Ambiguidade morfológica

O último grupo de fatores linguísticos selecionado pelo programa estatístico como significativo foi o de *ambiguidade morfológica*, cujos resultados estão expostos na Tabela 14.

Tabela 14. Ambiguidade morfológica

Grupo de fatores	%Total	Freq.	%	PR
Ambiguidade Morfológica				
Ambígua	16,9	619/1280	48,4	.55
Não ambígua	83,1	2565/6315	40,6	.49
Amplitude		3184/7595	41,9	6
<i>Input</i> .405			<i>Log likelihood</i> = - 4619.338	<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

A baixa amplitude dessa variável (6) indica que ela não tem um efeito significativo para o uso pronominal como no estudo da terceira pessoa na variedade de Caracas (MARTÍNEZ, 2012), ainda que seja um fator já testado em muitos trabalhos sobre o sujeito, com resultados significativos (SILVA-CORVALÁN, 1982; BENTIVOGLIO, 1987a; HURTADO, 2001; 2005b; PÉREZ, 2010; CERRÓN-PALOMINO, 2014; LASTRA; MARTÍN, 2015; MICHNOWICZ, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Vejamos o comportamento do PPS com formas morfológicas ambíguas (89) no exemplo seguinte:

- (89) **yo se la apoyaría** con mucho (guhto) gusto/ mientras que **yo (ehté) esté** vivo/ (pokque) porque si yo/ (pokque) porque si yo (vacilación) me muero paila/ o mientras **yo (ehté) esté** vivo yo: en (cuakquiera) cualquier cosa que (loh) **los pueda (ayudá) ayudar yo/** yo (vacilación) (loh) los ayudo. [CA-01-11H]

Algumas dessas pesquisas testam, além da ambiguidade morfológica, a ambiguidade textual, ao demonstrar que, quando a ambiguidade também é textual a frequência de uso do sujeito pronominal é muito mais elevada (HURTADO, 2001; 2005b), embora, em outras pesquisas, os dois tipos também favoreçam o uso (BENTIVOGLIO, 1987a; PÉREZ, 2010). Os casos de ambiguidade textual sempre são bastante baixos, como atestam Lastra e Martín (2015), por exemplo, sobre a variedade da Cidade de México. Essa escassez de ocorrências conduz à utilização de unicamente dois fatores em um amplo número de trabalhos.

Em relação às formas morfológicas não ambíguas, vários trabalhos revelam também que podem ter um efeito neutro (MARTÍNEZ, 2012; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). O peso relativo de baixa significância para ambiguidade afetando a expressão pronominal (.55) e as evidências de outros trabalhos que mostram um efeito mais claro desse fator, quando se analisa a ambiguidade morfológica e textual, juntamente, leva a pensar que ele está interagindo com outros fatores, o que será testado no item 4.2.8.

4.2.7 Comparação da hierarquia de fatores significativos com Barranquilla

Outro tipo de comparação com o espanhol de Barranquilla que é importante estabelecer é uma hierarquia dos fatores, que é útil para determinar mais especificamente as semelhanças e diferenças na gramática variável dos falantes (TAGLIAMONTE, 2002). A Tabela 15 apresenta essa comparação.

Tabela 15. Hierarquia de fatores do Caribe colombiano e Barranquilla (OROZCO, 2015)¹⁰⁹

Grupo de fatores	Caribe, Col (presente estudo)		Barranquilla (OROZCO, 2015)	
	%	PR	%	PR
Pessoa e especificidade do sujeito				
Tú -esp	57,8	.66	32,5 ¹¹⁰	.52
Yo	48,0	.57	44,5	.62
Ud	54,3	.56		
Tú +esp	48,8	.53		
Ustedes	52,5	.52	38,9 ¹¹¹	
Él, Ella	36,3	.45	39,6 ¹¹²	.62
Ellos, Ellas:	27,4	.36	9,9	.21 ¹¹³
Nosotros	26,0	.29	12,8	
Correferencialidade				
Não correferencialidade	56,1	.66	45,9	.67
Correferencialidade parcial	38,7	.48	32,3	.50
Correferencialidade com o sujeito anterior	29,3	.35	24,2	.35
Paralelismo formal				
Pronome	50,0	.62	44,0	.62
Frase nominal	43,0	.48	38,0	.49
Nulo	31,4	.43	28,3	.45
Grupo de fatores				
Tempo verbal				
Imperfeito de indicativo	44,3	.56	38,5	.59
Presente de indicativo	46,3	.56	38,0	.50
Condicional y Futuro	50	.54		
Formas perfeitas	0.412	.52	36,5 ¹¹⁴	.51
Pretérito	0.343	.44	29,4	.46
Subjuntivo	0.346	.42	20,7 ¹¹⁵	.32

¹⁰⁹ Somente os grupos de fatores que resultaram significativos em ambos os estudos são retomados. É importante lembrar que, como já foi dito, nos resultados discutidos aqui, alguns grupos de fatores não coincidem nem na configuração dos mesmos nem no número de fatores; além disso, alguns grupos foram utilizados em um trabalho e não em outro, em razão do que serão excluídos. Por fim, os valores dos pesos relativos do *tempo, modo e aspecto verbal* são os obtidos com base em novas rodadas no programa Rbrul (JOHNSON, 2019) tentando buscar uma similaridade deste estudo com o trabalho de Barranquilla (2015); reconfigurando os grupos de fatores *tempo, modo e morfologia verbal*, que estão separados nos resultados gerais, e os que são significativos segundo o programa Goldvarb X.

¹¹⁰ O autor não faz diferenças entre o tú [+/-esp].

¹¹¹ Esse fator apresenta clara concordância percentual com os resultados deste trabalho, mas foi agrupado com as formas plurais na análise de regressão logística pela baixa frequência.

¹¹² Orozco (2015) agrupa nesse fator os pronomes de terceira do singular com o pronome no-específico *uno*.

¹¹³ O valor deste peso relativo (.21) em Barranquilla, agrupa todas as pessoas do plural, mesmo tendo claras diferenças percentuais entre elas.

¹¹⁴ Orozco (2015) amalgama as formas perfeitas compostas de indicativo com o condicional.

Classe semântica do verbo				
Dicendi	50,5	.56	38,0	.57
Mental	47,4	.51	46,0	.52
	41,9		34,2	

Fonte: elaborada pelo autor

Como já foi dito na análise independente de cada grupo de fatores, e como é possível perceber mais claramente nos dados da Tabela 15, existem mais similaridades que diferenças entre o dialeto de Barranquilla, no estudo de Orozco (2015), e o do Caribe colombiano investigado neste trabalho, como esperado, pelo fato de Barranquilla fazer parte do Caribe colombiano. Os resultados deste trabalho confirmam que, em relação ao fenômeno estudado, o dialeto do caribe colombiano parece ser muito coeso em razão dos grupos de fatores significativos e a hierarquia dos mesmos como consequência da amplitude, mas também pelos pesos relativos de cada fator, mesmo em face das diferenças metodológicas entre os dois trabalhos.

As diferenças mais marcadas, além dos grupos de fatores diferentes em ambas as pesquisas e de certos fatores diferentes dentro do mesmo grupo, residem nos fatores do grupo *pessoa e número*, provavelmente em razão dos critérios de codificação e de outras decisões metodológicas. Apenas para dar exemplo, em Barranquilla (OROZCO, 2015), a primeira e a terceira pessoas do singular favorecem com o mesmo peso relativo, .62, a expressão pronominal, embora, embora visto da perspectiva da média percentual, a primeira (44,5%) favoreça mais a expressão do pronome que a terceira (39,6%); é por essa razão que essa variável ocupa a primeira posição como grupo significativo na variedade de Barranquilla. A segunda pessoa aparece nessa variedade na terceira posição das formas singulares, quase com um efeito neutro (.52). Assim, os resultados se diferenciam não só em função da ordem dos fatores do singular, mas também da atuação da terceira pessoa do singular, que não se mostra significativa para a expressão pronominal neste trabalho.

As possíveis explicações para essas diferenças são as seguintes: a) em Barranquilla a terceira do singular *él, ella*, se agrupa com o pronome indefinido *uno*, o que, seguramente, contribui para o aumento do percentual da terceira pessoa e conseqüentemente de seu peso relativo. De fato, uma análise prévia, incluindo também *uno*, comprovou que pronome indefinido é o fator que mais favorece a expressão do pronome sujeito com um peso relativo

¹¹⁵ Orozco (2105) amalgama as formas do subjuntivo e imperativo com as de futuro de indicativo. Também neste trabalho, o futuro perifrástico não favorece a expressão do PPS (39,6%; 44).

muito representativo de .80 nos diferentes subcorpus deste trabalho (PÉREZ; CAMACHO, 2019).

Outros trabalhos sobre as variedades americanas que incluem o pronome *uno* nesse grupo de fatores revelam também que é esse fator que mais favorece a expressão pronominal e de tal maneira que pode ter um efeito categórico (CAMERON, 1992; MARTÍNEZ, 2011; LASTRA; MARTÍN, 2015), motivando em certas situações a agrupá-lo com a terceira pessoa do singular como no trabalho sobre a variedade da Cidade de México (LASTRA; MARTÍN, 2015), onde também a terceira pessoa do singular aparece com o segundo peso relativo mais significativo depois da primeira do singular.

No que diz respeito à segunda pessoa do singular, não só há evidentes diferenças percentuais e de pesos relativos entre a variedade de Barranquilla e a do Caribe colombiano, mas também chega a inverterem-se as posições na hierarquia dos fatores mais significativos com os resultados de Barranquilla apresentando um distanciamento expressivo dos dialetos do Caribe antilhano, do qual ele mesmo faz parte (OROZCO; GUY, 2008; OROZCO, 2015). Uma análise mais detalhada, no entanto, permite revelar as razões das diferenças com os dados deste trabalho e com o dialeto espanhol caribenho.

Em primeiro lugar, Orozco (2015) não apresenta diferenças entre pronomes tú [+/- específico], o que seguramente vai motivar as diferenças quantitativas já demonstradas aqui; em segundo lugar, o trabalho de Barranquilla inclui as orações imperativas, um fator que o mesmo autor nesse trabalho, e num trabalho prévio com os mesmos dados (OROZCO; GUY, 2008), demonstrou ser desfavorecedor da expressão pronominal, estando seguramente a maioria dessas orações imperativas na segunda do singular como é o uso mais frequente no espanhol. Essas duas decisões metodológicas determinam não só as diferenças dos resultados desse trabalho com os do Caribe colombiano, mas também com os dados do dialeto caribenho em geral. São essas diferenças que motivam as diferenças de pesos relativos e naturalmente da ordem de importância desses fatores.

Os outros grupos de fatores que recebem uma codificação quase idêntica, não só ocupam as mesmas posições, mas também têm pesos relativos praticamente idênticos. Basta contrastar grupos de fatores como correferencialidade ou paralelismo formal para comprovar a simetria entre peso e ordem de fatores. Também as variedades mostram muita similaridade em relação aos fatores que menos favorecem a expressão pronominal. Essa similaridade permite afirmar, conforme vários argumentos levantados até aqui, que as variedades caribenhas colombianas se comportam como um mesmo dialeto em relação a grupos e a fatores linguísticos significativos que contribuem para a expressão pronominal do sujeito,

como foi parcialmente demonstrado em um estudo prévio com os dados desta pesquisa (PÉREZ; CAMACHO, 2019).

4.2.8 Interação entre variáveis independentes linguísticas

Uma vez examinados os principais efeitos de cada variável independentemente, discute-se agora como algumas variáveis independentes interagem entre si para influenciar a expressão variável de sujeito, com a clara finalidade de se obter uma visão mais detalhada e diferenciada de como está determinada a escolha de uso pronominal pelos falantes. As interações¹¹⁶ resultaram de cruzamentos rodados pelo programa Goldvarb X (SANKOFF; TAGLIAMONTE; SMITH, 2005) e também de árvores de inferência condicional realizadas com o programa *Language Variation Suite* (SCRIVNER; DÍAZ-CAMPOS 2016).

Nas árvores de inferência condicional, como pode se observar no Gráfico 3, os fatores significativos se ramificam até chegar aos nódulos terminais onde se podem perceber, visual e quantitativamente, os fatores mais e menos influentes. As divisões nas estruturas ramificadas mostradas nas árvores indicam instâncias em que contextos específicos para uma dada variável exibem frequências pronominais significativamente diferentes e como essa variável interage com outras. Por exemplo, o Gráfico 3, em que interagem os grupos correferencialidade e paralelismo formal, mostra como a variável correferência (nódulo 1) se cinde em dois ramos, quando há e quando não há correferencialidade com o sujeito e quando há mudança de turno, de modo a interagirem esses fatores com o grupo paralelismo formal. O grupo paralelismo formal (nódulo 2 e nódulo 5) se ramifica, por sua vez, em nulo/outro para a esquerda e em PPS para a direita. É possível fazer a leitura do gráfico de cima para baixo ou de baixo para cima. Uma das leituras possíveis é a seguinte: quando se cruza o grupo de fatores correferencialidade com o grupo paralelismo formal, é a correferencialidade com o sujeito anterior o fator que favorece a expressão em menor escala, e os fatores que mais favorecem são não correferencialidade e mudança do turno. Quando houver correferencialidade e, simultaneamente, o sujeito anterior for nulo ou não pronominal, será ainda menor o favorecimento de PPS. Pelo contrário, o nódulo 7 mostra os cruzamentos que mais favorecem a expressão do PPS, ou seja, não ser correferente com o sujeito anterior ou ocorrer uma mudança de turno num contexto em que a oração anterior já tenha um PPS.

¹¹⁶ As interações mais significativas são preditas pelo programa Rbrul (2019).

As interações particulares testadas baseiam-se em postulados da literatura sobre PPS, que chamaram a atenção para a necessidade de que elas fossem submetidas a uma pesquisa sistemática (BENTIVOGLIO, 1987a; CAMERON, 1992; HURTADO, 2001; FLORES-FERRÁN, 2004a, OROZCO, 2015); baseiam-se também nos resultados obtidos ao longo desta pesquisa e nos resultados das diferentes variáveis selecionadas pelos programas como significativas.

4.2.8.1 Correferencialidade e paralelismo formal

Com relação ao cruzamento dos grupos de fatores correferencialidade e paralelismo formal, a Tabela 16 e a árvore de inferência condicional do Gráfico 3 permitem ver as interações mais significativas entre certos fatores.

