

PREDICATE CONSTITUENT ORDER  
VARIATION IN THE ANDES: A  
COMPARATIVE ANALYSIS OF ANDEAN  
SPANISH AND BILINGUAL CUSCO  
QUECHUA

by

SARAH ELIZABETH HUBBEL

(Under the Direction of Chad Howe)

ABSTRACT

This dissertation examines predicate constituent order variation (OV/VO) in the varieties of Andean Spanish and Quechua spoken in Cusco, Peru. The investigated languages exhibit mirror-image ordering typologies— Andean Spanish is a VO-dominant language and Quechua an OV-dominant language. It has been argued that non-canonical structures are more frequent in each of these languages as a result of language contact and that OV/VO variation is conditioned by both social factors, e.g., socioeconomic status, and linguistic factors, e.g., information structure (Kalt & Geary, 2021; Muntendam, 2009; Muysken, 1984; F. A. Ocampo & Klee, 1995; Sánchez, 2003). Cross-linguistic studies on word order variation have determined that other factors, like animacy (Tanaka et al., 2011), definiteness (van Bergen & de Swart, 2009, 2010), and weight (Faghiri & Samvelian, 2020; Yamashita & Chang, 2001; Yao, 2018) also govern word order variation in several languages and syntactic domains.

In the present study, I use naturalistic spoken data elicited via Sociolinguistic Interviews to determine the distribution of OV/VO orders in Andean Spanish and bilingual Cusco Quechua. Overall, I find that non-canonical OV order occurs at a rate of 11.0% in Andean Spanish, though bilinguals employ OV order more frequently than monolinguals, 13.2% and 7.1% respectively. In bilingual Cusco Quechua, non-canonical VO order occurs at a higher rate, 31.8%. An examination of the extralinguistic variables reveals an inverse correlation between age and VO order frequency for both Andean Spanish and Quechua bilinguals, which suggests a syntactic change in progress in which OV/VO variation is 1) moving toward the prescriptive Spanish norm in Andean Spanish

and 2) moving away from OV-dominant order in Quechua. Through a cross-linguistic comparison of the linguistic predictors motivated by Poplack and Levey (2010), I uncover two convergent structures in bilingual Andean Spanish and Quechua— OV order with demonstrative and universally-quantified pronominal objects (e.g., *eso, todo*) and VO order with explicit subjects. I argue that, taken together, the structural change in progress and existence of convergent structures constitute evidence of bidirectional contact-induced change in the Quechua-Spanish contact situation regarding syntactic variation within the predicate.

INDEX WORDS: [word order, contact-induced change, predicate constituent order variation, language typology, Andean Spanish, Quechua]

PREDICATE CONSTITUENT ORDER VARIATION IN THE ANDES:  
A COMPARATIVE ANALYSIS OF ANDEAN SPANISH AND  
BILINGUAL CUSCO QUECHUA

by

SARAH ELIZABETH HUBBEL

B.A., Calvin University, 2014

M.A., University of Georgia, 2016

A Dissertation Submitted to the Graduate Faculty of the  
University of Georgia in Partial Fulfillment of the Requirements for the  
Degree.

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

©2024  
Sarah Elizabeth Hubbel  
All Rights Reserved

PREDICATE CONSTITUENT ORDER VARIATION IN THE ANDES:  
A COMPARATIVE ANALYSIS OF ANDEAN SPANISH AND  
BILINGUAL CUSCO QUECHUA

by

SARAH ELIZABETH HUBBEL

Major Professor: Chad Howe

Committee: Timothy Gupton  
Pilar Chamorro

Electronic Version Approved:

Ron Walcott  
Dean of the Graduate School  
The University of Georgia  
May 2024

# DEDICATION

To the lovely participants of this study.  
Thank you for sharing your stories with me.

# ACKNOWLEDGMENTS

As I near the end of the dissertation writing process, I realize that the submission of this document marks the culmination of a ten-year graduate career at the University of Georgia. During this time, I have crossed paths with numerous individuals that have contributed positively to my academic and personal journey that I must acknowledge here. Though obtaining a PhD is a complicated endeavor with many ups and many downs, the individuals who I encountered along the way made it all worth it.

First, I would like to extend my gratitude to my major professor, Dr. Chad Howe. You are the very first person that I met in Georgia over a decade ago on a prospective student visit, during which you informed me of the nascent Quechua program at UGA and consequently inspired a life-long interest in indigenous languages. Over the years, I have been struck by your humility, intelligence, humor, and sincerity. Thank you not only for the intellectual stimulation you have provided through brainstorming sessions and insightful feedback, but more importantly, for your enthusiastic support of my transition into motherhood, which you have consistently viewed as an asset, not a hindrance.

Thank you to my committee members, Dr. Timothy Gupton and Dr. Pilar Chamorro. The work that I present here has been molded not only by your direct feedback but also through your fascinating courses and general academic mentorship over the years. Thank you to the professors that have deepened my interest in linguistics through their own enthusiasm and expertise: Dr. Diana Ranson, Dr. Sarah Blackwell, Dr. Margaret Lubbers Quesada, Dr. Gary Baker, Dr. Peggy Renwick, and Dr. Mark Wenthe.

To the individuals who participated in the data collection process and without whom this project, quite literally, would not have been possible: this work is dedicated to you. It was the greatest privilege to be entrusted with your personal stories. Your vulnerability is inspiring. Thank you to the academic institutions that supported my Quechua education, Centro Tinku and ACUPARI, and the fantastic educators that are affiliated with them. In particular, to Aliciacha and Gina, masters of experiential learning: thank you for taking me far and wide to learn Quechua and giving me an inside view of the Andean cosmovision.

Thank you to Rosa, Felsie, Vlad, Mateo, and Esteban for your generous hospitality during my stays in Cusco. I hope to repay the favor one day. Thank you Elena, Señor Juancito, Mama Julia, Pilar, Gollita, and Yovana for your friendship and for making Cusco feel like home.

To the Ling Ladies, Bethany, Camila, Madeline, and Jihee: what is left to say? The beautiful community that we built together strengthened us to weather the storm of writing our dissertations during a global pandemic while becoming mothers. We met as graduate students who relocated very far from our respective homes, but we became a family. I admire the amazing women, researchers, educators, and mothers you all are. I feel so honored to know you and your beautiful families.

Thank you to my dear friend, Lizzie Baker. You were my first friend in Georgia, and I never could have anticipated how deep our friendship would become. Watching you kick ass in your own career has often given me the push I needed to get back on track. Thank you for your love and support, which you have so freely given me over the years.

To my fellow ROMLers, Evelyn Autry, Kate Bove, Diogo Cosme, Chase Fulford, Racheal Fulford, Jorge García-Granados, María González, David Hair, Darío Lizancos Robles, Caitlin Samples, Catherine Simpson, Kerry Steinberg, Jessica Tiegs, Martin Ward, and others I have likely failed to list here: thank you for making graduate school enjoyable. Thank you to my future colleagues at the University of North Georgia, particularly Kristi Hislope, for your words of encouragement in the final days. Go Nighthawks!

Thank you to my family; travelling to Michigan for holidays and summers to spend time together provided a much-needed respite from an endless cycle of reading, writing, and editing. Mom, thank you for exemplifying what it means to be a hard-working, driven woman. Dad, thank you for the comic relief and for always being unapologetically yourself. Brooke, thank you for being my best friend and for bringing Elliot Sue into this world (with the help of Jake, of course). Thank you to my in-laws, Kelly, Britt, Katie, Burt, Landry, Cooper, Taylor, Ryan, and Declan, who never missed an opportunity to encourage me and celebrate big moments. A special thanks to Kelly— your generous offer to care for Ophelia has been indispensable in finishing this dissertation.

The biggest thanks goes to my husband, Carson. We met just weeks after I defended my prospectus, and you have loyally stuck by my side ever since. I cannot imagine it has been easy to live with a person perpetually working on their dissertation, but you have done so with love and kindness. You have taken on more than your fair share to ensure that I finish this damn degree, especially in the childcare department, and I am endlessly grateful for your sacrifices. I

am beyond ready to experience life with you on the other side and begin a new chapter with our little family in Dahlonega.

Finally, to my daughter, Ophelia: being your mother is the single greatest accomplishment of my life. Your presence on this earth put writing a dissertation into perspective in a way that nothing else could. Thank you for making me laugh every single day. I saved the best for last, boo bear.

# CONTENTS

<b>Acknowledgments</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Objectives of the present study . . . . .	1
1.2 Theories of language contact . . . . .	5
1.3 The Quechua-Spanish contact situation . . . . .	14
1.4 A note on terminology . . . . .	29
1.5 Outline of the dissertation . . . . .	30
<b>2 Word Order</b>	<b>33</b>
2.1 Introduction . . . . .	33
2.2 Word order typology . . . . .	35
2.3 Language contact . . . . .	42
2.4 Linguistic predictors of constituent order variation . . . . .	47
2.5 Summary . . . . .	73
<b>3 Word Order in Spanish and Quechua</b>	<b>75</b>
3.1 Introduction . . . . .	75
3.2 Word order in Spanish . . . . .	76
3.3 Word order in Andean Spanish . . . . .	80
3.4 Word order in Quechua . . . . .	104
3.5 Summary . . . . .	122
<b>4 Methodology</b>	<b>125</b>
4.1 Introduction . . . . .	125
4.2 Participants . . . . .	126
4.3 Data collection procedure . . . . .	136

4.4	Coding the data . . . . .	141
4.5	Research questions . . . . .	175
4.6	Summary . . . . .	177
<b>5</b>	<b>Analysis: Andean Spanish</b>	<b>179</b>
5.1	Introduction . . . . .	179
5.2	General results . . . . .	180
5.3	Monolingual Andean Spanish results . . . . .	183
5.4	Bilingual Andean Spanish results . . . . .	194
5.5	Summary . . . . .	212
<b>6</b>	<b>Analysis: Bilingual Cusco Quechua</b>	<b>214</b>
6.1	Introduction . . . . .	214
6.2	Housekeeping . . . . .	215
6.3	General results . . . . .	219
6.4	Analysis of extralinguistic variables . . . . .	220
6.5	Analysis of linguistic variables . . . . .	228
6.6	Summary . . . . .	237
<b>7</b>	<b>Discussion</b>	<b>239</b>
7.1	Introduction . . . . .	239
7.2	Research Question 1 . . . . .	239
7.3	Research Question 2 . . . . .	244
7.4	Research Question 3 . . . . .	256
7.5	Research Question 4 . . . . .	259
7.6	Research Question 5 . . . . .	276
7.7	Summary . . . . .	289
<b>8</b>	<b>Conclusion</b>	<b>293</b>
8.1	Introduction . . . . .	293
8.2	Contributions . . . . .	293
8.3	Implications of potential typological shift . . . . .	300
8.4	Other considerations and future directions . . . . .	304
8.5	Summary and final remarks . . . . .	307
	<b>Appendices</b>	<b>311</b>
	<b>Bibliography</b>	<b>325</b>

# LIST OF FIGURES

1.1	Distribution of Quechua and Aymara in the mid-20th century (Adelaar, 2004, p. 169) . . . . .	15
4.1	Map of the Department of Cusco (Instituto Nacional de Estadística e Informática, 2017, p. 15) . . . . .	127
4.2	Density plot of participant age . . . . .	130
4.3	Density plot of monolingual and bilingual participant age . . . . .	132
4.4	Educational system of Peru . . . . .	134
4.5	Data collection procedure flowchart . . . . .	136
4.6	Distribution of BLP scores of bilingual participants . . . . .	138
4.7	Boxplot of BLP score ranges across first language . . . . .	140
5.1	Rate of OV order (%) across language profile of Andean Spanish speakers . . . . .	182
5.2	Barplot of OV/VO order across monolingual participants . . . . .	184
5.3	Density plot of OV/VO order across absolute weight for monolingual group . . . . .	188
5.4	Density plot of OV/VO order across relative weight for monolingual group . . . . .	189
5.5	Visualization of best-fit logistic regression model for monolingual group . . . . .	192
5.6	Definiteness-polarity interaction in monolingual group . . . . .	193
5.7	OV/VO proportions across bilingual group participants . . . . .	195
5.8	Density plot of OV/VO order across age for bilingual group . . . . .	199
5.9	Density plot of OV/VO order across BLP score for bilingual group . . . . .	201
5.10	Density plot of OV/VO order across relative weight for bilingual group . . . . .	207
5.11	Density plot of OV/VO order across absolute weight for bilingual group . . . . .	208

5.12	Visualization of best-fit mixed-effects model of linguistic predictors for bilingual group . . . . .	210
5.13	Definiteness-polarity interaction in bilingual group . . . . .	211
5.14	Definiteness-animacy interaction in bilingual group . . . . .	211
6.1	OV order rate per participants in Quechua dataset . . . . .	220
6.2	Density plot of OV/VO order across age . . . . .	224
6.3	Density plot of OV/VO order across BLP score in Quechua dataset . . . . .	225
6.4	Probability of VO order across age in Quechua dataset . . . . .	227
6.5	Probability of VO order across first language (L1) in Quechua dataset . . . . .	228
6.6	Density plot of OV/VO order across <b>relative</b> weight for Quechua dataset . . . . .	233
6.7	Density plot of OV/VO order across <b>absolute</b> weight for Quechua dataset . . . . .	233
6.8	Visualization of best-fit mixed-effects model of linguistic predictors in Quechua dataset . . . . .	235
6.9	Probability of VO order across absolute weight in Quechua dataset . . . . .	237
7.1	Scatterplot of age and OV order frequency in bilingual Andean Spanish group . . . . .	245
7.2	Scatterplot of age and OV order frequency in bilingual Cusco Quechua . . . . .	260
7.3	Age of Quechua participants across educational level . . . . .	262
7.4	Age of Quechua participants across first language . . . . .	263
7.5	Distribution of age across <i>ta</i> -marking in Quechua . . . . .	269
7.6	Correlation between OV/VO order rates of Quechua and Andean Spanish corpora for bilinguals . . . . .	277

# LIST OF TABLES

1.1	Language profile of Peruvian population (5 years or older) from 1940 to 2017 (partially adapted from Klee and Lynch (2009, p. 133)	21
2.1	Dominant Order of Clausal Constituents in Greenberg (1963) and WALS (Dryer, 2013c)	37
2.2	Dominant Order of Predicate Constituents in Greenberg (1963) and WALS (Dryer, 2013b)	42
2.3	Selected Examples of Clausal Word Order Shift	43
2.4	The Givenness Hierarchy, adapted from Gundel et al. (1993)	51
3.1	Word order distribution in three syntactic domains across age of L2 Spanish bilinguals (adapted from Luján et al., 1984, p. 359)	82
3.2	OV/VO order distribution across five groups of adult Andean Spanish speakers (adapted from Muysken, 1984, p. 114)	84
3.3	OV/VO order distribution in Andean Spanish across two groups of speakers from Calca, Peru (adapted from F. A. Ocampo and Klee, 1995, p. 77)	88
3.4	OV/VO order distribution across three groups Andean Spanish speakers from Calca, Peru (adapted from Klee, 1996, p. 80)	90
3.5	OV/VO distribution in Andean Spanish word order studies	94
3.6	OV/VO distribution in Quechua word order studies	115
3.7	Order of topicalized and focalized constituents	118
4.1	Demographic information of participants	129
4.2	Age distribution by decade in participant sample and Cusco population	131
4.3	Level of education among participant sample and Cusco population	134
4.4	Bilingual participants' linguistic characteristics	139
4.5	Independent variables of the present study	174

5.1	Frequency of clausal constituent orders in Andean Spanish dataset . . . . .	180
5.2	OV/VO order across language profile of Andean Spanish speakers . . . . .	181
5.3	OV/VO order across monolingual participants . . . . .	183
5.4	OV/VO order across extralinguistic variables for monolingual group . . . . .	185
5.5	Descriptive statistics of age for monolingual group . . . . .	185
5.6	OV/VO order across linguistic variables for monolingual group	186
5.7	Descriptive statistics of weight for monolingual group . . . . .	188
5.8	Results of best-fit mixed-effects model for monolingual group	190
5.9	Distribution of OV/VO order rates per bilingual participant . . . . .	194
5.10	OV/VO order across extralinguistic variables for bilingual group	196
5.11	Barplots of OV/VO order across extralinguistic variables for bilingual group . . . . .	197
5.12	Descriptive statistics of age (in years) for bilingual group . . . . .	198
5.13	Descriptive statistics of BLP score for bilingual group . . . . .	200
5.14	Results of best-fit mixed-effects model of extralinguistic predictors for bilingual group . . . . .	202
5.15	Probability of VO order across age (left) and BLP score (right) for bilingual group . . . . .	203
5.16	OV/VO order across linguistic variables for bilingual group . . . . .	204
5.17	Barplots of OV/VO order across linguistic variables for bilingual group . . . . .	205
5.18	Descriptive statistics of relative and absolute weight for bilingual group . . . . .	206
5.19	Results of best-fit mixed-effects model of linguistic predictors for bilingual group . . . . .	209
6.1	OV/VO order across object type in Quechua dataset . . . . .	218
6.2	Frequency of clausal constituent orders in Quechua dataset . . . . .	219
6.3	OV/VO order across extralinguistic variables in Quechua dataset	221
6.4	Barplots of OV/VO proportions across extralinguistic variables in Quechua dataset . . . . .	222
6.5	Descriptive statistics for continuous extralinguistic variables in Quechua dataset . . . . .	223
6.6	Results of best-fit mixed-effects model of extralinguistic predictors in Quechua dataset . . . . .	226
6.7	OV/VO order across linguistic variables in Quechua dataset . . . . .	229

6.8	Barplots of OV/VO order across linguistic variables in Quechua dataset . . . . .	230
6.9	Descriptive statistics of weight for Quechua dataset . . . . .	232
6.10	Results of best-fit mixed-effects model of linguistic predictors for Quechua dataset . . . . .	235
7.1	OV/VO order across Andean Spanish groups . . . . .	240
7.2	OV/VO order in Andean Spanish word order studies . . . . .	241
7.3	Average OV/VO order in Andean Spanish word order studies . . . . .	242
7.4	Average OV/VO distribution in monolingual Spanish word order studies . . . . .	244
7.5	Distribution of clausal constituent orders in Andean Spanish . . . . .	252
7.6	Distribution of information status of the object across definiteness . . . . .	255
7.7	OV/VO distribution in Quechua word order studies . . . . .	257
7.8	Distribution of clausal constituent orders in Quechua data . . . . .	258
7.9	OV/VO order across constituents marked for topic and focus . . . . .	266
7.10	OV/VO order across argument type for L1 Quechua and simultaneous bilinguals . . . . .	275
7.11	OV/VO order for Andean Spanish and Quechua . . . . .	276
7.12	Comparison of significant predictors of predicate constituent order variation across monolingual and bilingual Andean Spanish groups . . . . .	279
7.13	Comparison of significant predictors of OV/VO variation across Quechua and Andean Spanish bilinguals . . . . .	281
7.14	Distribution of OV/VO order across pronominal category in Andean Spanish and Quechua . . . . .	283
7.15	Clausal constituent order in Andean Spanish and Quechua . . . . .	287

# CHAPTER I

## INTRODUCTION

### 1.1 Objectives of the present study

The purpose of this dissertation is to investigate the **variable order of predicate constituents**— object and verb— in the varieties of Quechua and Andean Spanish spoken in Cusco, Peru. I apply the theoretical frameworks offered by variationist sociolinguistics, language contact, and linguistic typology to effectively examine this phenomenon. Typologically speaking, these languages exhibit mirror-image ordering preferences; that is, the dominant predicate constituent order is VO in Andean Spanish and OV in Quechua. Due to the typological incongruity in the dominant orders of object and verb, it may be hypothesized that linguistic contact will result in variation or change in this syntactic domain for one or both language systems. Theories of language contact provide a framework for interpreting the directionality and types of contact-induced change that may transpire when language varieties enter into contact. Sociolinguistic insights suggest that variation within a particular linguistic structure (e.g., variable order of predicate constituents) is governed by both social factors (e.g., age and sex) and linguistic factors (e.g., definiteness).

To repeat, prescriptively speaking, Andean Spanish (like Spanish) is classified as a VO-dominant language and Quechua an OV-dominant language (Cerrón-Palomino, 1987a; Dryer & Haspelmath, 2013; Zagona, 2002). However, it has been reported that order in this syntactic domain is variable in both language varieties. In other words, non-canonical OV order is attested in Andean Spanish and non-canonical VO order is attested in Quechua. The examples below illustrate canonical (a) and non-canonical (b) structures in Andean Spanish (1) and Quechua (2). These examples are extracted from the corpora of the present study, and the participant label is given in brackets in the translation line.

(1) **Andean Spanish**

- a. *iba a Juliaca, de Juliaca me iba hacia Quillabamba, a la selva... entonces de Quillabamba, de la selva, yo transportaba fruta... a Juliaca*

‘I would go to Juliaca, from Juliaca I would go toward Quillabamba, to the jungle... then from Quillabamba, from the jungle, I would transport **fruit...** to Juliaca’ [P25]

- b. *por cesaria di al bebé... **parto normal** yo quería pero no... no se pudo*

‘I gave birth by c-section... I wanted a **normal birth**, but it couldn’t... be done’ [P37]

(2) **Quechua**

- a. *huq p’unchay wayk’u-q señora achbu-sqa... ‘ay  
one day cook-AG woman approach-PST2 oh  
machu-cha hamu-y,’ **qhuña-n-ta**  
old.man-DIM come-IMP snot-POSS.3SG-ACC  
picha-ru-sqa falda-wan  
clean.off-EXH-PST2 skirt-INS*

‘one day, a female cook approached [him]... “oh, come here little old man,” [and she] cleaned off **his snot** with [her] skirt’ [P27]

- b. *qayna wata celebra-ra-ni, eh, chay-pi  
last year celebrate-PST-1SG eh that-LOC  
ruwa-ka-mu-ra-yku **pacha-manka-ta**  
make-REFL-DIR-PST-IPL.EXCL earth-pot-ACC*

‘last year I celebrated, eh, there we made [ourselves] **an underground stove**’ [P18]

In (1a), the object, *fruta*, ‘fruit’, is collocated after the verb, *transportaba*, ‘would transport’, yielding canonical VO order. On the contrary, the object in (1b), *parto normal*, ‘normal birth’, is positioned before the verb, *quería*, ‘wanted’, which results in non-canonical OV order. In the Quechua examples, the object, *qhuñanta*, ‘his snot’, in (2a) is positioned before the verb, *picharusqa*, ‘wiped’, which reflects canonical OV order in Quechua. Example (2b), however,

depicts non-canonical VO order, as the object, *pachamankata*, ‘underground stove’, is collocated in post-verbal position. These examples highlight that 1) Andean Spanish and Quechua exhibit mirror-image word order typologies and 2) non-canonical structures are attested in both languages.

Non-canonical syntactic configurations exemplified in (1b) are reportedly realized at an elevated rate in Andean Spanish compared to non-contact varieties of Spanish. A wide range of rates corresponding to OV order have been reported for Andean Spanish, with the highest values nearing 40% for certain groups of speakers (F. A. Ocampo & Klee, 1995). This value is substantially higher than the range of OV order rates cited for non-contact varieties of Spanish, approximately 0% - 11% (Givón, 2001; F. A. Ocampo, 1992, 1995; Puerma Bonilla, 2019). Predicate constituent order variation does not pattern homogeneously across all speakers of Andean Spanish, rather it is conditioned by a number of social and linguistic variables, like socioeconomic status and information structure (Camacho, 1999; Klee, 1996; Klee et al., 2011; Muntendam, 2008, 2009, 2010, 2013; Muysken, 1984; F. A. Ocampo & Klee, 1995).

Quantitative analyses on predicate constituent order variation in Quechua offer an even wider range of non-canonical word order rates. For instance, in some regional varieties, like Lamas Quechua, VO order may actually be considered the dominant order on the basis of the relative infrequency of canonical OV order (Sánchez, 2003). Conversely, in the varieties of Quechua spoken in Chuquisaca, Bolivia and Cusco, Peru, though syntactic variation exists, OV remains the dominant order (Kalt & Geary, 2021). Notwithstanding their opposing conclusions regarding OV/VO proportions, these studies converge on the assertion that variation in this syntactic domain is conditioned by certain linguistic features, particularly variable accusative case-marking in Quechua. Other analyses of word order variation in Quechua examine the effect of different linguistic factors, like information structure (Muntendam, 2015; Muysken, 1995; Sánchez, 2010).

Over the past three decades, these studies (and others to be cited later) have contributed indispensable insights on the topic of the present study, like the observation that Andean Spanish speakers of a lower socioeconomic status tend to realize OV order more frequently than those of a higher socioeconomic status and the confirmation of a robust relationship between information structure and word order variation in both varieties, to name a couple examples. Nevertheless, several unanswered questions remain. The present study endeavors to enhance the existing literature on predicate constituent order variation in the Andes by addressing three perceived gaps concerning 1) the methodological design, 2) the integration of cross-linguistic research on constituent order, and

3) the approach to statistical analysis. I detail my procedure for attending to these matters in the following paragraphs.

First, in the present study, semi-structured sociolinguistic interviews elicit speech data that resemble the natural, conversational speech of native Andean Spanish and Quechua speakers in Cusco, Peru. Several of the aforementioned studies that have investigated predicate constituent order variation in these language varieties have employed relatively structured elicitation tasks (e.g., closed question-answer pairs, story-telling tasks), which may prompt somewhat contrived or context-specific speech. The speech samples examined in the present study approximate natural speech because the context in which data is collected resembles a conversation between interlocutors. Furthermore, I collect speech samples in both Andean Spanish and Quechua to directly compare the distribution of predicate constituent orders in each variety at the level of the individual. To my knowledge, the present study is the first on predicate constituent order variation to directly compare speech data from both languages with the objective of appropriately assessing the possibility of contact-induced language change. Lastly, I account for methodological discordance across past studies by evaluating sentential polarity and subject expression as independent linguistic predictors. I elaborate on these methodological adjustments in Chapter 3 and Chapter 4.

Second, several linguistic features are considered as potential predictors of predicate constituent order variation in the present study due to their prevalence in prior cross-linguistic research on variation in a variety of syntactic domains. In particular, I consider the effect of animacy, definiteness, and grammatical weight. The inclusion of these variables yields novel results, as the primary linguistic predictors of OV/VO variation included in previous analyses on Andean Spanish and Quechua have been limited to the informational status of the object and subject expression.<sup>1</sup> I further enhance the statistical analyses by examining additional extralinguistic and linguistic factors, some of which have been considered in previous studies (e.g., sex and morphological case marking) and some of which are lacking from the literature (e.g., age and sentential polarity).

Lastly, the data of the present study are partitioned into three groups—monolingual Andean Spanish, bilingual Andean Spanish, and bilingual Cusco Quechua—prior to statistical analysis. By examining the conditioning effect of the independent variables on predicate constituent order variation in all three groups, it is possible to satisfy the proposed criteria for properly establishing contact-induced change, which involves both horizontal and vertical comparisons to rule out alternative explanations, like internal mechanisms of language change. The term **horizontal** refers to a three-way comparison across the con-

<sup>1</sup> As mentioned earlier, the effect of variable morphological case marking has also been considered, but this feature applies only to Quechua analyses.

tact variety, a non-contact variety, and the source language; the latter term, **vertical**, refers to a comparison between a current and earlier version of the contact variety (Poplack & Levey, 2010; Thomason, 2010). Past studies on predicate constituent order variation in both languages have asserted that the elevated frequency of non-canonical word order in the Andes as compared to non-contact varieties are the result of language contact in the region.<sup>2</sup> However, analogizing a variable structure to contact-induced change requires a comparative analysis that is permitted by the methodological structure of the present study.

Below I advance five research questions that guide the present study:

### **Research questions**

1. What are the overall proportions of OV/VO orders in monolingual and bilingual Andean Spanish in Cusco, Peru? Do monolingual and bilingual speakers differ in their distribution of predicate constituent orders?
2. Which extralinguistic and linguistic factors condition OV/VO variation in monolingual and bilingual Andean Spanish?
3. What is the overall proportion of OV/VO orders in bilingual Cusco Quechua?
4. Which extralinguistic and linguistic factors condition OV/VO variation in bilingual Cusco Quechua?
5. How do the factors selected as predictors of predicate constituent order variation in Andean Spanish compare to those selected for Quechua? What might this comparison suggest about contact-induced language change in this syntactic domain?

In the paragraphs that follow, I contemplate select seminal theories on the outcomes of language contact in order to situate the Quechua-Spanish contact situation in the Andes within a broader context and provide the necessary theoretical background to satisfactorily respond to the fifth research question.

## **1.2 Theories of language contact**

### **1.2.1 Mechanisms of contact-induced change**

Prior to the pioneering work of Weinreich (1953) and Haugen (1953), which involved the integration of social, psychological, and linguistic processes in establishing a theory of language contact, many historical linguists dismissed the

<sup>2</sup> Most literature on Andean Spanish argue that transfer from Quechua to Andean Spanish should be considered indirect transfer. I will elaborate on this argument in Chapter 3.

role of external mechanisms in favor of language-internal ones to account for diachronic linguistic change (Meillet, 1921; Müller, 1875). In opposition to this stance, Thomason and Kaufman (1988) argue that language contact is in fact a fundamental catalyst of linguistic change. Consequently, their work is primarily preoccupied with establishing a framework of contact-induced language change. To begin, the authors differentiate two primary transference patterns—borrowing and substratum influence.<sup>3</sup> In their words, “**borrowing** is the incorporation of foreign features into a groups’ native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated feature” (p. 37, emphasis mine). By contrast, “**substratum interference** is a subtype of interference that results from imperfect group learning during a process of language shift... the errors made by the members of the shifting group in speaking the TL<sup>4</sup> then spread to the TL as a whole when they are imitated by the speakers of that language” (pp. 38-39, emphasis mine).<sup>5</sup> In general, transfer via the borrowing mechanism prevails in situations of **language maintenance** (i.e., the contact languages are preserved) while substratum interference predominates in situations of **language shift** (i.e., a language displaces another language).<sup>6</sup> Though the aforementioned mechanisms of contact-induced change generally correspond to these two contact situation types in a one-to-one fashion, borrowing and substratum interference are not mutually exclusive, meaning they may function in parallel in a given contact situation. Citing Yiddish-English contact as an example, Thomason and Kaufman (1988) contend that, among bilingual speakers, Yiddish is shaped via borrowing from English and English is molded by substratum influence from Yiddish. This configuration is likely the result of a shift in progress from Yiddish to English with an interim period of bilingualism (i.e., language maintenance).

According to Thomason and Kaufman (1988), borrowing and substratum influence are associated with discrete linguistic outcomes. The authors maintain that borrowing adheres to a chronological process in which lexical items are inserted into the target language first to be followed by the adoption of structural features later, provided that the contact situation is characterized by long-standing, widespread bilingualism. Bilingualism is a necessary condition for structural borrowing because an individual or group must be sufficiently proficient in the second language so as to incorporate its structural (e.g., phonological and morphosyntactic) elements. Conversely, substratum influence typically promotes structural change prior to the transmission of lexical items. The paucity of loanwords in situations of language shift is justified by the following reasoning:

<sup>3</sup> Thomason and Kaufman (1988) use the term ‘interference’, which I have replaced with the terms ‘influence’ and ‘transfer’, as these are more widely accepted in contemporary writing. When quoting the authors, however, I retain their original terminology.

<sup>4</sup> target language

<sup>5</sup> Thomason and Kaufman (1988) define *substratum*, and the related terms, *superstratum* and *adstratum*, in the following manner: “superstratum languages are typically those of victorious invaders who then shift to the language of the conquered people; substratum languages are those of conquered, or at least socio-politically subordinate, indigenous populations and immigrants; and adstratum languages are those of invaded or invader groups that are neither dominant nor subordinate in the contact situation” (p. 116).

If the speakers' goal is to give up their native language and speak some other language instead, vocabulary is the first part of the TL they will need, so it is the first part they will learn. They will probably keep their own native-language words for only things the TL has no words for: foods and other cultural items, and (if the TL speakers are invaders from elsewhere) names for local animals plants, and so forth (Thomason & Kaufman, 1988, p. 39).

Returning to the Yiddish-English contact situation mentioned above, the authors report that English interference in Yiddish is characterized by a high degree of lexical borrowing but little structural borrowing. However, the variety of English spoken by these bilinguals is characterized by only a moderate degree of lexical transference but a high degree of structural influence. The divergent linguistic outcomes in each contact language may be explained by the disparate mechanisms applied— borrowing in Yiddish and substratum influence in English.

Contrary to Thomason and Kaufman's view that structural change may proceed in situations of language shift in the absence of lexical borrowing, Winford (2003) argues that "structural change is practically always mediated by lexical transfer" (p. 61). He asserts that structural components cannot be borrowed in their own right and must, in fact, be transported through other linguistic processes. According to Winford, structural diffusion is attributable to three mechanisms— "direct borrowing of structural elements (quite rare), indirect structural diffusion via lexical borrowing (fairly common), and indirect structural transfer via SL<sup>7</sup> agentivity in situations of bilingualism and mutual accommodation between linguistic groups" — though clearly only the latter two processes are considered probable mechanisms of contact-induced change. (p. 63). These three mechanisms all beget the commensurate outcome of structural convergence, which is defined as the process whereby "previous differences in grammar between them [two languages] are reduced or eliminated either because one adopts structural features from the other as a replacement for its own, or because both adopt an identical compromise between their conflicting structures" (p. 63).

The first mechanism Winford (2003) lists loosely corresponds to structural transfer via borrowing, as defined by Thomason and Kaufman (1988), and the last mechanism to substratum influence. To exemplify the second listed mechanism of contact-induced change, indirect structural diffusion via lexical borrowing, I consult Silva-Corvalán's work on Los Angeles Spanish— a variety of Spanish in contact with English. Silva-Corvalán (1998) argues that, what appears on the surface to be syntactic transfer may actually be "lexico-syntactic

<sup>6</sup> There are other types of contact situations, like area convergence (i.e., Sprachbunds), language attrition or death, and new language creation (e.g., bilingual mixed languages, creoles, and pidgins) (Thomason & Kaufman, 1988; Winford, 2003). However, I will limit the discussion here to language maintenance and language shift since they are most pertinent to the contact situation under investigation in the present study.