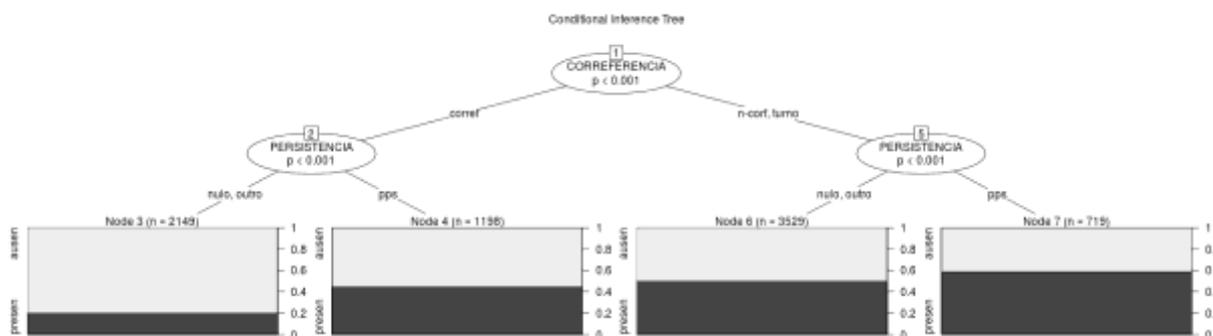
Tabela 16. Cruzamento de correferencialidade¹¹⁷ e paralelismo formal

		PPS	%	Nulo	%	Outro	%	Σ	%
Corref.	Presen	537	<u>45</u>	338	20	105	23	980	29
	Ausen	661	55	1357	80	349	77	2367	71
	Σ	1198		1695		454		3347	
Não Corref.	Presen	377	57	466	53	1030	50	1873	52
	Ausen	280	43	412	<u>47</u>	1036	50	1728	48
	Σ	657		878		2066		3601	
Σ	Presen	914	49	804	31	1135	45	2853	41
	Ausen	941	51	1769	69	1385	55	4095	59
	Σ	1855		2573		2520		6948	

Fonte: elaborada pelo autor

¹¹⁷ Para cruzamento da tabela foram excluídos os dados da mudança de turno e amalgamados os fatores correferencialidade parcial e não correferencialidade em não correferencialidade (também no gráfico) por interessar só a forma para a análise do paralelismo formal (deve-se ler *persistência*).

Gráfico 3. Árvore de inferência condicional para correferencialidade e paralelismo formal



Fonte: elaborada pelo autor

As diferentes rodadas dos programas estatísticos já haviam mostrado a relevância dos dois grupos de fatores aqui enfocados, que mantêm os valores significativos para a expressão do sujeito. Quando interagem, a não correferencialidade e a expressão do PPS na oração precedente, a oração em análise mostra um índice de PPS de 57%. Pelo contrário, a situação em que há correferencialidade com o sujeito da oração anterior é a mais desfavorecedora de PPS da oração em análise (20%); essa é a interação mais significativa. Dito em outros termos, sujeito nulo leva a sujeito nulo em caso de o referente ser o mesmo (80%). De alguma maneira, os dados corroboram o princípio de paralelismo formal de pronome levar a pronome e nulo levar a nulo.

Deve-se ressaltar que a paralelismo formal de pronome sujeito não só ocorre com sujeitos expressos diferentes, mas também com sujeitos que compartilham o mesmo referente (45%), como se o efeito da mudança da referência começasse a estender-se a contextos correferenciais com os mesmos sujeitos, embora análises mais detalhadas sobre pessoa parecem indicar que as que favorecem a paralelismo formal nesses contextos são a primeira e a segunda do singular, justamente as pessoas que mais favorecem a expressão pronominal em pesos relativos.

Também vale a pena ressaltar que, mesmo em situações de mudança de referência, o percentual de sujeitos nulos é relativamente elevado quando o sujeito anterior for nulo (47%), como se formas nulas favorecessem nulas em contextos não correferenciais. Uma vez mais, porém, análises mais detalhadas mostram que o percentual mais elevado de sujeitos nulos nesses contextos ocorre com as pessoas do plural (*nosotros, ellos*), o que demonstra que a primeira e a segunda pessoa do singular se comportam diferentemente das pessoas do plural quando interagem correferencialidade, paralelismo formal e pessoa.

4.2.8.2 Correferencialidade e ambiguidade

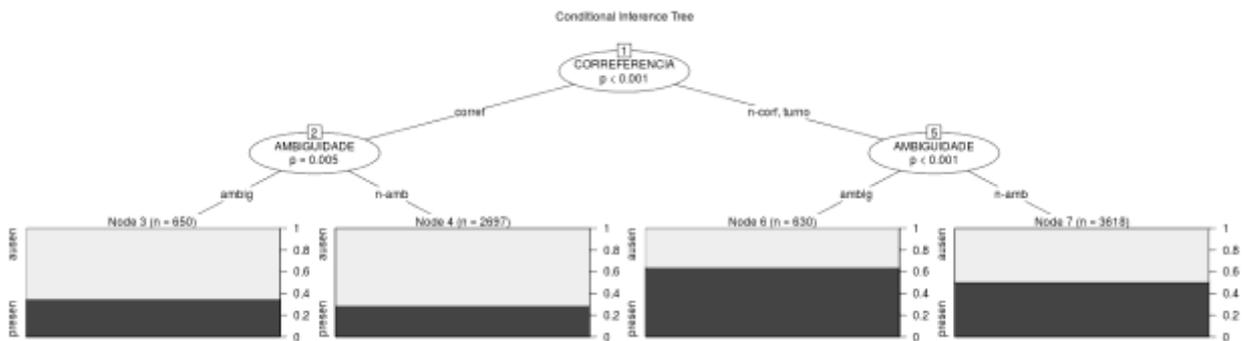
Uma variável que não se mostra fortemente significativa na análise de regressão logística é ambiguidade, cujo efeito na expressão do sujeito pronominal se desvela mais significativamente quando cruzada com a variável correferencialidade como mostram a Tabela 17 e o Gráfico 4.

Tabela 17. Cruzamento de correferencialidade¹¹⁸ e ambiguidade

		Corref.	%	NãoCorf.	%	Corf.Par	%	Σ	%
-Ambíguo	Presença	758	28	1264	54	256	37	2278	40
	Ausência	1939	72	1092	46	434	63	3465	60
	Σ	2697		2356		690		5743	
+Ambíguo	Presença	222	34	280	71	73	46	575	48
	Ausência	428	66	115	29	87	54	630	52
	Σ	650		395		160		1205	
Σ	Presen	980	29	1544	56	329	39	2853	41
	Ausen	2367	71	1207	44	521	61	4095	59
	Σ	3347		2751		850		6948	

Fonte: elaborada pelo autor

Gráfico 4. Árvore de inferência condicional para correferencialidade e ambiguidade



Fonte: elaborada pelo autor

¹¹⁸ Do cruzamento da tabela foram excluídos os dados da mudança de turno.

A Tabela 17 e o Gráfico 4 mostram que a ambiguidade morfológica (+ambíguo) favorece mais a expressão pronominal que as correspondentes formas não ambíguas (-ambíguo), mas, quando não houver correferencialidade, a expressão pronominal das formas morfológicas ambíguas aumenta consideravelmente (71%), situação que contrasta com a presença de pronome para formas ambíguas nas situações de correferencialidade (34%). Esse resultado indica que os fatores não correferencialidade e forma verbal ambígua interagem significativamente mostrando um efeito evidente na expressão do sujeito para resolver qualquer dúvida sobre o referente. O fato de as formas ambíguas não terem um efeito favorecedor nas situações de correferencialidade total implica que o contexto contribui para resolver a dúvida do referente. Esses resultados corroboram os de outras pesquisas que tenderam para valores fortemente significativos para ambiguidade morfológica, nas situações em que a ambiguidade é também textual (BENTIVOGLIO, 1987a; HURTADO, 2001, 2005b; PÉREZ, 2010). Em termos gerais, pode-se dizer que a ambiguidade só favorece significativamente a expressão do PPS em contexto de não correferencialidade, confirmando as conclusões de Cameron (1993) que afirma que esse fator interage claramente com a correferencialidade, sendo dela dependente, e não um grupo de fator independente restringindo a expressão pronominal.

Por outro lado, quando não houver ambiguidade morfológica e também correferencialidade, aumenta consideravelmente o índice de expressão pronominal(54%), o que demonstra o efeito da mudança do referente na expressão pronominal.

4.2.8.3 Pessoa e tempo

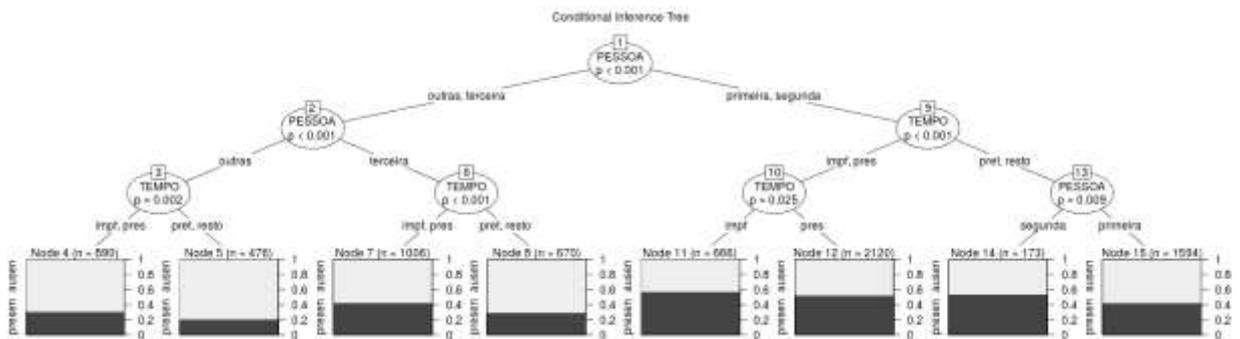
Com relação a como interagem a pessoa e o tempo, pode se observar na Tabela 18 e no Gráfico 5.

Tabela 18. Cruzamento de pessoa e tempo

		1S	%	2 ¹¹⁹	%	3S	%	Plural	%	Σ	%
Presente do Indicativo	Presença	908	51	229	51	219	41	180	31	1536	46
	Ausência	885	49	221	49	320	59	409	69	1835	54
	Σ	1793		450		539		589		3371	
Imperfeito do Indicativo	Presença	377	56	26	70	214	39	105	27	722	44
	Ausência	302	44	11	30	332	61	281	73	926	56
	Σ	679		37		546		386		1648	
Outros	Presença	616	41	54	52	176	30	80	20	926	36
	Ausência	874	59	50	48	415	70	311	80	1650	64
	Σ	1490		104		591		391		2576	
Σ	Presença	1901	48	309	52	609	36	365	27	3184	42
	Ausência	2061	52	282	48	1067	64	1001	73	4411	58
	Σ	3962		591		1676		1366		7595	

Fonte: elaborada pelo autor

Gráfico 5. Árvore de inferência condicional para pessoa e tempo



Fonte: elaborada pelo autor

O cruzamento entre pessoa e tempo acima mostra que o imperfeito favorece a expressão pronominal, especialmente na primeira do singular (56%) e as segundas pessoas (70%), mais que em outros tempos; essa relação mostra haver uma interação com ambiguidade morfológica, levando-se em conta que as pessoas do singular (*yo, usted e él*) e as do plural (*ustedes e ellos*) do imperfeito têm a mesma flexão verbal, o que contribui para a ambiguidade do referente. No entanto, a terceira pessoa do singular e as formas do plural

¹¹⁹ Aqui também foi agrupada a segunda do plural *ustedes*

(*nosotros* e *ellos*) não favorecem a presença do pronome no tempo imperfeito, pois, como mostram os dados gerais deste trabalho, essas pessoas não favorecem a expressão pronominal.

Em outros tempos (52%) e no presente (51%), a segunda pessoa novamente favorece muito discretamente a expressão pronominal, mas análises detalhadas desse grupo de fatores mostra que são *usted* e *ustedes* as pessoas que contribuem para esses elevados percentuais, o que novamente aponta para uma interação com ambigüidade, uma vez que essas pessoas compartilham a mesma flexão verbal das terceiras pessoas do singular e plural, respectivamente, em todos os tempos.

4.2.8.3 Correferencialidade e Pessoa

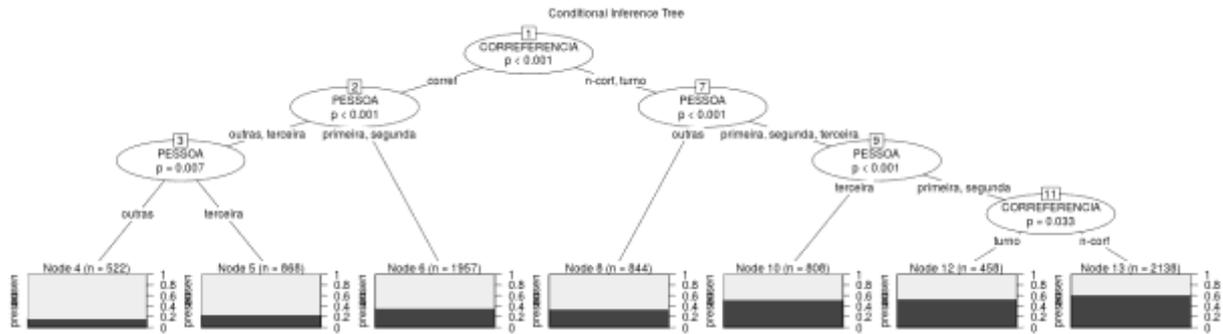
Por fim, vejamos o comportamento da variável pessoa em cruzamento com correferencialidade como se apresenta na Tabela 19 e no Gráfico 6.

Tabela 19. Cruzamento de correferencialidade e pessoa

		1S	%	2 ¹²⁰	%	3S	%	Plural	%	Σ	%
Correferência	Presença	634	36	63	33	198	23	85	16	980	29
	Ausência	1133	64	127	67	670	77	437	84	2367	71
	Σ	1767		190		868		522		3347	
Não Correferência	Presença	871	63	193	66	261	56	219	36	1544	56
	Ausência	506	37	101	34	203	44	397	64	1207	44
	Σ	1377		294		464		616		2751	
Correferência Parcial	Presença	173	43	32	48	89	37	35	24	329	39
	Ausência	228	57	34	52	149	63	110	76	521	61
	Σ	401		66		238		145		850	
Σ	Presença	1678	47	288	52	548	35	339	26	2853	41
	Ausência	1867	53	262	48	1022	65	944	74	4095	59
	Σ	3545		550		1570		1283		6948	

Fonte: elaborada pelo autor

¹²⁰ Aqui também foi agrupada a segunda do plural *ustedes*.

Gráfico 6. Árvore de inferência condicional para correferencialidade¹²¹ e pessoa

Fonte: elaborada pelo autor

Com exceção da primeira e da terceira do plural, todas as demais pessoas favorecem a expressão do sujeito em situação de não correferencialidade, mesmo a terceira pessoa do singular (56%) que não favorece a expressão pronominal nos resultados gerais nem em interação com outros fatores. Nenhuma das pessoas favorece a expressão pronominal quando houver correferência com o sujeito da oração precedente, embora a primeira do singular (36%) e a segunda pessoa (33%) tendam a ter seus índices aumentados como se dá na interação com outros fatores. Esses resultados de favorecimento e não favorecimento da expressão pronominal sugerem a necessidade de tratamento separado para a análise do efeito desses grupos e do efeito de cada pessoa para ser possível detectar o efeito mais evidente de cada qual na expressão do pronome. Por ora, as interações das variáveis auxiliam no entendimento do comportamento da expressão pronominal; além disso, a análise de grupos de fatores separados auxilia a entender melhor como um grupo de fatores tem interações multiplex, como se dá com correferencialidade.

Passemos, agora, à análise, na seção seguinte, dos grupos de fatores externos ou sociais.

4.3 O efeito das variáveis sociais no uso do pronome sujeito no Caribe colombiano

A significância limitada das variáveis sociais na expressão do PPS é uma observação constante na literatura linguística na variedade de língua espanhola (SILVA-CORVALÁN, 2001; OROZCO; GUY 2008, OROZCO, 2015; LASTRA; MARTÍN, 2015). Nesta pesquisa, só aplicamos quatro variáveis externas básicas para examinar o efeito delas na ausência ou presença de PPS: *grau de escolaridade*, *faixa etária*, *sexo/gênero* e *variedade dialetal*. Todas essas variáveis são pré-estratificadas no âmbito do próprio PRESEEA e elas se têm mostrado

¹²¹ Do cruzamento da tabela foram excluídos os dados da mudança de turno.

significativas em inúmeros trabalhos variacionistas sobre diversas variáveis dependentes no contexto hispânico (LASTRA; MARTÍN, 2015) e, também, nas investigações sobre os *corpora* utilizados nesta pesquisa (RODRÍGUEZ CADENA, 2009a, 2009b, 2010, 2012).

Como mostra a Tabela 20, o grupo de fatores *variedade dialetal* foi a única variável externa não selecionada como significativa, o que é um indicador relevante de que as variedades analisadas se comportam como um mesmo dialeto.

Tabela 20. Grupo de fatores externos significativos para a presença do PPS no Caribe colombiano, ordenados pela amplitude, a partir de 7595 dados.

Grupo de fatores	%Total	Freq.	%	PR não centrado	PR centrado
Sexo/gênero					
Mulheres	50,8	1726/3859	44,7	.54	.54
Homens	49,2	1458/3736	39,0	.46	.46
<i>Amplitude</i>					8
Grau de escolaridade					
Baixo e Médio ¹²²	66,1	2143/5019	42,7	.51	.52
Alto	33,9	1041/2576	40,4	.47	.48
<i>Amplitude</i>					4
Faixa etária					
55 e mais anos	33,4	1115/2537	43,9	.52	.52
35 a 54 anos	33,1	1067/2516	42,4	.50	.50
20 a 34 anos	33,5	1002/2542	39,4	.48	.48
<i>Amplitude</i>					4
		3184/7595	41,9		
<i>Input</i> .405				<i>Log likelihood</i> = -4619.338	<i>Significância</i> = 0.031
<i>Input</i> .365				<i>Log likelihood</i> = -4619.347	<i>Significância</i> = 0.031

Fonte: elaborada pelo autor

Na sequência, discute-se cada grupo de fatores externos.

¹²² Médio (.51; 1118/2594; 43,1%; 34,2% total); Baixo (.52; 1025/2425; 42,3%; 31,9% total).

4.3.1 Sexo/gênero

Sexo/gênero foi a variável externa com o grau maior de força estatística como se pode observar no valor da amplitude, 8, superior das demais variáveis sociais e as variáveis *morfologia verbal e ambiguidade morfológica*. Os resultados estão expostos na Tabela 21.

Tabela 21- Sexo/gênero

Grupo de fatores	%Total	Freq.	%	PR
Sexo/gênero				
Mulheres	50,8	1726/3859	44,7	.54
Homens	49,2	1458/3736	39,0	.46
Amplitude				8
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood = -4619.338</i>		<i>Significância = 0.031</i>

Fonte: elaborada pelo autor

Como mostra a Tabela 21, ser mulher é um fator que favorece discretamente, é verdade, a presença do PPS e ser homem desfavorece, também discretamente, a presença do PPS. Os resultados para Barranquilla, de Orozco (2015), são muito semelhantes aos desta pesquisa. Em testes preliminares sobre a variável *sexo*, ele descobriu que as mulheres favorecem sutilmente a expressão do PPS com um peso relativo de .53 e os homens desfavorecem sutilmente a presença do PPS com um peso relativo de .46. Descobertas preliminares do mesmo autor revelam também uma interação entre faixa etária e sexo.

Para obter uma visão adicional das forças sociais que restringem a expressão do PPS, Orozco (2015) explorou os efeitos combinados da faixa etária e de sexo como uma restrição complexa, controlando cuidadosamente a distribuição das frequências e pesos relativos. Desse procedimento, emergiram algumas diferenças interessantes. Os resultados mostram que, enquanto mulheres nascidas antes de 1960 favorecem a presença de PPS com um peso estatístico de .55, os homens nascidos após 1960 desfavorecem a presença de PPS com um peso relativo de .45. Por outro lado, o comportamento verbal de mulheres nascidas depois de 1960 e de homens nascidos antes dessa data tem um efeito neutro sobre a variável com pesos relativos de .50 e .49, respectivamente. Na variedade de Barranquilla, mulheres idosas se diferenciam claramente do resto da comunidade pelo uso significativo de pronomes expressos e se opõem diretamente aos homens mais jovens, cujo uso desfavorece a expressão

pronominal. Os resultados de ambas as pesquisas mostram haver um comportamento geral bastante homogêneo no tocante a essa variável no litoral do Caribe colombiano.

Levando-se em conta, agora, tão somente os índices percentuais, é possível perceber que outros estudos corroboram o favorecimento no comportamento feminino de presença do PPS. Na variedade de Cidade do México (LASTRA; MARTÍN, 2015), por exemplo, as mulheres usam PPS ligeiramente com um índice que se situa acima do atingido pelos homens (23,6% vs 19,8%). Há convergência de resultados também nos dados gerais da variedade de Nova Iorque (30% vs 25%) e nos dos continentais recém-chegados a NY (29% vs. 22%) (OTHEGUY; ZENTELLA 2012).

Com base nesses resultados, é importante discutir o papel da mulher no uso da expressão pronominal do sujeito. Labov (2001) analisa o papel das mulheres na variação e na mudança linguística para reformular o que denominou “paradoxo do gênero” (Gender Paradox), que está relacionado com o prestígio e a inovação.

Quando se trata de mudanças com consciência social, de cima (*changes from above*), segundo Labov (2001), as mulheres utilizam mais as formas de prestígio do que os homens; mas quando se trata de mudanças sem consciência social, de baixo (*changes from below*), são as mulheres que usam mais frequentemente as formas inovadoras, o seja, apresentam comportamento inovador e menos conformista (*nonconforming*) com as normas que não são abertamente prescritas.

Nesse sentido, os resultados deste trabalho, mesmo com as reservas motivadas pelos índices de frequência, parecem indicar que as mulheres estariam liderando a mudança nesse sentido, o que encontra uma explicação quando se considera que a expressão do pronome, mesmo tendo o uso recomendado em certos contextos na tradição gramatical, não é uma variante estigmatizada na ampla comunidade de fala hispânica. Segundo Labov (2001), uma análise mais adequada das mudanças lideradas pelas mulheres deve relacionar a variável gênero com o status social, variável que não foi possível analisar nesta pesquisa e que poderia contribuir melhor para compreender os dados à luz dos postulados de Labov.

Sustentados nos resultados de Orozco (2015) e de outros pesquisadores, que fornecem evidência de interação entre variáveis sociais quer em estudos sobre o PPS quer em outros fenômenos analisados, é preciso testar nos dados desta pesquisa esse tipo de interação para saber também se se aplica a eles. Não é uma previsão improvável, considerando-se que as outras variáveis externas, selecionadas pelo programa estatístico Goldvarb X como significativas, faixa etária e grau de escolaridade, têm alguns fatores que não favorecem a

expressão pronominal, o que pode ser um indício de que estejam em interação com sexo/gênero.

4.3.2 Grau de escolaridade

O Grau de escolaridade é a segunda variável externa selecionada pelo programa estatístico como significativa, embora sua baixa amplitude de 4 como efeito sobre o uso pronominal. A Tabela 22 mostra os resultados, mas com o amalgamento dos fatores médio e baixo, por não apresentarem significativas diferenças nem percentual nem probabilisticamente.

Tabela 22- Grau de escolaridade

Grupo de fatores	%Total	Freq.	%	PR
Grau de escolaridade				
Baixo e Médio ¹²³	66,1	2143/5019	42,7	.51
Alto	33,9	1041/2576	40,4	.47
Amplitude				4
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood</i> = -4619.338 <i>Significância</i> = 0.031		

Fonte: elaborada pelo autor

Não se registra nenhum trabalho sobre uma variedade monolíngue que revele influência do grau de escolaridade na presença ou ausência do pronome sujeito. Os dados deste trabalho também não revelam uma influência significativa em qualquer sentido, só uma influência moderada do uso manifestado pelos falantes com elevado grau de escolaridade com tendência para a ausência do pronome, como se o contato com a variedade padrão, muito mais forte na formação superior estivesse gerando certo controle sobre o uso. Só os dados da variedade da Cidade de México trazem índices percentuais para grau de escolaridade, mesmo que não tenha sido selecionado como significativo: baixo (24.2%, 161/666), médio (20.9%, 157/752) e alto (20.1%, 125/622). Esses resultados sugerem um reagrupamento diferente do executado neste trabalho, de modo que é o baixo grau separadamente que mais contribui para a presença da forma pronominal. Mesmo assim, esse reagrupamento também não deu motivos para que essa variável fosse selecionada como significativa (LASTRA; MARTÍN, 2015).

¹²³ Médio (.51; 1118/2594; 43,1%; 34,2% total); Baixo (.52; 1025/2425; 42,3%; 31,9% total).

Esse grupo de fatores, mesmo uma amplitude baixa, parece interagir com outros grupos de fatores sociais. O estudo da variedade de Granada já demonstrou que faixa etária e grau de escolaridade interagem na análise detalhada do fator *especificidade*: os mais jovens com grau superior de escolaridade são os que mais favorecem a ausência pronominal nas formas não específicas, o que se diferencia claramente do comportamento de jovens dos outros dos níveis com um comportamento similar (MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016).

4.3.3 Faixa etária

Por fim, a *faixa etária* é o outro grupo de fatores selecionado pelo programa Golvarb X como significativo para interagir nos cálculos, apesar de estar a amplitude 4 muito distante das obtidas em outras pesquisas que também selecionaram essa variável como significativa (OROZCO; GUY, 2008; CLAES, 2011; LASTRA; MARTÍN, 2015,); os dados deste trabalho, no entanto, dispõem de três ou mais vezes do número de ocorrências. A Tabela 23 mostra os resultados para esta variável.

Tabela 23- Faixa etária

Grupo de fatores	%Total	Freq.	%	PR
Faixa etária				
55 e mais anos	33,4	1115/2537	43,9	.52
35 a 54 anos	33,1	1067/2516	42,4	.50
20 a 34 anos	33,5	1002/2542	39,4	.48
Amplitude				4
		3184/7595	41,9	
<i>Input</i> .405		<i>Log likelihood = -4619.338</i>		<i>Significância = 0.031</i>

Fonte: elaborada pelo autor

Esses dados confirmam, com moderação, os obtidos em algumas investigações (HURTADO, 2001; OROZCO; GUY, 2008; CLAES, 2011; LASTRA; MARTÍN, 2015, MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016). Todos esses trabalhos demonstram que gerações mais idosas favorecem a presença do sujeito e as mais jovens não a favorecem, com exceção do trabalho sobre a variedade de Porto Rico: esse estudo mostra que, se, por um lado, os jovens não favorecem a presença pronominal, são os informantes de faixas etárias intermediárias que mais a favorecem (CLAES, 2011).

O estudo de Orozco e Guy (2008) sobre Barranquilla constatou que os adolescentes¹²⁴ desfavorecem o uso de sujeitos explícitos, mas que vão aumentando conforme aumenta a idade dos falantes. Esse resultado leva os autores a afirmar que se as inovações linguísticas são promovidas pelos falantes mais jovens. Parece, então, que os falantes adolescentes estariam atuando como inovadores de uma possível mudança em progresso, sendo que a frequência mais baixa de PPS estaria sendo promovida pelos adolescentes. Todavia, os autores reconhecem que esse comportamento poderia ser reflexo do maior acesso a educação dos mais jovens quando comparados com as gerações anteriores.

Vale a pena ressaltar esses argumentos sobre o efeito da idade em Barranquilla, porque a gramática normativa do espanhol recomenda não exprimir o pronome sujeito na maioria dos casos, exceto por questões de ambiguidade, contraste e ênfase, sendo, portanto, a norma prescrita do espanhol predominantemente de sujeitos nulos. Nesse caso seriam as frequências mais altas que constituiriam um uso inovador, o que pode ser sustentado pelos resultados de diferentes pesquisas sobre variedades do espanhol que demonstram ser o uso de PPS relativamente baixo em comparação com a frequência mais baixa dos adolescentes de Barranquilla. A orientação do comportamento dos jovens estaria seguindo para a direção regressiva de uma forma mais conservadora e não inovadora. Como os mesmos autores assinalam, esse comportamento poderia ser o resultado de uma interação desse grupo de fatores com o grau de escolaridade o que mais poderia revelar as razões desse comportamento, que não foi aprofundada por eles.

De fato o trabalho de Orozco (2015) sobre a mesma variedade, com os mesmos dados, mas com alguns ajustes metodológicos, não seleciona idade como um fator significativo, mas, postula que essa variável interage com o sexo/gênero¹²⁵, com resultados um pouco diferentes aos de Orozco e Guy (2008), que continuam questionando sobre como esse grupo de fatores determina o uso pronominal nessa variedade. Nesse trabalho de 2015, Orozco mantém outro argumento que Orozco e Guy (2008) tinham fornecido para explicar o fato dos adolescentes de Barranquilla empregarem uma forma mais ‘inovadora’ ou de baixo uso pronominal, indo na contramão do que acontece no espanhol caribenho, que poderia sofrer influência do chamado espanhol *cachaco* ou continental. Considera-se que esse argumento carece ainda de muita sustentação por duas razões:

¹²⁴ Essa faixa etária não corresponde à que se examina neste trabalho e com muitos outros que analisam o uso do PPS em variedades monolíngues.

¹²⁵ Ao proceder ao cruzamento entre faixa etária e sexo, Orozco (2015) constata que são as mulheres mais idosas que favorecem a expressão pronominal e os homens mais jovens, a ausência. Os percentuais e os pesos relativos não indicam uma mudança em curso.

- (i) até agora nenhum trabalho faz um estudo comparativo que mostre a influência dessa outra variedade colombiana na variedade caribenha nacional, nem no que respeita a outros fenômenos linguísticos estudados nem especificamente sobre a expressão pronominal, além da carência de trabalhos sobre o PPS em variedades “cachaças”¹²⁶ monolíngues;
- (ii) os estudos sobre PPS nas variedades do caribe antilhano também não mostram muita significância das variáveis sociais; particularmente a variável faixa etária, quando é significativa, não tem um claro comportamento de mudança em progresso (CLAES, 2011)¹²⁷.

Concorda-se mais com Orozco (2015) que são necessários estudos mais aprofundados sobre os efeitos das faixas etárias e de outras variáveis sociais, mas acredita-se, com base no amplo número de trabalhos sobre o PPS e nos resultados deste trabalho, que as categorias sociais amplas que frequentemente foram utilizadas não permitem ver a influência desses fatores sociais na expressão pronominal e que se requerem fazer análises mais detalhadas para saber como é que o fenômeno interage com fatores de tipo social e contextual. Os percentuais dos dados do grupo de fatores faixa etária deste trabalho e o comportamento dele em outros trabalhos, mesmo tendo sido selecionado como significativo, não permite inferir que aponta para uma mudança em progresso, o que se tornaria mais evidente quando se fizer um cruzamento desta variável, com o sexo/gênero, por exemplo.