<sup>7</sup> source language

calques” in which “an English word is matched up with a Spanish word which incorporates semantic elements and subcategorization and selectional restrictions from English” (p. 231). To illustrate this point, she provides the following examples from LA Spanish (3) and General Spanish (4):

(3) *Y tu carro que compraste, ¿cómo te gusta?*

‘and your car that [you] bought, how to-you pleases?’

(Silva-Corvalán, 1998, p. 232)

(4) a. *Y el carro que compraste, ¿te gusta?*

‘and your car that [you] bought, do you like it?’

b. *¿Cómo te gusta el café?*

‘How do you like your coffee?’

(Silva-Corvalán, 1998, p. 232)

In non-contact varieties of Spanish, the interrogative pronoun *cómo*, ‘how’, is not grammatically acceptable in the context exemplified in (3), though this use of *cómo* is perfectly acceptable in LA Spanish. Alternatively, the closed interrogative phrase is the prescriptive structure required in General Spanish, as illustrated in (4a). However, Silva-Corvalán (1998) argues that the structure “*cómo X gustar Y*” does not constitute an example of syntactic borrowing because this structure already exists in other contexts in general Spanish, which is illustrated in (4b). What has been borrowed is another definition of the word *cómo* after the English model, in which the interrogative word ‘how’ may also mean “to what extent” (p.232). Ultimately, on the basis of this example and others like it, Silva-Corvalán concludes that (3) is a case of lexical borrowing.

Additionally, ostensible structural transfer may be attributable to a relaxation of the semantic and pragmatic restrictions that govern the behavior of a particular structure in lieu of direct structural transfer. For instance, in non-contact varieties of Spanish, the subject of intransitive verbs may be collocated in either pre-verbal or post-verbal position (i.e., SV or VS order) depending largely on the informational status of the subject.<sup>8</sup> However, in bilingual Los Angeles Spanish, Silva-Corvalán (1994) observes SV configurations that are both more frequent and agnostic to the aforementioned pragmatic constraints. Because SV order is already available in Spanish, Silva-Corvalán (1994) ultimately attributes the discordant distribution of SV/Vs orders in Los Angeles Spanish to a disregard of pragmatic rules, not syntactic transfer.

<sup>8</sup> The accusativity of the verb also impacts the respective order of the subject and verb (Bosque & Gutiérrez-Rexach, 2008).

To summarize, the contact situation type governs the applied mechanism of contact-induced change— borrowing in situations of language maintenance and substratum influence in situations of language shift. The primary difference between the proposed mechanisms of contact-induced change espoused by Thomason and Kaufman (1988) on the one hand and Winford (2003) and Silva-Corvalán (1994, 1998, 2008) on the other, is their perspective on structural borrowing. For Thomason and Kaufman (1988), there are few restrictions on structural borrowing.<sup>9</sup> Winford and Silva-Corvalán, however, argue that structural borrowing cannot occur in its own right and must be incited by lexical borrowing or other indirect processes, like the loosening of semantic and/or pragmatic restrictions.<sup>10</sup>

### 1.2.2 Agents of contact-induced change

Irrespective of the dissenting views on the mechanisms responsible for contact-induced change, there is a general consensus regarding the locus of language change— the individual. For instance, Hickey (2010) states, “it is convenient shorthand to claim, for example, that language A borrowed from language B. However, this is already an abstraction as the appearance of borrowings in a speech community can only be the result of actions by individual members of this community” (p. 8). Additionally, Penny (2000) parallels language change to a chain reaction in which an innovative feature begins with a single individual, who spreads this feature through face-to-face contact<sup>11</sup> to other individuals who then imitate the innovative feature. The original innovation is considered a contact-induced change only when a speech community has incorporated the feature into their shared linguistic code (Milroy, 1992; Poplack & Levey, 2010).

To be more precise, contact-induced change commences with the bilingual (or multilingual) individual (Weinreich, 1953; Winford, 2003), whose proficiency in two or more linguistic systems facilitates the conception of nascent innovations. Recall that, for Thomason and Kaufman (1988), bilingualism is a necessary condition for more intense (i.e., structural) linguistic transfer under the borrowing mechanism of contact-induced change. This is because, in borrowing situations, the agents of change are the target (i.e., recipient) language speakers (van Coetsem, 1988), who must have sufficient proficiency in the source language to identify potential borrowings. Conversely, in situations of language shift, structural change is promulgated via the agentivity of source language speakers. That is, the learner versions of shifting speakers, which are developed through a process of group second language acquisition, are imitated by native speakers of the recipient language (Winford, 2003). Simply put, in situations of language shift (i.e., substratum influence) the agents of change

<sup>9</sup> It is important to note, however, that Thomason and Kaufman (1988) do propose that lexical borrowing precedes structural borrowing when the borrowing mechanism is responsible for contact-induced change.

<sup>10</sup> To summarize her position on structural borrowing, Silva-Corvalán (1998) writes: “mine is an argument against the existence of syntactic *borrowing* and not against syntactic change resulting from lexical borrowing or from the transfer of discourse or pragmatic constraints from another language” (p. 226).

<sup>11</sup> Penny (2000) notes that linguistic change may be promulgated through mediums engendered by modern technological advancements (e.g., the internet), though he ultimately argues that face-to-face contact remains the primary avenue through which language change is spread.

are bilinguals dominant in the source language, and in situations of language maintenance (i.e., borrowing) the agents of change are bilinguals dominant in the recipient language.

### 1.2.3 Sociohistorical aspects of language contact

An evaluation of the sociohistorical context of a contact situation is indispensable in interpreting the linguistic outcomes that result. In fact, Thomason and Kaufman (1988) assert, “it is the sociolinguistic history of the speakers, and not the structure of their language, that is the *primary determinant* of the linguistic outcome of language contact” (p. 35, emphasis mine). The authors propose that factors like the duration of the contact situation, the relative prestige of the contact languages, pervasiveness of sociopolitical and cultural pressure, speaker attitudes, and the size of the target language and substratum populations both characterize a given contact situation and determine the directionality and type of contact-induced change. In their framework, Thomason and Kaufman conflate many of these individual determinants of the sociolinguistic context (but most crucially duration and level of bilingualism) into a single variable, **intensity of contact**. They claim that intensity of contact<sup>12</sup> interacts with borrowing (i.e., language maintenance) and substratum influence (i.e., language shift) unequally.

<sup>12</sup> It is beyond the scope of Thomason and Kaufman’s work to offer a predictive model of contact-induced change based on the numerous sociohistorical factors that characterize a contact situation. They state, “we do not believe that the information now available [on social circumstances] permits broad generalizations to be made about their effects on the linguistic outcomes of contact situations” (p. 46).

For situations of language maintenance, the authors propose a Borrowing Scale that includes two interrelated axes— intensity of contact and borrowed structures. The intensity of contact is categorized using five discrete labels— 1) casual contact, 2) slightly more intense contact, 3) more intense contact, 4) strong cultural pressure, and 5) very strong cultural pressure— that correspond chronologically to five types of borrowed structures— 1) lexical borrowing only, 2) slight structural borrowing, 3) slightly more structural borrowing, 4) moderate structural borrowing, and 5) heavy structural borrowing (pp. 74-76). Only in the context of language maintenance is intense transfer determined by a long duration of contact and widespread bilingualism. In situations of language shift, on the other hand, a longer shift duration (i.e., several generations) and extensive bilingualism are associated with a lesser likelihood of structural influence because this duration of time permits “perfect” second-language acquisition and complete bilingualism amongst source language speakers (p. 47). Furthermore, disparities in the size of the source language and recipient language speaker groups shape linguistic outcomes. For instance, “if an entire large population shifts to the language of a much smaller group of conquering invaders over one or two generations, the shifting speakers are unlikely to be fully bilingual in the TL before they abandon their native language, so we can expect to find

extensive substratum interference” (p. 120). In essence, the intensity of contact yields opposing outcomes in situations of language maintenance versus language shift.

Similarly, Winford (2003) identifies various social motivations that pertain to distinct types of contact situations. For example, in situations of “distant contact,” “need and prestige” are cited as the two primary motivators of lexical borrowing (p. 37). The former motivation applies when new labels are required to refer to “new things, places, and concepts” (Weinreich, 1953, p. 56), which is evident in the cross-linguistic diffusion of contemporary English terms describing concepts integral to the recent technological revolution involving internet-based innovations, like the names of social media platforms (e.g., Facebook, TikTok, Instagram). In bilingual communities, the social motivations for borrowing are more numerous. These motivations include cultural pressure, language attitudes, degree of social solidarity, socioeconomic status of speaker groups, geographical distribution of speaker groups, and degree of support by governmental institutions (e.g., state-sponsored bilingual education and language policies), among others (Appel & Muysken, 2005, pp. 33–38). In situations of structural convergence, many of these motivations apply, in addition to others like the relative size of contact groups, power dynamics, frequency of interaction between groups, directionality of shift, and degree of stability (Winford, 2003, p. 90).

#### **1.2.4 Contact-induced word order change**

In addition to the sociohistorical context of a contact situation, the typological configuration of the contact languages determines, at least to some degree, what can be transferred from the source language to the recipient language (Silva-Corvalán, 1994, 1998, 2008; Thomason, 2010; Thomason & Kaufman, 1988; Winford, 2003).<sup>13</sup> Broadly speaking, there is a positive relationship between typological similarity and likelihood of linguistic influence, which is captured in one of Winford’s proposed “morphological constraints” on borrowing— “the greater congruence between morphological structures across languages in contact, the greater ease of borrowing” (p. 93). With respect to contact-induced word order change, Thomason and Kaufman (1988) invoke the concept of functional congruence to account for the prevalence of clausal word order borrowing that transpires in language contact situations:

The evidence we have collected does not support the often implicit assumption, in the literature on word order change, that word order patterns constitute a fundamental “deep” structural feature

<sup>13</sup> Thomason (2010) clarifies that the degree of typological distance “helps to predict the kinds of interference that may occur,” but does not necessarily motivate contact-induced change per se (p. 40).

relatively impervious to foreign influence. On the contrary, word order seems to be the easiest sort of syntactic feature to borrow or to acquire via language shift. This is not surprising when it is considered from the viewpoint of functional congruence: dominant SVO and SOV word order patterns, after all, typically perform the same basic syntactic function— identification of subject and object by their position relative to each other and to the verb (pp. 54-55).

According to Thomason and Kaufman's Borrowing Scale, impacts on word order begin at level 2 (slightly more intense contact, slight structural borrowing) and become more extensive by level 4 (strong cultural pressure, moderate structural borrowing) in situations of language maintenance. In shifting situations, word order changes also constitute moderate to heavy transference. Thomason and Kaufman (1988) describe word order changes as a type of syntactic transfer that "involve[s] simple replacements, neither simplifying nor complicating" (p.129), which explains the reported ease with which word order patterns evolve via language contact in comparison to other syntactic changes that require more intense contact (Thomason, 2001). Thus, contact-induced word order change appears to be the exception to the relative "impermeability" of syntactic systems (Silva-Corvalán, 1993).

Furthermore, syntactic flexibility in a recipient language is attributed to promoting contact-induced word order change, "as more flexible languages are more likely to have convergent structures with their contact languages" (Levshina et al., 2023, p. 847). By way of example, if language A (recipient) is an SOV-dominant language that is in contact with SVO-dominant language B (source), and if word order in language A is at least marginally flexible, it is likely that SVO is an existing word order pattern in language A prior to contact. The existence of this parallel structure may facilitate the transfer of SVO order in the process of borrowing or shift. In this case, the outcome of language contact may be a change in the frequency of a particular structure. According to Heine (2008, p. 55), "what frequently happens is that speakers draw on a minor use pattern— one that has a more marginal status, being used rarely and/or only in specific contexts only to build a new major use pattern by increasing the frequency of use and extending the range of contexts in which it may occur." Thus, contact-induced word order change may simply refer to the increased prevalence of an existing word order configuration of the recipient language, not necessarily the introduction of this pattern into the recipient language for the first time. However, I clarify in the next section that linguistic variation does not necessarily constitute contact-induced change.

### 1.2.5 Establishing contact-induced change

Though variation is an inherent property of all languages (Levshina et al., 2023), certain variable structures may be attributable, at least partially, to language contact. To first establish that a particular linguistic structure constitutes a change at all, a comparison between the modern contact variety and an earlier stage of the contact variety is obligatory (Poplack & Levey, 2010). According to Poplack and Levey, it is not sufficient to utilize the prescriptive norm offered by language grammars (or similar resources) as a point of comparison, given that these norms may not reflect the manner in which the language is (or was) actually spoken. To exemplify an appropriate vertical comparison, Otheguy et al. (2007) compare the speech of older and younger generations in considering the incidence of overt pronouns in New York Spanish to determine that younger speakers employ overt pronouns at a rate that surpasses that of the older speakers. To establish that the confirmed change is contact-induced, this vertical comparison must be accompanied by a horizontal comparison, which considers the behavior of the relevant feature in the source language and a non-contact variety of the recipient language, to conclusively rule out the possibility that language-internal mechanisms are solely responsible for the change (Poplack & Levey, 2010; Thomason, 2001). Poplack and Levey (2010) offer the following criteria for establishing contact-induced change:

A candidate for contact-induced change in a contact variety is present in the presumed source variety and either 1) absent in the pre-contact or non-contact variety, or 2) if present (e.g., through interlingual coincidence), is not conditioned in the same way as in the source, and 3) can also be shown to parallel in some non-trivial way the behavior of a counterpart feature in the source (p. 398).

Satisfying the above criteria necessarily requires statistical analyses that consider both the “trajectory of the variant use over time” and the conditioning effect of independent predictors (e.g., linguistic and social factors) on variation of the relevant feature in both contact languages (p. 394).<sup>14</sup> To illustrate the application of these criteria, Poplack and Levey (2010) review the case of relative marker expression in Quebec English. Though null relative markers<sup>15</sup> are ungrammatical in prescriptive French, they are permitted in Quebec French. Through a comparative analysis of “mainstream” English and French, it is found that null relative markers are most strongly conditioned by discrete predictors in each language—intervening material and phonological environment, respectively. Though overall, relative marker expression in Quebec English is constrained by similar factors that govern mainstream English, the young, highly

<sup>14</sup> Specifically, the authors recommend a multivariate analysis like *variable rule analysis*.

<sup>15</sup> The following is an example of a null relative marker in English provided by Poplack and Levey (2010, p. 407): *I was always known as Carrot Head and Freckle Face and a few other very vulgar names Ø they called me.*

bilingual population (who are typically the agents of contact-induced change) diverge from older, less bilingual speakers regarding the effect of the phonological environment, which parallels mainstream French. However, upon taking a closer look at younger mainstream English speakers as well, phonological environment is also found to condition relative marker expression within this group. Thus, though on the surface variable relative marker expression appears to be a linguistic site of contact-induced change, variation in this domain is actually attributable to an internal change in English. Importantly, in this analysis, variable relative marker expression fails to be categorized as a contact-induced change in Quebec English because, even though the authors establish that this structure is **present in the source language** (i.e., Quebec French) and **parallel to the counterpart feature in the source** (i.e., both are constrained by phonological environment), this feature does not meet the criteria of being **absent in the pre-contact variety or not being conditioned in the same way in the pre-contact variety** (i.e., variable relative marker expression is a feature of young, mainstream English that is also governed by the phonological environment). Ultimately, Poplack and Levey (2010) argue that these types of comparisons and statistical analyses are compulsory to verify that a variable linguistic structure is indeed the result of language contact and not other language internal processes.

### 1.3 The Quechua-Spanish contact situation

In §1.2.3, it was determined that a thorough understanding of the sociohistorical context that characterizes a given contact situation is integral to effectively interpret the resultant linguistic outcomes. Consequently, in this section I offer an overview of both the history of Quechua-Spanish contact and the linguistic outcomes engendered by this contact situation, including the formation of a contact variety, Andean Spanish. To begin, I clarify what is meant by the term ‘Quechua’.

#### 1.3.1 What is Quechua?

Historically speaking, Quechua is the language of the Incan empire, which is spoken by an estimated 6-12 million people in the present day depending on the source (Escobar, 2011; Hornberger & Coronel-Molina, 2004; Klee & Lynch, 2009; Mannheim, 1991). Linguistically speaking, the term ‘Quechua’ refers to the language family comprised of many dialectal varieties spoken across six countries in South America— Colombia, Ecuador, Peru, Bolivia, Chile, and

Argentina. The Quechua language family is often partitioned into two main dialectal groups— Quechua I and Quechua II— though the labels applied to these groups vary by author (Cerrón-Palomino, 1987a; Mannheim, 1991; Parker, 1963; Torero, 1964, 1974). There is a general consensus among these historical Quechua linguists, however, that the regional varieties of Quechua spoken in central Peru (Quechua I) are distinct from those spoken in the northern and southern peripheries (Quechua II), to the point of mutual unintelligibility (Escobar, 2011; Hornberger & Coronel-Molina, 2004). The approximate distribution of Quechua I and II are illustrated in Figure 1.1 below.



Figure 1.1: Distribution of Quechua and Aymara in the mid-20th century (Ade-laar, 2004, p. 169)

Torero (1964, 1974) further partitions Quechua II into three groups— A, B, and C— which correspond to specific geographical regions where these varieties are spoken. Quechua II-C, which is the most dissimilar from Quechua I, encompasses the varieties spoken in southern Peru, like Ayacuchano and Cusqueño Quechua, in addition to Bolivian and Argentinian varieties. Considering only the Peruvian Quechua II-C dialects, Mannheim (1991, p. 5) applies the label ‘Southern Peruvian Quechua’ to the regional variety spoken in the “contiguous highlands of Apurímac, Arequipa, Ayacucho, Cusco, Huancavelica, and Puno,” which is believed to most closely resemble the Incan *lingua franca*. At his time of writing, Mannheim (1991) estimated that this dialect had a speaker base of about 2 million.<sup>16</sup> Delimiting further, Cusco-Collao Quechua is the dialect of Southern Peruvian Quechua spoken in Arequipa, Cusco, and Puno, which is distinguished from Ayacucho-Chanka Quechua (another dialect of Southern Peruvian Quechua) by several phonological features.<sup>17</sup> Cusco-Collao Quechua is the regional variety investigated in the present study, which will be referred to simply as ‘Cusco Quechua’ in accordance with previous studies on word order in Quechua that label the variety using the corresponding place-name (Kalt & Geary, 2021; Sánchez, 2003).

<sup>16</sup> According to the 2017 Peruvian Census, the sum of individuals (5 years of age or older) that identify Quechua as their maternal language in these departments is nearly 2.2 million (Instituto Nacional de Estadística e Informática, 2017).

<sup>17</sup> For instance, there are three types of stop consonants— ejective, aspirate, and plain— in Cusco-Collao Quechua, while Ayacucho-Chanka Quechua has only plain stop consonants (Mannheim, 1991, p. 15).

<sup>18</sup> The historical sketch of the rise of the Incan empire provided here is informed by Lanning (1967).

### 1.3.2 Sociohistorical context

Now I turn to a brief overview of the history of Quechua-Spanish contact spanning from the rise of the Incan empire to the present day.

#### Pre-Spanish contact: Quechua in the Incan empire

Following the fall of the Huari and Tiahuanco empires during the Late Intermediate Period (1000AD – 1476AD), the southern highlands of modern-day Peru were devoid of a consolidated power.<sup>18</sup> The social organization of this region was characterized by many discrete, small tribal states, one of which were the Incas. Though feuding among tribal states was commonplace, the Incas differed from their autonomous neighbor groups in that they desired to incorporate conquered neighboring territories into their own state. By way of this mentality, the Incas swiftly became the strongest military power in the region. During the expansion of the Incan state beginning in the early 15th century, the Incas encountered considerable resistance from other tribal groups (e.g., the Chancas in the Apurímac Basin); however, over the next 25 years, the Incas would conquer the Titicaca, Apurímac, Mantaro, and Urubamba Basins. In the 1460’s, the Inca military power split into two branches, one which organized and strengthened the administrative system of the capital in Cusco, and another

which expanded the empire to the north through present-day Cajamarca, Quito, and the nearby coastal region, ultimately overthrowing the robust Chimú Empire. Like the smaller tribal states before them, the Chimú Empire was subjugated by the Incas through the military efforts of Topa Inca, son of Pachacuti, who also acquired territories in the south coast of Peru, highlands of Bolivia, northwestern Argentina, and northern Chile by 1471. The expansion of the Inca Empire (i.e., Tawantinsuyu) culminated in 1483, with the incorporation of small regions in both northern Ecuador and its southern coast by Huayna Capac, son of Topa Inca. The population of the Inca Empire at its height is estimated at approximately 6,000,000 people.

The primary urban center, Cusco, was the administrative and religious center of the Tawantinsuyu, where only the Emperor and upper nobility resided. The rest of the population inhabited either villages surrounding the capital or small towns<sup>19</sup> scattered throughout the empire. These towns were mainly comprised of commoners, like farmers, herders, or fisherman, depending on the topographical features of the area, though *curacas*— local or regional government officials— were also dispersed throughout rural towns. *Curacas* were considered members of the lower nobility whose primary role was to facilitate the state-controlled redistributive economy between the capital and the local, autonomous communities. Though the Incas retained their conquered territory, there was little emphasis on urbanization, preferring instead a rural mode of living.

<sup>19</sup> Lanning (1967) asserts that these small towns likely did not exceed 1,000 inhabitants.

The extent to which the Incas imposed linguistic dominance throughout the empire is debated. For instance, Lanning (1967) claims that “the new emperors imposed the Incan language (usually called Quechua) to facilitate trade and administration” (p. 13). Similarly, Klee and Lynch (2009) state that the Incans had begun to prohibit the use of the languages of subjugated populations in favor of Quechua shortly before the arrival of the Spanish (p. 129). Mannheim (1991), however, describes the linguistic landscape of the Incan empire quite differently:

...[the Inkas] were able to annex their populations through a form of indirect rule, in which the relationships between local communities and the state were mediated by local ethnic lords. The state maintained, and perhaps even promoted, cultural and linguistic diversity among the conquered populations. Although local administrators were required to learn the Inka lingua franca, there is no evidence that the Inkas imposed their language and culture on the conquered peoples on a mass scale (p. 18).

In fact, this *laissez-faire* linguistic policy was reportedly instituted in the early days of the Incan Empire expansion as a means of keeping record of their subjugated populations.

“Each local group was associated with a way of speaking, as well as a mythological place of origin, a style of dressing and braiding the hair, and pattern of cranial deformation. According to the indigenous chronicler Santa Cruz Pachacuti Yamqui (c. 1613), the legendary founder of the Inka dynasty, Manco Qhapaq, ordered that the clothing and speech of each people be different so that its place of origin would be easily identifiable” (Mannheim, 1991, p. 50).

According to Mannheim (1991), linguistic diversity flourished in all corners of the Tawantinsuyu, including within the communities located near the capital. In fact, the linguistic expansion of Quechua across the conquered territories of the Incan Empire would proceed only after the arrival of the Spanish less than a century after the consolidation of Incan power.

### **Early Spanish contact**

<sup>20</sup> The reported year of the Spanish arrival ranges from 1531 to 1533, depending on the source (Klee & Lynch, 2009; Lanning, 1967; Mannheim, 1991).

The Spaniards, led by Francisco Pizarro, arrived to modern-day Peru in 1532.<sup>20</sup> Though the Incan military was highly organized and greatly outnumbered the Spanish troops, the Spanish were able to conquer the entire empire in about six years, effectively driving the Incan government into exile in the mountains (Lanning, 1967, p. 171). Lanning cites four primary reasons for the speedy transition of power. First, the Spanish arrived only five years after the death of Huayna Capac, the penultimate Inca emperor, who failed to choose a clear successor to the throne prior to his death. This indecision sparked a full-blown civil war between his two sons—Huascar and Atahualpa—meaning the empire was already fragmenting at the time of the Spanish arrival. Second, many of the conquered tribal states were indignant about their subjugation by the Inca Empire, and the Spanish strategically formed alliances with such populations. In fact, Lanning (1967, p. 171) describes the Spanish army prior to the arrival of Spanish reinforcements as an “Indian army under Spanish control.” Third, the Incas were already succumbing to the diseases introduced by the Europeans, like small pox and measles. In fact, it is attested that these epidemics are what took the life of Huayna Capac five years prior to the arrival of Francisco Pizarro and his troops. Fourth, the diplomatic tactics used by the Spanish differed from those of the Incas. When conquering new territory, the Incas tended to announce their arrival and intentions to the target tribal state, effectively frightening them

into submission. The Spanish, however, employed more manipulative tactics that were unheard of at this time, like hiding their intentions under “the guise of diplomatic mission” (p. 171). From this point forward, I will focus on the history of Quechua-Spanish contact in modern-day Peru, as this is both where the majority of contemporary Quechua speakers are concentrated and the geographical locus of investigation in the present study.

As the Spanish gained a sense of the linguistic landscape of the Incan Empire throughout their conquest, they were reportedly astonished by the lack of homogeneity and interpreted the “linguistic diversity of native South America as a sign of the degree of barbarianism of its inhabitants” (Mannheim, 1991, p. 36). From the Spanish perspective, the lack of linguistic domination was a great failure of the Incan Empire. Paradoxically, it was the Spanish administration that imposed Quechua across the territories recently secured by the Incans, swiftly displacing the existing local vernaculars and tribal cultures (Klee & Lynch, 2009; Lanning, 1967). Initially, Quechua was adopted as the *lingua franca* for communication within administrative, political, and religious domains for some time prior to the enforcement of Hispanization efforts. Though Quechua enjoyed a certain degree of prestige during this time, Mannheim (1991, p. 62) argues that Quechua should nevertheless be regarded as an oppressed language because, since initial Spanish contact, those making decisions about the fate of Quechua have not historically been its own speakers.

The next several centuries of Quechua-Spanish contact were characterized by contrasting policies and interests.<sup>21</sup> The official policy espoused by the Spanish Crown was the prioritization of religious and linguistic conversion through Hispanization (Heath & Laprade, 1982). For instance, in 1634, from across the ocean Felipe IV declared that all conquered indigenous people must acquire Spanish and be indoctrinated in the Christian religion. However, this decree was followed by the Quechua Golden Age (1680-1770), in which Quechua continued to extend to the corners of the former Incan Empire, and Quechua literature and theatre flourished (Klee & Lynch, 2009; Mannheim, 1991). In 1770, a stricter linguistic policy was decreed by Carlos III, which required the complete abandonment of indigenous languages in favor of Castilian monolingualism. Again, despite the Crown’s mandates, linguistic assimilation did not ensue. These policies were not enforced in the New World in order to maintain the social and economic distance that had emerged between the urban, Spanish-speaking elite and the rural, Quechua-speaking lower class. To this end, Quechua speakers were denied access to education in Spanish to ensure their continued ignorance of the prestige language, consequently strengthening the sociopolitical stronghold of the Spanish-speaking upper class.

<sup>21</sup> See Mannheim (1991, pp.34-35) for an elaboration of these policies.

Quechua monolingualism among the rural lower class persisted well into the 20th century (Escobar, 2011; Klee & Lynch, 2009). Even after Peru gained its independence from the Spanish Crown in 1820, the local elite— *hacendados* (landowners)—continued to limit community access to Spanish in order to strengthen the superiority of their social position as intermediaries between indigenous campesinos and government officials (Klee & Lynch, 2009). Though sparse bilingualism emerged in the colonial period among select local residents, like the children of indigenous elites, religious workers, and medical center assistants (Mannheim, 1991), Escobar (2011) describes bilingualism in this region as an “individual phenomenon” until the 1950s. At this time, “the role of central government strengthened in the provinces, the caciques lost their power, and an intense campaign of Hispanization of the indigenous populations spread throughout the south Andes” (Klee & Lynch, 2009, p. 132, translation mine). This campaign garnered more success than its predecessors, likely due to the parallel operative social processes of globalization, modernization, democratization, and urbanization (Escobar, 2011; Hornberger & Coronel-Molina, 2004; Klee & Lynch, 2009).

### Contemporary Quechua-Spanish contact

Peruvian census data corroborates the view that a shift from Quechua monolingualism to Quechua-Spanish bilingualism and Spanish monolingualism is a fairly recent phenomenon. Table 1.1 below provides census data concerning the linguistic profile of the Peruvian population (over 5 years of age) spanning from 1940 to the most recent census conducted in 2017 (Instituto Nacional de Estadística e Informática, 2017). Unfortunately, census data from 1940, 1961, and 1981 can not be directly compared to the more recent data from 1993, 2007, and 2017 due to the lack of consistency regarding question phrasing.<sup>22</sup> In the earlier set of years, separate figures were reported for Quechua monolinguals, Spanish monolinguals, and Quechua-Spanish bilinguals. In the table below, the label ‘Quechua’ for these earlier years refers only to the proportion of Quechua monolinguals. Beginning in 1993, census takers were asked to report their maternal language (i.e., the language with which they learned to speak), and they were unable to provide more than one language in their response. For these later years, the label ‘Quechua’ conflates Quechua monolinguals and Quechua-Spanish bilinguals. Thus, the data corresponding to these years obscures the linguistic reality of Peru.<sup>23</sup>

<sup>22</sup> See Hornberger and Coronel-Molina (2004) for more information on census data inconsistencies and inaccuracies.

<sup>23</sup> Because a small proportion of Peruvian census-takers speak languages other than Quechua and Spanish (e.g., Aymara), the percentages provided in Table 1.1 do not add up to 100% in each column.

	1940	1961	1981	1993	2007	2017
Spanish	47%	60%	73%	80.3%	83.9%	82.6%
Quechua	31%	17%	8%	16.6%	13.2%	13.9%
Bilingual (Q-S)	16%	16%	14%	–	–	–
<b>Total Quechua</b>	<b>47%</b>	<b>33%</b>	<b>24%</b>	<b>16.6%</b>	<b>13.2%</b>	<b>13.9%</b>

Table 1.1: Language profile of Peruvian population (5 years or older) from 1940 to 2017 (partially adapted from Klee and Lynch (2009, p. 133))

Notwithstanding the methodological incongruities of the Peruvian censuses, it is clear that there has been a sharp decline in the Quechua-speaking population from the mid-20th century to the present day. In 1940, 31% of the population identified as Quechua monolinguals and 16% identified as Quechua-Spanish bilinguals, meaning half (47%) of the total population of Peru at this time was Quechua-speaking. Approximately 20 years later, this figure dropped to a third of the population, and then to a quarter another 20 years later. Thus, in less than two generations, the proportion of the Quechua-speaking population decreased by nearly 50%. Another important trend to note from the census data is the relative stability of the Quechua-speaking population in recent years. Though from 1993 to 2007, there was approximately a 20% decrease in the Quechua-speaking population, from 2007 to 2017, there was very little change in the Quechua-speaking population, proportionally speaking. Though it is unclear what proportion of the Quechua-speaking population is bilingual versus monolingual in the census data collected from 1993 onward, based on the social and demographic trends detailed in the coming paragraphs, I assume that the bilingual Quechua-speaking population has remained stable at the expense of the monolingual population.

In Andean departments, the contemporary presence of Quechua speakers is more notable. In Cusco, for example, 55.3% of the population claimed Quechua as their maternal language, compared to 41.8% who selected Spanish as their maternal language (Instituto Nacional de Estadística e Informática, 2017). In two provinces within the department of Cusco, Canas and Paruro, nearly all residents identify Quechua as their maternal language (91.4% and 91.3%, respectively).

Despite the prevalence of Quechua speakers in Andean communities in the present day, many scholars describe the Andean linguistic situation as diglossic; that is, Spanish is the prestige language that dominates in formal sectors (e.g., government, politics, education), while Quechua is the stigmatized language of a minoritized population (Cerrón-Palomino, 1987b; Hornberger & Coronel-Molina, 2004). Mannheim (1991) employs the term ‘linguistic hegemony’ in

describing the Quechua-Spanish contact situation to emphasize the pervasive dominance of the Spanish language over Quechua in all bilingual encounters. He writes,

Social domination is a lived reality every day in the lives of every single Quechua speaker. It is reproduced in different ways in every act of speaking, be it in Spanish or Quechua; in the choices that bilinguals make to speak Quechua; in the gnawing closure that it places on the lives of Quechua monolinguals; in the silence of Quechua-speaking children in a Spanish-speaking classroom; in the pride with which a Quechua peasant *writes* his name; in the shame of a military draftee who cannot understand the Spanish spoken around him; in crude bilingual puns used in the marketplace; in the insults shouted at a monolingual Spanish speaker who cannot understand them, since that is the only way that she can be insulted with impunity, and in countless other situations (Mannheim, 1991, p. 81).

One of the more extreme interpretations of the status of Quechua is its classification as an endangered language on the basis of current linguistic ideologies and practices (Hornberger & Coronel-Molina, 2004).<sup>24</sup> In recent decades, cultural and linguistic inter-generational transmission has been interrupted, as Quechua-speaking parents prioritize Spanish acquisition on their children's behalf. Since upward socioeconomic mobility can be achieved only through knowing Spanish, Quechua speakers feel pressured to abandon their mother language and assimilate in order to experience the economic and educational opportunities that are otherwise lacking for Quechua speakers. In fact, the Quechua-speaking groups that most zealously support bilingual education are young, economically-secure bilinguals, suggesting that support for the inter-generational transmission of Quechua is offered primarily by those who can afford to do so.