Com base nos resultados de Cidade de México, também Lastra e Martín (2015), sugerem discussões mais detalhadas dos cruzamentos entre faixa etária e sexo/gênero ou grau de escolaridade, por considerarem que esses grupos de fatores interagem.

4.3.4 Interação entre variáveis independentes sexo/gênero e faixa etária

Considerando-se, com base em outros estudos, que as variáveis sociais, podem interagir (OROZCO, 2015; LASTRA; MARTÍN, 2015; MANJÓN-CABEZA; POSE; SÁNCHEZ, 2016), procedeu-se à execução de algumas rodadas com as variáveis mais significativas. Várias rodadas dos dados em diferentes programas mostram que os grupos de

¹²⁶ Os dados que sobre o que parece ser “uma” variedade “cachaca” colombiana são os trabalhos de Hurtado (2001, 2005a,b) e Otheguy, Zentella e Livert (2007) de colombianos em Miami e Nova Iorque, respectivamente, e numa situação de contato linguístico o que não garante o comportamento dos falantes monolíngues.

¹²⁷ Ainda que não foi selecionada como significativa, Alfaraz (2015) mostra que os índices percentuais na variedade da República Dominicana são menores nos jovens comparativamente ao comportamento de as outras gerações.

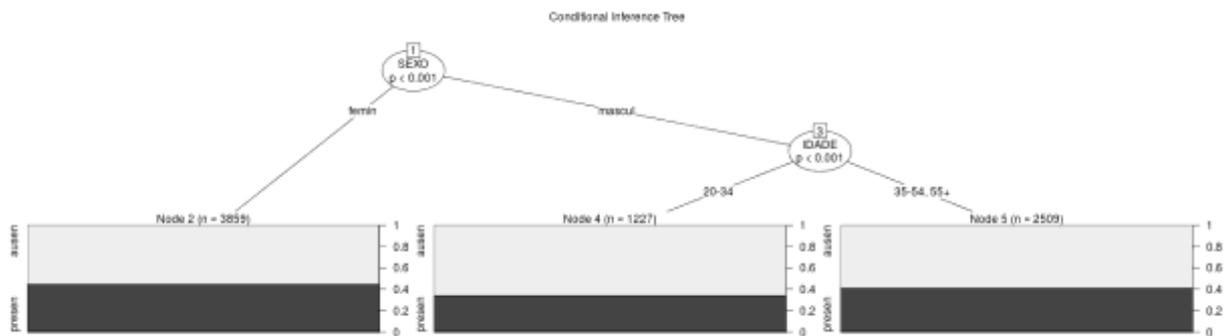
fatores externos com potencial para apresentar algum tipo de interação são sexo/gênero e faixa etária¹²⁸, cujos resultados são apresentados na Tabela 24 e no Gráfico 7.

Tabela 24. Cruzamento de sexo/gênero e faixa etária

		20-34	%	35-54	%	55+	%	Σ	%
Mulheres	Presença	582	44	543	44	601	46	1726	45
	Ausência	733	56	703	56	697	54	2133	55
	Σ	1315		1246		1298		3859	
Homens	Presença	420	34	524	41	514	41	1458	39
	Ausência	807	66	746	59	725	59	2278	61
	Σ	1227		1270		1239		3736	
Σ	Presença	1002	39	1067	42	1115	44	3184	42
	Ausência	1540	61	1449	58	1422	56	4411	58
	Σ	2542		2516		2537		7595	

Fonte: elaborada pelo autor

Gráfico 7. Árvore de inferência condicional para sexo/gênero e faixa etária



Fonte: elaborada pelo autor

Os resultados da Tabela 24 e do Gráfico 7 são consistentes com os resultados de Barranquilla (OROZCO, 2015), que mostram que as mulheres mais idosas favorecem a expressão pronominal (46%) em contraposição aos homens mais jovens (34%); ou seja, os resultados de Barranquilla podem generalizar-se para todo o Caribe colombiano.

¹²⁸ O grau de escolaridade não mostrou interação nas predições de Rbrul(2019) nem mostrou significância nos cruzamentos com as outras variáveis sociais. Ela só é retomada quando tem um percentual significativo para ressaltar esse fato.

De fato, um teste separado por cidade reitera serem as mulheres as que mais favorecem a expressão pronominal em todas as variedades. Análises mais detalhadas, em que interagem os três grupos de fatores externos, mostram que ser mulher idosa de grau de escolaridade baixo e médio é o fator que mais favorece a expressão do PPS e ser homem mais jovem do grau superior é o fator que menos favorece a expressão.

A baixa amplitude e conseqüentemente a pequena diferença nos valores dos fatores sugerem estudos mais aprofundados sobre os efeitos sociais na expressão pronominal, especialmente porque categorias amplas podem mascarar os efeitos que exercem grupos sociais mais específicos. Alfaraz (2015), por exemplo, postula que sexo/gênero é significativo na variedade da República Dominicana, mas quando cruzado com idade não encontra um padrão regular de comportamento. Nos dados aqui analisados, essa variável parece ter um padrão um pouco mais consistente com um único grupo bastante diferenciado, os homens jovens. Como assinala também Alfaraz (2015), análises futuras com mais participantes desse grupo de idade ajudarão a compreender melhor se a distribuição quantitativa é específica desta amostra ou representativa de toda a comunidade.

Manjón-Cabeza, Pose e Sánchez (2016), por exemplo, relatam descobertas relevantes sobre a atuação de jovens na expressão pronominal de segunda pessoa do singular. Análises mais detalhadas do fenômeno permitiram-lhes descobrir que o comportamento dos jovens que desfavorece a expressão pronominal na segunda do singular não específico [*tú-esp*], além de serem esses jovens com ensino superior os que menos favorecem a expressão pronominal. Essas situações de pesquisa mostram claramente que análises mais aprofundadas dos fatores sociais podem revelar comportamentos mais específicos e definitivos em relação à expressão pronominal que permanecem obscuras no estudo de categorias amplas.

A baixa significância da amplitude dos grupos de fatores sociais neste estudo e em um grande número de trabalhos, quando não excluídos do modelo estatístico como não significativos, leva a refletir se são apenas linguísticos os fatores que contribuem efetivamente para a expressão pronominal ou se as macrocategorias sociais predeterminadas na pesquisa variacionista é que não têm suficiente poder explanatório para mostrar covariação com o fenômeno sintático aqui investigado.

Os efeitos pouco distintivos desses fatores levaram à análise das médias percentuais para cada sujeito e há sujeitos que, obviamente, favorecem ou desfavorecem significativamente a expressão pronominal, distanciados da média geral e da média de seus respectivos grupos. Por isso, acolhendo sugestões para incluir o indivíduo como um grupo de fatores aleatório, agora que os programas estatísticos fazem análises mistas, procedeu-se à

execução de várias rodadas no programa Rbrul para verificar como seria o efeito do indivíduo como variável aleatória na expressão pronominal.

Em nenhuma dessas rodadas, todavia, o programa mostrou convergência, o que não permitiu fazer a análise por indivíduo. Assim mesmo, a Tabela 30 (no Apêndice B) mostra os percentuais de expressão pronominal por indivíduo, que possibilitarão análises mais aprofundadas no futuro. Analisar, então, as características sociais dos indivíduos que se distanciam da média geral e da média de seus respectivos grupos e executar um exame mais detalhado das características das entrevistas e dos temas nelas tratados poderiam constituir tarefas auxiliares para também ajudar a entender melhor as características sociais e contextuais que contribuem para a presença ou ausência do PPS.

4.4 A posposição do PPS no Caribe colombiano: fatores que condicionam seu uso

4.4.1 Distribuição da posição PPS no espanhol caribenho colombiano e em outras variedades

Os estudos variacionistas em geral não fazem diferença entre os pronomes expressos na posição pré-verbal e na posição pós-verbal, que entendem como a expressão da mesma variante; embora alguns trabalhos tenham chegado à conclusão de que os grupos de fatores analisados, que contribuem para a expressão de PPS, também atuam indistintamente na posposição, com a ressalva, todavia, de que têm baixíssima frequência os casos de posposição em vários dialetos.

A relação entre posposição e expressão pronominal tem permitido a formulação de algumas hipóteses ainda pouco testadas e analisadas com profundidade nos trabalhos variacionistas sobre as variedades do espanhol, provavelmente em razão desse baixo número de ocorrências ou mesmo em razão de não haver relevância teórica para a empreitada. Há uma hipótese, todavia, que, enfocando a relação entre os índices percentuais de expressão e a posição do sujeito pronominal, tem a ver com a mudança que conduziria à ordem fixa sujeito-verbo-objeto. Essa hipótese, que foi investigada com alguma frequência nas variedades caribenhas, por serem justamente as que se acham num estágio mais avançado nesse processo de mudança, não foi explorada ainda nas pesquisas sobre a variedade caribenha colombiana, certamente em consequência do baixo número de pesquisas sobre a sintaxe do espanhol colombiano em geral, como mostrado nos capítulos 2 e 3.

Com base em seus próprios trabalhos sobre a expressão pronominal e em outras pesquisas precedentes, Cameron (1993) propôs uma hipótese sociodialetal, também já

mencionada anteriormente, mas que vale a pena repetir aqui: se um dialeto A tem um percentual mais baixo de expressão de PPS do que o do dialeto B, então o dialeto A teria um percentual mais elevado de sujeitos pospostos do que o do dialeto B. Se o dialeto B tem um percentual mais elevado de expressão de PPS do que o do dialeto A, então o dialeto B teria um percentual menos elevado de sujeitos pospostos do que o do dialeto A. Para examinar essa questão, estabelecemos um levantamento da posição do sujeito pronominal na oração, conforme exposto na Tabela 25.

Tabela 25. Percentual de pronomes antepostos e pospostos ao verbo

Pronomes	N	%
Anteposto	3015	94,7
Posposto	169	5,3
Total	3184	100

Fonte: elaborada pelo autor

Como se pode observar na Tabela 25, a incidência percentual de pronomes pospostos supera ligeiramente a marca dos 5%, o que é bastante baixo, se comparada com o mesmo parâmetro em outras variedades, cuja distribuição pode ser observada nos dados da Tabela 26.

Tabela 26. Índices percentuais de pronomes e sua posição

	Percentual PPS	Anteposto	Posposto
Porto Rico (MORALES, 1982) ¹²⁹	53	81	19
Barranquilla (OROZCO; GUY, 2008)	35,4	96,1	3,9
Caribe colombiano	41,9	94,7	5,3
Oeste de Los Angeles, EUA (SILVA-CORVALÁN, 1977) ¹³⁰	39	65	35
Yucatán (MICHNOWICZ, 2015)	19,7	88	12
Cidade de México (LASTRA; MARTÍN, 2015)	21,7	85	14,4
Iquique, noroeste Chile (VAN ESBROECK, 2014)	25,9	74	26,1

Fonte: elaborada pelo autor

¹²⁹ Índices percentuais para o sujeito (incluindo o lexical). São tomados de Cameron (1993, p.321-322).

¹³⁰ Índices percentuais para o sujeito (incluindo o lexical). São tomados de Cameron (1993, p.321-322).

Na direção contrária do que argumentam Lastra e Martín (2015)¹³¹, os trabalhos que enfocam a expressão e posição pronominal não refutam a hipótese de Cameron (1993); pelo contrário, como mostra a Tabela 26, os resultados quantitativos a confirmam. É igualmente, um indício adicional de que o dialeto colombiano tem um comportamento mais próximo do espanhol caribenho, em relação ao sujeito pronominal para além das simples índices percentuais de expressão, como ocorre com outros fenômenos linguísticos que são também compartilhados.

4.4.2 Variáveis linguísticas e extralinguísticas que favorecem a posposição do sujeito pronominal

Os resultados da tabela 25 mostra um total de 3.184 ocorrências de pronomes expressos, das quais só 169 ocorrências são de posposição, o equivalente a 5,3%. A Tabela 27, por sua vez, mostra que fatores linguísticos e extralinguísticos contribuem para a posposição do sujeito pronominal. Mesmo que os casos de posposição sejam quantitativamente limitados para dar lugar a conclusões mais definitivas, os resultados permitem detectar certas tendências, embora outros precisem de análises mais aprofundadas e com um número maior de dados, ou deva assumir a perspectiva de levar em consideração o sujeito variável e não só o pronominal, para se obterem mais ocorrências.

¹³¹ Ver discussão no item 3.2.3.1

Tabela 27. Grupos de fatores significativos para a Posposição do PPS no Caribe colombiano

Grupo de fatores	%Total	Freq.	%	PR não centrado	PR centrado
Tipo de frase					
Interrogativa com e sem Q	1.4	14/43	32.6	.91	.88
Afirmativa	86.8	151/2765	5.5	.54	.47
Negativa	11.8	4/376	1.1	.17	.13
Amplitude				.74	
Pessoa e número do sujeito					
Nosotros	5.6	26/177	14.7	.76	.75
Tú +esp	4.4	11/140	7.9	.53	.52
Yo	59.7	96/1901	5.0	.51	.50
Ellos, Ellas:	5.9	8/188	4.3	.47	.46
Él, Ella	19.1	22/609	3.6	.42	.41
Tú –esp, Ud e Uds ¹³²	5.3	6/169	3.6	.35	.34
Amplitude				.41	
Tempo verbal					
Condicional e futuro perifrástico ¹³³	2.9	12/93	12.9	.75	.71
Presente	48.5	81/1536	5.3	.52	.47
Pretérito	25.8	42/818	5.1	.47	.42
Imperfeito	22.8	34/722	4.7	.45	.40
Amplitude				.30	
Dialeto					
Valledupar	32.2	80/1024	7.8	.62	.59
BA e CA ¹³⁴	67.8	89/2160	4.1	.44	.41
Amplitude				.18	
Grau de escolaridade					
Baixo	32.2	75/1025	7.3	.61	.58
Médio e Alto ¹³⁵	67.8	94/2159	4.4	.45	.41
Amplitude				.16	
		169/3184	5.3		
<i>Input .040</i>	<i>Log likelihood = -601.613</i>		<i>Significância = 0.015</i>	Não centrado	
<i>Input .084</i>	<i>Log likelihood = -601.613</i>		<i>Significância = 0.015</i>	Centrado	

Fonte: elaborada pelo autor

¹³² Tú [-esp] (1/78; **1,3%**; 2,4% total); usted [+ esp] (1/53; **1,9%**; 1,7% total); usted [-esp] (3/17; **17,6%**; 0,5%); ustedes (1/21; **4,8%**; 0,7% total)

¹³³ Condicional (4/19; **21,1%**; 0,6% total); futuro perifrástico (8/74; **10,8%**; 2,3%) e futuro (não tem caso de posposição).