The modern process of urbanization may also threaten the status of the Quechua language. In the late 20th and early 21st centuries, the rate of urbanization has completely altered the demographic composition of Peru. In 1940, about a third of Peruvians (35%) inhabited urban areas, while the remaining proportion (65%) inhabited rural areas. However, in 2017, the grand majority of Peruvians, 82.4%, were concentrated in urban areas (Instituto Nacional de Estadística e Informática, 2017). Upon migrating to larger cities, Quechua speakers tend to shift quickly to Spanish (Klee & Caravedo, 2005, 2006; Klee et al., 2011), as knowing the prestige language is required to access many jobs

<sup>24</sup> The authors clarify that Quechua varieties spoken in Colombia and Argentina have a higher risk of extinction than the varieties spoken in Ecuador, Peru, and Bolivia.

and superior educational opportunities, particularly beyond primary-level education. Similarly, through the process of globalization and the expansion of communication networks, rural Quechua speakers have become more aware of the socioeconomic gaps that separate them from their urban-living, Spanish-speaking countrymen. Furthermore, though processes like modernization and democratization engender the positive consequence of alliance formation between minoritized groups and international human rights organizations that work to diminish these gaps, these social processes also bring about a fierce threat to cultural and linguistic transmission through a stronger dependency on the majority culture (Hornberger & Coronel-Molina, 2004, p. 29). In order to reverse the trajectory of Quechua in the 21st century, Quechua must begin to enter into prestigious domains (e.g., literary spaces) and be adopted by people of the upper echelons of society to send the message that “Quechua is a language worth knowing and using” (Hornberger & Coronel-Molina, 2004, p. 53).

Escobar (2011) offers a more optimistic perspective on the current status of Quechua. She emphasizes that, though Quechua monolingualism may be on the decline, Quechua-Spanish bilingualism appears to be stabilizing. Moreover, she reports that 41% of the indigenous population is 18 years of age and younger, pointing to a “robust young indigenous language-speaking population” (p. 336). She also rejects the notion that urbanization is a death sentence for Quechua, highlighting instead the growth of a “new urban Andean identity” that is defined by the integration of Quechua into popular music genres (e.g., rock), the growth of migrant social networks, the use of Quechua in public domains, and the weakening of negative social values associated with Andean Spanish. She even ventures to claim that “bilingualism without diglossia seems to be incipiently emerging in the Andean region” (p. 337). Perhaps because processes like modernization and globalization increase the degree to which individuals are exposed to diversity (cultural, linguistic, etc.), stigmatization is diminishing, and along with it, fear of the unknown.

Bilingual education is another domain that has witnessed significant improvements in the 21st century. Hornberger and Coronel-Molina (2004, p. 46) note that “when the Peruvian officialization of Quechua in 1975 mandated that Spanish-speaking children be taught Quechua in school, the outcry of limeños was such that the provision was quickly abandoned.” Nearly two decades later, the Peruvian constitution of 1993 acknowledged the pluricultural and plurilinguistic nature of the country by both recognizing indigenous languages as official in the regions where they predominate and guaranteeing access to bilingual education (Hornberger & Coronel-Molina, 2004). First, a transitional bilingual education curriculum was implemented, which prioritized cultural

assimilation and the displacement of indigenous languages in favor of Spanish (Escobar, 2011). However, the focus of bilingual education shifted in Peru in the early 21st century with the passing of the *Ley para la Educación Bilingüe Intercultural* ('Law for Intercultural Bilingual Education'). In opposition to transitional bilingual education, Intercultural Bilingual Education (IBE) emphasized the maintenance of indigenous languages and the active participation of indigenous communities in curriculum development. For example, Article 2 of this bill reads:

*El Plan Nacional de Educación Bilingüe Intercultural deberá incorporar, la visión y el conocimiento indígenas. La educación para los pueblos indígenas debe ser igual en calidad, eficiencia, accesibilidad y en todos los demás aspectos previstos para la población en general. El estado garantiza el derecho de los pueblos indígenas a participar en la administración de los sistemas e instituciones estatales de educación bilingüe intercultural, así como en los centros y programas de preparación de maestros bilingües interculturales* ("Ley para la educación bilingüe intercultural", 2002).

The National Plan of Intercultural Bilingual Education should incorporate the indigenous vision and knowledge. Education for indigenous communities should be equal in quality, efficiency, accessibility, and all other expected aspects for the general population. The state guarantees the right of indigenous communities to participate in the administration of the state systems and institutions of intercultural bilingual education, as in the preparatory centers and programs for intercultural bilingual teachers (translation mine).

According to the National Registry of Educational Institutions of Intercultural Bilingual Education (*Registro Nacional de Instituciones Educativas EIB*) of Peru, in 2023, 51 language varieties were represented IBE institutions, and over 1.2 million students were enrolled in IBE programs at the initial, primary, and secondary levels (Ministerio de Educación, 2023). Though the Peruvian government has become a more active participant in the maintenance of indigenous languages, there are still considerable shortcomings in the IBE system. Porter and Morrison (2024) note that, in Peru, only "38% of primary-school-aged Indigenous children have access to IBE" (p. 9). Furthermore, IBE does not guarantee equal academic outcomes as evidenced by the reported testing gaps between indigenous and non-indigenous students. These gaps may be a result of ongoing challenges, like the lack of classroom monitoring, improper

implementation of IBE curriculum, and poor instructor quality and retention. Additionally, IBE educators may not be appropriately trained due to a current lack of resources required to effectively implement the policies that have been put forth. In short, while there is room for improvement, the attitudes and efforts of the Peruvian government (e.g., the creation of indigenous language educational materials) reflect their ambition to revitalize and maintain the Quechua language (Hornberger & Coronel-Molina, 2004).

To summarize the Quechua-Spanish contact situation, prior to the arrival of the Spanish, linguistic diversity flourished in the Inca empire, with Quechua spoken by the Incans and local vernaculars spoken by recently conquered tribes. After their arrival, the linguistic landscape of this region was characterized by two successive phases of language displacement (Mannheim, 1991). First, Quechua (and Aymara) displaced local vernaculars and served as the *lingua franca* in administrative and religious spheres. Then, several centuries later, Spanish began to rapidly displace Quechua via Hispanization efforts mediated by modern social processes like urbanization, democratization, modernization, and globalization. During these periods of language shift, there was considerable social and linguistic separation between the urban, Spanish-speaking elite and the rural, Quechua-speaking peasantry. After the initial prestige of Quechua waned approximately a century following initial contact, bilingualism emerged as an individual phenomenon available only to a select few in the interest of maintaining socioeconomic distance between the two groups. The sociolinguistic relationship between Spanish and Quechua is often characterized as diglossic, though in recent years, there has been a concerted effort to move Quechua into prestigious spaces, like educational institutions, governmental organizations, mass media, and literary spheres. Though census data corroborates the claim that Quechua monolingualism has been supplanted by Spanish monolingualism, there is evidence of a stable and robust Quechua-Spanish bilingual population in Peru in the early 21st century.

Based on the sociohistorical context described above, I speculate that the Quechua-Spanish contact situation may be simultaneously considered an example of language shift and language maintenance. The shifting component of the contact situation began about 2-3 generations ago, when a relatively large group of monolingual Quechua speakers (31% of the Peruvian population in 1940) began acquiring Spanish rapidly to better position themselves and their children in the prevailing socioeconomic hierarchy. These factors, taken together with the inaccessibility of formal education among rural Quechua speakers at the time, suggest that substratum influence (i.e., group L2 acquisition) is likely the mechanism responsible for contact-induced change in the recipient language,

Spanish. The learner varieties of the original shifting speakers, which may be considered the incipient version of Andean Spanish, then became the new linguistic target for the next generation, and so on. The ongoing shift from Spanish to Quechua, however, has been interrupted by a period of stable bilingualism in recent decades. The existence of a robust bilingual population (along with a small degree of Quechua monolingualism that perpetuates in more isolated communities) suggests that the contemporary Quechua-Spanish contact situation may be currently classified as one of language maintenance. It remains to be seen if this maintenance of Quechua via bilingualism is merely an intermediary phase on the path to complete shift or the final destination.

If my conceptualization of the Quechua-Spanish contact situation is correct, contact-induced language change may be expected to proceed in each language under divergent mechanisms of transfer— substratum influence in Andean Spanish and borrowing in Quechua. In line with the framework of contact-induced change advanced by Thomason and Kaufman (1988), I hypothesize that, initially, Quechua-like structures were more pervasive in Andean Spanish due to the considerable size of the shifting group, speed of language shift, socioeconomic pressure, and inaccessibility of the target language. However, as Quechua-Spanish bilingualism began to stabilize, and dialect contact with non-contact varieties of Spanish became commonplace due to processes like globalization and modernization, the prevalence of Quechua-like features in Andean Spanish began to deteriorate. Conversely, because Quechua has been maintained in recent decades, mainly through bilingualism, I theorize that Spanish borrowings in Quechua have increased in number and kind (i.e., structural borrowing), given that long-standing bilingualism is typically characterized by more intense borrowing. I return to these postulations in the discussion and conclusion chapters.

### **1.3.3 What is Andean Spanish?**

Andean Spanish is a contact variety of Spanish molded by contact with Quechua. Though a single label is applied to this variety, Andean Spanish is far from monolithic. Consider the description of Andean Spanish offered by Escobar (2011) below:

In defining [Andean Spanish] as a dialect, we are acknowledging that it has native speakers who grew up speaking Spanish in this region. These native speakers include both monolingual speakers as well as what is called 2L1 bilinguals in language-acquisition studies, that is, individuals who have two native languages— Quechua

and Spanish. However, since this region is also home to Andean languages, native speakers of Andean Spanish are in daily contact with second-language speakers of Andean Spanish, who are native speakers of Quechua (or Aymara). Due to migration patterns within these countries, all types of speakers of Andean Spanish are also found outside of the Andes. This makes it necessary for researchers to carefully consider the sociolinguistic background of their subjects when studying Andean Spanish (p. 327).

Though Escobar (2011) recognizes that Andean Spanish speakers may be either monolingual or bilingual, she distinguishes 2L1 bilinguals from L1 Quechua bilinguals who acquire Andean Spanish as a second language, and are thus L2 speakers of Spanish. In this quote, dialect mixing is also acknowledged as a factor shaping Andean Spanish, especially in modern times. As mentioned in the previous section, the process of urbanization has accelerated contact between Andean Spanish and non-contact, prestige dialects, which likely engenders accommodation, and ultimately, a potential loss of dialectal variation (Penny, 2000). For example, regarding the speech of Andean migrants in Lima, it has been observed that some phonological and syntactic features of Andean Spanish (e.g., palatal lateral [ʎ] and frequent pre-verbal objects) are displaced by Limeño features within a generation, but other morphosyntactic and semantic features (e.g., simplification of the object clitic system and novel uses of perfect aspect) are maintained (Caravedo & Klee, 2012; Klee & Caravedo, 2005, 2006; Klee et al., 2011; Rojas-Sosa, 2008).

Furthermore, the variety of Andean Spanish spoken in a particular community is molded by its distinct speakers. According to Muysken (1984), “two speech communities may have different vernacular varieties, while they result from the interaction and contact of the same L1 and L2” (p. 103). Within a bilingual speech community, speakers may exhibit varying levels of proficiency in the L2, and it is up to the community to decide (subconsciously) which of the innovative, non-standard features put forth by these L2 speakers will become adopted in subsequent generations. It is by this process that the non-standard, learner features of the interlanguage<sup>25</sup> are integrated into the speech of L1 speakers of the contact variety. Considering the extant literature on word order variation in Andean Spanish alone, the extent to which an Andean Spanish speaker employs non-contact features is further compounded by individual social factors like their sex, socioeconomic status, and linguistic profile.<sup>26</sup> In the paragraphs that follow, I offer a sampling of the linguistic features that characterize Andean Spanish in addition to select outcomes of contact in the reverse direction, from Spanish to Quechua.

<sup>25</sup> The term ‘interlanguage’, originally coined by Selinker (1972), refers to the innumerable iterations of the learner varieties of a target language molded through the process of second language acquisition.

<sup>26</sup> By linguistic profile I mean the number of languages that are known by an individual, in addition to the sequence in which these languages (if more than one) were acquired.

### 1.3.4 Linguistic outcomes of Quechua-Spanish contact

The reported linguistic outcomes of Quechua-Spanish contact are numerous. Though the relevant literature accounts for linguistic outcomes in both contact languages, the majority have tended to the novel structures of Andean Spanish.<sup>27</sup> In Andean Spanish, all linguistic subsystems have presumably been affected by contact with Quechua. In the phonological system, Andean Spanish is characterized by unstressed vowel reduction and devoicing (Delforge, 2008, 2012), maintenance of the palatal lateral [ʎ] (Klee & Caravedo, 2006), assibilation of /r/ (Adelaar, 2004; Escobar, 2011; Klee et al., 2011), and intervocalic voicing of /s/ (Klee et al., 2023). In the morphosyntactic system, some notable features include simplification and reorganization of the complex clitic system (Klee, 1990; Klee & Caravedo, 2005; Mayer & Sánchez, 2016; Paredes & Valdez, 2008), clitic doubling (Zdrojewski & Sánchez, 2014), and elevated rates of pre-verbal constituents, including adverbial and prepositional phrases (Escobar, 2011). A well-documented syntactic-semantic feature of Andean Spanish is the re-purposing of perfect constructions to convey epistemic notions like evidentiality (Bateman McDonald, 2022; Escobar & Roy, 2022; Faller, 2007; García Tesoro & Jang, 2018; Howe, 2018; Klee & Ocampo, 1995; Manley, 2007; Sánchez, 2004). The novel use of the discourse marker *pues* is cited in the discursive-pragmatic domain (Brody, 1995; Zavala, 2001). Finally, a handful of Quechua terms, particularly cultural loanwords (e.g., flora, fauna, agriculture, folklore), have been borrowed into Andean Spanish, typically with phonetic assimilation (Adelaar, 2004). The above list of features is by no means exhaustive; it is meant to highlight the significance of Quechua-Spanish contact in the formation of the contact variety, Andean Spanish, and to recognize essential and recent works on Andean Spanish.

Some novel features of Andean Spanish are attributed to incomplete acquisition of Spanish (i.e., group language shift or second language acquisition), while others are considered conventional examples of contact-induced change. These features may be distinguished from one another vis-à-vis the linguistic profile of the speaker that employs them (Escobar, 2011; Klee & Lynch, 2009). To be more precise, if a particular feature is present only in the speech of L2 Andean Spanish speakers, these features may be considered part of an “interlanguage sociolect” (Escobar, 2011, p. 327), but not necessarily of the contact variety, (e.g. lack of gender and number agreement). The Andean Spanish features employed by monolinguals and 2L1 (simultaneous) bilinguals may be considered genuine features of Andean Spanish according to Escobar (2011) and Klee and Lynch (2009). It bears mentioning that not all reported features of Andean Spanish are present or as frequent in all regional varieties of Andean Spanish, which may

<sup>27</sup> See Escobar (2011) and Klee and Lynch (2009) for a comprehensive list of Andean Spanish characteristics.

be a reflection of the corresponding sociohistorical context. For example, in a recent study, O'Rourke (2023) finds that, in the variety of Andean Spanish in contact with Kichwa of the Ecuadorian Amazon, the palatal lateral [ʎ] exhibits signs of depalatalization and delateralization— phenomena that have not (yet) been attested in the variety of Andean Spanish spoken in Cusco.

As previously mentioned, the effects of language contact on the linguistic subsystems of Quechua are more sparsely reported in the previous literature. The most widely researched contact phenomena include changes to the prosodic subsystem of Quechua (Muntendam, 2015; Muntendam & Torreira, 2016; O'Rourke, 2009) and word order variation (Kalt & Geary, 2021; Muntendam, 2015; Muysken, 1995; Sánchez, 2003, 2010). Many of these listed studies emphasize the interaction between these subsystems (prosodic and syntactic) and information structure. Other studies report contact-induced changes in the phonological (Pasquale, 2002, 2009), morphological (Muysken, 2012), and semantic (Kalt, 2021; Sánchez, 2004) subsystems of Quechua.

## 1.4 A note on terminology

Before proceeding, I would like to clarify a frequent term employed in the remaining chapters— **constituent order**. This clarification is necessary given that, in many places, the term 'word order' is used to refer to the order of linguistic elements within a particular domain. In essence, the terms "word order" and "constituent order" in the forthcoming contexts are synonymous. Consider the reflections of Comrie (1989) below:

Although we retain the term word order typology, which has become established for referring to this area of typology, it should be noted that, strictly speaking, we are concerned not so much with the order of words as with the order of constituents, i.e., it would be more correct to speak of constituent order. On the one hand, in saying, for instance, that a given language has subject-verb-object basic word order, it is irrelevant whether the constituents referred to consist of one or more words, so that this characterization applies equally to *John hit Mary* and to *the rogue elephant with the missing tusk attacked the hunter who had just noticed that his rifle was unloaded* (pp. 86-87).

Though I primarily refer to the structure under investigation as variable predicate **constituent order**, the terms "word order" and "constituent order"

are used interchangeably in the present study, and both refer to the relative order of elements, which are each comprised of one or more words.

## **1.5 Outline of the dissertation**

This introductory chapter has established the primary linguistic phenomenon under investigation in the present study— predicate constituent order variation in Andean Spanish and bilingual Cusco Quechua. Additionally, I have provided here an overview of contact-induced change in situations of language contact along with details surrounding the historical sociolinguistic context of Quechua-Spanish contact to prepare the reader for the integration of past research on these topics and the results of the present study, particularly in the discussion and conclusion chapters. What follows is an outline of the remaining chapters.

In Chapter 2, I summarize the relevant literature on word order. I begin with an overview of word order as a metric of typological classification of the world's languages in order to situate the languages under investigation within a wider context. Moreover, I describe the conventional methods of determining dominant order, as these methods will be invoked in later chapters to discuss the possibility and implications of typological shift based on word order patterns in the contact languages. The majority of this chapter is reserved for an overview of the factors that have been cited as predictors of word order variation in array of languages— animacy, definiteness, informational status, and weight, in addition to language contact itself. Ultimately, this information serves as a justification for the inclusion of these factors as potential independent predictors of predicate constituent order variation in the present study, especially animacy, definiteness, and weight, as these cross-linguistic predictors of syntactic variation have not yet been considered in the extant literature on word order in Andean Spanish and Quechua.

The syntactic behavior of predicate constituents in Andean Spanish and Quechua specifically is the topic of Chapter 3. As a point of comparison, I first confirm that VO is the prescriptive order of predicate constituents in non-contact varieties of Spanish and establish a baseline of OV/VO variation in the absence of contact. I then review the literature on predicate constituent order variation in Andean Spanish, which largely falls into two categories regarding the applied framework— quantitative (i.e., variationist) and syntactic (i.e., generative). Here, I verify that OV order is more frequent in Andean Spanish than in non-contact varieties of Spanish and that higher rates of non-canonical word order in Andean Spanish have been attributed to several social and linguistic

factors, like socioeconomic status and information structure. Next, I examine word order in Quechua and present the detail properties of direct objects in Quechua, like morphological accusative case-marking, null objects, and object agreement. Crucially, I identify the factors that contribute to variable word order in Quechua according to past studies, like variable accusative case-marking, information structure, and the sociolinguistic profile of the Quechua speaker.

Chapter 4 is dedicated to an overview of the methodological architecture that guides the present study. After restating the research questions, I describe the participants of the study, who fall into two main groups— monolinguals (Andean Spanish) and bilinguals (Andean Spanish and Cusco Quechua). I compare the demographic characteristics of the participant group to the Cusco population from which it is drawn to address the representativeness of the sample. Subsequently, I chronicle the data collection procedure, detailing the primary instruments— the Language Background Questionnaire and the Sociolinguistic Interview. The spontaneous speech elicited by the latter instrument is the basis for the construction of three corpora— monolingual Andean Spanish, bilingual Andean Spanish, and bilingual Cusco Quechua— which are searched for all instances of OV or VO order that fall within the circumscribed envelope of variation. The dependent and independent variables (extralinguistic and linguistic) are described and exemplified in the remainder of this chapter.

The following two chapters present the results of the statistical analyses conducted on the data extracted from the aforementioned corpora. Chapter 5 contains the respective findings of the Andean Spanish corpora, beginning with a report of the general results of the combined bilingual and monolingual groups and promptly moving to the results of each group separately. The results of the statistical analyses of the bilingual Cusco Quechua corpus are housed in Chapter 6. Both chapters are furnished with the overall proportions of OV/VO orders, a brief description of inter-participant variation, descriptive statistics on the distribution of OV and VO orders across all evaluated extralinguistic and linguistic predictors, in addition to binomial logistic regression analyses, which determine the factors that significantly condition predicate constituent order variation.

In Chapter 7, I discuss the results presented in the previous two chapters in accordance with the advanced research questions. In comparing the monolingual and bilingual Andean Spanish results to previous studies, I determine that predicate constituent order variation in monolingual Andean Spanish patterns closely with that of non-contact varieties of Andean Spanish and that the rate of OV order among bilingual Andean Spanish speakers is substantially lower than the rates reported by past studies. Through an age analysis, I observe a

positive correlation between the rate of OV order and the age of the speaker, which indicates a change in progress. I conclude that this change in progress is largely responsible for the discordant rates of OV order reported in the present study versus past studies on this topic. In reviewing the results of the logistic regression analyses, I observe that pre-verbal objects are governed by the same linguistic factors in monolingual and bilingual Andean Spanish, and I discuss the implications of this finding for establishing contact-induced change.

In responding to the research questions on predicate constituent order variation in Quechua, I note the high rate of non-canonical VO order and inter-participant variation. Once again, I argue that elevated rates of non-canonical order constitute a change in progress based on the correlation between age and OV order frequency. I compare the change in progress in the bilingual Andean Spanish sample with the bilingual Cusco Quechua sample, ultimately observing an opposing directionality in this change with respect to adherence to the prescriptive norm over time. I contextualize these divergent trends through the application of the above-described mechanisms of contact-induced change in discussing the linguistic outcomes of Quechua-Spanish contact. In comparing the results of the logistic regression analyses across the three groups, I conclude that some linguistic predictors have a language-specific effect on predicate constituent order variation, while others govern this syntactic variation in an equivalent manner across contact languages. Regarding the latter type of predictor, I highlight two convergent structures in bilingual Andean Spanish and Quechua— pre-verbal pronominal objects and VO order with explicit subjects. I theorize that these structures have opposite origins: the former structure originates in Quechua and the latter structure in Andean Spanish.

The final chapter highlights the novel contributions of the present study, identifies the potential implications of these findings, and proposes directions for future scholarship. The chapter closes with a cohesive summary of the dissertation project.

# CHAPTER 2

## WORD ORDER

### 2.1 Introduction

The primary purpose of this chapter is to introduce the topic of constituent order from a typological and variationist perspective through a review of the relevant literature to set the stage for the present study. As mentioned in Chapter 1, the typological perspective proves indispensable because, ultimately, I will compare the syntactic behavior of two languages with mirror-image ordering typologies. To highlight the typological differences between the two languages, and, consequently, the potential implications of these structural points of departure for contact-induced language change, it is useful to first situate each of these languages vis-à-vis the broad typological labels that have been traditionally applied in classifying languages. The reasons for including the variationist perspective are more obvious—each language variety examined in the present study exhibits variable word order at the clausal level, which has been shaped by long-standing language contact in addition to other linguistic and extralinguistic variables, like information structure and socioeconomic status, respectively.

Section 2.2 provides a general overview of the broad structural properties that are considered by typologists in categorizing a language variety. This section includes a comparison of the findings of Greenberg (1963) and Dryer and Haspelmath (2013) (i.e., WALS database) regarding the general distribution of the dominant order of clausal constituents across the languages sampled. I then discuss the various proposed metrics for determining dominant word order. Next, I write briefly about the relationship between dominant word order and other structural properties, including morphological type, to establish cross-linguistic correlates of clausal constituent order. Lastly, considering that both Andean Spanish and Quechua allow null subjects, I focus on the order of pred-

icate constituents— object and verb— as this is the locus of variation in the present study.

I reserve §2.3 for the effect of language contact on clausal constituent order. First, I introduce the notion that, though syntactic structures may be relatively impermeable, word order appears to be a syntactic structure that is more susceptible to cross-linguistic influence. I support this idea by reviewing cases of contact-induced word order shift as documented in Thomason and Kaufman (1988) and Heine (2008). Next, I highlight several studies that assess another possible syntactic outcome of language contact— an increase in word order variation.

In §2.4, I present the syntactic, semantic, and pragmatic properties that have been evaluated as internal predictors of word order variation for an array of languages. These include animacy, definiteness, weight, and information status. Ultimately, these variables will be considered as independent linguistic predictors in the present study. In the last subsection of §2.4, I review the observed correlations between these predictors, and address the theory advanced by Hawkins (1994) that weight is the only relevant predictor of word order variation (and that other considered predictors are merely epiphenomenal) by noting how others have received this idea.

It bears mentioning that, though the present study addresses word order variation at the clausal level, in this chapter I review select literature that investigates word order variation in other domains (e.g., the genitive phrase). This is particularly true of §2.4, as several studies that consider the presented linguistic features as possible predictors of word order variation (e.g., animacy) do not investigate word order at the clausal level. Still, I assume that insights provided by these studies may shed light on predicate constituent order variation. The statistical analyses in chapter 5 and chapter 6 will put these assumptions to the test.

Lastly, it is important to note that this chapter is restricted to a typological analysis of word order and the cross-linguistic factors that effect syntactic variation. I do not address the relevant particularities of Andean Spanish and Quechua in great depth here, though occasional reference will be made to these languages throughout the chapter to highlight the pertinence of certain findings to the present study. Predicate constituent order variation in these two languages specifically will be addressed in greater detail in the coming chapter.

## 2.2 Word order typology

Traditionally, the typological classification of a particular language variety relies on the structural properties that characterize it. In his foundational work, Greenberg (1963) proposes a basic order typology, which categorizes language varieties in accordance with three criteria— i) the relative order of clausal constituents— subject, object, and verb— in declarative sentences, ii) the order of a noun and its qualifying adjective, and iii) adposition type (i.e., prepositions versus postpositions). Greenberg documents the basic order typology of 30 language varieties,<sup>28</sup> which he argues are diverse enough, both typologically- and geographically- speaking, to draw tentative, but meaningful conclusions formalized as language ‘universals’. These language ‘universals’ may or may not be implicational. For instance, (5) illustrates a non-implicational universal, which reports the observation that the subject holds a syntactically privileged position in most languages.

### (5) Universal 1

In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object

(Greenberg, 1963, p. 76)

An implicational universal, exemplified in (6), reports a dependent relationship between two criteria, like the relationship between the order of clausal constituents and adposition type.

### (6) Universal 4

With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.

(Greenberg, 1963, p. 77)

Determining the typological classification of a language variety is not only an essential step in observing cross-linguistic tendencies, but also provides a ‘jumping-off’ point for situations of language contact. Specifically, formulating predictions regarding the potential outcomes of language contact necessarily requires a knowledge of the basic order typology, and other relevant structural properties, of each language in contact. In contact situations where a particular structural property is similar across all languages, there is no reason to assume, *a priori*, that contact between these languages should result in major structural change. However, where there is a typological, structural mismatch between contact languages,<sup>29</sup> it may be reasonably predicted that the relevant structural property of one or more of the languages is altered as a result of contact.

For instance, in investigating the possible syntactic outcomes of Spanish-English contact, presumably, one would not expect that the basic order of

Basque, Berber (African),  
Burmese, Burushaski,  
Chibcha, Finnish (European),  
Fulani, Greek,  
Guarani (American Indian),  
Hebrew, Hindi, Italian,  
Kannada, Japanese, Loritja  
(Oceanian), Malay (Asian),  
Maori, Masai, Maya, Norwegian,  
Nubian, Quechua, Serbian,  
Songhai, Swahili, Thai,  
Turkish, Welsh, Yoruba,  
Zapotec

<sup>29</sup> Sánchez (2004) notes that structural change, or convergence, may also occur in situations of language contact when one of the languages lacks a certain functional feature that is present in the other language. She refers to this process as the Functional Convergence Hypothesis (p. 147).

<sup>30</sup> This generalization does not apply to verbs like *gustar*, for which there is a ‘theme’ subject that typically appears in post-verbal position and an ‘experiencer’ indirect object that appears in pre-verbal position. In fact, Silva-Corvalán (1998) has documented structures like *Yo gusto eso* by English-dominant Spanish-speakers in Los Angeles. Here, the verb *gustar* is structurally modeled after the verb *to like* in English, which is a prototypical transitive verb (Winford, 2003).

<sup>31</sup> Like in other Romance languages, in some contexts, the order AN is semantically more felicitous in Spanish. However, because the majority of descriptive adjectives follow the noun, Spanish is considered a language with dominant NA order.

<sup>32</sup> It should be noted that heavy adjectival phrases in English may exhibit non-canonical NA order, e.g., *people fluent in three languages* (Comrie, 1989, p. 90).

<sup>33</sup> WALS defines a feature as “a structural property of language that describes one aspect of cross-linguistic diversity” (Dryer & Haspelmath, 2013).

<sup>34</sup> Dryer clarifies that the subject denotes ‘more agent-like elements’ and the object denotes ‘more patient-like elements’ (Dryer, 2013c).

clausal constituents would vary or change in either language as a result of contact, because SVO is the dominant order of declarative sentences in both Spanish and English.<sup>30</sup> On the other hand, one could hypothesize that the order of a noun and its qualifying adjective may be affected by Spanish-English contact, as noun-adjective (NA) is the dominant order<sup>31</sup> in Spanish (Zagona, 2002), and the opposite order, adjective-noun (AN), is dominant in English (Dryer & Haspelmath, 2013).<sup>32</sup>

In comparing Andean Spanish and Southern Peruvian Quechua— the two language varieties examined in the present study— a structural mismatch in the relative order of clausal constituents is observed. That is, Andean Spanish is an SVO-dominant language and Southern Peruvian Quechua an SOV-dominant language. So, these two languages have mirror-image ordering preferences with respect to the relative position of the object and the verb within a predicate. To contextualize the typology of these specific languages, in the following subsections, I review the available statistical information on the order of clausal constituents, and more specifically, the order of predicate constituents, in the languages of the world.

## 2.2.1 Order of clausal constituents

### Distribution of clausal constituent orders across languages

The World Atlas of Language Structures (WALS) hosts an online database that compiles typological information found in the descriptive materials (e.g., grammars) of over 2,000 language varieties (Dryer & Haspelmath, 2013). This database provides a sample of languages that far exceeds that of Greenberg (1963). Ostensibly, a database with a larger sample is more representative of the population it is drawn from, per the law of large numbers. Though the WALS database evaluates the sample of language varieties for 144 features,<sup>33</sup> only the features “order of subject, object and verb” (i.e., order of clausal constituents) and “order of object and verb” (i.e., order of predicate constituents) will be presented here, as they are the structures examined in the present study (Dryer & Haspelmath, 2013).

Regarding the order of clausal constituents (i.e., the relative position of the subject (S), object (O)<sup>34</sup>, and verb (V)), a language variety may exhibit one of seven possible arrangements— SVO, SOV, VSO, VOS, OSV, OVS, and no dominant order. The proportions of dominant clausal constituent orders across all sampled languages reported in Greenberg (1963) and WALS (Dryer, 2013c) are compared side-by-side in Table 2.1.

Value	Greenberg (1963)	WALS (2013)
subject - object - verb (SOV)	11 (36.7%)	564 (41.0%)
subject - verb - object (SVO)	13 (43.3%)	488 (35.5%)
verb - subject - object (VSO)	6 (20.0%)	95 (6.9%)
verb - object - subject (VOS)	–	25 (1.8%)
object - verb - subject (OVS)	–	11 (0.8%)
object - subject - verb (OSV)	–	4 (0.3%)
lacking a dominant order	–	189 (13.7%)
<b>Total</b>	<b>30 (100.0%)</b>	<b>1,376 (100.0%)</b>

Table 2.1: Dominant Order of Clausal Constituents in Greenberg (1963) and WALS (Dryer, 2013c)

In both samples, language varieties with subject-initial orders (e.g., SOV, SVO) far outnumber those with orders in which the subject is the second or third constituent. In Greenberg’s sample, 24 of the 30 language varieties evaluated exhibit subject-initial order. The proportion of languages with subject-initial orders reported in the much larger WALS sample (Dryer, 2013c) is strikingly similar to Greenberg’s— 76.5% and 80.0% respectively. However, there are some notable differences between the two samples.

First, in the sample considered by Greenberg (1963), only three dominant orders— SOV, SVO, and VSO— are attested, whereas in the WALS sample (Dryer, 2013c), all orders are accounted for, including VOS, OSV and OVS. Though languages that exhibit these object-subject orders are highly infrequent compared to those with dominant subject-object order, they still make up approximately 3% of the languages sampled in WALS. Second, WALS provides a seventh possible configuration of clausal constituents labelled “lacking a dominant order,” which Greenberg does not consider. Third, the proportion of SOV languages is slightly lower in Greenberg’s sample compared to WALS. Overall, though, roughly half of all subject-initial orders are SOV and the other half SVO. As it happens, the two languages investigated in the present study— Quechua and Andean Spanish— fall into these two categories respectively.

### Determining dominant order

It should be mentioned that, though Greenberg (1963) does not define the term ‘dominant’ order, it is clear that he does not assume that dominant order is equivalent to the only possible order. He writes, “all VSO languages apparently have alternative basic orders among which SVO always figures” (Greenberg, 1963, p. 77). In fact, most languages permit more than one configuration of

clausal constituents (Dryer, 2010). In the WALS database, the dominant order of some languages are determined using the following metric:

For some languages, the classification of a language in this atlas is based on actual text counts. The rule of thumb employed [in determining dominant order] is that if text counts reveal one order of a pair of elements to be more than twice as common as the other order, then that order is considered dominant, while if the frequency of the two orders is such that the more frequent order is less than twice as common as the other, the language is treated as lacking a dominant order for that pair of elements. For sets of three elements, one order is considered dominant if text counts reveal it to be more than twice as common as the next most frequent order; if no order has this property, then the language is treated as lacking a dominant order for that set of elements. Of course, unless one examines a large number and broad variety of texts, one cannot be sure that differences in frequency may not occasionally reflect the idiosyncratic properties of a particular set of texts. It is likely that in some cases, further text counts would lead to classifying a language differently (Dryer, 2013a).

Let's apply this rule of thumb to a theoretical language. Say there is a language that permits three orders— VSO, SOV, SVO— and that, in a text in this language with 100 tokens containing all clausal constituents there are 50 instances of VSO, 40 of SOV, and 10 of SVO. According to the aforementioned process of determining dominant order, this language would be categorized as lacking a dominant word order, as the most frequent order, VSO, is not twice as frequent as the next most frequent order, SOV.