¹³⁴ Barranquilla, BA (47/1100; 4,3%; 34,5% total); Cartagena, CA (42/1060; 4,0%; 33,3% total)

¹³⁵ Alto (49/1041; 4,7%; 32,7% total); Médio (45/1118; 4,0%; 35,1% total)

Em contraste com as hipóteses levantadas em vários estudos sobre a posposição pronominal, os resultados deste trabalho não se mostram significativos como o fazem os resultados desses estudos em relação a diversos grupos de fatores, como a mudança na referência, ambiguidade morfológica, pessoa primeira e terceira do singular, entre outras (SILVA-CORVALÁN, 2001; HURTADO, 2001; LASTRA; MARTÍN, 2015). Apenas o fator baixo grau de escolaridade teve um comportamento similar com os resultados obtidos da variedade da Cidade de México (LASTRA; MARTÍN, 2015). No entanto, uma hipótese viável está no fato de a não inclusão do pronome *uno* refletir de alguma forma as diferenças dos resultados da variedade da Cidade de México (LASTRA; MARTÍN, 2015); com efeito, em razão da morfologia ambígua desse pronome, ele foi agrupado nesse estudo com os de terceira do singular, sendo ainda categórico e pós-verbal em quase todos os casos. Há um aspecto em que estes resultados parecem convergir com o dos poucos estudos recentes que analisam certos fatores favorecedores da posposição pronominal: o de que o fator primeira pessoa do plural (*nosotros*) é o que mais favorece a posposição pronominal (LASTRA, MARTÍN, 2015; ORTIZ LÓPEZ; DAUPHINAIS; APONTE, 2018).

Diferentemente de outras investigações, mostram-se significativos nos resultados deste trabalho os índices de posposição do sujeito pronominal, os dos fatores *orações interrogativas* (.91), *as pessoas nosotros* (.76) e *tú[+esp]* (.53), *tempos condicional e futuro perifrástico* (.75), *subvariedade de Valledupar* (.62) e *grau baixo de escolaridade* (.61). Como ilustração de pronomes em posição pós-verbal, o exemplo (90) representa uma oração interrogativa, o exemplo (91), no contexto de primeira pessoa do plural, (92), no de segunda do singular específica e (93), no contexto de tempo condicional.

- (90) 584 A: ¿para dónde **vives tú**? (BA-55-11H)
- (91) eso es como: la doctrina 636 de lo que **llamamos nosotros** / Jesúsolo / que se apoyan / en un discurso de 637 Pedro / (BA-11-32M)
- (92) pero la yuquita mía esa la sembró fue mingo torres pa' (para) mi pa' (para) mi juje (mujer) y mihsijo (mis hijos) así que en el nombre de Dios **no bay (vas) tu a roba (arrobar)** aquí <risas=R/todos> [VA-31-32H]
- (93) esos dos compositores / los 642 **pondría yo** en el mismo nivel // o sea hablando de los de los (vacilación) 643 compositores / de allá de la provincia ¿no? (BA-37-22H)

Surpreende que *ambiguidade morfológica e textual*, que é o fator mais recorrente nas variedades enfocadas por outros estudos, não tenha sido selecionada como significativa; é isso

que explica por que a primeira e a terceira do singular, assim como os tempos imperfeitos, também constituem fatores não significativos.

A posição do pronome nas interrogativas mostra ainda que os falantes da variedade manifestam preferência, em termos probabilísticos, para essa forma, que, embora seja considerada canônica e, portanto, conservadora, ainda coexiste com as formas inovadoras¹³⁶. Também é importante destacar que o condicional é o único tempo verbal com morfologia ambígua que favorece a posposição; e que os fatores externos variedade dialetal e baixo grau de escolaridade parecem interagir entre si e com outros fatores. Por fim, no contexto de orações negativas, há uma rejeição quase categórica da posposição pronominal.

¹³⁶ O falante do exemplo (90) também utilizou a seguinte pergunta: 580 A: ¿oye **tú** dónde vives?.(BA-55-11H)

5. CONSIDERAÇÕES FINAIS

Para finalizar este trabalho, faz-se aqui um balanço do que foi proposto como objetivos, das hipóteses e perguntas de pesquisa e de como os resultados discutidos no capítulo de análise as confirmam ou as rejeitam.

Na introdução tomou-se o princípio de que é variável a expressão do pronome pessoal no sujeito em termos de expressão ou ausência, uma variável que é, para a maioria dos pesquisadores, fortemente correlacionada com o paradigma verbal do espanhol: quando o verbo contém uma marca explícita da pessoa, o PPS pode ser explícito ou nulo na maioria dos contextos, mas quando a forma verbal se mostra ambígua entre a primeira e a terceira, deveria haver naturalmente uma preferência pela manifestação de pronome no sujeito.

Viu-se que, como a tradição gramatical considera o espanhol uma língua de variação livre, seus seguidores explicam a manifestação de PPS com um recurso para indicar ênfase, contraste e desambiguar o referente nos tempos e modos em que pode apresentar ambiguidade.

Já no contexto da pesquisa linguística, uma teoria formal, como a de Princípios e Parâmetros (CHOMSKY, 1986) postula ser o espanhol uma língua de sujeito nulo, que, portanto, admite sujeitos vazios nas orações finitas com marcação na flexão verbal, em razão de a teoria gerativa admitir que a flexão verbal é um componente da categoria funcional “concordância” (AGR); é com essa categoria que é possível identificar o sujeito foneticamente nulo.

Por outro lado, viu-se também que diferentes estudos, especificamente os concernentes à variedade caribenha antilhana, trazem evidência clara das características das línguas de sujeito nulo (pro-drop), que podem converter-se para uma variedade que, ora contém sujeito obrigatório (non pro-drop), ora sujeito misto (ORTIZ LÓPEZ, 2016). É nesse âmbito que atuam os estudos variacionistas que apontam para os grupos de fatores relevantes que motivam ora a ausência ora a manifestação explícita de PPS.

Foi justamente dos pressupostos sociolinguísticos de ser língua inerentemente variável e de ser a variação existente sistematicamente governada por grupos de fatores de natureza interna e externa (LABOV, [1972] 2008) que se assentaram as bases desta pesquisa. E foi dentro dessas bases, que se estabeleceu como objetivo deste trabalho a análise da variável *expressão do pronome pessoal sujeito* no espanhol falado no Caribe colombiano, limitado às

subvariedades de Cartagena, de Barranquilla e de Valledupar, com a finalidade de especificar quais são os fatores linguísticos e extralinguísticos que motivam essa variação.

Esse objetivo geral foi desmembrado nos objetivos específicos de: i) descrever a distribuição dos pronomes sujeitos no espanhol falado no Caribe colombiano; ii) determinar a influência de fatores linguísticos e sociais na expressão do pronome sujeito na variedade objeto de estudo; iii) e determinar a coerência intradialetal em relação ao fenômeno linguístico focado e contrastar o comportamento da variedade caribenha da Colômbia com o de outras variedades do espanhol já estudadas.

Uma justificativa especial que se estabeleceu para empreender esta investigação foi a de que a ampla bibliografia existente sobre o tema não contempla, com o mesmo interesse que motivou a análise de outras variedades do espanhol, trabalhos extensivos sobre as variedades linguísticas colombianas. As abordagens existentes se limitam aos trabalhos de Orozco e Guy (2008), Orozco (2015), com os mesmos dados, Travis (2007) e Travis e Torres (2012), com os mesmos dados, destacando-se que apenas os estudos de Orozco e Guy (2008) e Orozco (2015) se concentram na análise de todos os pronomes sujeitos do espanhol de Barranquilla.

Todos esses trabalhos testam uma ou várias hipóteses interligadas, que têm despertado o interesse dos sociolinguistas: i) o uso do pronome expreso se dá por razões de ênfase ou ambiguidade; ii) a mudança da referência e a mudança do turno conversacional contribuem para a expressão pronominal; iii) a semântica do verbo favorece a presença dos pronomes sujeitos; iv) os pronomes expressos na posição de sujeito são mais frequentes nas variedades caribenhas e na chilena do que na da América continental e na península ibérica; v) a razão pela qual acontece essa frequência elevada nas variedades caribenhas é a supressão do *-s* e do *-n* no final de verbos e a posição quase obrigatória da ordem SVO; vi) o contato de algumas variedades do espanhol com outras línguas, especialmente com o inglês, provavelmente contribui para a maior expressão de PPS.

O que determinou esta pesquisa foi aplicar às variedades do Caribe colombiano praticamente as mesmas hipóteses e perguntas de pesquisa para dar uma contribuição ampla ao uso dos pronomes sujeitos no sentido de desvelar as semelhanças e/ou as diferenças entre os subdialetos do Caribe e em relação a estes e outros dialetos de língua espanhola na América e na Espanha; em caso de os grupos de fatores transcenderem as diferenças dialetais, essa situação seria postulada como um comportamento geral da língua espanhola.

Com base no exposto, este trabalho se debruçou sobre resposta possível às seguintes perguntas de pesquisa: (1) Qual é a distribuição da presença/ausência dos pronomes sujeitos no espanhol do Caribe colombiano? (2) Dos grupos de fatores linguísticos e extralinguísticos

investigados, quais são os que influenciam significativamente a expressão dos pronomes sujeitos no espanhol do Caribe colombiano? (3) A variedade caribenha colombiana mostra similaridades ou diferenças com o espanhol caribenho, em relação a percentuais de pronomes expressos, posição do pronome, fatores significativos? (4) A variedade estudada, comparada com outras do espanhol, mostra semelhanças ou diferenças nos índices percentuais de expressão dos pronomes sujeitos e nos grupos de fatores significativos?

Em resposta à pergunta (1) sobre a distribuição da presença/ausência dos pronomes sujeitos no espanhol do Caribe colombiano, os índices percentuais de expressão pronominal de 41,9% , e os baixos índices de posposição pronominal de 5,3% mostram que a variedade estudada faz parte do espanhol caribenho e se diferencia claramente dos outros dialetos do espanhol americano e peninsular. Além desses critérios também outros fatores começam a definir um comportamento bastante similar nas variedades caribenhas, a saber, o fato de culminar nas duas pessoas do discurso o maior grau de incidência de expressão pronominal e também o maior grau de favorecimento em termos de pesos relativos; nesse caso, a segunda pessoa sobrepuja a primeira, especialmente o pronome tú [-esp], como se pode verificar nos dados da Tabela 28, que estabelece uma comparação geral entre os resultados desta pesquisa e os de outros trabalhos já realizados.

Tabela 28 - Comportamento da expressão do PPS nas variedades caribenhas com foco na segunda do singular

	Caribe, Col (presente estudo)	Porto Rico (CAMERON , 1993, p. 325)	Porto Rico (ÁVILA;JI MENEZ, 1996) ¹³⁷	Porto Rico (ABREU 2012, p.4) Monolíngu es	Rep. Dominicana (MARTÍNEZ SÁENZ, 2011, p. 332)	Cuba (ALFA RAZ, 2018, p. 87)			
Grupo de fatores	%	PR	%	PR	%	%	PR	PR	
Pessoa e especificidade do sujeito									
Yo	48,0	.57	50	.59	53	50	59,8	.59	.65
Tú +esp	48,8	.53	48	.51	59	56	81,9	.82¹³⁸	.77¹³⁹
Tú -esp	57,8	.66	69	.72	63	54			
Ud ¹⁴⁰	54,3	.56				60			
Él, Ella	36,3	.45	39	.47	38	31	44,3	.44	.54
Nosotros	26,0	.29	15	.15	16	19	22,9	.17	.36
Ustedes	52,5	.52				41			
Ellos, Ellas:	27,4	.36	25	.26	24	10	20,2	.19	.20
						38	53,2		35,4

Fonte: elaborada pelo autor

A Tabela 28 mostra quantitativamente que a variedade caribenha colombiana do espanhol está em convergência com uma das principais características morfossintáticas do espanhol caribenho antilhano, ressaltadas por Ortiz López (2016): as elevadas frequências de PPS, especialmente os pronomes específicos singulares (*yo*, *tú*) e os não específicos (*uno*, *tú*), o que fica claramente evidenciada nos diferentes pesquisas dos dialetos antilhanos.

A resposta à pergunta (2), sobre os grupos de fatores linguísticos e extralinguísticos que exercem influência mais significativa na expressão dos pronomes sujeitos no espanhol do Caribe colombiano aponta para a seguinte hierarquia como as restrições que favorecem a expressão de PPS, compartilhando esses grupos de fatores com investigações prévias sobre um subdialeto dos estudados, o de Barranquilla: pessoa, número e especificidade do sujeito, correferencialidade, paralelismo formal, modo verbal, tempo verbal, progressividade, classe semântica do verbo, sexo/gênero, morfologia verbal, ambiguidade morfológica, grau de escolaridade, faixa etária.

¹³⁷ Os dados de Ávila-Jiménez (1996) foram tomados de Flores-Ferran (2004a, p. 59)

¹³⁸ Só oferece pesos relativos para a segunda em geral, mas ela registra para o *tú* [+esp] (88,7%) e para o *tú* [-esp] (78,1%), observe-se que ambos percentuais estão muito por encima da primeira pessoa do singular (p.330).

¹³⁹ Alfaraz (2018) não faz distinção entre as formas específicas e não específicas da segunda do singular

¹⁴⁰ *Usted* [+esp] (53/97; 54,6%; 1,3 total) e *usted* [-esp](17/32; 53,1%; 0,4 total)

Assim, os grupos de fatores linguísticos mais significativos na variedade estudada tiveram quase a mesma amplitude, ou ordem de importância, segundo o programa estatístico, e os fatores com maior frequência mostraram bastante convergência em termos de percentuais e de pesos relativos, o que permite deduzir a existência de uma unidade dialetal consistente das variedades com base na análise desse fenômeno.

Também o grupo de fatores sociais significativo mostra ter os mesmos efeitos em termos de amplitude (range) e pesos relativos. O fato de se ter excluído o grupo de fatores dialeto, as rodadas individuais das variedades, que selecionaram os mesmos grupos de fatores significativos e uma pesquisa de natureza contrastiva, feita preliminarmente, sobre algumas variáveis linguísticas (PEREZ; CAMACHO, 2019), apontaram para a existência de uma forte coerência interna da variedade caribenha colombiana em relação ao fenômeno estudado, que permitiu confirmar, portanto, a classificação dialetal das variedades colombianas (MONTES, 1982).

Grupos de fatores sem relevância aparente em termos de amplitude e frequência, como ambiguidade, passam a mostrar um grau elevado de significância nas situações de interação com outros fatores como o de mudança da referência. Essa expressividade demonstra que são os fatores de natureza textual e pragmática que convergem para ressaltar a relevância que não é possível detectar numa análise isolada da variável.

Por seu lado, a interação das variáveis sociais e de certos grupos de fatores contribuindo para a expressão do pronome permite inferir que essas macrocategorias provavelmente não revelam, com a clareza necessária, como características sociais e contextuais contribuem para a expressão pronominal. É possível essa seja a principal razão pela qual os fatores sociais não se mostraram significativos em muitas pesquisas sobre o PPS, ou mesmo a principal razão pela qual um só grupo tenha sido selecionado como significativo pelos programas estatísticos. Nos dados da variedade aqui analisada, os efeitos sociais não revelam claras diferenças geracionais para deduzir uma mudança em progresso na distribuição por tempo aparente. Há um indicio de que a expressão do sujeito poderia constituir uma instância de variação estável, cuja distribuição de gênero (uma pequena preferência entre as mulheres) sugere que pode ter prestígio encoberto. Também é importante deduzir com base em Alfaraz (2015), que as variáveis sociais definidas com base em um modelo de contexto anglo-americano poderiam não ser tão apropriadas para explicar a variação e mudança em contextos latino-americanos, particularmente na variação morfossintática.