In some languages, constituent order variability is not necessarily random, rather the distribution of a certain orders is dependent on other syntactic considerations. For example, in French, full noun phrases appear in post-verbal position (7a), but pronominal clitics precede the main verb (7b). In general, French is considered SVO-dominant due to the uniform behavior of pronominal clitics.

(7) a. *Le garçon a vu la jeune fille.*  
S            V        O

‘The boy has seen the girl.’

b. *Le garçon l'a vue.*

**S O V**

‘The boy has seen her.’

(Comrie, 1989, p. 89)

German and a few other languages are considered languages with two primary alternating orders, which are ultimately categorized as ‘lacking a dominant order’ in the WALS database. Because German is a V<sub>2</sub> language, SVO order is the typical order of clausal constituents when the main verb is the inflected verb, and there is no auxiliary verb present (8a); however, where an auxiliary appears, the main verbs shifts to the end of the clause, yielding SOV order (Dryer, 2013c). Consider (8):<sup>35</sup>

(8) a. *Der Leber trink-t das Wasser*

DEF teacher drink-3 SG DEF water

**S V O**

‘The teacher is drinking the water.’

b. *Der Leber ha-t das Wasser getrunken*

DEF teacher have-3 SG DEF water drink:PAST.PTCP

**S Aux O V**

‘The teacher has drunk the water.’

(Dryer, 2013c)

<sup>35</sup> When necessary, glossed examples have been adapted so that the glossing abbreviations of the example correspond to those listed in Appendix A.

According to the WALS database, 67 of the documented languages exhibit two primary alternating orders, the majority of which exhibit both SVO and SOV order ( $n=29$ ).

The dominant order of other languages in the WALS database are determined based on claims in a reference grammar or similar source that a particular order is “pragmatically neutral” (Dryer, 2013a), meaning there is a lack of a “special effect” like topicalization or contrastive focus (Dryer, 2010, p. 76). Where there is a discordance between pragmatic neutrality and frequency, the latter is prioritized for this particular sample of languages. The three mentioned criteria for determining dominant order— statistical frequency, distribution (as exemplified in (7) and (8)), and pragmatic neutrality— generally converge on the same order (Dryer, 2010).

More recently, Levshina et al. (2023) have argued for a gradient approach to word order that considers both a “presumption of variability” and a “presumption of gradience” (p. 826). The former presumption means that constituent order variability is not only possible, but expected in all languages. The

latter presumption corresponds to the idea that languages vary in degree, not kind. For instance, a select SOV-dominant language may exhibit this order in all declarative sentences, whereas another SOV-dominant language may exhibit this order only in affirmative declarative sentences, but not negative declarative sentences. For Levshina et al. (2023), a gradient approach to word order minimally requires that the proportion of attested constituent orders are reported for a language. For instance, the authors provide the frequency of subject-object order across 30 languages using online news corpora to demonstrate that these languages fall on a continuum from “rigid languages,” like English and Danish, in which subject-object order is nearly categorical, to “flexible languages,” like Lithuanian, which exhibits subject-object order at a frequency of  $\approx 60\%$  (pp. 831-833). They observe that there is “no clear boundary between the two types, so it is difficult to choose the most appropriate cut-off point” (p. 833). Ultimately, the authors argue that categorical labels should be accompanied by quantitative data, like the proportions of the available syntactic configurations within a particular domain (e.g., percentage of SV and VS orders in intransitive sentences).

### Structural correlates of word order typology

Lastly, it has been observed that clausal constituent order roughly corresponds with other structural properties cross-linguistically. These correspondences include tendencies like head-initial languages (i.e., VO-type) are typically prepositional, exhibit noun-genitive order in possessive phrases, and employ noun-adjective order in adjectival phrases, among other properties. The opposite tendencies have been observed for head-final (i.e., OV-type) languages. That is, they are mostly postpositional (as mentioned in (6)), and exhibit genitive-noun and adjective-noun order (Comrie, 1989; Dryer, 2013b; Greenberg, 1963). Of course, these tendencies do not apply to every single language,<sup>36</sup> and there is often some ambiguity in classifying a language using these conventions. However, these tendencies are useful for an initial characterization of the investigated languages in contact to understand where they align structurally and where they diverge.

It has also been observed that head-initial, head-final, and free languages<sup>37</sup> roughly correspond to distinct morphological types. Baker (1996, p. 4) notes:

... there is a sense that some languages are syntactically oriented, whereas others are morphologically oriented. Within the more morphologically oriented languages, one can identify two types: head-marking languages and dependent-marking languages, to use

<sup>36</sup> In fact, Comrie refers to the observation that NA order is prevalent in head-final languages “an embarrassment to this generalization” (Comrie (1989):95).

<sup>37</sup> That is, languages with no dominant order.

Nicols' (1986) terms. Roughly speaking, head-marking languages make extensive use of agreement morphology on, say, the verb to express linguistic relationships, whereas dependent-marking languages use Case morphology on noun phrases to express comparable relationships. On the other hand, following Greenberg (1963), linguists often categorize languages as being head initial (i.e., SVO or VSO), head final (SOV) or having free word order. Moreover, there are some rough correlations between these two typological schemes. *It is well known that head-final languages generally have well-developed systems of grammatical Case, that is, they are dependent-marking languages.* Head-marking languages, on the other hand, very often have free word order, with dependents either preceding or following the head (Nicols 1992:ch.3). *This leaves SVO languages as the most likely to not have well-developed Case or agreement morphology* (emphasis mine).

Baker's schema will be applied to Spanish and Quechua in the following chapter with the objective of comparing the overarching syntactic and morphological structure of the two languages in contact.

### 2.2.2 Order of predicate constituents

Both languages under investigation in the present study permit null subjects, presumably because the verb is inflected with a subject agreement particle that allows for easy identification of the subject, even in the absence of an explicit referent.<sup>38</sup> Because both Quechua and Spanish are null-subject languages, evaluating the order of clausal constituents for only those transitive predicates that contain an explicit subject would severely limit the number of tokens included in the analysis, especially considering that sentences with null subjects account for the majority of the predicates in the corpus of the present study. Thus, the relative order of **predicate constituents**— the object and verb— is examined in the present study. The distribution of the dominant order of predicate constituents for all languages evaluated in Greenberg (1963) and WALS (Dryer, 2013b) are presented in Table 2.2 below.<sup>39</sup>

<sup>38</sup> This is not to say that languages without subject agreement particles or rich inflectional verb morphology may not also permit null subjects.

<sup>39</sup> The proportions for the Greenberg sample were determined by simply conflating object-verb orders (SOV, OVS, and OSV) into the category *OV-type orders* and verb-object orders (SVO, VSO, and VOS) into the category *VO-type orders*.

Value	Greenberg (1963)	WALS (2013)
object- verb (OV-type orders)	11 (36.7%)	712 (46.9%)
verb - object (VO-type orders)	19 (63.3%)	705 (46.4%)
Both orders with neither dominant	–	101 (6.7%)
<b>Total</b>	<b>30 (100.0%)</b>	<b>1,518 (100.0%)</b>

Table 2.2: Dominant Order of Predicate Constituents in Greenberg (1963) and WALS (Dryer, 2013b)

It is clear from Table 2.2 above that Greenberg’s sample overestimates the proportion of VO-type languages when compared to the WALS sample (Dryer, 2013b). According to the WALS database, about half of all languages with a dominant order exhibit OV order (46.9%) and the other half VO order (46.4%). Note that only lexical objects, not pronominal objects, are considered in determining dominant order in this domain. So, revisiting (7), only sentences like (7a) are included, and those like (7b) are excluded.

As is the case for the order of clausal constituents, WALS also accounts for a third possibility with respect to predicate constituent order— ‘both orders with neither dominant’. This category accounts for a lower proportion of languages than its counterpart, ‘lacking a dominant order’, for clausal constituents, as illustrated in Table 2.1. Dryer (2013b) mentions, “there are a number of languages which are shown as languages lacking a dominant order on Map 81A,<sup>40</sup> but which are classifiable as OV or VO on this map... the most common sub-type of such languages consists of languages in which SVO is a common order in transitive clauses, but where VSO or VOS (or both) is also common.” This suggests that languages tend to either exhibit OV or VO order, but may be more flexible regarding the position of the subject. The small proportion of languages that lack a dominant order of predicate constituents either exhibit flexible word order in general (e.g., Ngandi, an Australian language spoken in the Northern Territory) or employ each order (OV and VO) in distinct syntactic contexts (e.g., German, see (8)).

Now I turn to a discussion of the observed effect of language contact on clausal and predicate constituent order.

### 2.3 Language contact

As reported in Chapter 1, language contact may yield grammatical change in any linguistic subsystem in the involved contact languages. Syntactic systems have been regarded as particularly resistant to change; however, this generalization

<sup>40</sup> “Map 81A” refers to a map that depicts the geographical distribution of clausal constituent order of the sampled languages.

may not apply to word order, especially in contact settings where there is little typological distance between contact languages or there exists a certain degree of bilingualism (Thomason, 2001; Thomason & Kaufman, 1988). Thomason & Kauffman write:

The evidence we have collected does not support the often implicit assumption, in the literature on word order change, that word order patterns constitute a fundamental “deep” structural feature relatively impervious to foreign influence. On the contrary, word order seems to be the easiest sort of syntactic feature to borrow or to acquire via language shift. This is not surprising when it is considered from the viewpoint of functional congruence: dominant SVO and SOV word order patterns, after all, typically perform the same basic syntactic function— identification of subject and object by their position relative to each other and to the verb (Thomason & Kaufman, 1988, p. 55).

Table 2.3 below provides a list of languages presented in Thomason and Kaufman (1988, p. 55) and Heine (2008, pp. 34–35) that have shifted from one basic word order to another due to contact with a language or group of genetically related languages. The entries in Table 2.3 illustrate that, typically, languages shift from SOV to SVO order or vice versa, likely due to the ubiquity of these two orders. Recall that 76.5% of the world’s languages are either SOV or SVO dominant, according to Dryer and Haspelmath (2013).

<b>Recipient</b>	<b>Contact</b>	<b>WO Before</b>	<b>WO After</b>
Finnish	Indo-European languages	SOV	SVO
Ma’a	Bantu	SOV	SVO
Romani	Balkan languages	SOV	SVO/VSO
Austronesian languages	Papuan	SVO	SOV
Takia (Western Oceanic)	Waskia (Papaun)	SVO	SOV
Amharic	Cushitic languages	SVO/VSO	SOV
Akkadian	Sumerian	VSO	SOV
Wutun (Chinese)	Tibetan	SVO	verb-final

Table 2.3: Selected Examples of Clausal Word Order Shift

In the process of contact-induced structural change, there is an interim period in which the structure native to the recipient language and the structure furnished by the source language co-exist as grammatical alternatives of one another (Poplack et al., 2011). These variants may coexist indefinitely, depending

on the nature of the contact situation (Winford, 2003). In essence, word order in a language may either 1) shift completely or 2) exhibit a higher degree of syntactic variation as a result of language contact. The latter result is substantiated by the findings of the studies summarized below.

To begin, Fortescue (1993, p. 285) finds that word order in some Eskimo language varieties, which he asserts predominantly exhibit SOV order in pragmatically neutral contexts, has been “loosened up” as a result of contact with English, an SVO language. His observation is made on the basis of a text analysis of traditional Eskimo legends in Canadian and Alaskan varieties of Eskimo compared to West Greenlandic— an Inuit dialect of Eskimo spoken in Greenland. He claims that Danish contact with the latter variety has not influenced syntactic ordering at the clausal level to the same degree that English has affected select Canadian and Alaskan Eskimo varieties. He hypothesizes that the effect of English may be made even clearer through an analysis of the spoken register, as contact exists primarily in this space. Though Fortescue concedes that he cannot completely account for the differential effect of Danish and English on Eskimo languages varieties, he does cite the “growing bilingualism and English monolingualism in the larger settlements” as well as the ever-present influence of television as possible explanations (p. 286).

Contact-induced word order variation has also been reported in other syntactic domains. For instance, Bancu (2017) investigates contact-induced syntactic transfer within the verb phrase in the variety of Transylvanian Saxon spoken in Viscri, Romania: a small, multi-ethnic Romanian village where Transylvanian Saxons emigrated after a Romanian revolution in 1989. The remaining speakers of Transylvanian Saxon<sup>41</sup> occupying this community report varying degrees of TrSax-German<sup>42</sup> and TrSax-Romanian bilingualism in addition to trilingualism. In this community, German may be considered the ‘high language’ (Penny, 2000, p. 32), which is circumscribed to educational and religious domains, Romanian the *lingua franca*, and Transylvanian Saxon the endangered native language of only 15 inhabitants.

In Viscri Saxon, the order of elements within ‘two verb clusters’— complex verb phrases containing a finite verb (i.e., an auxiliary (Aux) or modal (M) verb) and a non-finite participle— in subordinate clauses is variable. Historically, only Aux/M-V order was permitted in Transylvanian Saxon; however, presently, both Aux/M-V and V-Aux/M orders are available in Viscri Saxon.<sup>43</sup> The structure Aux/M-V is also the grammatical order of constituents in subordinate clause two-verb clusters in Romanian, and the opposite order, V-Aux/M, is the preferred order in German. Example (9a) illustrates Aux/M-V order and

<sup>41</sup> The author dubs this contact variety ‘Viscri Saxon’.

<sup>42</sup> ‘TrSax’ is an abbreviation of Transylvanian Saxon used by the author.

<sup>43</sup> According to the author, corpus data suggests that, by the 1960s, both of these orders were attested.

(9b) exemplifies V-Aux/M order in Viscri Saxon. The finite verb (e.g., auxiliary or modal) is underlined and the participle is bolded.

- (9) a. (...) *wuat mar hun                      **ge-breich-t***  
           that we have.IPL.PRS P T C P-need-P T C P

‘(We brought everything) that we needed.’

- b. (...) *dat ech sa     **verkeuf-t** hun*  
           that I    them sell-P C P T have.ISG.PRS

‘(I went to the market) so that I sell them.’

(Bancu, 2017, p. 16)

Natural speech data was elicited from six native speakers of Viscri Saxon using a sociolinguistic interview. Four of these participants are TrSax-German-Romanian trilinguals, one is a TrSax-German bilingual and the other is a TrSax-Romanian bilingual. Bancu (2017) finds that, overall, there is a nearly even distribution of Aux/M-V and V-Aux/M orders in the dataset, though the latter order is slightly more frequent. Importantly, Bancu observes that, though there is some inter-speaker variation, trilinguals use both orders with similar frequency; however, the TrSax-German bilingual employs German-like V-Aux/M order nearly categorically, and the TrSax-Romanian bilingual employs Romanian-like Aux/M order nearly categorically. From these results, the author concludes that V-Aux/M is likely a syntactic innovation of German bilinguals and trilinguals that has been transferred into the Viscri Saxon variety. She also states that the effect of Romanian on Transylvanian Saxon is one of maintenance. She speculates that, if contact between Romanian and Transylvanian Saxon was less intense, perhaps two-verb clusters would swiftly give way to the German model. So, variation in subordinate clause two-verb clusters in Viscri Saxon has been both caused and maintained by intimate contact among the three languages.

The genitive phrase is another syntactic domain in which contact-induced word order variation has been observed. Naccarato et al. (2021) find that in Daghestanian Russian— a contact variety of Russian spoken in Daghestan that is influenced by a heterogeneous substratum of 40 indigenous languages— there is a high frequency of “left genitives,” i.e., GEN-N order (p. 389). The opposite configuration, N-GEN, is the canonical and more frequent order in monolingual Russian, though GEN-N order occurs sparingly and only with a particular intonation pattern and information status configuration. The preferred order of monolingual Russian, N-GEN, is illustrated in (10a)<sup>44</sup> and the non-canonical order permitted in Daghestan Russian, GEN-N is exemplified in (10b).<sup>45</sup>

<sup>44</sup> ‘RNC’ is the abbreviation for *Russian National Corpus*.

<sup>45</sup> ‘DagRus’ is the abbreviation for *Corpus of Russian spoken in Daghestan*.

- (10) a. *Eščë mat' moej svekrovi byla*  
 also mother my.F.GEN mother-in-law.GEN COP.PST.SG.F  
*krepostnaja*  
 serf.F  
 'And also **my mother-in-law's** mother was a serf.' [RNC, Memo-  
 ries]
- b. *Ona vsju žizn' ne ljubila deduški*  
 3SG.F all.F.ACC life.ACC NEG love.PST.SG.F grandfather.GEN  
*sestru*  
 sister.ACC  
 'In her entire life she never loved **(her) grandfather's** sister.' [Da-  
 gRus, L1 Avar]
- (Naccarato et al., 2021, p. 389)

Using the *Corpus of Russian spoken in Daghestan*, which compiles inter-views with rural, L2 speakers of Russian, Naccarato et al. (2021) find that bilin-gual speakers of Daghestanian Russian employ non-canonical GEN-N order more frequently (21.4%) than Russian monolinguals (2.8%). The authors con-sider both sociolinguistic (e.g., L1 of the speaker, place of birth, year of birth, gender, and level of education) and linguistic variables (e.g., lexical class, refer-entiality, individuation, length, and givenness) as possible predictors of word order variation in the genitive phrase for both language varieties. Regarding Daghestanian Russian, Naccarato et al. (2021) find that all linguistic factors are selected as significant predictors of variation. Specifically, genitive constituents (GEN) that are human, denote kinship or proper names, definite, singular, multiple words (i.e., relatively heavier), and given have a higher likelihood of presenting the non-canonical GEN-N configuration. Furthermore, the head of the genitive phrase (N) is likely to appear after the genitive constituent when it denotes a kinship term and is comprised of a single word.

It is important to mention that, though left genitives are more frequent among bilingual speakers, the authors report that “the strongest predictors<sup>46</sup> of left genitives are the same in both varieties” (Naccarato et al., 2021, p. 402). Regarding the sociolinguistic variables, it was determined that gender is the only significant predictor of genitive phrase variation, with women having a higher likelihood of employing left genitives. Ultimately, the authors conclude that bilingual speakers of Daghestanian Russian produce a nonstandard structure, left genitives, more often than their monolingual counterparts as a result of individual contact with indigenous languages that standardly employ GEN-N order.

<sup>46</sup> Both the lexical class of the head and the lexical class of the genitive modifier condition word order varia-tion for monolinguals and bilinguals.

Though Naccarato et al. (2021) attribute variation in this syntactic domain to language contact, as summarized above, they also consider the possible conditioning effect of linguistic elements, like givenness, animacy (i.e., ‘lexical class’), definiteness (i.e., ‘referentiality’), and weight (i.e., ‘length’). Other studies not concerned with language contact have evaluated these and other linguistic features as possible predictors of syntactic variation in an array of languages. I turn now to this body of literature, organizing the discussion around each considered linguistic predictor.

## 2.4 Linguistic predictors of constituent order variation

In this section, I review the cross-linguistic literature on the linguistic properties that condition word order variation, including semantic (*animacy* and *definiteness*), syntactic (the *weight* of constituents), and discourse-pragmatic properties (*information status*). I begin with animacy.

### 2.4.1 Animacy

**Animacy** is a conceptual property alluding to the degree of sentience of an entity. Animacy may be conceptualized as a continuous category that, broadly speaking, ranges from human (i.e., most animate) to inanimate (i.e., least animate) (Croft, 2012, p. 130). Proposed hierarchies capture the animacy gradient through discrete labels that cover partial ranges of this continuum. The simplest of these hierarchies partition animacy into a binary feature—animate versus inanimate. Contrarily, more nuanced hierarchies incorporate notions like anthropocentrism, egocentrism, and referentiality, in addition to animacy (Sorlin & Gardelle, 2018). For instance, the extended animacy hierarchy in (11) is an amalgamation of three discrete hierarchies: person (first, second < third), referentiality (pronoun < proper name < common noun) and animacy (human < animate < inanimate). The hierarchy labels are organized from most animate to least animate.

- (11) **Extended Animacy Hierarchy:** first/second person pronouns < third person pronouns < human common noun < nonhuman animate common noun < inanimate common noun (Croft, 2012, p. 130)

Comrie (1989) argues that, though animacy itself is a cognitive property, it is a linguistically relevant one, because it affects grammatical constructions cross-linguistically (p. 185). For instance, the grammatical phenomenon of differential

object marking, in which some objects are morphologically marked and others are left unmarked, may depend on the animacy of the object referent (Kagan, 2020). In Spanish, human accusative objects are typically marked with the prepositional particle *a*, while inanimate objects receive no such marking.<sup>47</sup> In the examples adapted from Comrie (1989) below, the inanimate accusative object, *el carro*, ‘the car’, is unmarked (12a), but the human object in (12b), *la empleada*, ‘the employee’, is preceded by the differential object marker, *a*.

<sup>47</sup> Differential object marking in Spanish also relies on definiteness; that is, marked objects are interpreted as specific and unmarked objects, non-specific. So, in the example *El director busca (a) un empleado*, ‘The manager looks for an employee’, the marked object refers to a uniquely identifiable employee and the unmarked object to an unknown, unidentifiable employee (Comrie, 1989, p. 134).

- (12) a. *El director busca el carro.*  
 The manager looks.for the car  
 ‘The manager is looking for the car.’
- b. *El director busca a la empleada.*  
 The manager looks.for **DOM** the.F employee.  
 ‘The manager is looking for the employee.’
- (Comrie, 1989, p. 134)

Importantly, it has also been observed that the lexicosemantic property of animacy may affect word order (Branigan et al., 2008). Tanaka et al. (2011) explain that there are two primary ways in which animacy and word order interact: 1) the assignment of grammatical function and 2) the ordering of elements within a syntactic domain. Regarding the former, it is postulated that animate referents are more likely to be assigned functions higher up in grammatical case hierarchies (e.g., nominative, accusative, dative), which promote them to earlier, more prominent positions in the sentence (Levshina et al., 2023). This is because animacy is presumably related to the notion of conceptual accessibility, that is, “the ease with which the mental representation of some potential referent can be activated in or retrieved from memory” (J. K. Bock & Warren, 1985, p. 50). Animate referents are typically considered more conceptually accessible because they are central to human thought, and thus more easily retrieved (McDonald et al., 1993).

According to J. K. Bock and Warren (1985), the conceptual accessibility of an entity is determined by its predicability, that is, “the range of conceptual relations into which an entity could enter” (Branigan et al., 2008, p. 174). Thus, the reason that animacy may well be considered a graded category (i.e., not one reducible to animate versus inanimate) is because predicability is also a continuous category. Consider the comparison below:

... a human being is highly predictable because he or she can enter into many relations (e.g., growing, eating, sleeping, talking,

ironing, and arguing). Spiders can enter into fewer relations (e.g., growing, eating, sleeping, but not talking, ironing, or arguing), and clouds still fewer (Branigan et al., 2008, p. 174).

According to the above quote, it is the degree of predicability of animate entities that make them more conceptually accessible than inanimate entities.

The tendency to elevate animate referents to prominent syntactic positions is supported by McDonald et al. (1993). The authors find that, in recalling contextualized sentences, speakers prefer to collocate animate entities before inanimate entities in transitive predicates, even if this requires passivization in the recalling process. For instance, a passive sentence with an animate subject (e.g., *The students were frightened by the sound*) is recalled in lieu of an active sentence with an inanimate subject (e.g., *The sound frightened the students*) (p. 198). Again, this preference is born of an innate human ability to access animate entities from memory over inanimate entities.

Concerning the more direct effect of animacy on syntactic ordering (Tanaka et al., 2011), van Bergen and de Swart (2010) find that, in Dutch, animacy influences *scrambling*— a constructional choice to collocate the direct object either before or after an adverb. In a corpus<sup>48</sup> analysis, the authors find that animacy is a significant predictor of scrambling. Specifically, animate proper noun objects tend to ‘scramble’ to a non-canonical position (i.e., NP-AdvP), whereas inanimate proper noun objects tend to remain unscrambled (i.e., AdvP-NP).

<sup>48</sup> *The authors use the Corpus Gesproken Nederlands (Spoken Dutch Corpus).*

Tanaka et al. (2011) uncover a similar effect of animacy on syntactic sequencing in Japanese. In a sentence-recall task, the authors find that participants tend to recall presented OSV sentences as SOV when the subject is animate and the object is inanimate. In other words, there is a statistically significant preference to place an animate entity before an inanimate entity in Japanese. The findings of Tanaka et al. (2011) and van Bergen and de Swart (2010) confirm that animacy affects ordering preferences such that more conceptually accessible (i.e., animate) referents tend to be realized earlier in a phrase (Levshina et al., 2023). What remains to be examined for Andean Spanish and Quechua is whether predicate constituent order variation is also affected by the animacy of the object referent, that is, if an animate object appears in an earlier, pre-verbal position at a higher frequency than inanimate objects. This possibility will be explored in the present study.

While the preference to place animate referents in an early or prominent position is observed cross-linguistically, which referents are considered animate may not be as universal. Most native speakers of American English, for example, would likely consider only living things (e.g., humans, monkeys, and butterflies) animate referents. In Andean culture, however, all natural entities

(e.g., plants, rivers, and rocks) are said to have a spirit and are thus conceptually animate (Salas Cruz, 1993). Therefore, some entities that are typically considered inanimate from a Western viewpoint may indeed be regarded as relatively more predicable by certain non-Western cosmovisions. Referring back to the observation of Branigan et al. (2008) quoted above, from an Andean perspective, spiders and clouds (as well as mountains, land, water, etc.) may indeed be able to talk and argue. Impressionistically speaking, this is supported by the abundant Quechua folklore that involves interactions between animal species and other entities of the natural world.

### 2.4.2 Definiteness

**Definiteness** is a semantic notion that has traditionally been characterized by properties like the degree of familiarity, identifiability, or individuability of a nominal referent (Birner, 2013). For instance, a definite referent is one that the interlocutors are familiar with, is uniquely identifiable, and is individuable from other entities in the discourse.<sup>49</sup> The opposite properties hold for indefinite referents. Birner affirms that definiteness may be distinguished from indefiniteness through the use of certain grammatical expressions. For definites, these expressions include definite articles, pronouns, demonstratives, possessives, and NPs with particular exhaustive quantifiers (e.g., *all* and *every*). Often, NPs lacking articles are considered indefinite, as are NPs accompanied by indefinite articles and other non-exhaustive quantifiers (e.g., *any* and *some*) (Birner, 2013, pp. 121–122).

Definiteness is also inextricably linked to the pragmatic notion of givenness, as referents that are uniquely identifiable, individuable, and familiar have often been previously introduced into the discourse, or may be inferred by the hearer (Birner, 2013). Gundel (1988) is also attuned to the relationship between information structure and definiteness, as indicated by the *Topic-Identifiability Condition* she advances:

(13) *Topic-Identifiability Condition*

An expression, E, can successfully refer to a topic, T, iff<sup>50</sup> E is of a form that allows the addressee to uniquely identify T. (Gundel, 1988, p. 214)

Relatedly, the accessibility of the referent has been proposed to be related to definiteness. Gundel et al. (1993) propose a *Givenness Hierarchy*, which provides cognitive statuses that correspond directly to the referring expression employed in discourse:

<sup>49</sup> Birner states that definiteness may also be established when there is “no relevant basis for individuating” the referent (p. 126).

<sup>50</sup> if and only if

in focus	>	activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
<i>it</i>		<i>this/that,</i> <i>this N</i>		<i>that N</i>		<i>the N</i>		indefinite <i>this N</i>		<i>a N</i>

Table 2.4: The Givenness Hierarchy, adapted from Gundel et al. (1993)

This hierarchy inherently proposes that the relationship between accessibility and definiteness is not binary, rather gradient. That is, the ‘most’ definite of definite expressions, pronominal *it*, is restricted for use when the referent is in focus in the discourse. In other words, the listener is not able to infer what is meant by the word *it* if the referent is not center stage in the ongoing discourse. Definite expressions accompanied by the definite article, *the*, are still definite, as they are uniquely identifiable, but perhaps less accessible than other definite expressions that are in focus, activated, or familiar.

Birner (2013) concedes that definiteness is a somewhat slippery notion, as all definitions of this concept, even taken together, do not account for particular instances in which grammatical definites do not take a definite reading. Consider (14):

- (14) a. Please answer *the phone*.  
 b. Please pass *the milk*.  
 c. Place that book back on *the shelf*. (Birner, 2013, p. 127)

The italicized referents in (14) are termed “weak definites” (Birner, 2013, p. 126). Though these entities are accompanied by a definite article, they are not familiar, unique, or individuable in the discourse. This is evidenced by the fact that the above examples are semantically felicitous in situations where there is more than one of the italicized entities. Similarly, there are cases of “false definites” (p. 129), particularly with the demonstrative determiner *this*, as illustrated in (15):

- (15) What can happen is a hangup such as Rocky Smith ran into, as the independent hauler was traversing Chicago with a load of machinery that just had to get to a factory by morning. “There was *this truck* in front of me carrying giant steel coils, and potholes all over the place,” he remembers. (Birner, 2013, p. 128)

In the above example, *this truck* does not refer to an entity that is familiar, nor uniquely identifiable, nor individuable from other trucks in the discourse.

It is clear that the speaker knows exactly which truck he is referring to, but it is not one that the listener is capable of identifying. In light of these examples, Birner concludes:

In short, while uniqueness, individuability, and familiarity are all relevant concepts that can explain some of the data, and while it is certainly true that different types of definite NPs mark different cognitive status in some sense, none of the accounts thus far proposed seems able to account for the full range of definiteness in English (much less cross-linguistically). There is still a great deal of research to be done on the subject of definiteness... (Birner, 2013, p. 130).

Expanding on Birner's assertion that cross-linguistic literature on definiteness is lacking, it is important to note that the presented syntactic strategies for marking definiteness, which primarily include the use of grammatical expressions like articles, do not apply to all languages. So, how can definiteness be determined in languages that are not equipped with definite and indefinite articles, like Quechua? Mannheim et al. (2010) offers some insight on this matter by exploring how the concepts of genericity and specificity are grammatically expressed through an investigation of L1 acquisition among Quechua-speaking children. Though the correspondence between generics and indefinites or specifics and definites is not one-to-one (i.e., definite nouns can be generic and indefinite nouns can be specific) there is considerable overlap between these concepts. Mannheim et al. (2010) assert that, in Quechua, an agglutinative language, specific referents are not marked by one particular suffix, rather a combination of nominal and verbal suffixes, including aspectual, evidential, and pragmatic suffixes. On the other hand, generic is the default interpretation of a referent in the absence of these types of suffixes. Consider (16) below:

- (16) a. *Waka q'achu-ta mibu-n*  
 bovine forage-ACC eat-3SG  
 'Bovines eat forage.'
- b. *Waka-qa q'achu-ta-n mibu-sha-n*  
 Bovine-TOP forage-ACC-EVIDI eat-PROG-3SG  
*wata-na-n-pi*  
 tie.material-NMLZ-POSS.3SG-LOC  
 'The cow is eating forage at its hitching post.'  
 (Mannheim et al., 2010, p. 2)

The only suffixes that appear in the transitive sentence in (16a) are those that are grammatically obligatory— an accusative marker on the object and a subject-agreement suffix on the verb. Because generic is the default interpretation in the absence of certain semantic-pragmatic suffixes in Quechua, according to Mannheim et al. (2010), the subject in (16a) is considered generic and likely indefinite. That is, there is no particular cow that can be uniquely identified or individuated by the interlocutors. Instead, the statement conveys a more general behavior of all cows: they eat forage. Contrarily, in (16b), a topic suffix (-*qa*) appears on the subject, an evidential suffix<sup>51</sup> on the object (-*n*), an aspectual suffix on the verb (-*sha*), and a locative phrase that specifies where the action performed by the subject is carried out (*watananpi*). Taken together, it is clear that the subject, *waka*, ‘cow’, is both specific and definite, meaning the cow is individuable, identifiable, and familiar to the interlocutors. As is clear from this example, evaluating the definiteness (or specificity) of a referent in languages that lack determiners or another grammatical mechanism solely reserved for indicating definiteness is a nuanced endeavor.

Similarly, Kalt and Geary (2021) state that definiteness and specificity may also be determined by the presence of a qualitative adjective, like *batun llama*, ‘[the] big llama’. Though historically Quechua lacks both definite and indefinite determiners, it is interesting to note that, after prolonged contact with Spanish (a language with both definite and indefinite determiners), it appears that the adjective *buk*, ‘one, another, other’, functions as an indefinite determiner in some varieties contemporary Quechua (Kalt & Geary, 2021; Sánchez, 2003). Moreover, Sánchez (2003) suggests that the emergence of an indefinite determiner, in addition to a drop in the use of the accusative case marker, is associated with a shift from SOV to SVO order in Ulcumayo and Lamas Quechua.

Sánchez (2003) is not alone in identifying a link between definiteness and word order variation. In contemporary Dutch, van Bergen and de Swart (2009, 2010) find that the definiteness of an object referent is the most accurate predictor of scrambling. Recall that scrambling is a constructional choice to collocate the direct object either before or after the adverbial phrase in modern Dutch. In the canonical, unscrambled configuration, illustrated in (17a), the adverb precedes the direct object NP. In the non-canonical configuration, the object and adverb may scramble, yielding NP-AdvP order (17b). The adverbial expression is underlined and the object is bolded in the examples below.

<sup>51</sup> This particular suffix means that the speaker has first-hand evidence of an event.