Também as macrocategorias de outras variáveis linguísticas, como a semântica do predicado verbal, parecem desacreditar a avaliação de que fatores verdadeiramente favorecem

a expressão pronominal e que subcategorias permitiriam descrever mais adequadamente o comportamento semântico na expressão do pronome sujeito; nesse aspecto específico, é provável que a contribuição mais efetiva se derive do ADESSE, ou seja, seria mais efetivo analisar a frequência verbal complementada por essa variável como já admitido por alguns pesquisadores que começaram a testar essas possibilidades.

Por fim, os resultados obtidos mostram uma resposta positiva às perguntas (3) e (4) sobre possíveis semelhanças e diferenças entre a variedade estudada, o espanhol caribenho e outras variedades do espanhol em termos de frequência de expressão dos pronomes sujeitos e aos grupos de fatores significativos. Com efeito, os resultados obtidos apontaram não somente para a existência de semelhanças com o espanhol caribenho em relação aos percentuais e de posição do sujeito pronominal, mas também em relação aos grupos de fatores mais significativos, e a fatores específicos que parecem ser mais recorrentes nos dialetos caribenhos. Ressalte-se ainda que os resultados também indicaram considerável convergência com o comportamento dos diferentes dialetos do espanhol, o que parece um indício significativo de superação de limites dialetais.

Desse modo, a expressão do pronome nas variedades do Caribe colombiano aponta para uma atenuação do parâmetro *pro-drop*, que parece governar o espanhol peninsular em geral, o que permite também rejeitar o princípio postulado pela tradição gramatical em língua espanhola, que, fiando-se na rigidez do parâmetro *pro-drop*, reafirma razões mais estilísticas que morfossintáticas para as situações de expressão do pronome sujeito.

Para fechar este balanço, pode-se assegurar que os objetivos da pesquisa foram satisfatoriamente cumpridos, uma avaliação subjetiva, que, oxalá, venha a convergir harmonicamente com a interpretação dos leitores.

REFERÊNCIAS

- ABREU, L. Subject Pronoun Expression and Priming Effects among Bilingual Speakers of Puerto Rican Spanish. *In: GEESLIN, K.; DÍAZ-CAMPOS M. (ed.). **Selected Pro-ceedings of the 14th Hispanic Linguistics Symposium***. Somerville, MA: Cascadilla Proceedings Project, 2012, p.1-8.
- ALARCOS LLORACH, E. **Gramática de la lengua española**. Madrid: Espasa Calpe, 1994.
- ALFARAZ, G. Variation of Overt and Null Subject Pronouns in the Spanish of Santo Domingo. *In: CARVALHO, A. M.; OROZCO R.; SHIN, N. (ed.), **Subject Pro-noun Expression in Spanish: A cross-dialectal perspective***. Washington DC: Georgetown University Press, 2015, p. 3-16.
- _____. The Variation of Subject Pronouns over Time in Cuban Spanish. *In: CUZA, A. (ed.). **Cuban Spanish Dialectology: Variation, Contact and Change***. Georgetown University Press, 2018, p.83-98.
- BALASCH, S. La conectividad discursiva en el discurso interactivo. *In: BRUHN, J.; VALENZUELA, E. (ed.), **Selected proceedings of the 10th Hispanic Linguistics Symposium***. Somerville, MA: Cascadilla Proceedings Project, 2008, p. 301-311.
- BARRENECHEA, A. M.; ALONSO, A. Los pronombres personales sujetos en el español de Buenos Aires. *In: LOPE BLANCH, J. M. (ed.). **Estudios sobre el español hablado en las principales ciudades de América***. Cidade de México: Universidad Nacional Autónoma de México, 1977, p. 333-349.
- BAYLEY, R.; CÁRDENAS, N. L.; TREVIÑO SCHOUTEN, B.; MARTIN VÉLEZ, C. SALAS. Spanish Dialect Contact in San Antonio, Texas: An Exploratory Study. *In: GEESLIN, K.; DÍAZ-CAMPOS, M. (ed.). **Selected Proceedings of the 14th Hispanic Linguistics Symposium***, Somerville, MA: Cascadilla Proceedings Project, 2012, p.48-60.
- BECERRA, S. **Fonología de las consonantes implosivas en el español urbano de Cartagena de Indias**. Bogotá: Instituto Caro y Cuervo, 1985.
- BENTIVOGLIO, P. **Los sujetos pronominales de primera persona en el habla de Caracas**. Caracas: Universidad Central de Venezuela, 1987a.
- _____. A variação nos estudos sintáticos. **Estudos Linguísticos (Anais de Seminários do GEL)**, Campinas, v. 14, p. 7-29, 1987b.
- _____. La variación sociosintáctica en español. In: II CONGRESO INTERNACIONAL DE LA LENGUA ESPAÑOLA, Valladolid, 2001, Centro Virtual Cervantes, versão eletrónica, 2001. Disponível em: http://congresosdelalengua.es/valladolid/ponencias/unidad_diversidad_del_espanol/1_la_norma_hispanica/bentivoglio_p.htm Acesso em: 2 fev. 2016.
- BENTIVOGLIO, P.; ORTIZ LÓPEZ, L.; SILVA-CORVALÁN, C. PRESEEA. **Proyecto para el estudio sociolingüístico del español de España y de América**. Guía PRESEEA para

la investigación lingüística, 2011. Disponível em:

<<http://preseea.linguas.net/MetodologC3ADa.aspx>>. Acesso em: 2 fev. 2016.

BLANCO CANALES, A. Presencia/ausencia de sujeto pronominal de primera persona en español. **Español Actual**. Revista de español vivo, Madri, n.72, p. 31-40, 1999.

BLAS ARROYO, J. L. **Sociolingüística del español**. Desarrollos y perspectivas en el estudio de la lengua española en contexto social. Madri: Cátedra, 2005.

BOSQUE, I.; DEMONTE, V. (dir.), **Gramática descriptiva de la lengua española**, Vol. 1, Madrid: Espasa, 1999.

BOSQUE, I.; GUTIÉRREZ REXACH, J. **Fundamentos de sintaxis formal**. Madrid: Akal, 2009.

CALDERÓN NOGUERA, D. **El español hablado en Valledupar**: materiales para su estudio. ProyectoPreseea -Valledupar- Co. (Não publicado). Valledupar: Universidad Popular del Cesar, 2005.

CAMACHO, R.G. O formal e o funcional na teoria variacionista. *In*: RONCARATI, C. ABRAÇADO, J. (org.). **Português brasileiro**: contato linguístico, heterogeneidade e história. Rio de Janeiro: 7letras, 2003, p. 55-65.

_____. Da linguística formal à linguística social. São Paulo: Parábola editorial, 2013.

CAMERON, R. **Pronominal and Null Subject Variation in Spanish**: Constraints, Dialects, and Functional Compensation. 1992. 362 f. Tese (Doutorado em Linguística) University of Pennsylvania, Philadelphia, 1992.

_____. Ambiguous agreement, functional compensation, and non-specific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. **Language Variation and Change**, USA, v.5, n.3, p. 305-334, 1993.

_____. A Proposed Explanation of the Specific/Nonspecific TU Constraint Ranking in Spanish. **UPen Working Papers in Linguistics**, Philadelphia, v.3, n.1, 1996.

CAMERON, R.; FLORES-FERRÁN, N. Perseveration of subject expression across regional dialects of Spanish. **Spanish in Context**, v. 1, n.1, p. 41-65, 2004.

CARDOZO, J. **Uso de formas verbales subjuntivas ra y se en el habla de Montería**. 2012. 90 f. Dissertação (Mestrado em Linguística) Universidade Nacional de Colombia, Facultad de Ciencias Humanas, Bogotá D. C, 2011. Disponível em:
<<http://www.bdigital.unal.edu.co/4262/1/448177.2011.pdf>> Acesso em: 2 maio 2012.

CARVALHO, A. M.; OROZCO, R.; SHIN, N. (ed.). **Subject Pronoun Expression in Spanish**. A Cross-Dialectal Perspective. Washington DC: Georgetown University Press, 2015.

CARVALHO, A. M.; CHILD, M. Subject Pronoun Expression in a Variety of Spanish in Contact with Portuguese. *In*: MICHNOWICZ, J. & DODSWORTH, R. (ed.). **Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics**. Somerville, MA: Cascadilla Proceedings Project, 2011, p. 14-25.

CERRÓN-PALOMINO, Á. Ser más PRO o menos PRO: variación en la expresión de sujeto pronominal en el castellano limeño. *Linguística*, Montevideo, v. 30, n. 1, p. 61-83, 2014.

CHOMSKY, N. **Knowledge of Language: its origin, nature and use**. New York: Praeger, 1986.

CIFUENTES, H. Presencia y ausencia del pronombre personal sujeto en el habla culta de Santiago de Chile. *Boletín de Filología*, Homenaje a Ambrosio Rabanales, Santiago de Chile, v. 31, n.2, p. 743-752, 1980-81.

CLAES, J. ¿Constituyen las Antillas y el Caribe continental una sola zona dialectal? Datos de la variable expresión del sujeto pronominal en San Juan de Puerto Rico y Barranquilla, Colombia. *Spanish in Context*, v. 8, n.2, p.191–212, 2011.

CONSTITUIÇÃO POLÍTICA DA REPÚBLICA DA COLÔMBIA, 1991.

CUARTAS LÓPEZ. L. **Marcadores discursivos en el habla de la ciudad de Cartagena de Indias**. 2011. 132 f. Dissertação (Mestrado em Linguística) Universidade Nacional de Colombia, Facultad de Ciencias Humanas, Bogotá D. C, 2011.

DANE (DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA DE COLOMBIA). Colombia una nación multicultural. Su diversidad étnica, 2007. Disponible em: <https://www.dane.gov.co/files/censo2005/etnia/sys/colombia_nacion.pdf> Acceso em 06 de abril de 2019.

DANE 2018. Disponible em: <<https://www.dane.gov.co/index.php/estadisticas-portal/tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivienda-2018/cuantos-somos>>. Acceso em: 06 abr. 2019.

DAVIDSON, B. ‘Pragmatic weight’ and Spanish subject pronouns: The pragmatic and discourse uses of tú and yo in spoken Madrid Spanish. *Journal of Pragmatics*, v 26, p.543-565, 1996.

DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. A. (org.). **Português Brasileiro: uma viagem diacrônica**. Campinas: Ed. Unicamp, 1993, p. 107-128.

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

ERKER, D.; GUY, G. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, EUA, v. 88, n. 3, p. 526-557, 2012.

FANJUL, A. P. Conhecendo assimetrias: a ocorrência de pronomes pessoais. In: FANJUL, A. P.; GONZÁLEZ, N. M. (org.). **Espanhol e português brasileiro: estudos comparados**. São Paulo: Parábola Editorial, 2014, p. 29-71.

FERNÁNDEZ SORIANO, O. Sobre el orden de palabras en español. *Dicenda*. Cuadernos de Filología Hispánica, Madrid, n. 11 p. 113-152, 1993.

_____. El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. In BOQUE, I.; DEMONTE, V. (ed.). **Gramática Descriptiva de la Lengua Española**. Vol. 1. Madri: Espasa Calpe, 1999, p.1209-1273.

FILE-MURIEL, R. J. A laboratory approach to s-lenition in the spanish of Barranquilla, Colombia. In: FILE-MURIEL, R. J.; OROZCO; R.I (ed.). **Colombian Varieties of Spanish**. Madri e Frankfurt am Main; Iberoamericana-Vervuert, 2012, p.127-140.

FLORES-FERRÁN, N. Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? **Language Variation and Change**, EUA, v. 16, n.1. 49-73, 2004a.

_____. La expresión del sujeto en el español de Nueva York: el factor de la perseverancia. **ASJU**, País Basco, v. 38, n. 1 p. 353-366, 2004b.

FLÓREZ, L. El atlas lingüístico-etnográfico de Colombia (ALEC): Nota Informativa. **Thesaurus**, BICC, v. 16, n.1, 77-125, 1961.

GARCÍA, E. C. Shifting variation. **Língua**, v. 67, n.2-3, p. 189-224, 1985.

GARCÍA-MIGUEL; J. ALBERTUZ, F. Verbs, Semantic Classes and Semantic Roles in the ADESSE project, In: ERK, K.; MELINGER, A. & WALDE, S. S. (ed.). **Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes**, Saarbrücken, 2005.

GARCÍA-MIGUEL, J. COSTAS; L. MARTÍNEZ, S. Diátesis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE, In: WOTJAK, G.; CUARTERO OTAL, J. (ed.). **Entre semántica léxica, teoría del léxico y sintaxis**. Frankfurt am Main: Peter Lang, 2005, p. 373-384.

GARCÍA SALIDO, M. Pronombres personales focalizados en español conversacional, 2011, p.35-45. Disponible em: <https://emono.unibas.ch/catalog/download/28/17/168-1>. Acceso em: 1 dez. 2018.

GILI GAYA, S. **Curso superior de sintaxis española**. 3. ed. Barcelona: Bibliograf, 1961.

GONZÁLEZ MAFUD, A.; PÉREZ RODRÍGUEZ, M. El Habla Culta de La Habana, In: CASTAÑER MARTÍN; R. M.; LAGÜÉNS GRACIA, V. (coord.). Estudios dedicados a José M^a Enguita Utrilla. Zaragoza: Insituto Fernando El Católico, CSIC, 2010, p.327-336

GUY, G.; ZILLES, A. **Sociolingüística quantitativa**. Instrumental de análise. São Paulo: Parábola, 2007.

HENRÍQUEZ UREÑA, P. Observaciones sobre el español en América. **Revista de Filología Española**, Madri, n. 8, p. 357-390, 1921.

HERNÁNDEZ ALONSO, C. **Gramática Funcional del Español**. 3. ed. Madrid: Gredos, 1996.

HERNÁNDEZ CAMPOY, J.; ALMEIDA, M. **Metodología de la investigación sociolingüística**. Málaga: Editorial Comares, 2005.

HOCHBERG, J. G. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language*, LSA, EUA, v. 62, n.3, p. 609-621, set. 1986.

HURTADO, L. M. **La variable expresión del sujeto en el español de los colombianos y colombo-americanos residentes en el condado de Miami-Dade**. 2001. 224 f. Tese (Doutorado em Filosofia) University of Florida, Gainesville, FL, 2001.

_____. El uso de tú, usted y uno en el español de los colombianos y colomboamericanos. *In*: ORTIZ LÓPEZ L.; LACORTE, M. (ed.). *Contactos y contextos lingüísticos: el español en los Estados Unidos y en contacto con otras lenguas*. Madrid/Frankfurt: Iberoamericana/Vervuert, 2005a, p.187-200.

_____. Condicionamientos sintáctico-semánticos de la expresión del sujeto en el español colombiano. *Hispania*, EUA, v. 88, n.2, p. 335-348, 2005b.

JOHNSON, D. E. Getting off the Goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, v. 3, n.1, p. 359–383, 2009.

LABOV, W. **Padrões sociolingüísticos**. Tradução de Marcos Bagno Maria Marta Pereira Scherre e Caroline Rodrigues de Oliveira. São Paulo: Parábola Editorial, 2008

_____. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Texas Sociolinguistic Working Paper*. v. 44. Austin: Southwest Development Educational Laboratory, 1978. , p. 6-22.

_____. **Principles of linguistic change**, v. 2: Social factors. Malden, MA: Blackwell, 2001.

LAFFORD, B. El nuevo conservadurismo en el Caribe hispánico: el habla de Cartagena, Colombia. *Boletín de la Academia Puertorriqueña de la Lengua Española* 8, v. 2, p.72-90, 1980.

LASTRA, Y.; MARTÍN-BUTRAGUEÑO, P. “Subject pronoun expression in oral Mexican Spanish”. *In*: CARVALHO, A. M.; OROZCO, R.; SHIN, N. (ed.). **Subject Pronoun Expression in Spanish**. A Cross-Dialectal Perspective. Washington DC: Georgetown University Press, 2015. (Preprinter)

LAVANDERA, B. Where does the sociolinguistic variable stop? *Language in Society, Great Britain*, v.7, n. 2, p. 171-183, 1978.

Ley 1381 (25 de enero de 2010). *Diário Oficial da República da Colômbia*. N.º 47.603. Bogotá, D. C.

LIMERICK, P. **Subject Expression in a Southeastern U.S. Mexican Community**. 2018. 214 f.. Tese (Doutorado em Filosofia) Graduate Faculty of The University of Georgia. Athens, Georgia, 2018.

LIPSKI, J. **El español de América**. Madri: Cátedra, 1994.

LÓPEZ MORALES, H. **Métodos de investigación lingüística**. Salamanca: Ediciones del Colegio de España, 1994.

MANJON-CABEZA CRUZ, A.; POSE FUREST, F.; SANCHEZ GARCÍA, F. Factores determinantes en la expresión del sujeto pronominal en el corpus preescolar de Granada. **Boletín de Filología**, Santiago de Chile, v. 51, n. 2, p. 181-207, dic. 2016.

MARTÍN BUTRAGUEÑO, P. Hacia una tipología de la variación gramatical en sociolingüística del español. **Nueva Revista de Folología Hispánica**, Ciudad de México, v. 42, p. 29-75, 1994

_____. **La expresión del sujeto pronominal en la Ciudad de México**: explorando la variación lingüística con efectos estadísticos fijos y con efectos mixtos, p. 1-34, (no prelo).

MARTÍNEZ LARA, J. A. **Estudio sociolingüístico del sujeto variable de tercera persona en el habla de Caracas 2004-2010**. 2012. 121 f. Dissertação (Mestrado em Linguística) Universidade Central de Venezuela, Caracas, 2012.

MARTÍNEZ-SANZ, C. **Null and overt subjects in a variable system**: The case of Dominican Spanish. 2011. 484 f. Tese (Doutorado em Espanhol) Faculty of Graduate and Postdoctoral Studies, University of Ottawa, Ottawa, Canada, 2011.

MEYER-HERMANN, R. Sobre algunas condiciones pragmáticas de la posición del sujeto en español. **Estudios de Lingüística de la Universidad de Alicante**, v. 6, p. 73–88, 1990.

_____. Sobre el uso del sujeto yo en el habla culta de Costa Rica. *In*: KOTSCHI T., ZIMMERMANN, K.; OSTERREICHER, W. (ed.). **El español hablado y la cultura oral en España e Iberoamérica**. Frankfurt am Main: Vervuert, 1996, pp. 279-301.

MICHNOWICZ, J. "Subject pronoun expression in Yucatan Spanish". *In*: CARVALHO, A. M.; OROZCO, R.; SHIN, N. (ed.). **Subject Pronoun Expression in Spanish**. A Cross-Dialectal Perspective. Washington DC: Georgetown University Press, 2015. (Preprinter)

MONTENEGRO, L. Un aspecto de la sintaxis: los pronombres personales sujetos en el habla estudiantil santiaguera. **Eme Eme**. Estudios Dominicanos. Santiago de los Caballeros, v.12, n.71, p. 3-18, maio/junho 1984.

MONTES GIRALDO, J.J. El español de Colombia. Propuesta de clasificación dialectal. **Thesaurus**, BICC, Bogotá, v. 37, n. 1, p. 23-92, 1982.

_____. La bipartición dialectal del español. **Boletín de Filología**. Homenaje a Rodolfo Oroz, Santiago de Chile, v.35, n. 1, p.317-331. 1995-1996.

_____. Colombia, *In*: ALVAR, M. (ed.). **Manual de dialectología hispánica**. El español de América. Barcelona: Ariel, 1996, p. 134-145.

MONTES GIRALDO, J.J. et al. **El español hablado en Bogotá**: análisis previo de su estratificación social. Bogotá: ICC, 1998.

MORALES, A. La perspectiva dinámica oracional en el español de Puerto Rico, *In*: ALBA, O. (ed.). **El español del Caribe**: Ponencias del VI Simposio de Dialectología. Santiago de los Caballeros: Universidad Católica Madre y Maestra, 1982, p.203-219.

_____. Algunos aspectos de gramática en contacto: la expresión del sujeto en el español de Puerto Rico. **Anuario de Letras**. Lingüística y Filología, Cidade de México, v. 24, p.71-85, 1986

_____. Procesos discursivos del español de Puerto Rico, *In*: CONGRESO INTERNACIONAL DE LA LENGUA ESPAÑOLA. UNIDAD EN LA DIVERSIDAD LINGÜÍSTICA: La norma policéntrica del español, Cartagena, 2007. Disponible em: http://congresosdelalengua.es/cartagena/ponencias/seccion_3/31/morales_amparo.htm#n2. Acceso em: 2 fev. 2016.

MORENO FERNÁNDEZ, F. **Principios de sociolingüística y sociología del lenguaje**. Barcelona: Editorial Ariel, 1998.

_____. Metodología del “Proyecto para el estudio sociolingüístico del Español de España y América” (PRESEEA). **Lingüística**, Montevideo, v. 8, p. 257-285, 1996.

_____. **La lengua española en su geografía**. Manual de dialectología hispánica. 2. ed. Madrid: Arco/Libros, S.L., 2014.

MORENO FERNÁNDEZ, F.; OTERO ROTH, J., “Geografía del mundo hispánico”, MORENO FERNÁNDEZ, F.; OTERO ROTH, J. **Atlas de la lengua española en el mundo** 3. ed. Madri e Barcelona: editoriales Fundación Telefónica y Ariel S.A., 2016, p. 44-74.

MINISTERIO DE CULTURA DE COLOMBIA. **Lenguas Nativas y Criollas de Colombia**. Bogotá. Disponible em: < <http://www.mincultura.gov.co/areas/poblaciones/APP-de-lenguas-nativas/Paginas/default.aspx>.>. Acceso em: 06 abr. 2019.

OLMOS, A.; GÓMEZ, T.; PÉREZ, A. La variación sociolingüística de la /s/ implosiva en la comunidad de habla de Valledupar – Colombia, 2010. *In*: **Memorias del XXVI Congreso Nacional de Lingüística, Literatura y Semiótica**, Universidade Industrial de Santander, Bucaramanga, 22-24 de Septiembre de 2010. ISBN: 978-958-708-337-8. Disponible em: <<http://www.llibrototal.com/ltotal/ficha.jsp?idLibro=5392>>. Acceso em: 2 fev. 2016.

OROZCO, R. Distribution of Future Time Forms in Northern Colombian Spanish. In EDDINGTON, D. (ed.). **Selected Proceedings of the 7th Hispanic Linguistics Symposium**. Somerville: Cascadilla Proceedings Project, 2005, p.56-65.

_____. The Impact of Linguistic Constrains on the Expression of Futurity in the Spanish of New York Colombians. In: POTOWSKI, K.; CAMERON, R. (ed.). **Spanish in Contact: Policy, Social and Linguistic Inquiries**. Amsterdam/Philadelphia: John Benjamins, 2007a, p.311-327.

_____. Social Constraints on the Expression of Futurity in Spanish-Speaking Urban Communities. In HOLMQUIST, J.; LORENZINO, A.; SAYAHI, L. (eds.): **Selected Proceedings of the Third Workshop on Spanish Sociolinguistics**. Somerville: Cascadilla, 2007b, p.103-112.

_____. El castellano del Caribe colombiano a comienzos del siglo XXI. **Revista Internacional de Lingüística Iberoamericana (RILI)**, Madri, v.7, n.2, p. 95-113, 2009a.

_____. La influencia de factores sociales en la expresión del posesivo. **Lingüística**, Montevideo, v.22 , p.25-60, 2009b.

_____. Variation in the Expression of Possession in Costeño Spanish. **Spanish in Context** v.7, n.2, p.194-220, 2010.

_____. The expression of nominal possession in the Spanish of Colombians in New York City. In: FILE-MURIEL, R. J.; OROZCO, R.I (ed.). **Colombian Varieties of Spanish**. Madrid e Frankfurt am Main; Iberoamericana-Vervuert, 2012, p.205-234.

_____. Pronominal Variation in Colombian Costeño Spanish. In: CARVALHO, A. M.; OROZCO, R.; SHIN, N. (ed.). **Subject Pronoun Expression in Spanish**. A Cross-Dialectal Perspective. Washington DC: Georgetown University Press, 2015, p.17-37.

_____. Subject Pronoun Expression in Mexican Spanish: ¿Qué pasa en Xalapa? **Proc Ling Soc Amer**, Washington D.C., v. 1, art. 7, p.1-15. Disponível em: <http://dx.doi.org/10.3765/plsa.v1i0.3703>, 2016. Acesso em: fev. 2017.

_____; DÍAZ CAMPO, M. Dialectos del español de América: Colombia y Venezuela, In: GUTIÉRREZ-REXACH, J. (ed.). **Enciclopedia de Lingüística Hispánica**, Volumen 2. New York: Routledge, p.341-352, 2016.

_____; GUY, G. El uso variable de los pronombres sujetos: ¿Qué pasa en la costa Caribe colombiano? In: WESTMORELAND, M.; THOMAS, J.A. (ed.) **Selected Proceedings of the Fourth Workshop on Spanish Sociolinguistics**. Somerville, MA: Cascadilla, 2008, p.70-80.

ORTIZ LÓPEZ, L. Dialectos del español de América: Caribe Antillano (Morfosintaxis y Pragmática). In: GUTIÉRREZ-REXACH, J. (ed.). **Enciclopedia de Lingüística Hispánica**, Volumen 2. New York: Routledge, 2016, p. 316-329.

_____; DAUPHINAIS; A.; APONTE; H. Cuban Spanish: Is it a Null Subject Parameter dialect? In: CUZA, A. (ed.). **Cuban Spanish Dialectology: Variation, Contact and Change**. Georgetown University Press, 2018, p.97-117.

OTÁLORA, H. **Léxico del habla culta de Santafé de Bogotá**. Bogotá: Instituto Caro y Cuervo, 1997.

OTHEGUY, R.; ZENTELLA, A. C. **Spanish in New York**. Language Contact, Dialectal Leveling, and Structural Continuity. Oxford: Oxford University Press, 2012.

_____; HEIDRICK, I.; C. Coding Manual. In: OTHEGUY, R; ZENTELLA, A.C. **Spanish in New York**. Language Contact, Dialectal Leveling, and Structural Continuity. Oxford: Oxford University Press, 2012, p.225-273.

_____; LIVERT, D. Language and dialect contact in Spanish in New York: Towards the formation of a speech community”, **Language**, v. 83, n.4, p. 770-802, 2007.

PAREDES SILVA, V. L. **Cartas cariocas**. A variação do sujeito na escrita informal. 1988. Tese (Doutorado em Linguística) UFRJ, Rio de Janeiro, 1988.

PÉREZ BRABANDERE, V. **Los sujetos pronominales de primera persona en el español de Caracas 2004-2010**. 2010. 115 f. Dissertação (Mestrado em Linguística) Universidade Central de Venezuela, Caracas, 2010.

PÉREZ CÓRDOBA, A.; CAMACHO, R. A expressão do pronome pessoal sujeito no espanhol falado no Caribe colombiano. **Revista Estudos Linguísticos**, São Paulo, v. 48, n. 1, p. 404-424, abr. 2019.

POSIO, P. **Uso del pronombre personal sujeto de la primera persona del singular en español y portugués hablados**: factores semánticos y pragmáticos. 2008. 79 f. Dissertação (Mestrado em Filologia Hispânica) University of Helsinki, Helsinki, 2008.

PRADA-PÉREZ, A. de. **Subject expression in Minorcan Spanish: Consequences of contact with Catalan**. 2009. Tese (Doutorado em Filosofia) Pennsylvania State University, State College, PA, 2009.

QUESADA, M. A. **El español de América**. Cartago, Costa Rica: Editorial Tecnológica de Costa Rica, 2000.

QUIJADA BRINGTOWN, M. 2006. **Ø cantaba... ¿él, ella o yo? Sujetos de tercera persona en el habla de Caracas**. 2006. Dissertação (Mestrado em Linguística) Universidade Central de Venezuela, Caracas, 2006.

REAL ACADEMIA DE LA LENGUA. 1973. **Esbozo de una nueva gramática de la lengua española**. Madrid: Espasa Calpe.

REAL ACADEMIA ESPAÑOLA Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. **Nueva Gramática de La Lengua Española**. V. 3. Madrid: Espasa Libros, 2009.

RODRÍGUEZ Cadena, Y. Tendencias de variación sociolingüística en la Ciudad de Barranquilla. **Polifonía**, n.1. p. 109-134, 1997.

_____. Marcadores discursivos en el habla de Barranquilla. **Litterae**. Revista de la asociación de exalumnos del Seminario Andrés Bello del ICC, Bogotá, n. 8, p.197-220, 1999.

_____. El español del Caribe colombiano: Unidad y diversidad. **La casa de asterión. Revista trimestral de estudios literarios**. v. 4, n. 16, p.1-26, enero-febrero-marzo de 2004. ISSN: 0124 – 9282. Disponível em:
<<http://lacasadeasterionB.homestead.com/v4n16esp.html>>. Acesso em 15 jun. 2012.

_____. Estudio fonético del Español del Caribe. **Itaca, Revista del Lenguaje**, Valledupar, Año 2, v. 4, p.160-192, 2005.

_____. Colombia. In PALACIOS, A. (cord.). **El español en América**. Contactos lingüísticos en Hispanoamérica. Barcelona: Ariel, p. 135-160, 2008a.

RODRÍGUEZ Cadena, Y. (cord.) **El Habla de Barranquilla**: Materiales para su estudio. Tomo I. PRESEEA- Barranquilla. Universidad del Atlántico, GIESCA, 2008b.

_____. **Variación y cambio fónico en el Caribe colombiano: estudio sociolingüístico sobre la ciudad de Barranquilla**. 2009a. 535 f. Tese (Doutorado em Linguística) Colegio de México. Cidade de México D.F., 2009a.

_____.(cord.) **El Habla de Barranquilla**: Materiales para su estudio. Tomo II. PRESEEA- Barranquilla. Universidad del Atlántico, GIESCA, 2009b.

_____. (cord.) **El Habla de Barranquilla: Materiales para su estudio. Tomo III.** PRESEEA- Barranquilla. Universidad del Atlántico, GIESCA, 2010.

_____. Las investigaciones sociolingüísticas en Colombia. **Español Actual.** Revista de español vivo, Madri, n. 98, p. 71-96, 2012.