- (17) a. *Sonja heeft gisteren de kaas opgegeten.*  
 Sonja has yesterday the cheese eaten [unscrambled]
- b. *Sonja heeft de kaas gisteren opgegeten.*  
 Sonja has the cheese yesterday eaten [scrambled]
- ‘Sonja ate the cheese yesterday.’  
 (van Bergen & de Swart, 2010, p. 268)

Extracting all transitive sentences with a contiguous adverb and object from the *Corpus Gesproken Nederlands* (Spoken Dutch Corpus), van Bergen and de Swart (2009, 2010) find that the incidence of scrambling decreases (i.e., NP-AdvP order becomes less frequent) along the definiteness hierarchy from most definite to least definite. In earlier work, van Bergen and de Swart (2009) establish the following hierarchy: pronouns > proper nouns > definites > indefinites. The category *pronouns* includes pronominal clitics, personal pronouns, reflexive pronouns, and reciprocal pronouns; *proper nouns* include names of people, places, or institutional entities<sup>52</sup>; *definite* referents include those accompanied by definite articles, demonstratives, possessives, and universal quantifiers (e.g., *all, everything*); and *indefinite* referents include bare nominals, mass nouns, generic nouns, and existential quantifiers, as well as those accompanied by indefinite articles. In van Bergen and de Swart (2010), however, universally and existentially quantified pronouns are considered separately, as quantificational pronouns are “formally like pronouns, but functionally like noun phrases” (p. 272). In both studies, it was found that pronouns scramble obligatorily, and definite and indefinite referents remain unscrambled nearly categorically. When universally and existentially quantified pronouns are analyzed separately, however, it becomes apparent that not all pronouns exhibit the same behavior. The probability that universally quantified pronouns scramble is approximately 30%, falling between proper nouns and definite nouns, whereas existentially quantified pronouns rarely scramble, like definite and indefinite nouns.

Though van Bergen and de Swart (2009, 2010) take determining the relationship between definiteness and word order as their primary objective, other word order studies have considered definiteness as one of many possible predictors of word order variation. For instance, in their statistical analysis of SVO/S-*ba*-OV variation in Modern Chinese, Yao (2018) includes two independent variables corresponding to definiteness— whether the object constituent is a pronoun and whether the object contains a pronoun (p. 15)— in addition to other variables, like the weight of the object constituent.<sup>53</sup> Regarding definiteness, Yao finds that the *ba*-construction, which co-occurs with non-canonical SOV order, is significantly more frequent when the object constituent contains

<sup>52</sup> This is not a specification made by the authors, rather the types of proper nouns I observed in their provided examples.

<sup>53</sup> This study will be described further in the following section.

a pronoun compared to instances in which an object does not contain a referential, definite expression of this kind. Similarly, in investigating ordering preferences in Persian, Faghiri and Samvelian (2020) find that when an object NP is accompanied by *=rā*, a differential object marker that encodes definiteness,<sup>54</sup> non-canonical NP-PP-V order is preferred over canonical PP-NP-V order. This is not the case for indefinite, non-*rā*-marked objects, which tend to remain in canonical order. Finally, recall that in Daghestanian Russian, bilingual speakers employ GEN-N order in the genitive phrase, which is not the preferred order in monolingual Russian (Naccarato et al., 2021). The authors consider the variable definiteness<sup>55</sup> as an independent variable in their statistical model by coding the genitive noun as either definite, indefinite, or nonspecific. They find that GEN-N order is used primarily when the genitive noun is definite.

Interestingly, though each of these studies investigates a different language (e.g., Dutch, Chinese, Persian, and Russian) with distinct loci of variation (e.g., AdvP-NP, V-O, PP-NP, and N-GEN, respectively), these studies converge on the finding that more definite referents appear in non-canonical positions to the left, meaning that definite entities appear earlier in the investigated syntactic domain for all languages. This is true of both SOV languages, like Persian, and SVO languages, like Chinese. In the present study, definiteness will also be considered as a possible linguistic predictor of OV/VO variation in Andean Spanish and Quechua. This promises to be an interesting area of exploration because, if definite referents appear in non-canonical positions, this means that OV order will be more likely with definite objects in Spanish and VO order will be more likely with definite objects in Quechua. However, if definite expressions always appear earlier, then OV order will be preferred with definite objects in both languages, regardless of the canonical order.

### 2.4.3 Weight

Grammatical **weight**— the relative or absolute<sup>56</sup> length of constituents— is a factor that conditions word order variation cross-linguistically. According to Wasow (1997), there are several ways in which weight may be defined. In comparing three distinct methods of operationalizing weight— number of words, number of nodes, and number of phrasal nodes dominated by the head of the constituent— Wasow finds that these metrics strongly correlate with one another. Similarly, Szmrecsanyi (2004) uncovers a strong statistical relationship between word count, index of syntactic complexity, and node count as measurements of sentence length. The high correlation between different weight-based measurements suggests that various methods may be employed in studying weight-sensitive grammatical phenomena. Some studies cited in this section

<sup>54</sup> The authors note that definite NPs may also be formed through the use of definite determiners. That is, *=rā* is not the only means of indicating definiteness in Persian.

<sup>55</sup> Though the Naccarato et al. (2021) employ the term ‘referentiality’, it is clear they are measuring definiteness, as they state that genitive nouns are classified “according to whether they denote a definite, indefinite, or non-specific referent” (p. 394).

<sup>56</sup> Though Wasow (2002) asserts that weight should be conceptualized as a relative measurement, in other studies weight is evaluated as an absolute measurement, that is, the weight of only one constituent (Struik & Van Kemenade, 2018; Yao, 2018).

operationalize weight using the following metrics: number of characters (van Bergen & de Swart, 2010; Yao, 2018), number of words (Faghiri & Samvelian, 2020; Yamashita & Chang, 2001)<sup>57</sup>, and log base 2<sup>58</sup> of the number of letters (Struik & Van Kemenade, 2018).

The majority of the literature on weight-sensitive grammatical phenomena centers on English. Though English syntax tends to be quite rigid, there are instances in which a constituent may shift into a non-canonical position as a consequence of its ‘heaviness’ without jeopardizing the grammaticality of the phrase. In one well-studied alternation, *heavy NP shift*,<sup>59</sup> a relatively heavy argument NP is shifted after a prepositional phrase (PP) or other adjunct in the post-verbal space (Arnold et al., 2000; Stallings et al., 1998; Wasow, 1997, 2002). In the following example (18a), the accusative argument NP, ‘the wine we had ordered’, is positioned between the verb and adjunct PP, ‘to the table’, in canonical position. In (18b), however, the argument NP, which is relatively heavier than the PP, shifts to the right of the adjunct PP. The alternations provided in (18) are both grammatical.

- (18) a. The waiter brought *NP*[the wine we had ordered] *PP*[to the table].  
 b. The waiter brought *PP*[to the table] *NP*[the wine we had ordered].  
 (Arnold et al., 2000, p. 28)

### Short-before-long

In the configuration with a shifted NP (18b), the constituents of the linear sequence appear in increasing order. It has been observed that, in contexts where syntactic alternations are permitted, the preferred sequence tends to adhere to a short-before-long arrangement, meaning shorter constituents precede relatively longer ones. The terms **end-weight** (Quirk et al., 1972) and **short-before-long** (Arnold et al., 2000; Stallings et al., 1998; Wasow, 1997, 2002) have been applied to this tendency to order constituents on the basis of their syntactic weight.

English is not the only language that exhibits a short-before-long ordering preference. In fact, there is substantial cross-linguistic evidence that lighter elements appear in earlier syntactic positions.<sup>60</sup> For instance, scrambling in Dutch (17) is conditioned by the weight of the object NP (van Bergen & de Swart, 2010). The authors find that the median length, in characters, of scrambled objects (NP-AdvP) is lower than that of unscrambled objects (AdvP-NP). That is, a scrambled object, which precedes an adverb, is typically lighter than an unscrambled object, which is positioned after the adverb.

<sup>57</sup> More specifically, both studies consider constituents with an attributive or relative clause modifier (which increase the total number of words of a constituent) relatively ‘long’ in comparison to constituents lacking these elements.

<sup>58</sup> The authors justify this operationalization of weight by stating, “using the binary logarithm incorporates the idea that the effect of a difference between, say, a 3-letter object and 4-letter object is greater than [the difference] between a 25-letter word and 26-letter word” (p. 10).

<sup>59</sup> Particle movement and dative alternation are two other well-documented weight-sensitive syntactic constructions (Wasow, 1997).

Similarly, Heidinger (2013) finds that the relative syntactic weight of two constituents affects post-verbal word order in Spanish. Specifically, he investigates the variable order of two pairs of constituents— 1) a subject-oriented depictive (Dep) and direct object NP (DO) and 2) a subject-oriented depictive and locative PP (Loc). In a forced-choice task, participants of Cáceres, Spain select their preferred order between two sequencing options for each pair— Dep-DO or DO-Dep and Dep-Loc or Loc-Dep. To determine the effect of weight on order, Heidinger (2013) includes both a neutral condition that contains two constituents of similar weight and an increased-weight condition in which one constituent, particularly the direct object NP and the locative PP in their respective pairs, is substantially heavier than the other. Example (19) illustrates the possible alternations in the neutral condition, and (20) the alternations in the increased-weight conditions, for the subject-oriented depictive and locative PP pair. Heidinger (2013) notes that, though both orders are grammatical, there is a slight preference for the Dep-Loc (19b) order in the neutral condition.<sup>61</sup>

- (19) a. *Juan bailó [en su casa] [disfrazado].*  
 b. *Juan bailó [disfrazado] [en su casa].*  
 Juan danced disguised in his house  
 ‘Juan danced disguised in his house.’

(Heidinger, 2013, p. 176)

- (20) a. *Juan bailó [en la casa decorada por sus hermanas] [disfrazado].*  
 b. *Juan bailó [disfrazado] [en la casa decorada por sus hermanas].*  
 Juan danced disguised in the house decorated by his sisters  
 ‘Juan danced disguised in the house that has been decorated by his sisters.’

(Heidinger, 2013, p. 177)

The results of the forced-choice task confirm that in the increased-weight condition there is a significantly stronger preference for collocating the relatively heavier locative PP or object NP in clause-final position after the light, subject-oriented depictive. So, while both orders in (19), Loc-Dep and Dep-Loc, are virtually equally preferred, short-before-long ordering in (20b) is significantly preferred over long-before-short ordering in (20a).

<sup>60</sup> One particular phenomenon that runs counter to the short-before-long principle in head-initial languages is the variable placement of pronouns in Irish. Irish is a VSO language, and adjuncts typically collocate after the NP object. However, pronouns, particularly weak pronouns, may optionally appear to the right of other adjuncts, sometimes in clause-final position. The authors attribute this rightward movement of a light element to “rhythmic organization” (i.e., phonological restrictions). See Bennett et al. (2016) for a full analysis of the rightward displacement of weak pronouns in Irish.

<sup>61</sup> In the subject-oriented depictive and direct object NP condition, there is a stronger preference for DO-Dep over Dep-DO order in the neutral condition. The author suggests that, because the direct object NP is an argument, it preferably collocates closer to the verb. In the locative PP condition, on the other hand, both constituents are adjuncts, so the Dep-Loc preference is notably weaker than the DO-Dep preference in the neutral condition.

Both production-oriented and comprehension-oriented accounts have been proposed to explain a cross-linguistic short-before-long preference (Yao, 2018). **Production-oriented** (i.e., accessibility-oriented) accounts hinge on the notion that sentence planning is an online task, meaning a sentence is not fully formed in the mind of a speaker before utterance time, rather its production occurs incrementally (J. K. Bock & Warren, 1985; McDonald et al., 1993). Assuming grammatical restrictions are not violated, shorter constituents appear earlier in the phrase because they are more accessible to the speaker. Though accessibility has not been uniformly defined, typically, accessible referents are those that are discourse accessible (i.e., given), conceptually accessible (i.e., animate), and light (i.e., short). Thus, accessible referents are uttered first in online sentence production because they are more easily retrieved than heavy, non-discourse accessible referents, which require more “effort in planning and preparation” (Yao, 2018, p. 2).

In contrast, **comprehension-oriented** (i.e., parsing-based) accounts consider the demands on the working memory of a listener, not the speaker, in processing multiple constituents (Yao, 2018). These accounts assert that a sequence in which shorter constituents appear closer to the head of the phrase is easier to process. Among the comprehension-oriented accounts is the oft-cited Minimize Domain (MiD) principle<sup>62</sup> advanced by Hawkins (Hawkins, 1994, 2004)<sup>63</sup>. The MiD principle states that, “the human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed” (Hawkins, 2004, p. 31). The MiD principle predicts that within a particular domain<sup>64</sup> (e.g. a VP) constituents will be sequenced in a manner such that the head of the furthest daughter constituent (e.g., NP or PP) is as close as possible to head of the mother constituent (e.g., the verb of the VP) to facilitate processing. In head-initial languages, minimizing dependency distance results in a short-before-long ordering of constituents. For instance, the MiD principle predicts that the minimized linear sequence in (21b) is preferred over the alternative in (21a) because there are fewer words that intervene between the head of the mother constituent, *gave*, and the furthest daughter constituent— *to Mary* in (21a), and *the valuable book that was extremely difficult to find* in (21b).

- (21) a. *I* <sub>VP</sub>[ *gave* <sub>NP</sub>[ *the valuable book that was extremely difficult*  
           1          2 3          4 5 6 7          8  
           *to find*] <sub>PP</sub>[ *to Mary*]].  
           9 10          11

<sup>62</sup> Formerly the *Early Immediate Constituent* (EIC) Principle (Hawkins, 1994)

<sup>63</sup> *Dependency Length Minimization* (DLM) is another dependency-based minimization account proposed by Temperley (2007), which differs from MiD in the operationalization of distance. For a detailed comparison of the two models, see Faghiri and Samvelian (2020).

<sup>64</sup> See Hawkins (2004) for a more comprehensive account of the *phrasal combination domain* (PCD).

- b. I<sub>VP</sub>[ gave<sub>PP</sub>[ to<sub>Mary</sub>] NP[ the valuable book that was  
 extremely difficult to find]].
- 1                    2 3                    4

(Hawkins, 1994, p. 57)

Contrarily, in head-final languages, the MiD principle makes the opposite prediction; that is, that long-before-short ordering is preferred. Support for this prediction is provided in the subsection that follows.

### **Long-before-short**

Production-based accounts, which assert that the most accessible (i.e., given, animate, light) referents appear earlier in a phrase, should predict that all languages adhere to a short-before-long preference, regardless of head directionality, as the act of language production is universal, and thus, independent of the grammatical configuration of a language (Yamashita & Chang, 2001; Yao, 2018). Alternatively, according to parsing-based accounts, like the MiD principle (Hawkins, 1994, 2004), a short-before-long preference is the result of prioritizing linear proximity of the domain head and the head of the furthest daughter constituent. Consequently, for head-final languages (e.g., SOV-dominant), the MiD principle anticipates that the lightest daughter constituent is collocated immediately next to the domain head and that longer constituents are positioned relatively further away, i.e., earlier in the phrase. A long-before-short linearization therefore offers the maximally efficient sequence for the parser in a head-final language, according to comprehension-based accounts of word order.

Yamashita and Chang (2001) provide evidence of a long-before-short preference in Japanese, an SOV-dominant language equipped with a morphological case-marking system, which permits a high degree of word order variation. To determine the effect of weight on constituent order, specifically the relative order of the subject and object, the authors conduct a production experiment in which Japanese-speaking participants construct a sentence by sequencing three provided components— a subject, direct object, and verb. The components in each sequencing task are subject to one of three phrase-length conditions— all-short (22a), long-subject (22b), and long-object (22c). In the long subject condition, the subject includes modifier; in the long object condition, the object includes a modifier; and in the all-short condition, neither the subject nor the object are accompanied by a modifier. Each condition is exemplified below in (22):

- (22) a. *Keezi-ga hannin-o oikaketa.*  
 detective-NOM suspect-ACC chased  
 ‘The detective chased the suspect.’
- b. *Se-ga takakute gassiri sita keezi-ga hannin-o*  
 height-NOM tall and big-boned detective-NOM suspect-ACC  
*oikaketa.*  
 chased  
 “The detective who is tall and big-boned chased the suspect.”
- c. *Keezi-ga se-ga takakute gassiri sita hannin-o*  
 detective-NOM height-NOM tall and big-boned suspect-ACC  
*oikaketa.*  
 chased  
 “The detective chased the suspect who is tall and big boned.”  
 (Yamashita & Chang, 2001, B47)

Yamashita and Chang (2001) find that, though the dominant order (SOV) is generally preferred, the object has a higher likelihood of shifting to the left of the subject, yielding OSV order, in the long-object condition. The authors interpret this result to mean that weight-based shifts in Japanese are activated by a long-before-short preference rather than the short-before-long preference that is well documented in head-initial languages.

Though the MiD principle (Hawkins, 1994, 2004) predicts these results, Yamashita and Chang (2001) offer a new theory, the *conceptual-accessibility hypothesis*,<sup>65</sup> which suggests that languages are sensitive to different types of accessibility. Specifically, they reconceptualize accessibility as a bifaceted phenomena comprised of both **form** accessibility and **conceptual** accessibility. Form accessibility is a purely weight-based metric that refers simply to the number of characters, words, phrases, etc. of a constituent. Conceptual accessibility, on the other hand, refers to saliency. For instance, long constituents that contain more lexical elements and are semantically richer are more salient. They propose that Japanese, a head-final language with considerable syntactic freedom, is sensitive to accessibility conceived as a conceptually-based metric. As a result, in Japanese, lexically richer and more conceptually salient constituents precede lighter constituents. However, a language like English, which is head-initial and subject to stricter syntactic constraints, is more sensitive to form accessibility, so lighter constituents (i.e., constituents that are more accessible as a matter of their syntactic weight) precede constituents that are longer, and thus, less

<sup>65</sup> Yamashita and Chang’s hypothesis is given this label by Faghiri and Samvelian (2020).

form-accessible. In sum, they propose a cross-linguistic theory in which constituent order in the pre-verbal space is subject to semantic-pragmatic factors (long-before-short), and order in the post-verbal space is subject to form factors (short-before-long). This may explain why head-final and head-initial languages exhibit mirror-image ordering preferences.

Faghiri and Samvelian (2020) provide evidence of a long-before-short preference in Persian, an SOV-dominant language with mixed head direction (e.g., prepositional, NP objects precede the verb, clausal objects follow the verb). In a production-based experiment modeled after Yamashita and Chang (2001), participants order three provided constituents— a PP, object NP, and verb— after a prompted subject. The canonical order of these constituents in Persian is PP-NP-V. In order to determine the effects of weight on ordering, in the first condition of the experiment, half of the object NPs were modified by an attributive adjective and half were without modification. The results show that a modified NP object, which is syntactically heavier, is more likely to shift to the left of the PP, yielding non-canonical NP-PP-V order. In the second condition, a relative clause is added to the PP argument in half of the target stimuli. Similarly, the results of this condition show that weight affects ordering; that is, non-canonical NP-PP-V order is more frequent when PP remains unmodified by a relative clause and canonical order is preferred when the PP is modified by a relative clause. In sum, in Persian, there is a long-before-short preference that minimally pertains to PP-NP sequencing in the pre-verbal space.

Furthermore, that both an object NP with a single-word attributive modifier and a PP modified by a multi-word relative clause affect pre-verbal linearization in Persian lends support to the conceptual accessibility hypothesis advanced by Yamashita and Chang (2001). Faghiri and Samvelian (2020) assert that dependency-distance minimizing models, like the DiM (Hawkins, 1994, 2004) and DLM (Temperley, 2007), should not predict that one additional word in the object NP poses a large effect on weight-sensitive phenomena such as the variable placement of PP in Persian. The conceptual accessibility hypothesis, however, does make this prediction because an attributive modifier contributes to the saliency of the NP, as a modifier enhances its semantic richness and informativity.

Rosenbach (2005) presents evidence contrary to the conceptual accessibility hypothesis. Applied to the syntactic position of an English genitive phrase in a sentence, this hypothesis predicts that, for genitive phrases with a multi-word possessor and one-word possessum (e.g., *the energetic little puppy's tail*), the *s*-genitive construction should be preferred when the genitive phrase is the subject of the sentence (in pre-verbal position), and the *of*-genitive construction should

be preferred when the genitive phrase is the object of the sentence (in post-verbal position). This is because the former yields long-before-short order and the latter short-before-long. Rosenbach (2005) finds, however, that the syntactic position (i.e., pre-verbal or post-verbal) of the genitive phrase does not affect ordering preferences within the genitive phrase.

### Cross-domain phrasal shift

In the language-specific studies presented above, the investigated weight-sensitive grammatical phenomena are contained to either the post-verbal or pre-verbal domain and concern the relative weight of either two NPs (e.g., subject-object order in Japanese), or an NP and another constituent, like an AdvP (e.g., scrambling in Dutch) or a PP (e.g., heavy NP shift in English).<sup>66</sup> Yao (2018) argues, however, that both production-oriented and comprehension-oriented accounts of weight-related shifts may be extended to make predictions for **cross-domain phrasal shift**— the shifting of an NP across the head of the domain (i.e., the verb) from pre-verbal to post-verbal position, or vice versa. In other words, Yao (2018) proposes that the weight of a constituent may also condition OV/VO variation in a language. Yao assumes that the weight of the domain head remains constant and considers only the absolute weight of the NP in lieu of relative weight, in which the respective weights of two or more constituents are compared.

To ascertain the effect of the absolute weight of an NP on predicate constituent order variation, Yao (2018) investigates SVO/SOV alternation in Mandarin Chinese. Though Mandarin Chinese is an SVO-dominant language, the object NP may appear in pre-verbal position when it is preceded by the grammatical particle *ba*. Yao acknowledges that *ba* may indicate topicality, as the object NP that appears in the *ba*-construction must also be definite or specific; however, their objective is to determine whether weight is an additional, independent factor that determines SVO/S[*ba*-O]V alternation in contexts for which both the *ba*-construction and SVO are grammatical. Example (23) below illustrates that, typically, an object NP is collocated after the verb, as in (23a); however, in the *ba*-construction, the object NP may appear in pre-verbal position (23b). The verb constituent is underlined and the object constituent is in bold.

- (23) a. *tā fàng xià le nà běn shū.*  
           he put down ASP that CLA book

<sup>66</sup> The only exception to this is Heidinger (2013), who examines variable order of a subject-oriented depictive and locative PP, though these constituents both appear in the post-verbal space.

- b. *tā bǎ nà běn shū fàng xià le.*  
 he BA that CLA book put down ASP  
 ‘He put down that book.’

(Yao, 2018, p. 8)

Yao (2018) considers two transitive verbs, *fàng*, ‘to put’, and *nà*, ‘to take in one’s hands’, in the statistical analysis, which uncovers a significant quadratic relationship between constituent length of the object NP (i.e., number of characters) and the likelihood of the *ba*-construction (i.e., OV order). Yao writes that “the positive coefficient of the quadratic term... indicates that the effect of ObjLen<sup>67</sup> on *ba*-likelihood follows a U-shaped parabolic curve that opens upward. In other words, as ObjLen goes up, the likelihood of *ba* decreases at first and then increases” (p. 17). Though the value corresponding to the lowest part of the curve is higher for *fàng* ( $x = 5.15$ ) than for *nà* ( $x = 4.05$ ), the quadratic correlation between NP length and *ba*-likelihood applies to both verbs.

Yao (2018) concludes that this result may be explained by a preferential activation of the two types of accessibility proposed by Yamashita and Chang (2001) based on NP length. In other words, if the NP is light, conceptual accessibility is low, but form accessibility is high, which triggers the short-before-long [*ba*-O]V sequence. On the other hand, for a heavy NP, conceptual accessibility is high, which also triggers the [*ba*-O]V sequence, though in this case, the long-before-short principle is adhered to. When the NP is neither long nor short, neither form nor conceptual accessibility is activated, so the unmarked SVO construction without the *ba*-construction is preferred. That OV/VO alternation in Mandarin Chinese results in both short-before-long and long-before-short orders suggests that mirror-image ordering preferences in head-initial and head-final languages are just that— preferences, not rigid universals.

Similarly, Struik and Van Kemenade (2018) investigate the effect of grammatical weight on OV/VO variation in Old English.<sup>68</sup> The authors uncover a positive statistical correlation between the weight of the NP object and the likelihood of VO order; that is, as the weight of the NP object increases, so does VO order likelihood. This result means that heavier elements typically appear later in the clause, which is predicted by the end-weight principle (Quirk et al., 1972). This finding and those offered by Yao (2018) are of critical importance for the present study, because they are of the few that consider grammatical weight as a possible predictor of cross-domain phrasal shift, which is effectively synonymous to predicate constituent order variation. In the present study, the absolute weight of the object NP is also treated as a continuous, independent variable that may condition OV/VO variation in Andean Spanish and Quechua.

<sup>67</sup> object length

<sup>68</sup> This study will be presented in greater detail in the following subsection, as its primary objective is to determine the effect of givenness of the object referent on word order.

#### 2.4.4 Information status

**Information status** is a discourse-pragmatic notion that refers to “the extent to which the information represented by a constituent is known/given/familiar or unknown/new/unfamiliar” (Birner, 2013, p. 209). Though Prince (1981) argues for a more fine-grained conceptualization of information status,<sup>69</sup> it is conventionally regarded as a binary variable in which the label **given** is applied to entities that have already been introduced into the discourse and are thus known and familiar to the listener, and the label **new** is applied to entities that are introduced into the discourse for the first time and are thus unknown and unfamiliar (Birner, 2013). A large body of literature has been dedicated to understanding the relationship between word order and information status, though not always in these terms. Often, this relationship is discussed in terms of **information structure**, which refers to how information is packaged in a sentence on the basis of the information status of the relevant constituents. In the paragraphs that follow, I review the core terminology related to the concept of information structure. Ultimately, this discussion will provide a justification for the choice to analyze information status (given versus new) in the present study in lieu of more enigmatic notions, like *topic* and *focus*. To understand what is meant by information structure, it is crucial to first understand these italicized terms. I begin with the former.

<sup>69</sup> The information status taxonomy offered by Prince (1981) includes the labels brand-new unanchored, brand-new anchored, unused, noncontaining inferrable, containing inferrable, textually evoked, and situationally evoked.

#### **Information structure: Topic and focus**

Several features have been proposed to define the term **topic**, like givenness, familiarity, referentiality, and aboutness (i.e., what the sentence is about) (Birner, 2013; Lambrecht, 1994). Regarding the former, Gundel (1988) considers givenness a bifaceted concept that may be interpreted either relationally or referentially. Relational givenness is determined at the sentence level (i.e., the topic is given with respect to the comment made about the topic). On the other hand, referential givenness is determined by either the assumed familiarity of the interlocutors (à la Prince (1981)) or the activation of a particular referent in the discourse (à la Chafe (1976)). An activated referent is one that is currently being “attended to” by the interlocutors (Gundel, 1988, p. 212). Returning to the former conceptualization of referential givenness, Gundel formalizes the observed relationship between familiarity and topichood in the *Topic-Familiarity Condition* stated in (24) below:

(24) *Topic-Familiarity Condition*

An entity, E, can successfully serve as a topic, T, iff, both speaker and addressee have previous knowledge of or familiarity with E.

(Gundel, 1988, p. 212)

Furthermore, Lambrecht (1994) proposes the *Topic Acceptability* scale which ascertains that certain types of referents are more likely to be selected as topics. According to this scale, the acceptability of a topic correlates with its degree of givenness. For instance, at the top of the scale are active referents, which are available to the interlocutors at utterance time on account of their recent mentioning. These are followed by accessible referents (i.e., elements that were once active, but are currently deactivated and may be inferred by another active referent or the discursive situation), then inactive referents, brand-new anchored referents (e.g., *a woman I work with*), and brand-new unanchored referents.

Regarding the proposed dichotomous representations of information structure, some prioritize the topic of the sentence, conceptualizing a sentence as being partitioned into a topic and its comment. In the Topic-Comment structure formalized by Gundel (1988), the topic is the entity that “a speaker intends to increase the addressee’s knowledge about, request information about, or otherwise get the addressee to act with respect to,” and the comment is a predication that is “assessed relative to the topic” (p. 210). Other dichotomies, like Ground-Focus, prioritize the **focus** of the sentence. According to Vallduví and Engdahl (1996, p. 3), “the ground-focus articulation partitions the sentence into a noninformative, known, or expected part— the ground— and an informative, newsy, dominant, or contrary-to-expectation one— the focus.”<sup>70</sup> The term focus has also been understood as the “non-presupposed part of the sentence” (Zubizarreta, 1998, p. 1). In the view of Rochemont (1986), the ground material is “c-construable” (25), contrary to the focal material, which is not c-construable.

<sup>70</sup> See Vallduví and Engdahl (1996) and Casielles-Suárez (2004) for proposals on a trichotomous understanding of information structure that conflates the Topic-Comment and Ground-Focus dichotomies.

(25) A string P is c-construable in a discourse  $\delta$  if P has a semantic antecedent<sup>71</sup> in  $\delta$ . (Rochemont, 1986, p. 47)

<sup>71</sup> A semantic antecedent may have a direct or indirect interpretation. Indirect c-construable elements include personal pronouns, deictic expressions, verbs of appearance (e.g. appear, arrive, etc.), in addition to other information that can be easily accommodated by the listener (Rochemont, 1986).

The relationship between ground and focus is best illustrated with answer-question pairs, like those in (26) below:

- (26) a. What about the pipes? In what condition are they?  
The pipes are <sub>F</sub>[rusty].  
b. I have some rust remover. You have any rusty things?  
<sub>F</sub>[The pipes] are rusty.

- c. Why does the water from the tap come out brown?  
*F*[The pipes are rusty]. (Vallduví & Engdahl, 1996, p. 4)

<sup>72</sup> The relationship between prosody and focus will not be elaborated here. See Zubizarreta (1998) for a comprehensive analysis of prosody and focus in Germanic and Romance.

Though each answer exhibits the same syntactic structure, the corresponding prosodic pattern<sup>72</sup> changes based on the focal material. In (26a), the predicate adjective, *rusty*, receives prosodic stress, as it is the non-presupposed material communicated by the answer. The remaining part of the sentence, *the pipes are*, is considered the ground material because the existence of these pipes is established earlier in the discourse, namely by the question. In (26b), the prosodic stress is placed on the *the pipes*, which constitutes the focal material. Again, the remaining part of the sentence, *are rusty*, comprises the ground material because the idea that something is rusty is acknowledged by the question that elicits this answer. In the last example, (26c), the entire sentence is considered the focus. This example illustrates that, in some contexts, an entire sentence may be comprised of only focal material.

<sup>73</sup> See Gussenhoven (2007) for additional types of focus in English.

The examples above highlight two different types of foci, narrow (26a, 26b) and wide (26c), which correspond to a single constituent or a group of constituents respectively (Casielles-Suárez, 2004). Additionally, other types of foci have been identified including contrastive (Rochemont, 1986), emphatic (Zubizarreta, 1998), identificational (Kiss, 1998), and verum focus (Leonetti & Escandell-Vidal, 2009), among others.<sup>73</sup> Similarly, the term topic is not always applied uniformly. For example, Gundel (1988) employs the terms contrastive topic, old topic, and new topic to describe different types of topics. The term topicalization describes the syntactic process of preposing a topicalized constituent, like the object of the verb (Birner, 2013, p. 212). Birner (2013) also notes that a topic may have distinct iterations within an utterance. For instance, a topic may be determined at the level of the sentence or at the level of discourse. That is, in a discourse about a certain topic, there may be several subtopics at the sentence level. Moreover, even within a sentence, more than one entity may serve as a topic. This is clearly illustrated by the ability of the Quechua topic marker to appear on more than one constituent in a sentence, as exemplified in (27) below (Muysken, 1995; Sánchez, 2010).<sup>74</sup> In this example, the subject, *runa*, ‘man’, and object, *wasi*, ‘house’, are both accompanied by the morphological topic marker, *-qa*.

<sup>74</sup> Sánchez’s (2010) insights on topic and focus in Quechua will be discussed at length in the coming chapter.

- (27) *Runa-qa wasi-ta-qa ruwa-rqa-n.*  
 Man-TOP house-TOP built-PST-3SG

‘As for the man, he built the house.’

(Sánchez, 2010, p. 45)

In light of the polysemous nature of the terms *topic* and *focus*, the information status of an object constituent—that is, whether a referent is *given* or *new*—is considered in the present study as a means of determining the effect of information structure on word order. Discussing the effect of notions like topic and focus on word order better lends itself to a qualitative discourse analysis; however, the methodology applied in the present study involves a quantitative statistical analysis of a large number of tokens evaluated for an array of independent variables.

### The effect of information status on word order

It is well established that the information status of a referent may impact constituent ordering, even in a language that has relatively strict ordering preferences, like English (Birner, 2013). For instance, the below example illustrates that an object may be pre-posed with respect to the verb in a non-canonical position when it is given. Example (28)<sup>75</sup> is taken from an episode of the podcast *Armchair Expert*, in which the hosts, Dax Shepard and Monica Padman, interview psychotherapist, Esther Perel (Shepard & Padman, 2021). In this excerpt, Dax clarifies the reason behind his recent relapse.

<sup>75</sup> I have excluded conventional transcription notations that represent overlapping speech and laughter for the sake of clarity.

(28) **Dax:** It would also be incomplete to say... that I just... used because they [the prescription opioids] were in front of me and once I'm on them I make the, the decisions an addict makes. Rule number one in the big book of Alcoholics Anonymous is we can't afford to have resentments. We are not built to have resentments. They will make us use. And I had... some resentments that I wasn't acknowledging I had. So then became the process of like, I need to confront these...yeah

**Esther:** Of course, my inclination would be to say—

**Dax:** You.

**Esther:** What?

**Dax:** I was resentful about your successes and how well-spoken you are and credentialed—

**Esther:** Me?

**Dax:** Yes! You're the reason I relapsed.

**Monica:** She kinda believed you for a second.

**Esther:** For a second, for a second I thought, 'What is he talking about?'

**Dax:** Well, *my resentments* I'll keep private. I've worked through them with the people that it was necessary to work through them with.

In this example, Dax discusses the importance of addressing resentments as an addict. When he says that he will keep his resentments private toward the end of the selected fragment of the discourse, because the referent *resentments* has already been introduced, the italicized object, *my resentments*, is given. In this case, the information status of the referent permits non-canonical OSV order.

The tendency for given information to precede new information has been captured in a number of principles. Halliday (1967) advances the *Given-New Principle* (in the words of Birner (2013)) which states that given information appears earlier in the sentence than new information. Through an analysis of 30 languages, Gundel (1988) observes the cross-linguistic nature of this tendency and formalizes her observation with the *Given Before New Principle*, stated in (29).