ROMAINE, S. **El lenguaje en la sociedad: una introducción a la sociolingüística.** Barcelona: Ariel Lingüística, 1996.

RONA, J. P. **Geografía y morfología del voseo.** Porto Alegre, [s.n.], 1967.

SÁNCHEZ ARROBA, M. E. Compensación funcional y expresión de sujeto pronominal “Tú” en el español de Cuba. **Lengua y Sociedad**, Lima, v. 13 n.1, p. 131-154, 2013.

SANKOFF, D.; TAGLIAMONTE, S. A.; SMITH, E. **Goldvarb X: A variable rule application for Macintosh and Windows.** Toronto: Department of Linguistics, University of Toronto, 2005.

SCRIVNER, O; DÍAZ-CAMPOS; M. Language Variation Suite: A theoretical and methodological contribution for linguistic data analysis. *In: Proceedings of the Linguistic Society of America*, v. 1, art. 29, p.1-15, 2016.

SERRANO, M. J. Cronología del estudio de la variación sintáctica en español: pasado y presente. **Revista de Filología de la Universidad de la Laguna**, Islas Canarias, v. 27, p. 155-170, janeiro de 2009.

_____. Historia que ya es historia: evolución y actualidad del concepto y metodología de la variación sintáctica. **Boletín de Lingüística**, Caracas, v. 19, n. 28, p. 102-127, 2007.

_____. De la cognición al discurso: el efecto de la prominencia cognitiva y la informatividad textual en el estudio de la variación de los sujetos pronominales. **Estudios de Lingüística de la Universidad de Alicante**, Alicante, v. 27, p. 275-299, 2013a.

_____. Variación sociosituacional de la colocación del sujeto pronominal en textos conversacionales. **Spanish in Context**, v. 10, n. 2, p. 261-283, 2013b.

_____. Cognición y estilo comunicativo: el sujeto posverbal y el objeto sintáctico, *In: Estudios Filológicos*, Valdivia-Chile, v. 54, p. 139-156, 2014a

_____. El sujeto y la subjetividad: variación del pronombre yo en géneros textuales del español de Canarias. **Signos. Estudios de Lingüística**, Chile, v. 47, n. 85, p. 321-343, 2014b.

SHIN, N. L.; OTHEGUY, R. Shifting sensitivity to continuity of reference: Subject pronoun use in Spanish in New York City. *In: LACORTE, M. & LEEMAN, J. (ed.). Español en Estados Unidos y en otros contextos: Cuestiones sociolingüísticas, políticas y pedagógicas*, 111-136. Madrid: Iberoamericana, 2009, p. 111-136.

SILVA-CORVALÁN, C. Subject expression and placement in Mexican-American Spanish. *In: AMASTAE, J.; ELÍAS-OLIVARES, L. (ed.), Spanish in the United States. Sociolinguistic Aspects.* New York: Cambridge University Press, 1982, p.93-120.

_____. **Sociolingüística.** Teoría y análisis, Madrid, Alhambra, 1988.

_____. On the problem of meaning in sociolinguistic studies of syntactic variation. In: KASTOVSKY, D.; SZEWDEK, A. (ed.) **Linguistic across historical and geographical boundaries**. Berlin: Mouton, , 1991, p. 111-123..

_____. **Direcciones en los estudios sociolingüísticos de la lengua española**. La lengua española. Sociedad y enseñanza. Sevilla, 1992. Disponible em: <https://cvc.cervantes.es/obref/congresos/sevilla/sociedad/ponenc_silvac.htm>, Acceso em: 2 fev. 2016.

_____. **Sociolingüística y pragmática del español**. Washington D.E: Georgetown University Press, 2001.

_____. La expresión del Sujeto desde una perspectiva variacionista. In: CABRÉ, M.T.; LORENTE, M.; SOLÉ, E. **Lingüística aplicada**. Cicle de conferencies i seminaris 98-00. Barcelona: Universitat Pompeu Fabra, Intitut Universitari de Lingüística Aplicada, 2004, p. 61-75.

TAGLIAMONTE, S. Comparative sociolinguistics. In: CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (ed.). **The handbook of language variation and change**, Malden and Oxford: Blackwell, 2002, p.729-763.

_____. **Analysing Sociolinguistic Variation**. Cambridge: Cambridge University Press, 2006.

TAGLIAMONTE, S. A.; BAAYEN, R. H. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, v.24, n. 2, p.135-78, 2012.

TARALLO, F. **A pesquisa sociolingüística**. São Paulo: Ática, 1985.

TORRES CACOULLOS, R. Variation and Grammaticalization. In: DÍAZ-CAMPOS, M. (ed.). **The Handbook of Hispanic Sociolinguistics**. Oxford: Blackwell, 2011. p. 148-167.

TORRES CACOULLOS, R.; TRAVIS, C. E.. Foundations for the study of subject pro-noun expression in Spanish in contact with English. In: CARVALHO, A. M.; OROZCO, R.; SHIN, N. (ed.). **Subject Pronoun Expression in Spanish**. A Cross-Dialectal Perspective. Washington DC: Georgetown University Press, 2015. (Preprinter)

TRAVIS, C. E. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, USA, v. 19, n. 2, p.101-135, 2007.

TRAVIS, C. E.; TORRES CACOULLOS, R. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics*, v. 23, n.4, p. 711-748, 2012.

TRILLOS AMAYA, M. Español-L2 en el contexto multicultural y plurilingüe de la región Caribe. *Lingüística y Literatura*, Medellín, n. 73, p. 72-95, 2018.

_____. Ayer y hoy del Caribe de Colombia en sus lenguas. In: CASTILLO, A. (Comp.). **Respirando el Caribe**. Memorias de la Cátedra del Caribe Colombiano V.1. Bogotá: Observatorio del Caribe Colombiano, Ministerio de Cultura e Universidad del Atlántico, 2001.p. 152-180.

_____. Lingüística en la vida y para la vida. Estudios en el Caribe: balance 2005-2009. IRIARTE, P. (Edit.). **Respirando el Caribe. V.3.** Memorias del III encuentro de investigadores sobre el Caribe Colombiano. Bogotá: Observatorio del Caribe Colombiano, Colciencias, 2009. p. 185-214.

TRUDGILL, P. J.; HERNÁNDEZ CAMPOY, J. M. **Diccionario de sociolingüística.** Madrid: Editorial Gredos, 2007.

VÁSQUEZ CANTILLO, A. **Los marcadores discursivos en la comunidad de habla de Barranquilla.** 2006. Dissertação (Mestrado em Linguística Hispano-americana) Seminario Andrés Bello, ICC, Bogotá, 2006.

VÁSQUEZ CANTILLO, A.; CUARTAS LÓPEZ, L. **El habla de Cartegena de Indias: Materiales para su estudio.** PRESEEA-CARTAGENA. Cartagena de Indias-Colômbia: Alpha Editores, 2017.

ZAMORA MUNNÉ, J. C.; GUITART, J. M. **Dialectología hispanoamericana.** Teoría, descripción, historia. Salamanca: Ediciones Almar, 1982.

VAN ESBROECK, N. **La variable sujeto: ¿qué posición ocupa el español chileno dentro del espectro de la variación dialectal?** 2014. 89 f. Dissertação (Mestrado of Arts in de Taal- en Letterkunde: Spaans – Engels) Universidade Gent. Faculteit Letteren en Wijsbegeerte Vakgroep, 2014. Disponível em: <https://lib.ugent.be/fulltxt/RUG01/002/162/607/RUG01-002162607_2014_0001_AC.pdf>. Acesso em 2 fev. 2016.

APÊNDICES

APÊNDICE A. Frequência e percentual léxico.

Tabela 29. Frequência léxica de verbos utilizados 10 ou mais vezes

VERBO			
std.dev	intercept	ocorrências	%
	0.477	6280	0.428
PENSAR	1.106	68	0.735
CREER	0.778	110	0.618
SABER	0.716	234	0.585
ACEPTAR	0.614	17	0.706
PAGAR	0.586	23	0.652
DECIR	0.581	619	0.543
NACER	0.569	14	0.714
SER	0.532	499	0.531
PEDIR	0.478	31	0.581
TOMAR	0.47	39	0.564
AMAR	0.442	14	0.643
PREGUNTAR	0.44	18	0.611
VER	0.438	197	0.513
VIVIR	0.43	118	0.517
LEVANTAR	0.383	17	0.588
OIR	0.377	13	0.615
RECORDAR	0.367	42	0.524
QUERER	0.34	208	0.486
ENTRAR	0.328	18	0.556
ABRIR	0.308	12	0.583
ATENDER	0.308	12	0.583
MIRAR	0.299	10	0.6
VENIR	0.292	98	0.48
ACOSTAR	0.246	13	0.538
LLAMAR	0.213	27	0.481
COMPRAR	0.192	66	0.455
JUGAR	0.179	21	0.476
EMPEZAR	0.172	12	0.5
TIRAR	0.172	12	0.5
MUDAR	0.154	10	0.5
REGALAR	0.154	10	0.5
TERMINAR	0.135	29	0.448
DEDICAR	0.115	13	0.462
ESTAR	0.11	255	0.424
NECESITAR	0.108	25	0.44
COGER	0.0981	73	0.425

SEGUIR	0.0918	23	0.435
SENTIR	0.0836	81	0.42
TRABAJAR	0.0689	94	0.415
SALIR	0.0497	83	0.41
ESTUDIAR	0.049	56	0.411
DAR-	0.0478	29	0.414
PRESENTAR	0.0356	12	0.417
LLEGAR	0.0172	95	0.4
GANAR	0.0127	30	0.4
ESPERAR	0.00924	15	0.4
CONOCER	0.00915	88	0.398
AGRADECER	0.00724	10	0.4
VENDER	0.000785	43	0.395
TENER	-0.0238	632	0.389
CONSEGUIR	-0.0392	16	0.375
CASAR	-0.0468	37	0.378
TRATAR	-0.0477	24	0.375
RECIBIR	-0.0493	11	0.364
VIAJAR	-0.0493	11	0.364
SUFRIR	-0.0493	11	0.364
PODER	-0.0568	19	0.368
ENSEÑAR	-0.085	17	0.353
SACAR	-0.0947	36	0.361
AYUDAR	-0.0964	28	0.357
CONTAR	-0.0988	20	0.35
ENTENDER	-0.0988	20	0.35
HABLAR	-0.116	83	0.361
BAILAR	-0.117	15	0.333
PERDER	-0.139	21	0.333
APRENDER	-0.151	38	0.342
BUSCAR	-0.182	42	0.333
TRAER	-0.193	31	0.323
SENTAR	-0.193	11	0.273
HACER	-0.195	320	0.347
IR	-0.219	305	0.341
PARAR	-0.242	12	0.25
ACORDAR	-0.278	43	0.302
QUEDAR	-0.294	62	0.306
MORIR	-0.316	26	0.269
MANDAR	-0.319	23	0.261
CAER	-0.333	14	0.214
ECHAR	-0.353	24	0.25
COMPARTIR	-0.375	15	0.2
COMER	-0.385	12	0.167
REGRESAR	-0.385	12	0.167
PONER	-0.423	83	0.277

LLEVAR	-0.464	86	0.267
QUEMAR	-0.488	11	0.091
METER	-0.501	36	0.222
MATAR	-0.556	20	0.15
DAR	-0.558	156	0.256
PEGAR	-0.566	24	0.167
ENCONTRAR	-0.588	36	0.194
DURAR	-0.588	21	0.143
PASAR	-0.609	56	0.214
VOLVER	-0.678	20	0.1
DEJAR	-0.797	55	0.164
LEER	-0.806	29	0.103
model.basics	total.n	df	intercept
		6280	2 -0.426
		input.prob	grand.proportion
		0.395	0.428
model.fit	deviance	AIC	AICc
	8.403.804	8.407.804	8.407.806
Somers.Dxy.fixed	Somers.Dxy.total	R2.fixed	R2.total
0	0.266	0	0.0646

Fonte: elaborada pelo autor

APÊNDICE B. Frequência e percentual por informante.

Tabela 30. Percentual de pronome por informante

Informante	std.dev	intercept	ocorrências	%
		0.387	7595	0.419
	
(CA-05-13H)		1.01	135	0.711
(CA-34-22M)		0.714	144	0.625
(VA-30-22M)		0.568	146	0.582
(BA-01-31M)		0.474	146	0.555
(BA-71-13H)		0.452	108	0.556
(BA-37-22H)		0.431	140	0.543
(BA-65-13M)		0.335	138	0.514
(CA-27-22H)		0.311	142	0.507
(VA-18-31M)		0.289	148	0.5
(CA-42-33M)		0.289	148	0.5
(VA-42-13M)		0.278	149	0.497
(CA-06-13M)		0.277	145	0.497
(BA-49-11M)		0.254	145	0.49
(CA-36-23M)		0.231	147	0.483
(BA-27-21M)		0.225	137	0.482
(BA-15-32H)		0.222	150	0.48
(BA-19-33M)		0.197	133	0.474
(VA-27-22H)		0.182	143	0.469
(CA-46-32M)		0.167	140	0.464
(CA-20-21M)		0.162	147	0.463
(VA-45-23H)		0.162	147	0.463
(VA-47-23M)		0.147	144	0.458
(BA-55-11H)		0.118	149	0.45
(VA-50-33H)		0.0877	136	0.441
(VA-38-13H)		0.0521	137	0.431
(BA-59-12M)		0.042	145	0.428
(BA-41-23M)		0.039	150	0.427
(BA-22-33H)		0.0245	142	0.423
(CA-01-11H)		-0.00821	138	0.413
(BA-33-22M)		-0.0253	147	0.408
(BA-11-32M)		-0.0349	148	0.405
(CA-19-21H)		-0.042	129	0.403
(CA-02-11M)		-0.107	143	0.385
(CA-38-31M)		-0.125	150	0.38
(VA-11-21M)		-0.163	149	0.369
(CA-39-32H)		-0.166	144	0.368
(VA-05-11M)		-0.172	150	0.367
(VA-13-31H)		-0.179	140	0.364
(VA-34-32M)		-0.191	144	0.361

(BA-62-12H)	-0.191	117	0.359
(VA-19-12H)	-0.209	146	0.356
(VA-31-32H)	-0.232	143	0.35
(VA-53-33M)	-0.343	144	0.319
(BA-06-31H)	-0.36	149	0.315
(VA-24-12M)	-0.368	147	0.313
(CA-37-31H)	-0.397	135	0.304
(VA-07-21H)	-0.432	139	0.295
(CA-10-12M)	-0.47	85	0.271
(BA-30-21H)	-0.482	145	0.283
(CA-15-12H)	-0.509	145	0.276
(CA-23-23H)	-0.56	148	0.264
(CA-53-33H)	-0.576	136	0.257
(BA-48-23H)	-0.63	150	0.247
(VA-02-11H)	-0.67	103	0.223
<hr/>			
model.basics	total.n	df	intercept
	7595	2	-0.342
		input.prob	grand.
		0.415	proportion
			0.419
<hr/>			
model.fit	deviance	AIC	AICc
	10.168.012	10.172.012	10.172.013
<hr/>			
Somers.Dxy.	Somers.Dxy.	R2.fixed	R2.total
fixed	total		
0	0.229	0	0.0436

Fonte: elaborada pelo autor