(29) *Given Before New Principle*

State what is given before what is new in relation to it

(Gundel, 1988, p. 229)

Gundel also advances a second principle, stated in (30), that is sometimes in variance with the above principle. Though all of the studies summarized in the coming paragraphs lend support to the *Given Before New Principle*, it is important to bear in mind that the opposing principle may apply in certain discourse-pragmatic contexts and/or in particular languages.

(30) *First Things First Principle*

Provide the most important information first (Gundel, 1988, p. 229)

Recent studies in various languages have examined the effect of information status on word order variation in different syntactic domains. For instance, Struik and Van Kemenade (2018) investigate the relationship between givenness and OV/VO variation in Old English. The authors claim that VO is the default

word order of Old English, but that OV order may be triggered when the object referent is given. In their study, they use the Pentaset categories— ‘identity’, ‘inferred’, ‘assumed’, ‘new’, and ‘inert’— to delineate given from new referents (Komen, 2013). According to their binary partition, **given** referents are those corresponding to identity, inferred, and assumed categories in the Pentaset. The term *identity* refers to entities that have been previously mentioned in the discourse; *inferred* entities are those that have not been explicitly mentioned but may be reasonably inferred through a semantic relationship with another explicit referent; the *assumed* category refers to knowledge that is shared within and across cultures. **New** referents, on the other hand, are either those that are being introduced into the discourse for the first time or are bridging inferrables.<sup>76</sup> The authors exclude inert entities, like *death*, from the analysis due to their lack of referentiality.

Extracting all instances of referential direct objects in subclauses<sup>77</sup> from the *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (850-1050), the authors find that the information status of the object is indeed a significant predictor of OV/VO variation. Specifically, they find that the frequency of given objects is higher for tokens exhibiting OV order compared to VO order. In fact, the pre-verbal preference of given objects is nearly categorical (98%). The authors also report an effect of weight on ordering preferences, concluding that “pre-verbal word order is reserved for given information, while post-verbal order is mix of new and given objects, which are post-verbal for reasons of weight” (Struik & Van Kemenade, 2018, p. 14).

Similarly, Struik and Schoenmakers (2023) investigate the variable position of the object in Dutch, albeit through a diachronic lens. First, the authors note that the locus of word order variation has shifted over time in Dutch. In historical Dutch,<sup>78</sup> an object could occupy either the post-verbal or pre-verbal space; however, around the 16th century, the post-verbal option was lost for objects, which coincided with a loss of morphological case-marking and the grammaticalization of a definite determiner. Thus, in contemporary Dutch, OV/VO variation is not possible; however, object placement is still variable with respect to its position relative to the adverb— a phenomenon referred to as scrambling (see §2.4.2). Recall that, in the scrambled position the object NP precedes the adverbial phrase (NP-AdvP), and in the unscrambled position the adverbial phrase precedes the object (AdvP-NP).

Through a corpus analysis, the authors uncover a correlation between word order variation and information status of the object NP<sup>79</sup> in historical and contemporary Dutch. Information status of the object referent was encoded as a binary variable— given or new— following the same procedure applied by

<sup>76</sup> According to Birner (2006), bridging inferrables “relate a new entity to previously evoked information” (p. 32).

<sup>77</sup> The term subclause here refers to the predicate (i.e., the part of the clause that contains the object and the verb), not to subordinate clauses.

<sup>78</sup> The authors use the term ‘historical Dutch’ to refer to Middle Dutch, which spans from 1150-1500 and early New Dutch, which spans from 1500-1700.

<sup>79</sup> Pronominal objects are excluded from the analysis because they appear in pre-verbal and pre-adverbial position categorically and may actually be clitics.

Struik and Van Kemenade (2018). The authors find that, in historical Dutch, there was a strong preference for given objects to appear in pre-verbal position and for new objects to appear in post-verbal position. As VO order was lost, however, the adverb became the new locus of variation, and pragmatic mechanisms continued to “exploit” what the Dutch syntax provided (p. 663). In the 13th century, prior to the shift to OV, both given and new objects scrambled at similar rates. Over time, given objects continued to appear in pre-adverbial position, as they were able to do before the shift to OV order, but new objects began to appear in a pre-verbal but unscrambled position (AdvP-NP) to differentiate themselves from given objects.

In some of the studies presented in the previous sections on animacy, definiteness, and weight, a correlation between information status and word order variation has also been observed. For instance, Yao (2018) reports that when the object referent is mentioned previously, there is an increased probability of S[*ba*-O]V order in Mandarin Chinese. Furthermore, van Bergen and de Swart (2009) find that, though definite referents scramble at relatively low rates in general, anaphoric<sup>80</sup> definite referents scramble (i.e., appear before the adverb) at a rate of 22%, while non-anaphoric definite referents scramble at a rate of only 7%. Analogously, van Bergen and de Swart (2010) report that anaphoric proper nouns scramble more frequently than non-anaphoric proper nouns, 58% and 33% respectively. Concerning the behavior of new referents, Heidinger (2013) concludes that, in narrow information focus conditions in Spanish, the new, focalized constituent is likely to appear in final position.

The general findings reported here support the observed tendency for given referents to appear earlier and for new referents to appear later (i.e., given before new). The present study evaluates the effect of information status on predicate constituent order variation in both Andean Spanish and Quechua to determine if this cross-linguistic tendency holds for the language varieties under investigation.

#### 2.4.5 Discrete or epiphenomenal predictors?

Each of the linguistic covariates presented above are evaluated as individual predictors of word order variation in the cited studies. However, it is undeniable that these variables are interrelated. This has already been mentioned for definiteness and information status, as similar concepts, like familiarity and referentiality, are simultaneously employed to define these terms. The relationship between these variables is made exceedingly clear by the *Topic-Identifiability* (13) and *Topic-Familiarity* (24) conditions advanced by Gundel (1988). Levshina et al. (2023) make a similar observation regarding the relationship between

<sup>80</sup> An anaphoric referent is one which is co-referential with a previously introduced entity. A non-anaphoric referent is not co-referential, as there is no previously introduced entity that it may refer back to (Birner, 2013).

weight and information status, stating, “speakers tend to produce short and given phrases before long and new ones” (p. 838).

Overall, the majority of studies summarized in the previous sections find that relatively earlier positions in an investigated syntactic domain are often occupied by animate, definite, light, and given referents. Personal pronouns, like *me*, tend to exhibit all of these characteristics and, as such, may appear earlier in a given context. This is illustrated in the Andean Spanish example below (31), which is extracted from the corpus of the present study. The verb is underlined and the object is bolded.

(31) ...*con agua así nos echaba... mhm... después cuando he crecido poco a poco ya no... a mi papá también le decía yo... ‘Yo tengo que ir a trabajar... a **mi** ya no me ayuda’... así le decía*

‘he would throw water on us... mhm... later when I had grown up little by little I [couldn’t take it] any longer... and I said to my dad... “I have to go work... you don’t help **me** any more”... I told him like that’ [P28]

In this example, the pronominal object, *a mí*, ‘me’, appears in pre-verbal position, yielding non-canonical OV order. Note that this pronominal has the prototypical characteristics of a constituent that is likely to exhibit this non-prescriptive syntactic behavior; that is, it is definite, animate, light, and given.<sup>81</sup> In fact, it is difficult to conceptualize a situation in which *a mí* would not be construed as given, as this pronominal is deictic and refers to the speaker, who is always physically present for the discourse.

Perhaps the most antagonistic perspective on the interrelatedness of these predictors comes from Hawkins (1994), who argues that constituent weight is the prevailing predictor of word order variation (or instantiations of non-canonical word order), effectively dismissing the observed effect of other predictors, particularly information status and animacy, as merely epiphenomenal factors. He states, “I believe that animate entities will be shorter on aggregate than inanimate ones and will require less linguistic material for referent identification, much as given entities are shorter than new ones” (p. 424).

Nevertheless, the above studies have found that, in some cases, other factors are stronger predictors than weight, which directly contradicts Hawkins’ conclusions. For instance, van Bergen and de Swart (2010), who explore scrambling in Dutch, assert that the effect of definiteness and other variables cannot be reduced to weight alone. They conclude:

Despite the substantial correlation between these two variables [definiteness and weight], we have argued that such a reduction

<sup>81</sup> Of course, it is possible for a given referent to be inanimate, an indefinite referent to be light, a new referent to be animate, and so on.

[from the effect of both predictors to just grammatical weight] is not justified as a regression model based on definiteness was shown to out-perform a model with weight as the relevant predictor. In this way, we have added to the existing evidence that, contra Hawkins, effects of semantic/pragmatic features such as animacy, definiteness, and givenness cannot be considered epiphenomena of grammatical weight. (pp. 290-291)

In her study on the variable order of possessor and possessum within the genitive phrase in English, Rosenbach's (2005) primary objective is to determine whether weight and animacy function as factors independent of one another or if one is simply a proxy for the other. In English, the genitive phrase may be arranged as either an *s*-genitive or an *of*-genitive construction. In the *s*-genitive construction, the possessor precedes the possessum (e.g., *my daughter's favorite toy*). The order of the possessum and possessor are reversed in the *of*-genitive construction, and the preposition, *of*, intervenes between the two components (e.g., *the favorite toy of my daughter*). Rosenbach finds in both an experimental and corpus-based analysis that animacy and weight condition genitive construction preferences independently. Specifically, the *s*-genitive construction is the more frequent choice when the possessor is animate, indicating that animacy affects word order. Furthermore, there is a weight effect such that the likelihood of the *s*-genitive construction decreases as the number of premodifiers of the possessor increases (e.g., *Carson's smile* is preferred over *the smile of Carson*, but *the smile of my super sweet husband* is preferred over *my super sweet husband's smile*).<sup>82</sup> Importantly, Rosenbach (2005) finds that animacy is the stronger predictor of word order when the number of premodifiers of the possessor is less than three. When the number of premodifiers exceeds two, weight becomes the stronger predictor; however, genitive constructions with a one-word possessum and 3+ word possessor are relatively scarce in the consulted corpus. From Rosenbach's findings, it may be reasonably concluded that animacy is not merely an epiphenomenon of weight.

<sup>82</sup> In both the experimental and corpus study, the length of the possessum was held constant at one word.

Adhering to the methodologies of the studies cited above, in the present study, the weight, animacy, definiteness, and information status of the object constituent are all evaluated as independent predictors of predicate constituent order variation. It is up to the statistical analysis to determine if these variables condition predicate constituent order independently in Andean Spanish and Quechua.

## 2.5 Summary

Generally speaking, in this chapter I have provided an overview of word order through several linguistic lenses— language typology, contact linguistics, and language variation. To begin, in §2.2 I discussed how languages are typologically classified and the role that clausal constituent order plays in this classification. Consulting the language databases provided by Greenberg (1963) and Dryer and Haspelmath (2013), I observed that most languages of the world adopt a subject-initial dominant order, and that approximately half of these languages exhibit SOV order and the other half SVO order. These two basic typologies correspond to the two languages under investigation in the present study, Quechua and Andean Spanish, respectively. Because both of these languages also permit null subjects, I provided additional information on the order of predicate constituents— object and verb— across the world's languages, noting that approximately half of all languages with a dominant order exhibit OV order and the other half VO order (Dryer & Haspelmath, 2013).

In §2.3, I established that contact between languages with differing word order typologies may result in either a shift in basic word order typology (Heine, 2008; Thomason & Kaufman, 1988) or syntactic variation. Regarding the latter outcome, Bancu (2017) reports that contact between Transylvanian Saxon, German, and Romanian has yielded syntactic variation within the verb phrase (e.g., Aux/M-V versus V-Aux/M order), particularly for Transylvanian Saxon-German bilinguals, as these languages exhibit mirror image ordering preferences. Similar findings are reported for clausal word order in Eskimo varieties in contact with English and Danish (Fortescue, 1993) and constituent order within the genitive phrase in a contact variety of Russian (Naccarato et al., 2021).

Furthermore, I identified four linguistic predictors in §2.4 that have been observed to condition word order variation in other languages— animacy, definiteness, weight, and information status— such that animate, definite, and given referents tend to appear in earlier syntactic positions. Regarding weight, conflicting hypotheses have been advanced concerning its effect on syntactic structure. Production-based hypotheses, which argue that light constituents are preferred in earlier positions because they are easier to access, predict that short-before-long is a universal ordering principle that applies to all languages (J. K. Bock & Warren, 1985; McDonald et al., 1993). Comprehension-based hypotheses, on the other hand, contend that light constituents are collocated closer to the head of a domain to facilitate the listener's comprehension (Hawkins, 1994, 2004). Consequently, unlike head-initial languages, head-final languages should adhere to the opposite long-before-short principle. Evidence from Japanese and Persian

support this claim (Faghiri & Samvelian, 2020; Yamashita & Chang, 2001). Additionally, it has been observed that cross-domain phrasal shift (i.e., movement of a constituent across the verb) is conditioned by the absolute weight of the constituent (Yao, 2018). This finding is crucial to the present study given that the locus of variation is the predicate, which involves the respective position of the object and verb. What remains unclear from the previous literature is how weight should be defined in agglutinative languages, though it has been asserted that a variety of metrics for defining weight correlate with one another.

The information provided in this chapter will facilitate the typological classification of the two languages under investigation in the present study— Andean Spanish and Quechua— in the coming chapter. Furthermore, I establish precedent for hypothesizing that contact-induced structural variation occurs in the Quechua-Spanish contact situation by consulting studies that confirm this outcome in other contact situations involving languages with distinct word order typologies. Lastly, though information status has already been cited as a predictor of word order variation in Andean Spanish and Quechua, to my knowledge, the three remaining linguistic factors— animacy, definiteness, and weight— have not yet been evaluated for these languages. Because these predictors are considered in the analysis of the present study, they are examined here with the primary goal of understanding their effect on word order variation in other languages (i.e., definite referents appear earlier) to determine if the effect of these variables is universal or language-specific. Moreover, the cited studies that consider animacy, definiteness, and weight will serve as a reference in the process of operationalizing these variables in the methodology of the present study.

Now I turn to a discussion of word order, particularly predicate constituent order, in Andean Spanish and Quechua.

# CHAPTER 3

## WORD ORDER IN SPANISH AND QUECHUA

### 3.1 Introduction

In the previous chapter, I reviewed the relevant typological and variationist literature on constituent order variation, emphasizing the effect of particular linguistic features (e.g., definiteness) on word order in select languages. In this chapter, I narrow in on the two languages under investigation in the present study— Andean Spanish and Quechua— with the objectives of 1) establishing their prescriptive, canonical clausal word orders and 2) detailing the degree and nature of predicate constituent order variation in each language through consultation of the previous literature on this topic. In discussing syntactic variation, I will review only those studies that address the variable order of object and verb and exclude those that discuss variation of other two-constituent structures, like subject-verb, in accordance with the primary objective of the present study, which is to investigate variable predicate constituent order.

In §3.2, I briefly discuss word order in non-contact varieties of Spanish to first establish the basic word order of Spanish in the absence of language contact. The main goal of this section is to provide a baseline of predicate constituent order variation in more ‘standard’ varieties of Spanish for later comparison with Andean Spanish. To achieve this goal, I review the few quantitative studies that provide the distribution of OV and VO orders in non-contact varieties of Spanish through the consultation of written and spoken corpora. Furthermore, I acknowledge the influence of pragmatics, specifically information structure, on syntactic structure, but I save a more thorough discussion of this relationship for the following section.

Next, I present an exhaustive chronological literature review on predicate constituent order in Andean Spanish in §3.3. I begin in §3.3.1 with a review of the studies that provide quantitative data regarding the proportion of OV versus VO orders in several regional varieties of Andean Spanish. In this section, I also identify the sociolinguistic factors that have been found to affect syntactic variation in Andean Spanish, such as socioeconomic status and first language of the speaker, as well as certain linguistic variables, like subject expression. The following subsection, §3.3.2, is reserved for a review of the studies that adopt a more generative perspective of word order in Andean Spanish and accentuate the effect of information structure.

Lastly, in §3.4 I turn to word order in Quechua. I begin by establishing the canonical order of clausal constituents in §3.4.1 and present other relevant linguistic features of Quechua. Specifically, I present two important properties of direct objects in Quechua— morphological accusative case marking and the availability of null objects. Following this, I highlight the relatively high degree of variation in predicate constituent order in Quechua in §3.4.3 and cite the linguistic properties that reportedly govern constituent ordering, like information structure and variable morphological accusative case-marking.

In the final section of this chapter, §3.5, I recapitulate the three main sections on word order in non-contact Spanish, Andean Spanish, and Quechua and identify a few perceived gaps in the extant literature.

## 3.2 Word order in Spanish

Spanish is typologically classified as a language exhibiting dominant SVO order (Dryer & Haspelmath, 2013; Zagona, 2002). Clausal word order in Spanish is not inflexible, however, though the respective position of object and verb tends to be more fixed than the respective position of the subject and verb. Givón (2001) writes, “Spanish is as rigid a VO language as English, Jacaltec or Malagasy. However, the subject position is much less rigid, and is presumably sensitive to discourse-pragmatic considerations. The relative frequency of the SV vs. VS order in Spanish is subject to wide variation, pending historical period, dialect, genre, degree of formality, and literacy” (p. 273). The contrast between the flexible order of subject and verb and the rigid order of verb and object is illustrated through a simple quantitative analysis of VS/SV and OV/VO order variation in “the first 14 pages of *Don Quijote*” (Givón, 2001, p. 275). The results of Givón’s analysis demonstrate that non-canonical VS order appears in approximately a quarter of clauses that minimally consist of a verb and subject, while OV order occurs in 2.2% of clauses containing a verb and object.

Consequently, though Spanish is typically classified as exhibiting dominant SVO order, VSO order is considered a frequent alternative (Muntendam, 2009; Puerma Bonilla, 2019). For instance, Zubizarreta (1998) asserts that “both the VSO and SVO structures of MS<sup>83</sup> are compatible with a focus-neutral interpretation, in which case NS<sup>84</sup> falls on the last constituent” (p. 125).<sup>85</sup> Furthermore, in the World Atlas of Language Structures (WALS) online, Spanish is listed as exhibiting dominant SVO and VO order, but lacking a dominant order with respect to the syntactic configuration of subject and verb (Dryer & Haspelmath, 2013).

A recent corpus analysis by Puerma Bonilla (2019) confirms the rigidity of OV order in contemporary Spanish. Analyzing four written corpora representing four distinct prose types (literary, historiographic, legal, and scientific) that include works in Mexican and Peninsular varieties of Spanish, Puerma Bonilla (2019) finds that less than 1% of all transitive constructions with an explicit object and verb are arranged in non-canonical OV order.<sup>86</sup> Additionally, the author finds that the presence or absence of a subject does not exercise a statistical effect on the relative position of the predicate constituents. Puerma Bonilla (2019) concludes that the syntax of predicate constituents in contemporary, written Spanish is quite inflexible. He concedes that pragmatics may permit verb-object inversion in scant cases but ultimately argues that “this theoretical possibility remains restricted by syntactic probability” (p. 65, my translation).

F. A. Ocampo (1992) elaborates on the specific pragmatic contexts that trigger five possible word orders in Spanish: SVO, OVS, OSV, VSO, VOS. He provides a qualitative and quantitative analysis of the influence of pragmatics on word order using natural speech data from 30 native Spanish speakers from La Plata, Argentina— a non-contact variety.<sup>87</sup> F. A. Ocampo (1992) finds that canonical SVO order is employed categorically when the primary pragmatic function of the utterance is to convey information, regardless of the information status of the subject or object as either new or given (32).

(32) *El tipo se pone jabón*  
**S        V        O**

‘The guy lathers his face’

(F. A. Ocampo, 1992, p. 293)

Canonical word order, SVO, is also reserved for pragmatic contexts in which the focus of contrast<sup>88</sup> is the subject of the sentence. However, when the object is the focus of contrast, the syntactic outcome is OVS order with primary stress on the object. In (33), for example, ‘Francisco’ is the focus that contrasts with

<sup>83</sup> Modern Spanish

<sup>84</sup> nuclear stress

<sup>85</sup> It is clear from the previous literature that both syntactic and prosodic strategies are employed in information packaging in Spanish. Unfortunately, evaluating the effect of prosody on word order variation is outside the scope of the present study. See Zubizarreta (1998) for more on the relationship between information structure, syntax, and prosody.

<sup>86</sup> Puerma Bonilla (2019) excludes clitic, sentential, and coordinated objects from his analysis. Pronominal clitics are standardly excluded from Spanish word order studies because of their unique syntactic behavior that differentiates them from NP objects (e.g., obligatory pre-verbal collocation with simple VPs comprised of only one inflected verb).

<sup>87</sup> The constructions analyzed by F. A. Ocampo (1992) include only those with a subject, object, and verb in a main clause, in which the object is a lexical noun phrase, not a personal pronoun. Additionally, he excludes clauses containing other constituents, like an adverbial phrase.

<sup>88</sup> F. A. Ocampo (1992) defines the focus of contrast as “the constituent that stands in opposition to a closed number of alternatives, members of the same semantic set” (p. 296).

the contextually salient referent, ‘Leticia’, which is a possible alternative to the actual direct object.

(33) *No Francisco dice el tipo*  
**O            V    S**

‘No, Francisco says the guy’

(F. A. Ocampo, 1992, p. 297)

The same order, OVS, is also preferred when the object is the focus, which is best illustrated by a question-answer pair. The utterance exemplified in (34) is immediately preceded by the question *¿Cómo se llama?*, ‘What is it called?’. The object, *Caracalla*, is considered the focus because it is the non-presupposed material that directly answers the question being asked about the name of the town.

(34) *Caracalla se llama el pueblo*  
**O            V    S**

‘Caracalla is the name of the village’

(F. A. Ocampo, 1992, p. 298)

<sup>89</sup> F. A. Ocampo (1992) adopts the definition of topic offered by Lambrecht (1994): “what the sentence is about” (p. 298).

Object topics<sup>89</sup>, like foci, appear in pre-verbal position. In (35), the object topic, *ese cartelito*, ‘this bumper sticker’, precedes the verb of the sentence. Interestingly, there is one clause in the spoken corpus in which the subject is the focus of contrast and the object is the topic. The resulting order of constituents with this information structure configuration is OSV, which seems to suggest that topics have a stronger propensity for fronting than foci. I will return to this observation in the following section on Andean Spanish.

(35) *Y ese cartelito lo tienen los autos estos*  
**O            V    S**

‘And this bumper sticker these cars have it’

(F. A. Ocampo, 1992, p. 299)

The final two pragmatic contexts presented—contrary to expectation and comment focalization—result in verb initial orders: VSO and VOS, respectively. The former context refers to situations in which either the relayed information is unanticipated or the speaker clarifies an inaccurate supposition of their interlocutor. For instance, in (36), the speaker clarifies that the wine will stain the tablecloth, not that the wine spill is harmless, as the interlocutor alleges directly before the exemplified utterance.

(36) *No pero lo mancha el vino al mantel*

**V S O**

‘No, but wine stains the tablecloth’

(F. A. Ocampo, 1992, p. 301)

Finally, comment focalization refers to an information structure configuration in which the predicate— both the verb and the object— is the focus of the sentence. Syntactically speaking, the predicate precedes the subject topic about which it comments, as exemplified in (37).

(37) *No tiene ningún incentivo la pobre René*

**V O S**

‘Poor René has no incentive’

(F. A. Ocampo, 1992, p. 302)

To summarize, F. A. Ocampo (1992) concludes that three pragmatic contexts— object focus of contrast, focal object, and topic object— result in verb-object inversion.<sup>90</sup> Quantitatively speaking, these three pragmatic situations account for 15 of the 143 utterances extracted from the spoken corpus, meaning that OV order occurs at a rate of 10.5% in this sample of Rioplatense Spanish. This is considerably higher than the incidence of OV order reported in the analyses of written works conducted by Puerma Bonilla (2019) and Givón (2001), <1% and 3% respectively.

In later work, F. A. Ocampo (1995) reports the distribution of the relative order of several two-constituent constructions in spoken Spanish (e.g., OV/VO and V-PP/PP-V) across the pragmatic contexts detailed above. His data come from a spoken corpus comprised of 9 hours of conversation with 21 middle-class, native speakers of Rioplatense Spanish. Isolating the results of the relative order of verb and object, F. A. Ocampo (1995) finds that 93.1% of transitive predicates are configured in canonical VO order and the remaining 6.9% in non-canonical OV order.<sup>91</sup> That the proportion of OV order is relatively lower in the absence of a subject than in the presence of an explicit subject (6.9% versus 10.5%) suggests that subject expression may affect predicate constituent order in Spanish such that the presence of an explicit subject permits a higher degree of syntactic variation. This observation will be tested in the present study by evaluating subject expression as an independent variable.

It is important to note that F. A. Ocampo (1992, 1995) insinuates that there is a one-to-one relationship between pragmatic context and clausal word order. That is, SVO (or VO) is the only order available when the utterance conveys information, and OVS (or OV) the only order available when the object is the

<sup>90</sup> According to F. A. Ocampo (1992), these three pragmatic contexts yield OVS order. Verb-final OSV order results when the object is the topic of the clause and the subject is the focus of contrast, which is observed only once in this dataset.

<sup>91</sup> To clarify, the term non-canonical here does not mean ungrammatical, rather not the dominant order, which is VO in Spanish. I continue to use the term ‘non-canonical’ in this manner throughout the remainder of this document.

focus or focus of contrast. However, the spoken corpus of Andean Spanish that I consult in the present study provides evidence to the contrary. Consider (38) below:

(38) a. *¿Cuántos años tenías cuando fuiste?*

‘How old were you when you left?’

b. *Tendría unos veinte años.*

**V            O**

‘I would have been about 20 years old.’ [P17]

In this exchange between the interviewer (38a) and native speaker (38b), the age of the speaker is the focal constituent, as it is the non-presupposed information that directly responds to the question ‘How old were you when you left?’. The response is configured in canonical VO order, despite the fact that the object, *unos veinte años*, ‘about 20 years’, is the focus of the utterance. This example suggests that a certain pragmatic context may be conveyed by more than one syntactic configuration, which runs contrary to Ocampo’s assertion that each pragmatic function corresponds to a particular word order in a one-to-one fashion.<sup>92</sup>

In summation, OV order occurs quite infrequently in non-contact varieties of Spanish, at a rate ranging from less than 1% to approximately 11% depending on the register, regional variety, subject expression,<sup>93</sup> and information structure (i.e., pragmatic function) (Givón, 2001; F. A. Ocampo, 1992, 1995; Puerma Bonilla, 2019). Recall that the purpose of this section is to establish a baseline of predicate constituent order variation in non-contact Spanish for comparison with Andean Spanish. I now turn to a review of the extant literature on word order in this contact variety.

### 3.3 Word order in Andean Spanish

As established above, VO is the dominant order of predicate constituents in Spanish, though a small degree of syntactic variation is permitted and ostensibly triggered by certain pragmatic contexts. It has been observed that the frequency of non-canonical OV order in Andean Spanish is higher than in non-contact varieties of Spanish. Presumably, a higher degree of predicate constituent order variation is the result of contact with Quechua, which allegedly exhibits OV-dominant order. According to Klee (1996, p. 78), word order “transfer should be readily evident [in Andean Spanish], since typologically Quechua is a post-

<sup>92</sup> Though the provided example comes from Andean Spanish, a contact variety, acceptability judgment task results on question-answer pairs completed by speakers of ‘Standard’ Spanish also suggest that VO is an acceptable syntactic structure in response to a question eliciting narrow focus on the object (Muntendam, 2009).

<sup>93</sup> The possible effect of an explicit versus null object is found by comparing the results of F. A. Ocampo (1992) and F. A. Ocampo (1995). That the subject affects ordering preferences in the predicate is not substantiated by Puerma Bonilla (2019).

positional non-rigid V-final language, while Spanish is a prepositional non-rigid V-medial language,” though there is a lack of consensus regarding the nature of this cross-linguistic transfer— that is, whether transfer is due to direct or indirect influence from Quechua.

In the paragraphs that follow, I partition a review of the existing literature on predicate constituent order variation in Andean Spanish into two sections. First, I summarize the studies that approach this topic from a quantitative, variationist perspective, highlighting the general proportions of OV/VO order reported by each study in addition to the linguistic and extralinguistic factors that are identified as predictors of predicate constituent order variation. In the second section, I review the literature that inspects predicate constituent order variation through a generative lens. These studies are situated at the syntax-pragmatics interface inasmuch as they primarily consider the interaction between information structure (e.g., topic and focus) and word order.

### 3.3.1 Variationist analyses

#### Luján et al. (1984)

One of the earliest quantitative analyses of word order in Andean Spanish by Luján et al. (1984) investigates the acquisition of word order by Quechua-Spanish bilingual children in Peru. The authors record elicited and spontaneous conversations<sup>94</sup> with three age-stratified groups of children— 5-year-olds, 7-year-olds, and 9-year-olds— residing in the Peruvian highland cities of Cusco, Ayacucho, Huaraz, and Puno. Because the parents of these children report low levels of proficiency in Spanish, it can be assumed that the youngest participants (5-year-olds) have the lowest exposure to formal Spanish, which increases as the child’s age increases as a byproduct of their enrollment in an education system in which Spanish is the exclusive instructional language.<sup>95</sup> The authors report the proportions of ordering pairs in three syntactic domains— the predicate (OV/VO), the genitive phrase (GN/NG), and the adjectival phrase (AN/NA)— for each age group. These figures are supplied in Table 3.1.

<sup>94</sup> In particular, the recorded contexts include “(1) an adult interviewing a child, (2) a child interacting with peers, and (3) a child interacting with peers and an adult interviewer” (p. 359). The authors do not provide word order trends for each context.

<sup>95</sup> As noted in Chapter 1, the prioritization of Intercultural Bilingual Education in recent decades has brought indigenous languages into classrooms that adhere to an IBE curriculum.

Word Orders	Ages		
	age 5	age 7	age 9
OV/VO	51% / 49%	40% / 60%	30% / 70%
GN/NG	63% / 37%	54% / 46%	36% / 64%
AN/NA	91% / 9%	60% / 40%	38% / 62%

Table 3.1: Word order distribution in three syntactic domains across age of L2 Spanish bilinguals (adapted from Luján et al., 1984, p. 359)

The non-prescriptive orders— OV (39a), GN (39b), and AN (39c)— are exemplified in (39)<sup>96</sup> below.

<sup>96</sup> The information in parentheses on the translation line of the glossed examples “refers to the region, consultant, and tape numbers” (Luján et al., 1984, p. 345).

- (39) a. *Volantín antes hacían.*  
somersault before they-did  
**O V**

‘They used to do somersaults.’ (C 49:80)

- b. *de una señora su frazada*  
of a woman her blanket  
**GEN N**

‘a woman’s blanket’ (H 16:24)

- c. *tu chiquito oveja*  
your small sheep  
**ADJ N**

‘your small sheep’ (P 22:31)

(Luján et al., 1984, p. 345)

Table 3.1 shows that, as age and exposure to formal instruction in Spanish increase, children move from using more Quechua-like structures— OV, GN, AN— in the early stage of acquisition to more Spanish-like structures in the final intermediate stage, which refers to the variety of acquired Spanish spoken by 9-year-olds in their sample (p. 361). There is also a clear sequence of acquisition represented by the following schema: VO > NG > NA. In other words, in the process of acquiring Spanish, clausal word order is the first syntactic domain acquired, followed by the genitive phrase, then the adjectival phrase. Regarding word order in the predicate specifically, the authors report that 5-year-olds employ both OV and VO structures at similar rates. Then, at 7 years of age, a preference for canonical Spanish VO order begins to emerge, which is strengthened by 9 years of age. At this age, however, children still exhibit variation in the

order of predicate constituents, with nearly a third of their clausal utterances configured in Quechua-like OV order.

Luján et al. (1984) suggest that this degree of variation may persist in the speech of child bilinguals because the linguistic target that they are moving toward also permits syntactic variation. Though the authors do not investigate OV/VO order proportions in the speech of adult bilinguals, considering the figure provided by Muysken (1984) for Ecuadorian adult bilinguals, they write, “extrapolating the trend shown by the Peruvian children beyond age 9, one arrives at figures similar to those characterizing adult bilingual speech of Ecuador. The proportion of OV phrases falls from approximately 40% (age 5) to 30% (at age 7) to 25% (at age 9)<sup>97</sup>—and, for the Ecuadorian adults, to 20%” (p. 366).

### **Muysken (1984)**

Consulting adult Andean Spanish speakers, Muysken (1984) investigates the variable order of several pairs of constituents in the speech of 54 Ecuadorian bilinguals and monolinguals from a “small *mestizo* market town” (p. 106), who are partitioned into five groups on the basis of their linguistic dominance and socioeconomic status: incipient bilingual, Quechua-dominant bilingual, Spanish-dominant bilingual, lower-class monolingual, and middle-class monolingual. Specifically, Muysken examines XV/VX order variation in the following constituent pairs listed in XV order: adverb-verb, prepositional phrase-verb, predicate-copula, object-verb, and complement-verb.<sup>98</sup> Overall, Muysken observes that XV structures are most frequent for incipient bilinguals, followed by Quechua-dominant bilinguals. Lower-class monolinguals employ more Quechua-like XV structures than do Spanish-dominant bilinguals, who employ XV structures at a rate similar to that reported for middle-class monolinguals.

The proportions corresponding to OV/VO order specifically are reported in Table 3.2. Unsurprisingly, non-canonical OV order is most frequent among incipient and Quechua-dominant bilinguals. As is the case for XV structures overall, the lowest rate of OV order corresponds to Spanish-dominant bilinguals, with lower- and middle-class monolinguals employing OV order at a comparable rate ( $\approx 13\%$ ). The percentages reported here are based on a small number of tokens; therefore, the statistical implication of these results should be interpreted with caution.

<sup>97</sup> These percentages represent the proportion of OV order in all predicates containing one of three identified word order patterns among Peruvian child bilinguals—OV, VO, and VOV. This is why the percentages in this quote differ slightly from the values presented in Table 3.1, which consider only the first two orders. The order VOV is exemplified by the following sentence: *Conozco a los cabritos conozco*, ‘I know the little goats, I know’ (p. 347).

<sup>98</sup> According to Muysken (1995), a complement is “a sentential complement, containing at least a verb in infinitival, tensed, participle, or gerund form” (p. 113).

Type of speaker	OV		VO	
	<i>n</i>	%	<i>n</i>	%
<i>Incipient Bilingual</i>	17	32.7%	35	67.3%
<i>Quechua-dominant Bilingual</i>	22	34.9%	41	65.1%
<i>Spanish-dominant Bilingual</i>	6	8.3%	66	91.7%
<i>Lower-class Monolingual</i>	6	12.8%	41	87.2%
<i>Middle-class Monolingual</i>	9	12.5%	63	87.5%
<b>Total</b>	<b>60</b>	<b>19.6%</b>	<b>246</b>	<b>80.3%</b>

Table 3.2: OV/VO order distribution across five groups of adult Andean Spanish speakers (adapted from Muysken, 1984, p. 114)

In addition to determining an effect of sociolinguistic factors on syntactic variation in Ecuadorian Spanish, Muysken (1984) finds that one linguistic factor—subject expression—also appears to govern ordering preferences. He writes:

Careful analysis of the data shows an additional constraint on word order variation in the sample: XV is far more frequent than SXV; i.e., the presence of a subject inhibits to some extent verb final order. This phenomena is due, I assume to the fact that a considerable amount of instances of SV order are not base-generated by a VP → X... V expansion rule, as in Quechua, but rather result from the preposing of X to clause initial position, a stylistic device available in all varieties of Spanish. Since Spanish subjects are often absent, X preposing results in a sequence superficially like an XV clause. Speakers imitate XV structures as it were, by preposing verbal complements, but the degree to which they do this is subject to sociolinguistic stratification. Since with a subject present the results of the preposing rule would be XSV rather than... XV, no superficially similar structure results. X preposing, while perfectly possible, is not often employed in that context. The fact that actual SOV structures are absent in the whole sample studied, and SXV structures rather rare, suggests that the use of verb final structures in various interlanguage stages is only very indirectly due to Quechua influence: the verb final VP expansion rule is not involved, but rather the overextension of the preposing rule, from the earliest stage for which we have data (pp. 114-115).

To summarize, Muysken asserts that the absence of SOV order in the dataset is not due to direct syntactic transfer from Quechua to Andean Spanish, rather

to an increase in the application of a pre-verbal preposing mechanism, which is already available in non-contact Spanish and motivated by specific pragmatic contexts. This observation leads Muysken to the conclusion that the high incidence of XV structures in Andean Spanish is due to indirect influence from Quechua.

This conclusion hinges on the underlying assumption that SOV is the dominant order of clausal constituents in Quechua, because in order for this structure to be available for transfer, presumably it should be salient in the source language. However, previous studies have found that SOV order is also somewhat infrequent in bilingual spoken Quechua. For instance, Hubbel (2023) reports that, in bilingual Cusco Quechua, SOV order accounts for only 32.2% of instances in which there is an explicit subject, object, and verb. Furthermore, Sánchez (2003) finds that SOV order occurs at a rate of only 5.7% in Lamas Quechua and 9.9% in Ulcumayo Quechua among child bilinguals;<sup>99</sup> in fact, SVO is the most frequent configuration in both varieties, occurring at a rate of 29.6% in Ulcumayo Quechua and 51.2% in Lamas Quechua (p. 102).<sup>100</sup> Considering these studies, there is evidence to suggest that SOV order in contemporary bilingual Quechua is not the most frequent order, at least in some regional varieties. If SOV order is relatively infrequent in contemporary Quechua, presumably the order should not be salient enough to be transferred into Spanish. Thus, the lack of SOV structures in Muysken’s sample may be in part due to syntactic considerations, like preposing, but also due to the lack of its prevalence in the source language. These observations, which point to possible syntactic convergence among bilinguals, will be explored in greater depth in Chapter 7.

### F. A. Ocampo and Klee (1995)

F. A. Ocampo and Klee (1995) also conduct a variationist analysis of OV/VO variation among Andean Spanish speakers of Calca, Peru. The authors compare two subject groups in their analysis— town elites and the lower urban class.<sup>101</sup> The three town elites included in this study are college-educated individuals whose first language is Spanish. The authors describe this group of speakers as “bilingual to some degree,” but the actual level of Quechua proficiency that these speakers exhibit is unclear (p. 74). The town elites “set the cultural and linguistic standard for the city as a whole” (p. 74). The five individuals of the lower urban class included in this study are native speakers of Quechua whose linguistic shift to speaking primarily Spanish coincided with a geographic shift from their respective rural, native communities to the more urban locale of Calca.

<sup>99</sup> The breakdown of sentential word order in this study includes sentences with all three clausal constituents, in addition to sentences with two constituents (e.g., SV/VS and OV/VO) and one constituent (i.e., explicit verb only).

<sup>100</sup> Of course, it is important to recognize that the infrequency of SOV order in these Quechua varieties may also reflect contact-induced structural change resulting from contact with Andean Spanish, an SVO-dominant language.

<sup>101</sup> In the presentation of their data, F. A. Ocampo and Klee (1995) apply the labels ‘professionals’ and ‘lower group’ respectively.

Regarding the inclusion criteria applied in compiling their data set, the authors include only direct object NPs that can be replaced by the clitics *lo, la, los, las* (i.e., they exclude personal pronouns) that appear in affirmative, main clauses with no additional constituents. This criteria effectively eliminates all utterances that contain additional constituents, like subjects, adverbial phrases, prepositional phrases, among others.

Generally speaking, the authors find that the lower urban class employs non-canonical OV order more frequently (37.6%) than the town elites (12.9%). Though town elites certainly employ canonical OV order less frequently than the lower group, their rate of OV order exceeds that which corresponds to speakers of non-contact Spanish varieties, like Rioplatense Spanish, which is approximately 7% (F. A. Ocampo, 1995).

In addition to uncovering the distribution of OV and VO orders in this regional variety, one of the authors' primary objectives is to evaluate the effect of information structure on OV/VO variation. Consequently, F. A. Ocampo and Klee (1995) report the percentage of new and given object referents in the dataset. Overall, they find that new referents are more frequent than given referents and that "most new NP referents result in the direct objects' placement in post-verbal position when the construction only conveys information (informational word order)" (p.77). However, the distribution of new and given referents across OV and VO orders is not provided. Perhaps the most interesting claim from this investigation is that, in addition to the pragmatic functions that have already been established as correlates of OV order, like focus of contrast on the object, the authors find that OV order also corresponds to other "discourse situations"—repetition, summary, agreement, and explanation (p. 77). The term *repetition* describes discursive situations in which the speaker repeats a statement that they have uttered previously (40a), *summary* refers to an utterance that summarizes previous discourse (40b), *agreement* refers to instances in which the speaker agrees with their interlocutor (40c), and *explanation* describes instances in which the speaker identifies the reason for an occurrence (40d). Each of these discourse situations is exemplified in (40) below. The object is bolded and the verb is underlined.

(40) a. *Comen todo. **Todo** se comen*

‘[They] eat everything. **Everything** [they] eat up.’

(F. A. Ocampo & Klee, 1995, p. 78)

- b. *tamién del plátano verde se hace ya tamién este... tacacho lo que se dice como segundo como el arroz y el otro plátano se hace hervir. Y eso... hay unos palos, así como parés, especial para que cocinemos en la cocina y con eso se machuca su agua, too, exprimiéndolo con sal de lo que ha hervido. **Eso lo machucas** y en sartén en la, el el aderezo lo preparas bien preparado*

‘also from the green plantain one also makes this... tacacho which is what they call the second course, like the rice and the other plantains are made to boil. And that... there are some special sticks, like parés, to cook in the kitchen and with that the water is removed, squeezing it with salt from what has boiled. You crush that and in the frying pan in the seasoning you prepare it well.’

(F. A. Ocampo & Klee, 1995, p. 78)

- c. *E: Entonces, tienes una herida ahí, una cicatriz*  
*J: Sí. **Cicatriz** tengo.*

E: ‘Then you have a wound there, a scar’

J: ‘Yes. I have a scar.’

(F. A. Ocampo & Klee, 1995, pp. 78–79)

- d. *teníamos un mono Martín lo que se llama, era un chiquitito nomás [...] Un día mi mamá cuando mi hermanito estaba chiquito, estaba preparando la leche y la taza mi mamá así lo ha puesto a la mesa? y el mono había corrido al rato la taza ya lo estaba lamiendo [...] Muy satanás es el mono. Sí, de ahí mi mamá otra gente vino de afuera lo ha regalao. **Demasiado travesuras** hacía.*

‘we used to have a monkey Martin that was his name, he was a fairly little one [...] One day when my brother was little, my mother was preparing milk and my mother put the cup on the table like this? and the monkey tipped the cup after a while and was licking it. [...] The monkey was very devilish. Yes, after that some people came and my mother gave him away. He used to get into too much mischief.’

(F. A. Ocampo & Klee, 1995, p. 79)

F. A. Ocampo and Klee (1995) find that these discourse situations account for the majority of tokens realized in OV order (34/53, 64.2%) for the lower group participants. On the other hand, the town elites use the discourse situations

*repetition* and *agreement* sparingly; the majority of utterances exhibiting OV order in this group (6/11, 54.5%) are motivated by pragmatic situations that also govern syntactic variation in non-contact varieties of Spanish, like contrary to expectation. The distribution of OV/VO orders across the reported pragmatic functions, including the discourse situations proposed by F. A. Ocampo and Klee (1995), for both professionals and the lower group is provided in (Table 3.3).

<i>Pragmatic Function</i>	<b>Professionals</b>		<b>Lower Group</b>	
	<i>Tokens</i>	<i>%</i>	<i>Tokens</i>	<i>%</i>
V DO Conveying information	74	87.06	88	62.41
DO V Contrary to expectation	2	2.35	1	0.71
Focus of contrast	2	2.35	2	1.42
Focal constituent	-	-	2	1.42
Topic	2	2.35	4	2.84
<b>Repetition</b>	<b>2</b>	<b>2.35</b>	<b>11</b>	<b>7.80</b>
<b>Summary</b>	-	-	<b>11</b>	<b>7.80</b>
<b>Agreement</b>	<b>1</b>	<b>1.18</b>	<b>8</b>	<b>5.67</b>
<b>Explanation</b>	-	-	<b>4</b>	<b>2.84</b>
Unclear cases	2	2.35	10	7.09
<b>Total</b>	<b>85</b>	<b>100.0%</b>	<b>141</b>	<b>100.0%</b>

Table 3.3: OV/VO order distribution in Andean Spanish across two groups of speakers from Calca, Peru (adapted from F. A. Ocampo and Klee, 1995, p. 77)

Though the qualitative discourse analysis offered by F. A. Ocampo and Klee (1995) to describe the discursive circumstances that motivate non-canonical OV order is quite interesting, the implications of their conclusions are debatable. First, by concluding that “OV word order is extended to new discourse situations” for the lower group, they imply that these discourse situations do not arise as frequently in the speech of professionals. That is, professionals do not repeat, summarize, agree, or explain as often in their discourse. This does not seem probable as communicative acts like offering an explanation are not specific to L1 speakers of Quechua; rather, they are universal concepts that speakers of any language communicate in ordinary conversation. It is likely that professionals and speakers of non-contact Spanish do, in fact, make use of these discourse situations, but they do so in a syntactically prescriptive manner, that is, using VO order.

Furthermore, the data supplied in Table 3.3 implies that VO order is used only to convey information in Andean Spanish. It is perplexing that only OV order would be motivated by so many distinct pragmatic strategies that cannot be

expressed using canonical VO order. Moreover, the pragmatic function of “conveying information” and the new discourse situations advanced by the authors do not appear to be mutually exclusive, especially given the ambiguous definition provided for informational word order: “the ordering of a construction which only conveys information without any special overtones” (F. A. Ocampo & Klee, 1995, p. 75). For instance, in example (40d), the highlighted predicate, *demasiado travesuras hacía*, ‘[he] used to get into too much mischief’, is interpreted by the authors as an explanation that justifies the decision of the speaker’s mother to give the monkey away. However, there is no reason, based on the provided definition of informational word order, that this statement could not also be interpreted as conveying information, as the speaker is describing the general behavior of the monkey.

An additional observation concerning the new discourse situations proposed by F. A. Ocampo and Klee (1995) is that the information status of the NP object in each of these situations is necessarily given. Regarding *repetition*, if a constituent is repeated, this means that the constituent has been uttered previously, as is the case in (40a) in which the object, *todo*, is repeated directly after it appears in VO order in the first mention of the constituent. The same idea applies to *agreement*, the only difference being that the object constituent is uttered by another interlocutor first, but the constituent is still available in the immediate discursive context, as illustrated by the object, *cicatriz*, ‘scar’, in (40c). In (40b) the demonstrative pronoun, *eso*, ‘that’, is discourse deictic, meaning it co-referential with the prepared culinary entity that is described in the discourse immediately preceding the pronominal object constituent. The givenness of the object constituent in (40d), *demasiado travesuras*, ‘too much mischief’, is less clear; however, though *travesuras* is not a referent that has been mentioned explicitly in the previous discourse, it may be inferred from the preceding discourse, which describes the shenanigans of the monkey.

### **Klee (1996)**

Applying the methodology established in Muysken (1984), Klee (1996) offers an analysis of XV/VX variation among three groups of Andean Spanish speakers from Calca, Peru— the lower urban class, middle class, and professionals. The descriptions of the lower urban class and professional group supplied by F. A. Ocampo and Klee (1995) also apply to the respective groups in Klee’s (1996) analysis, and the middle class subjects are described as individuals who completed secondary education and currently have a skilled or semi-skilled occupation. In the subset of 20 participants included in the analysis, the lower group is composed primarily of Quechua-dominant and balanced bilinguals,

while the middle class and professional groups are comprised of balanced and Spanish-dominant bilinguals. Klee (1996) reports that there is no significant difference between the middle class and professional group regarding XV/VX variation; however, a significant difference emerges between the lower group and the combined middle class and professional groups such that the lower group employs XV structures more frequently. Though there is no significant difference between groups regarding OV/VO order variation in particular, as shown in Table 3.4 below, the general trend is that the rate of OV order decreases as Spanish proficiency and socioeconomic status increase.

Speaker group	OV		VO	
	<i>n</i>	%	<i>n</i>	%
<i>Lower group</i>	52	22.6%	178	77.4%
<i>Middle group</i>	42	17.9%	192	82.1%
<i>Professionals</i>	44	15.2%	246	84.8%
<b>Total</b>	<b>138</b>	<b>18.3%</b>	<b>616</b>	<b>81.7%</b>

Table 3.4: OV/VO order distribution across three groups Andean Spanish speakers from Calca, Peru (adapted from Klee, 1996, p. 80)

Klee (1996) observes that, though the professional group employs OV order at the lowest rate compared to the lower and middle groups, their usage of OV order still exceeds the general OV order rate reported for the non-contact variety of Spanish spoken in Buenos Aires, Argentina— 6% (p. 80).

Using the same participant group, Klee (1996) conducts an additional analysis in which she applies the methodology described in F. A. Ocampo and Klee (1995) to identify the distribution of OV order across various pragmatic functions. She finds that both the professionals and middle group employ informational word order (VO) in the majority of instances (91.9% and 95.2% respectively), but for the lower group, a relatively high incidence of OV order, 36.3%, is motivated by the four discourse situations presented above— repetition, summary, agreement, and explanation.

### Muntendam (2009)

Though Muntendam (2009) applies a generative framework to investigate the syntax-pragmatics interface, she also provides select descriptive statistics on OV/VO variation in Bolivian and Ecuadorian varieties of Andean Spanish using a spoken corpus comprised of naturalistic data from 16 Bolivian adult bilinguals and 17 Ecuadorian bilinguals.<sup>102</sup> Regarding the order of clausal constituents—

<sup>102</sup> Bolivian subjects are simultaneous bilinguals, and Ecuadorian subjects are sequential bilinguals who acquired Spanish around four or five years of age (Muntendam, 2009, pp. 54–55).

subject, object, and verb—SVO is the most frequent order in the corpus (78.8%), though VOS (9.7%) and OVS (7.6%) are also somewhat frequent in comparison to other orders (p. 161). When considering only the order of verb and object, Muntendam (2009) finds that VO order occurs at a rate of 81.5% and non-canonical OV order at a rate of 18.5% (see Table 3.5 for the distribution of OV/VO orders for each group). These proportions are quite similar to the average figures reported by Klee (1996) and Muysken (1984).

To determine the effect of select sociolinguistic factors on word order, Muntendam (2009) partitions the Bolivian and Ecuadorian participants into four groups—non-professional women, professional women, non-professional men, and professional men. Concerning the Bolivian participants, Muntendam finds that OV order is most frequent among non-professional women (32.5%) in comparison to all other groups, who employ OV order in approximately 12-13% of predicates. A similar finding is reported for Ecuadorian participants, though both non-professional women and non-professional men employ OV order relatively frequently (31.2% and 22.1% respectively). For Bolivian participants, there is a significant effect of both sex and educational level such that women employ OV order more frequently than men, and non-professionals employ OV order more frequently than professionals. For Ecuadorian subjects, only the educational level is a significant predictor of OV/VO variation.

### **Klee et al. (2011)**

Klee et al. (2011) examines word order variation in Lima, Peru with the objective of determining the linguistic outcomes of dialect contact. In the six decades preceding the publication of this study, Lima had experienced a substantial population growth to which Andean migrants have contributed considerably. The authors note that, upon their arrival to the city, monolingual indigenous speakers tend to shift to Spanish rapidly. The primary objectives of Klee et al. (2011) are to 1) determine if these Quechua-speaking migrants continue to “retain characteristics typical of rural Andean varieties” in the process of language shift and 2) if the Spanish spoken by Andean migrants has influenced the local Limeño Spanish via dialect contact (p. 7). In previous investigations on Limeño Spanish, the authors found that Andean phonological features have given way to Limeño features (e.g., loss of the palatal lateral, [ʎ]), but that some morphosyntactic features (e.g., simplification of the pronominal clitic system) have remained in their speech and exist alongside acquired Limeño features.

The participants of this study include first-generation, 1.5-generation, second-generation, and third-generation migrants residing in “shantytowns inhabited by Andean migrants” (p. 11).<sup>103</sup> Adhering to the methodology of F. A. Ocampo

<sup>103</sup> First- and 1.5-generation migrants are both born in Andean provinces, but the former group moved to Lima after 12 years of age and the latter group before 12 years of age.

<sup>104</sup> The authors include only predicates comprised of a lexical NP object and verb in affirmative, main clauses unaccompanied by other constituents (e.g., a subject).

<sup>105</sup> All but one participant are native speakers of Quechua. The L1 of this participant is Aymara.

and Klee (1995),<sup>104</sup> the authors find that, overall, 90.5% of all predicates in their sample exhibit VO order and 9.5% exhibit OV order. The authors report a significant effect of the first language such that participants with an Indigenous L1<sup>105</sup> employ OV order more frequently than L1 Spanish speakers (18.9% versus 5.6%). Within the L1 indigenous group, they uncover a significant difference for the sex of the speaker as well—OV order is more frequent among males than females (16.3% versus 5.6%). The authors attribute this discrepancy to sex-related differences in employment opportunities in Lima. They write, “the types of work available to male and female migrants may provide varying degrees of close contact with traditional Limeño speakers, which affects the type of Spanish they are exposed to” (p. 27). For example, one female participant recounts her experience as a nanny for a Limeño family as one in which she had intimate contact with the Limeño dialect. On the other hand, a male participant recalls working in various labor-intensive industries where he was surrounded by family members and other Andean migrants. The lack of close contact with speakers of the regional standard may permit the maintenance of Andean features in male speech, like a higher frequency of inverted predicate constituent order.

The authors also consider the possible influence of pragmatics on word order and report the OV/VO proportions for all pragmatic functions and discourse situations<sup>106</sup> evaluated in F. A. Ocampo and Klee (1995). The authors find that all pragmatic functions and discourse situations favor OV order except for the pragmatic function of conveying information.<sup>107</sup> Furthermore, they find that indigenous L1 participants “used an equal or greater percentage of OV word order” for all pragmatic functions compared to L1 Spanish speakers (p. 21). An important departure from the analysis of F. A. Ocampo and Klee (1995) is that not all instances of VO order are categorized as conveying information in Klee et al. (2011)—in fact, both OV and VO orders are attested for all pragmatic functions and discourse situations. The authors conclude that “for both groups, information word order<sup>108</sup> heavily favors VO word order, while other pragmatic functions generally favor OV order” (p. 23). However, regarding frequency of OV order, they assert that “Andean norms disappear in the second generation,” (p. 28) which is evidenced by the low proportion of OV order present in the speech of second-generation migrants, 3.6%.

Surprisingly, the authors do not find a significant effect of information status of the object NP; that is, whether the object is given or new. Considering my earlier observation that object referents within discursive contexts involving agreement, explanation, repetition, or summary are necessarily given, this is an unexpected finding. In the data provided by Klee et al. (2011), topics in particular highly favor OV order; descriptively speaking, 92.3% of topical objects

<sup>106</sup> To review, these include conveying information, contrary to expectation, focus of contrast, focal element, topic, agreement, explanation, repetition, and summary. There were no instances of contrary to expectation in the data.

<sup>107</sup> One exception is that OV order is not favored in discourse situations involving agreement for L1 Spanish speakers. For an example of agreement, refer back to (40c).

<sup>108</sup> The term information word order refers to the word order that is employed to “merely convey information” (Klee et al., 2011, p. 16).

appeared in OV order. Though focus of contrast constituents, which are new referents, also highly favor OV order, tokens of this type account for only 6 of the total 556 tokens of the dataset. An example of topic and focus of contrast are supplied in (41) and (42) respectively.<sup>109</sup>

<sup>109</sup> The number and letter combinations given in parentheses identify the participant who provided the example.

- (41) — *Y con la otra se casó, y con mi mamá no... y con mi mamá ya tenía cuatro hijos.*  
 — *Ah sí ¿no?*  
 — *Y a **mi mamá** la dejó. Imagínese. (13CC)*

— ‘And he married the other woman, but not my mother... and he already had four children with my mother’

— ‘Oh, really?’

— ‘And PREP **my mother** CL [he] left. Imagine that.’

(Klee et al., 2011, p. 21)

- (42) — *Hacia la misa pues, y esos siempre tenía y el padre tenía este su, ¿cómo se llama?*  
 — *Un barrilito.*  
 — *Su, su vino.*  
 — *Su vino, sí.*  
 — *No, **su vino** tenía guardado, pa’ hacer la misa tenía el vino guardado. (22BG)*

— ‘He was saying mass, and he always and the priest always had this, what do you call it?’

— ‘A little cask.’

— ‘His, his wine.’

— ‘His wine, yes.’

— ‘No, **his wine** [he] had stored away, to say mass he had the wine stored away.’

(Klee et al., 2011, p. 21)

In (41), the object, *mi mamá*, ‘my mom’, is considered the focus of contrast because it is being contrasted against the alternative referent, *la otra*, ‘the other woman’. The object constituent in (42) is the topic of the sentence because it is an activated referent in the discourse, which is known to the interlocutors and is currently being commented on. The objects in each of these examples are collocated in pre-verbal position due to the marked (i.e., non-informational)

pragmatic function they serve. In spite of the observed tendency for all non-informational pragmatic contexts to correlate with relatively higher rates of OV order, Klee et al. (2011) caution the reader against drawing firm conclusions from the statistical analysis due to the “small number of tokens in some cells” (p. 24).

### Comparing variationist analyses

The OV/VO proportions reported in all studies summarized in this section are repeated in chronological order in Table 3.5<sup>110</sup> below to facilitate cross-study comparisons.

<sup>110</sup> The two sets of results corresponding to the analyses conducted by Klee (1996) are listed separately. The methodological differences between the studies of Muysken (1984) and F. A. Ocampo and Klee (1995) account for these disparate results. One major methodological difference is that F. A. Ocampo and Klee (1995) exclude tokens with additional constituents (e.g., adverbial phrases), while Muysken (1984) does not limit his dataset in this way.

Region	Demographics	OV (%)	VO (%)	Study
Peru	age 5	51.0%	49.0%	Luján et al. (1984)
	age 7	40.0%	60.0%	
	age 9	30.0%	70.0%	
Ecuador	adult bilinguals	20.0%	80.0%	Muysken (1984)
Calca, Peru	lower group	37.6%	62.4%	Ocampo & Klee (1995)
	professionals	12.9%	87.1%	
Calca, Peru	lower	22.6%	77.4%	Klee (1996) applying methods of Muysken (1984)
	middle	17.9%	82.1%	
	professional	15.2%	84.8%	
Calca, Peru	lower	36.3%	63.7%	Klee (1996) applying methods of Ocampo & Klee (1995)
	middle	4.8%	95.2%	
	professional	8.1%	91.9%	
Tarata, Bolivia	sim. bilinguals	17.4%	82.6%	Muntendam 2009
Juncal, Ecuador	seq. bilinguals	19.9%	80.1%	
Lima, Peru	Indigenous L1	18.9%	81.1%	Klee et al. (2011)
	Spanish L1	5.6%	94.4%	

Table 3.5: OV/VO distribution in Andean Spanish word order studies

The table above demonstrates that a wide range of percentages are reported regarding the frequency of non-canonical OV order among speakers of Andean Spanish. Narrowing in on those studies that observe only the speech of adult bilinguals in an Andean region, the rate of OV order ranges from  $\approx 5\%$  to  $\approx 38\%$ , which is considerably higher than the range of OV order rates reported for non-contact varieties:  $\approx 1\%$  to  $\approx 11\%$ .

Several of the studies listed in Table 3.5 have established a conditional relationship between predicate constituent order variation and select sociolinguistic

factors, like sex, socioeconomic status, level of education, first language of the speaker, and linguistic dominance. F. A. Ocampo and Klee (1995) find, for instance, that Andean Spanish speakers of a lower socioeconomic status employ OV order nearly three times more frequently than those of a higher socioeconomic status. In some studies, more than one factor has been conflated to form groups that exhibit a bundle of demographic characteristics. For example, Muysken (1984) partitions his sample into five speaker groups with varying socioeconomic statuses and degrees of linguistic dominance in Spanish and Quechua. Similarly, F. A. Ocampo and Klee (1995) conflate socioeconomic status, level of education, and residential history to create two distinct participant groups— the lower group and professionals.

For the most part, the sociolinguistic factors listed above condition predicate constituent order variation in the same way. For instance, those with a relatively lower socioeconomic status, lower level of education, and higher linguistic proficiency in Quechua tend to employ OV order at higher rates than those that exhibit the opposite characteristics (Klee, 1996; Klee et al., 2011; Muntendam, 2009; Muysken, 1984; F. A. Ocampo, 1995). Nevertheless, there are conflicting findings regarding the effect of sex on syntactic variation. Muntendam (2009) finds, for example, that women from Tarata, Bolivia employ OV order more frequently than men,<sup>111</sup> but Klee et al. (2011) notice the opposite tendency in their participant group. This discrepancy is likely a reflection of the locales in which these studies were conducted. As mentioned earlier, in the non-Andean city of Lima, Klee et al. (2011) anecdotally observe that women tend to occupy positions that put them in contact with Limeño speakers (e.g., nannies), which may cause their speech to become more like the regional, coastal standard. On the other hand, men typically work alongside other male Andean migrants, and, thus, the linguistic norms of Andean Spanish are reinforced through both uninterrupted contact with speakers of a similar linguistic background and lack of exposure to the Limeño variety. Muntendam (2009) conducted her research in Tarata, Bolivia, which is a relatively rural Andean community. Based on personal knowledge of these types of communities, the women tend to remain in the community and work primarily as homemakers, vendors, and/or artisans, while the men tend to travel to more urban areas for work. Consequently, the same observations offered by Klee et al. (2011) may apply to this region, but in reverse— tight-knit, local communities of women reinforce Andean Spanish linguistic norms, and heightened exposure to the regional standard results in stronger influence of the prestige variety among men. This proposition is further bolstered by the fact that professional women in Bolivia employ OV order at a lower rate than non-professional women (Muntendam, 2009).

<sup>111</sup> This is true of the Bolivia group but not of the Ecuador group.

In the present study, I consider sex, age, level of education, residence, first language, and linguistic dominance as separate predictors to determine their individual effect on predicate constituent order variation. The operationalization of these variables sometimes diverges from the previous studies with the objective of providing a more nuanced statistical analysis. For example, linguistic dominance is a continuous variable in the present study. In the following chapter, I detail the methodology of the present study, which includes a description of each of the considered sociolinguistic predictors.

### 3.3.2 Syntactic analyses

In this subsection, I review the previous studies concerned with the syntax-pragmatics interface, particularly the relationship between information structure and pre-verbal objects in Andean Spanish.

Camacho (1999) investigates the significance of the syntax-pragmatics interface in the variety of Andean Spanish spoken by L1 Quechua bilinguals. He notes that Spanish and Quechua not only differ with respect to their basic word order typologies, but also regarding the expression of information structure notions, such as topic and focus. In Spanish, information structure is typically encoded via syntactic and prosodic strategies. Camacho notes, for example, that topical constituents may be fronted in Spanish. When these constituents refer to a specific entity, a resumptive pronoun accompanies the fronted topic, but not when the constituent refers to a generic entity. A fronted object lacking a resumptive pronoun is a syntactic structure that also corresponds to a focal object, though the fronted constituent must be prosodically stressed. Focused constituents may also remain *in situ*, provided that they are stressed. Consider the examples (43) below:<sup>112</sup>

<sup>112</sup> Topical and focal constituents are bolded. Capital letters denote prosodic stress.

- (43) a. **El libro**, lo trajeron ayer  
 The book CL brought yesterday  
 ‘**The book**, they brought it yesterday.’
- b. **Luces naturales**, no sé si tengo.  
 lights natural no know if have  
 ‘I don’t know if I have **natural lights**.’
- c. **El Libro**, trajeron ayer (no la carta).  
 The book they.brought yesterday (not the letter)  
 ‘It was **the book** they brought yesterday (not the letter).’  
 (Camacho, 1999, p. 118)

- d. *Los inGLEses arrestaron a Pinochet (no los españoles).*  
The British arrested to Pinochet (not the Spanish)

‘The British arrested Pinochet.’

(Camacho, 1999, p. 117)

In (43a), the topic of the sentence, *el libro*, ‘the book’, is both fronted and accompanied by a resumptive pronoun, meaning the topical referent is also specific. On the other hand, the object topic in (43b), *luces naturales*, ‘natural lights’, is a generic entity, as it is fronted without a resumptive pronoun. Similarly, the object in (43c), *el libro*, ‘the book’, is not accompanied by a resumptive pronoun; however, because the referent is specific and prosodically accentuated, it must receive a focus interpretation. Finally, in (43d), the accentuated subject, *los ingleses*, ‘the British’, must be the focus (i.e., non-presupposed, new information) of the sentence because it is prosodically accentuated.

Quechua makes use of morphological and syntactic strategies to express topic and focus. For instance, both topic and focus constituents are accompanied by their respective suffixes, which will be exemplified in the following section. Furthermore, focused constituents “must move to the vicinity of the verb” but topicalized constituents may remain *in situ* (Camacho, 1999, p. 120).<sup>113</sup> Thus, the syntactic realization of topic and focus is effectively the opposite in Quechua as compared to Spanish.

Camacho (1999) examines OV/VO variation in the Spanish of L1 Quechua, L2 Spanish bilinguals residing in Lima using spontaneous speech data elicited through interviews. He reports a “fairly high percentage of inverted direct objects,”<sup>114</sup> an example of which is provided in (44) below (p. 123). The verb is underlined and the object is bolded.

- (44) *Hay veces con las ropitas me ayudan, **ropitas usadas para mis hijitos***  
*me dan*

‘Sometimes they help me with clothes, **used clothes for my children**  
they give me’

(Camacho, 1999, p. 123)

Camacho (1999) argues that in Standard Spanish, examples like (44) are possible, only if the object is stressed and interpreted as contrastive focus because the referent is specific and lacks a resumptive pronoun. However, this is not the case in (44), as the object remains unstressed and is given, not new. Thus, the fronted constituent must be interpreted as topical information. Camacho also notices that the syntactic structure of this example reflects the syntactic structure expected for focused constituents in Quechua. Moreover, he posits that the

<sup>113</sup> In the following section, I present evidence from Sánchez (2010) that the syntactic position of topics in Quechua are quite flexible—they may be fronted, dislocated to the right, or remain *in situ*.

<sup>114</sup> Camacho provides the percentage of pre-verbal objects for each participant separately. The percentage of pre-verbal objects range from 0% to 15%.

relatively high incidence of pre-verbal objects without resumptive pronouns or prosodic stress in Andean Spanish is due to the fact that Quechua permits null objects. He writes, “since the learner’s system includes null objects, these will be present, yielding cases with alternative word orders which look ungrammatical in the target but only because the resumptive pronoun is null” (p. 129). Camacho ultimately concludes that native Quechua speakers transfer two syntactic parameters— null objects and focus fronting— to Spanish in the acquisition of the syntactic-pragmatic interface, yielding a increased frequency of OV order in the predicate.

**Muntendam (2008, 2009, 2010, 2013)**

Like Camacho, the objective of Muntendam’s work (2008, 2009, 2010, 2013) is to determine the nature of cross-linguistic transfer from Quechua to Andean Spanish regarding ordering preferences at the clausal level. In particular, she investigates whether the high rate of OV order in Andean Spanish is the result of syntactic or pragmatic transfer. To determine this, Muntendam identifies the underlying syntactic parameters at work in Quechua, Andean Spanish, and Standard Spanish. First, she observes that, in Standard Spanish, one pragmatic strategy associated with pre-verbal objects is focus fronting— the movement of non-presupposed material to the left edge of a syntactic domain. Muntendam then notes that, in Standard Spanish, two syntactic properties are concomitant of focus fronting— sensitivity to weak crossover effects and long-distance movement. The former is exemplified below:

- (45) [<sub>FA</sub> cada niño<sub>i</sub>] su<sub>i</sub> madre aprecia t<sub>i</sub>  
 [<sub>F</sub>To every child<sub>i</sub>] his<sub>i</sub> mother appreciates

‘His mother appreciates every child.’

(Muntendam, 2008, p. 49)

According to Muntendam (2008), in (45), “*cada niño*, ‘every child’, crosses a pronoun (*su*, ‘his’) with which it is co-indexed, but the pronoun does not c-command the trace, giving rise to weak crossover effects” (p. 49). Thus, the sentence in (45) is only marginally acceptable. Regarding long distance movement, when the object (or subject)<sup>115</sup> is fronted as a result of its focal status, long distance movement is possible in Spanish. This is exemplified in (46) below by the extraction and subsequent fronting of the object focus, *este libro*, ‘this book’, from a subordinate clause.

<sup>115</sup> VPs cannot be focalized via focus fronting.

- (46) [<sub>F</sub>*Este libro<sub>i</sub>*] [<sub>CP</sub> *creo que leyó Juan t<sub>i</sub>*]  
 [<sub>F</sub>This book] I.think that read Juan

‘I think Juan read this book.’

(Muntendam, 2008, p. 52)

Muntendam (2008, 2009, 2010, 2013) employs a sentence-judgment task to determine if both weak cross over effects arise with focus fronting and long distance movement is acceptable in Standard Spanish, Andean Spanish, and Quechua. Examples of the Quechua sentences that were judged by participants for weak crossover effects and long distance movement are respectively provided below in (47) and (48).

- (47) [<sub>F</sub>*Sapa wawa-ta<sub>i</sub>*] *mama-n<sub>i</sub>* *apa-mu-sba-n*  
 [<sub>F</sub>Every child-ACC<sub>i</sub>] mother-POSS.3SG<sub>i</sub> bring-DIR-PROG-3SG  
*yachaywasi-man.*  
 school-ALL.

‘His mother brings every child to school.’

(Muntendam, 2008, p. 50)

- (48) [<sub>F</sub>*Llama-s-ta<sub>i</sub>*] *warmi yuya-n* [<sub>CP</sub> *runa q’ati-sqa-n-ta*  
 [<sub>F</sub>llama-PL-ACC] woman think-3SG man take-NOM-3SG-ACC  
*t<sub>i</sub>*].

‘The woman thinks the man takes the llamas.’

(Muntendam, 2008, p. 53)

Regarding weak-crossover effects, Muntendam (2009, 2010, 2013) finds that 100% of the 12 monolingual Spanish-speaking participants exhibit weak cross-over effects in focus fronting. Of the 15 bilinguals judging sentences in Andean Spanish, 73.3% show weak crossover effects, but none of the 8 native Quechua speakers consulted show weak-crossover effects in the Quechua sentences.<sup>116</sup> Muntendam reports a similar pattern for long distance movement: half of all monolingual speakers of Standard Spanish and half of all bilinguals judging Andean Spanish sentences accept long distance movement of the object. A quarter (25%) of participants in the Standard Spanish group and a third (33.3%) of the Andean Spanish group also accept long-distance movement of the subject. Contrarily, none of the Quechua consultants accept long-distance movement of either the subject or the object. Overall, the results of both sentence-judgment

<sup>116</sup> In Muntendam (2008), she presents a subset of this data.

tasks reveal that, syntactically speaking, Andean Spanish patterns more closely with Standard Spanish than Quechua, given that the former pair of languages both show weak-crossover effects with focus fronting and a similar acceptance rate of long-distance movement of subjects and objects, while neither of these features pertain to Quechua.

Given these results, Muntendam concludes that Quechua contact has not resulted in syntactic transfer to Spanish, evidenced by the fact that the underlying syntactic parameters associated with focus fronting in Andean Spanish and Quechua are not the same. Instead, the high proportion of pre-verbal objects in Andean Spanish is the result of pragmatic transfer. She writes, “there are essentially two possibilities regarding the use of the pre-verbal object in Andean Spanish. the first is that it is not a focus strategy, i.e., the function has been lost. The second possibility is that it is a focus strategy, but that it is more generally used than in Standard Spanish, i.e., the function has been generalized. I tentatively adopt the second hypothesis, i.e., in Andean Spanish fronting of objects is not as restricted as in Standard Spanish” (Muntendam, 2008, p. 55). In later work, Muntendam substantiates this claim through a quantitative analysis of the spoken corpus she constructs using interviews with Ecuadorian and Bolivian bilinguals (Muntendam, 2009, 2010, 2013). She extracts all question-answer pairs that elicit focus on the object from the dataset and finds that in nearly half (45.8%) of these instances, pre-verbal order is employed, as exemplified in (49).

- (49) a. Researcher: *¿Cuántos años tiene usted?*  
 ‘How old are you?’
- b. Julio: *Cuarenta y cinco años tengo.*  
 ‘I am 45 years old.’ (Muntendam, 2009, p. 182)

The above example (49) is clearly one of focus fronting, as the object *cuarenta y cinco años*, ‘forty-five years’, is the focal or non-presupposed material of the sentence, and it appears in a fronted, pre-verbal position. According to Muntendam (2009), this pragmatic context is not commonly associated with OV order in Standard Spanish. Though the high rate of pre-verbal objects in responses to open questions that elicit focus on the object provide clear evidence of focus fronting in Andean Spanish, this subset of data only accounts for 48 of the total 2,350 tokens (2.0%) of OV/VO orders in Muntendam’s spoken corpus.

To better understand the relationship between focus and word order in Andean Spanish and Quechua, Muntendam (2009, 2010, 2013) employs a question-answer pair acceptability task in which subjects provide acceptability judgments for all word orders including an explicit subject, object and verb<sup>117</sup> that answer

<sup>117</sup> SVO, SOV, OSV, OVS, O-cl-VS, VSO, VOS

questions eliciting seven different focus conditions— sentence focus, focus on the subject, focus on the object, focus on the VP, contrastive focus on the subject, contrastive focus on the object, and contrastive focus on the VP. The questions are exemplified in (50), followed by the possible answers in seven different syntactic configurations that were judged by the subjects (51).

- (50) a. **Sentence focus**  
*¿Qué pasa?*  
'What happens?'
- b. **Focus on the subject**  
*¿Quién corta la sogá?*  
'Who cuts the rope?'
- c. **Focus on the object**  
*¿Qué corta el loro?*  
'What does the parrot cut?'
- d. **Focus on the VP**  
*¿Qué hace el loro?*  
'What does the parrot do?'
- e. **Contrastive focus on the subject**  
*¿El cóndor corta la sogá?*  
'Does the condor cut the rope?'
- f. **Contrastive focus on the object**  
*¿El loro corta la cola del zorro?*  
'Does the condor cut the fox's tail?'
- g. **Contrastive focus on the VP**  
*¿El loro coge la sogá?*  
'Does the condor grab the rope?'

(Muntendam, 2009, pp. 188–189)

- (51) a. El loro corta **la sogá**. (SVO)  
 ‘The parrot cuts the rope.’  
 b. El loro **la sogá** corta.’ (SOV)  
 c. **La sogá** el loro corta. (OSV)  
 d. **La sogá** corta el loro. (OVS)  
 e. **La sogá**, la corta el loro. (O-cl VS)  
 f. Corta el loro **la sogá**. (VSO)  
 g. Corta **la sogá** el loro. (VOS) (Muntendam, 2009, pp. 188–189)

Muntendam (2009, 2010, 2013) reports that both Standard Spanish and Andean Spanish subjects accept the prescriptive clausal order, SVO, in response to all questions, but SOV, OSV, and OVS orders are accepted more frequently by Andean Spanish subjects than Standard Spanish subjects for all focus conditions. In other words, pre-verbal objects are generally more acceptable in Andean Spanish than in Standard Spanish. Muntendam also finds that there is no correlation between focus and word order preferences in Andean Spanish; that is, there is a high rate of acceptance for all orders irrespective of the focus condition.

Similar results emerge in the Quechua analysis— acceptability rates for all orders are relatively high regardless of the focus condition. In comparing acceptance rates for answers with pre-verbal objects (SOV, OSV, OVS), Muntendam finds that speakers accept these orders in Andean Spanish and Quechua at comparable rates across all focus conditions (40%-48%), but for Standard Spanish subjects, acceptance rates depend on the focus condition. For instance, when there is focus or contrastive focus on the object, OV orders are accepted by speakers of Standard Spanish at a rate of 27%-28%; however, for all other focus conditions, OV order is accepted at a rate of about 11%-14%. Muntendam (2009) writes, “from these results, we can conclude that in Andean Spanish, fronting of objects is not as restricted as in Standard Spanish. In other words, OV word order has a more general use in Andean Spanish than in Standard Spanish” (p. 203).

It appears that the results of the acceptability judgment task somewhat contradict the notion that a higher proportion of pre-verbal objects in Andean Spanish is due only to focus fronting, especially given the result that OV orders are accepted often regardless of the focus condition. Furthermore, when there is narrow or contrastive focus on the subject, speakers of Andean Spanish accept OVS and O-clVS orders at higher rates than all other orders,<sup>118</sup> around 84% and 87% respectively. Returning to the corresponding focus conditions

<sup>118</sup> Excluding SVO order, which is accepted at a rate of 100% in all focus conditions.

in (5ob) and (5oe), when the subject is focalized, the object, *la sogá*, ‘the rope’, is necessarily given, as it appears in the question. This suggests that pre-verbal objects are most acceptable when their information status is given (i.e., they are topical). It is important to note that O-clVS order is also associated with a high acceptability rate when the object is the focused constituent (75.4%) or the focus of contrast (78.6%). Taken together, these results imply that pre-verbal position is reserved for both topics and foci in Andean Spanish.

The Standard Spanish participants also accept O-clVS at disproportionately high rates that mirror the acceptance rates of Andean Spanish speakers when the subject receives narrow or contrastive focus (86.7% and 88.3%). This finding is consistent with the assertion that “topicalization involves fronting of a constituent” (Camacho, 1999, p. 118), which means that an object topic’s given information status motivates non-canonical pre-verbal collocation in non-contact Spanish varieties. In line with Muysken’s (1984) proposal, it may be the case that an increased frequency of OV order in Andean Spanish is the result of the overextension of an existing Spanish preposing mechanism, which is triggered by certain pragmatic contexts, like topical or focal objects. Though Muntendam emphasizes the role of focus fronting in verb-object inversion, she concludes more generally that pre-verbal objects are less pragmatically restricted in Andean Spanish. She writes, “it can be concluded that Andean Spanish differs from non-Andean Spanish in that the restrictions on the interface between the syntactic system determining placement of constituents and the pragmatic system assigning given/new and topic/focus interpretations to these constituents differ between two varieties: pre-verbal placement of objects is pragmatically restricted in non-Andean Spanish, but not so in Andean Spanish, as a result of influence from Quechua” (Muntendam, 2013, p. 127). In short, Quechua contact weakens the pragmatic conditions that typically govern the pre-verbal positioning of objects in non-contact Spanish.

To summarize, both Camacho (1999) and Muntendam (2008, 2009, 2010, 2013) conclude that the elevated frequency of OV order observed in Andean Spanish is not due to direct syntactic transfer from Quechua but pragmatic transfer. For Camacho, this conclusion is supported by the syntactic similarities between focus fronting in Quechua and pre-verbal objects lacking a resumptive pronoun in Andean Spanish. For Muntendam, the finding that two underlying syntactic properties, weak crossover effects and long-distance movement, apply to pre-verbal objects in both Standard Spanish and Andean Spanish but not Quechua suggests that direct syntactic transfer is not the mechanism that yields a higher rate of OV order in Andean Spanish. Rather, Quechua influence is indirect and pragmatic; that is, contact with Quechua has caused the prag-

matic restrictions that motivate OV order to loosen, which is evidenced by the high acceptability ratings of all word orders in all focus conditions in Andean Spanish.

### 3.4 Word order in Quechua

<sup>119</sup> The majority of studies referenced in this section describe the linguistic properties of Quechua II-C varieties. For those studies that describe a different Quechua variety, I explicitly state the regional variety under investigation.

Now I turn to word order in Quechua,<sup>119</sup> again with a particular emphasis on the order of predicate constituents. I begin by establishing the canonical syntax of clausal constituents, which is followed by a description of other relevant morphosyntactic properties to clarify the typological identity of Quechua and prepare the reader for the many Quechua examples that appear in the current and coming chapters. I dedicate the next subsection to direct objects in Quechua, elaborating on the distribution of the accusative case suffix, *-ta*, and other grammatical properties of objects, like null objects and verbal object agreement marking. Next, I discuss predicate constituent order variation in Quechua and highlight the factors that reportedly condition this variation, like information structure and the omission of the accusative suffix. Finally, I present the results of a variationist analysis of word order, which uncovers statistical correlations between OV/VO variation and sociolinguistic and linguistic predictors, e.g., linguistic dominance of the bilingual speaker and subject expression respectively.

#### 3.4.1 Order of clausal constituents and other relevant properties

Quechua is considered an SOV-dominant language, though all alternative orders (i.e., SVO, VSO, VOS, OSV, and OVS) are also permitted with minimal semantic-pragmatic consequences (Cerrón-Palomino, 1987a; Cusihuamán, 1976, 2001; Salas Cruz, 1993). The syntactic flexibility observed in main clauses is not exhibited in subordinate clauses, which more strictly adhere to prescriptive OV order. Word order flexibility in Quechua may be a byproduct of its rich morphological system in which grammatical relationships between constituents are marked morphologically via suffixation. The following examples in (52), which are extracted from the spoken corpora consulted in the present study,<sup>120</sup> illustrate canonical SOV word order (52a), OV order in a subordinate clause (52b), and a non-canonical word order, VOS (52c). The direct object is bolded and the verb is underlined.

<sup>120</sup> The number accompanied by ‘P’ in brackets indicates the participant who uttered the statement. Participant information is detailed in chapter 4.

(52) a. **SOV**

*nishu-ta willaku-nku chay condenado-kuna-qa*  
very-ADV tell-3PL that condemned.being-PL-TOP  
*runa-ta mikbu-n-mi oveja-ta miku-n-mi*  
person-ACC eat-3SG-FOC sheep-ACC eat-3SG-FOC

‘they say a lot [that] those condemned beings eat people, [and they] eat sheep’ [P2]

b. **OV in a subordinate clause**

*trigo-ta-pas chay-ta bank’a-spa, mate-ta*  
wheat-ACC-COOR that-ACC toast-SUB mate-ACC  
*ruwa-yku chay-ta... pay-kuna-man qu-yku*  
make-IPL.EXCL that-ACC s/he-PL-DAT give-IPL.EXCL

‘and the wheat, toasting that, we make mate, that... we give to them [our kids]’ [P28]

c. **VSO**

*chay-pi bastante puma-kuna ka-sqa ka-ra-n antes...*  
that-LOC a.lot puma-PL be-PST2 be-PST-3SG before  
*ajá chaymanta-qa ah poco a poco mana eh mana*  
uh.huh then-TOP DM little by little NEG DM NEG  
*ka-pu-n-chu... tukukapu-n... sipi-pu-nku ima runa*  
be-REG-3SG-NEG end-3SG kill-REG-3PL and person  
*chay puma-kuna-ta chay desapareci-pu-ra-n mana*  
that puma-PL-ACC that disappear-REG-PST-3SG NEG  
*ka-n-ña-chu puma-pas*  
be-3SG-PFV-NEG puma-COOR

‘there had been a lot of pumas there before... uh huh, then, uh, little by little, uh, they were not... they were completely gone... and people killed those pumas, they disappeared completely, there are no longer any pumas’ [P22]

The above examples also illustrate other relevant morphosyntactic properties of Quechua that I will briefly summarize here. First, in Baker’s terms, Quechua is mostly a dependent-marking language, meaning it is head-final,<sup>121</sup> exhibits SOV order, and grammatical relationships are marked through the morphological strategy of case-marking (Baker, 1996). The head-final label applied to Quechua is further supported by the prescriptive order of two-constituent

<sup>121</sup> However, Kalt and Geary (2021) correctly point out that Quechua also exhibits some properties of head-marking languages, due to the fact that word order is quite flexible in Quechua.

structures in other syntactic domains, e.g., adjective-noun order in the adjectival phrase, possessor-possessed order in the genitive phrase, verb-auxiliary order in the verb phrase, and subordinate-matrix clause order at the sentential level (Cerrón-Palomino, 1987a). Furthermore, Quechua is a postpositional, agglutinative language, meaning spatial, temporal and other relationships are expressed through particles that appear after the relational noun phrase via suffixation.

Example (52a) highlights select dependent-marking properties of Quechua, particularly SOV order and its case-marking system. In this example, the object of the sentence, *runa*, ‘person’, is accompanied by the accusative suffix, *-ta*, and the verb, *mikhuy*, ‘to eat’, is inflected with the third person plural subject agreement particle, *-nku*. Note that the subject, *chay condenadokuna*, ‘those condemned people’, is not accompanied by a nominative suffix because Quechua is not equipped with a nominative case marker. Because all other arguments are marked, the unmarked constituent may be easily identified as the subject. The dative object is also case-marked, as exemplified in the last clause of (52b) in which *paykuna*, ‘they’, are the ones who receive the toasted wheat and mate; consequently, this constituent is marked with the dative suffix, *-man*.<sup>122</sup>

<sup>122</sup> There are two suffixes in Quechua that function as a dative suffix: *-man* and *-paq*. According to Salas Cruz (1993) the distribution of these suffixes is such that *-man* appears on “verbs of communication or transference” (p. 76, my translation), and *-paq* appears in all other contexts.

Example (52c) exemplifies other case-marking suffixes apart from those that mark the basic arguments of a predicate. For example, the locative case suffix, *-pi*, appears on the medial demonstrative, *chay*, ‘that’, to communicate the location where many pumas used to exist— *chaypi*, ‘there’. Additionally, the verb *desaparecipuran*, ‘they disappeared’ in (52c) illustrates that tense is expressed via affixation of a verbal particle between the verb root and the subject agreement particle; in this case, the past tense suffix, *-ra*, intervenes. Likewise, subordination is expressed morphologically in Quechua through the subordinating suffix, *-spa*. In (52c), the clause *chayta bank’aspa*, ‘toasting that’, is subordinate, and the main clause is the one that follows and contains the inflected verb, *mateta ruwayku*, ‘[we] make the tea’.

<sup>123</sup> As mentioned in previous discussions on the relationship between topic and focus and information status, a topic-marked entity is typically given.

Semantic-pragmatic notions, like topic, focus, and evidentiality, are also encoded in suffixes. This is highlighted in (52a) where the topic suffix, *-qa*, affixes to the subject, *chay condenadokuna*, ‘those condemned beings’, effectively identifying the subject as what the sentence is about.<sup>123</sup> Similarly, the suffix *-mi*, which appears on the verb, may be interpreted as a focus particle accentuating the non-presupposed information of the sentence, which, in this case, is that people are eaten by the *condenado*.<sup>124</sup> The relationship between information structure and word order in Quechua will be further discussed in §3.4.3.

<sup>124</sup> I interpret the focus particle in this example as having scope over the entire predicate, *runata mikhunmi*, because the act of ‘eating people’ is non-presupposed information.

### 3.4.2 The direct object in Quechua

#### *TA*-marking

As mentioned above, in Quechua, the accusative object of the sentence is prescriptively case-marked by the accusative suffix, *-ta*. In each example above (52), the bolded object of the sentence is affixed with *-ta*, effectively identifying this constituent as the accusative object.

Prior to detailing the distribution of accusative *-ta*, it is important to clarify that *-ta* serves an additional grammatical function as an adverbial marker. Consider (53) below:

- (53) a. *Allin-ta rima-sunchis.*  
good-**ADV** speak-FUT.1PL.INCL  
'We speak well.'
- b. *Q'uncha-ta-qa khayna-ta llut'a-na.*  
hearth-**ACC-TOP** thus-**ADV** cover-OBLG  
'The hearth must be covered like this.'

(Salas Cruz, 1993, p. 175)

In (53a), the adjective, *allin*, 'good', is converted into a manner adverbial through suffixation of *-ta* to describe the manner in which the subjects speak. Example (53b) illustrates an instance in which *-ta* appears on two constituents and serves a distinct grammatical function in each case. The first *ta*-marked entity, *q'unchataqa*, 'hearth', is the object of the sentence, and the second *ta*-marked entity, *khaynata*, 'thusly', is an adverbial expression that describes the manner in which the hearth is covered. Typically, the context is sufficient to discriminate the function of *-ta* as either an accusative case marker or an adverbial marker.

Returning to the accusative function of *-ta*, it bears mentioning that there are some contexts in which a *ta*-marked constituent would not be considered an accusative object in Spanish. For example, goal arguments of directional verbs are sometimes marked with *-ta* in Quechua (Kalt & Geary, 2021; Salas Cruz, 1993). Consider (54):

- (54) *Wasintun-cha llaqta-n-ta-chu ch'usa-rqa-n.*  
Washington-DIM town-POSS.3SG-ACC-Q travel-PST-3SG  
'Did Washington travel to his town?'

(Salas Cruz, 1993, p. 175)

<sup>125</sup> The suffix *-man* is synthetic; that is, it also marks dative objects.

Goal arguments may also be marked with the allative suffix *-man*,<sup>125</sup> as illustrated in (55).

- (55) *Kay yarqba-qa sara chakra-lla-man-mi ri-sha-n.*  
 This irrigation.canal-TOP corn field-LIM-ALL-EVIDI go-PROG-3SG  
 ‘This irrigation canal goes only to the corn field.’  
 (Salas Cruz, 1993, p. 77)

According to Salas Cruz (1993), both *-ta* and *-man* may be used “when the actor who executes the action is a person,” but *-ta* emphasizes the goal argument itself, and *-man* emphasizes the process of moving from one place to another (p. 76, my translation).

The suffix *-ta* may also appear on an infinitive verb within a complex verb phrase accompanied by certain inflected verbs, like *atiy*, ‘to be able to’; *munay*, ‘to want’; *qallariy*, ‘to begin’; *tukuy*, ‘to finish’, *yachay*, ‘to learn’; and *wanay*, ‘to learn from experience’ (Salas Cruz, 1993, p. 176). Consider (56) below:

- (56) *Sira-y-ta yacha-sha-ni-ña.*  
 sew-INF-ACC learn-PROG-1SG-PFV  
 ‘I am already learning to sew.’  
 (Salas Cruz, 1993, p. 177)

In (56), the uninflected verb, *siray*, ‘to sew’, is case-marked with the accusative morpheme, *-ta*, which implies that this constituent functions as the direct object of the verb *yachay*, ‘to learn’. In other words, it is the action of sewing that is being learned by the subject. Sánchez (2010) describes *ta*-marked infinitives as nominalized verbs as they exhibit both nominal and verbal morphosyntactic properties. That is, “they can be marked for case as DPs and at the same time they take internal complements that are case marked” (p. 16), as exemplified in (57) below:

- (57) *T’anta-ta ruwa-y-ta muna-ni.*  
 bread-ACC make-INF-ACC want-1SG  
 ‘I want to make bread.’  
 (Sánchez, 2010, p. 16)

In the above example, *ruwayta*, ‘to make’, which I call an **infinitive verb object**, functions as the accusative object of the inflected verb of the sentence, *munani*, ‘[I] want’. In other words, performing the action of ‘making’ is what is wanted by the subject. As Sánchez (2010) mentions, because *ruwayta* retains

some verbal functions, it permits a second accusative-marked entity, *t'antata*, 'bread'. Note that the object licensed by the infinitive verb object is not always case-marked, as illustrated in (58).

- (58) *Qulqi suchi-ku-y-ta ati-yman-chu.*  
 money send-REFL-INF-ACC can-COND.ISG-Q  
 'Could I send money?'  
 (Salas Cruz, 1993, p. 177)

In the above example, accusative case is overtly expressed on the infinitive verb object, *suchikuyta*, 'to send', but not on the NP object constituent, *qulqi*, 'money', which is licensed by the infinitive verb object.

Relatedly, *-ta* is also employed in conjunction with a nominalizing suffix, *-sqa* or *-na*, and a possessive suffix to form complement clauses, as exemplified below in (59).

- (59) a. *Huwan hamu-sqa-n-ta yacha-ni.*  
 Huwan come-NMLZ-POSS.3SG-ACC know-1SG  
 'I know that Huwan had come.'
- b. *Huwan hamu-na-n-ta yacha-ni.*  
 Huwan come-NMLZ-POSS.3SG-ACC know-1SG  
 'I know that Huwan had come.'
- (Sánchez, 2010, p. 17)

In both (59a) and (59b), the complement clause is formed via a sequence of suffixes on the subordinate verb including *-ta*, unlike complement clauses in Spanish, which are separated from the main clause with the lexical complementizer *que*, 'that'. Clausal complements like these are not included as object constituents in the analysis of the present study but are illustrated here to demonstrate the versatility of this suffix.

Though accusative objects are prescriptively marked with *-ta* in Quechua, it has been observed that *-ta* marking may not be invariably obligatory. For example, Muntendam (2015) finds that in the elicited speech of bilingual Quechua speakers from Cochabamba, Bolivia, *-ta* does not always appear on the direct object. In fact, she claims that overt *ta*-marking is in free variation with final stress on the object NP in non-utterance-final position. Because "primary stress falls on the penultimate syllable" for most words in Quechua,<sup>126</sup> final stress on an object is a marked stress pattern that may be sufficient for distinguishing an object from other grammatical elements of the sentence (Muntendam, 2015,

<sup>126</sup> There are exceptions to this norm, however, such as "words that are marked with an emphatic or exclamatory suffix attracting ultimate stress, e.g., *An-chacháy!*, 'Wow, it's hot!'" (Muntendam, 2015, p. 215).

p. 215). An example of prosodic object marking through final stress is illustrated in (60).

(60) *Runa-s-qa yunTA wata-sa-nku*  
man-PL-TOP yoke tie-PROG-3PL

‘The men are tying the yoke.’

(Muntendam, 2015, p. 231)

In the example above, *yunta*, ‘yoke’, is the accusative object of the verb *watay*, ‘to tie’, but it is not *ta*-marked. Instead, the final syllable is accentuated to indicate its status as the direct object; prescriptively speaking, the prosodic stress should fall on the first syllable of this word in the absence of other suffixes.

In an investigation of Quechua first-language acquisition in Chalhuanca, Peru, Courtney (1998) finds that one of the three child participants she consults, omits the accusative case marker in about a third of all instances where *ta*-marking is expected. Though *ta*-marking becomes more regular as he ages from 2 years and 5 months to 2 years and 10 months of age, Courtney notices that most *-ta* omissions occur when the direct object is *pelota*, ‘ball’, a Spanish borrowing ending in *-ta*. Sánchez (2003) also reports a high rate of *ta*-omission among Spanish loanword objects (66.1%) compared to *ta*-omission among Quechua objects (13.5%) in the speech of older children (8-13 years of age) from Ulcumayo, Peru.<sup>127</sup> To my knowledge, there are no studies on adult bilingual Quechua that identify a relationship between *ta*-marking and Spanish borrowings.

Lastly, Sánchez (2003) claims that the accusative marker *-ta* has begun to take on “some determiner-like properties” in Ulcumayo and Lamas Quechua (p. 100). Though traditionally Quechua does not have definite or indefinite determiners, in these varieties of contemporary Quechua, the numeral *buk/suk*, ‘one’, is emerging as an indefinite determiner. Definiteness, on the other hand, may be prescriptively indicated through the use of a demonstrative adjective, like *chay*, ‘that’. Sánchez concludes, “the co-occurrence of the dropping of *-ta* in Lamista Quechua with more extended use of indefinite and definite determiners indicates that in this variety, a new overt determiner system is emerging that replaces *-ta* in its determiner-like functions” (p. 101). Kalt and Geary (2021) also find that *buk* may function as an indefinite determiner in Chuquisaca Quechua,<sup>128</sup> but they do not provide evidence of a correlation between *ta*-omission and the presence of indefinite or definite determiners.

<sup>127</sup> In Sánchez’s sample of Lamas Quechua, however, only 6 direct objects of 312 total are Spanish borrowings, and, as a consequence, no meaningful conclusions can be drawn from such a small number of tokens.

<sup>128</sup> The same observation does not extend to Cusco Quechua.

## Null objects and object agreement

It is worth mentioning one final characteristic of objects in Quechua— the availability of null objects (Sánchez, 1999, 2003, 2010). According to Sánchez (1999), “null objects in Central Quechua are pronominal in nature and can be treated as instances of null pronouns (given the existence of overt strong forms) or as instances of null morphological markers (given the existence of overt morphology for first and second person direct objects)” (p. 235). The former type of null objects is illustrated by an excerpt from a conversation between two Ulcumayo Quechua speakers provided in (61) below:

- (61) a. *Chawra-qa rantika-yaa-ma-nki-man chay shutu-kuna-ta<sub>i</sub>.*  
 then-TOP sell-PL-OBJ.ISG-2SG-COND those drop-PL-ACC<sub>i</sub>  
 ‘Then could you sell me those drops?’
- b. *Imanuyyata kay-ta-qa<sub>i</sub> truraa-ku-shaq*  
 how this-ACC-TOP<sub>i</sub> put-REFL-FUT.ISG  
 ‘How do I apply these?’
- c. *Alita rura-shpti-qa, kima uura-ta-ra-n e<sub>i</sub> shuya-nki.*  
 well do-SUB-TOP three hour-ACC-REP-FOC e<sub>i</sub> put-2SG  
 ‘When they begin to work, put (them) on every three hours.’  
 (Sánchez, 2003, p. 37)

In this example, the object, *shutukunata*, ‘drops’, is introduced into to the discourse as a full NP in (61a), then referred to using a topic-marked demonstrative pronoun, *kaytaqa*, ‘those’ in (61b). In (61c), the object is null, which is grammatical because the referent is activated in the discourse. Thus, the interlocutors understand that the drops are the entity being applied every three hours, even in the absence of an explicit referent.

Third person human objects are also be null in Quechua with respect to morphological object agreement marking on the verb. In Cusco Quechua, an object agreement particle appears on the verb under two conditions: first, the object refers to a human entity, and, second, this human entity is a first or second person referent. For instance, when the object is the first person singular entity *ñuqa*, ‘I’, the object agreement morpheme *-wa* intervenes between the verb root and the subject agreement suffix, as exemplified in (62). Verbal agreement marking occurs on the verb when the object is either accusative (62a) or dative (62b). When the object is dative, object marking is obligatory, but the NP must be null (Kalt, 2002). Example (62a) from the dataset of the present study

<sup>129</sup> This example highlights another interesting verbal structure in contemporary bilingual Quechua, which is *gustar*.

illustrates that when the object is accusative, it may be explicit and marked on the verb simultaneously.<sup>129</sup>

- (62) a. *nuqa-qa kay Qusqu-man hamu-ra-ni... nuqa-ta*  
 I-TOP this Cusco-ALL come-PST-1SG I-ACC  
*suwa-wa-ra-nku... iskay runa-kuna... huq muqu-ta*  
 rob-OBJ.1SG-PST-3PL two person-PL some hill-ACC  
*ura-yu-sha-spa, suwa-wa-ra-nku*  
 descend-INT-PROG-SUB rob-OBJ.1SG-PST-3PL

‘I came to Cusco... they robbed me... two people... descending from a hill, they robbed me’ [P18]

- b. *(\*ñuqa-man) sara-ta qu-wa-n.*  
 (I-DAT) corn-ACC give-OBJ.1SG-3SG

‘He gives me the corn.’

(Kalt, 2002, p. 51)

On the contrary, there is no object agreement particle that corresponds to a third person object. In other words, when the object refers to a third person singular or plural human entity, object agreement marking on the verb is null, and the NP object may be either overt or null (Kalt, 2002; Sánchez, 2010). This is illustrated in (63) below:

- (63) *Mariya-m (pay-ta) riku-∅-n.*  
 Mariya-FOC/EVIDI (s/he-ACC) see-∅-3SG

‘Mariya sees (him/her).’

(Sánchez, 2010, p. 27)

In this example, the third person object pronominal, *payta*, ‘s/he’, may be explicit or null, but in either scenario, there is no agreement marking on the verb because the object is a third person entity.

### 3.4.3 Variable word order in Quechua

In this section, I turn to a discussion of word order variation in Quechua in which I highlight two linguistic variables that condition syntactic variation—variable *-ta* marking and information structure—and present the findings of a variationist analysis on predicate constituent order variation in bilingual Cusco Quechua.

## ***TA*-marking**

Though *ta*-marking is the prescriptive mechanism by which the accusative constituent of a predicate is labelled, it has been observed in some varieties of Quechua that *ta*-marking is not required by the syntax. For instance, in investigating functional interference and convergence in the speech of bilingual Quechua children from Lamas and Ulcumayo, Sánchez (2003) finds that the children of these regions do not consistently mark object constituents with *-ta*.<sup>130</sup> In Lamas Quechua, for example, 60.2% of all DPs<sup>131</sup> are not overtly marked with *-ta*. Sánchez concludes, “I take this to indicate that the Lamas Quechua morphological case system is evolving towards a system in which the accusative marker *-ta* is being dropped” (p. 96). In Ulcumayo, morphological accusative case-marking is more robust, but by no means compulsory, with 21.3% of object DPs lacking *-ta*.

Crucially, Sánchez (2003) observes that a reduction in morphological case-marking coincides with non-canonical clausal ordering preferences in Lamas and Ulcumayo Quechua. Once again, the frequency of the non-prescriptive variant, in this case SVO, is unexpectedly high in both varieties, but especially in Lamas Quechua. In this variety, 51.2% of all utterances exhibit SVO order.<sup>132</sup> In Ulcumayo Quechua, SVO order occurs at a rate of 29.6%, which is at least three times as frequent as all other orders with an overt subject and object. On the contrary, canonical SOV is exhibited in only 5.7% of all Lamas Quechua utterances and 9.9% of Ulcumayo Quechua utterances. Considering only the order of the explicit object and verb constituents by conflating all VO-type (e.g., SVO, VO, VSO, and VOS) and OV-type (e.g., SOV, OV, OVS, OSV) orders into two categories, 64.8% of all predicate constituents are configured in VO order in Ulcumayo Quechua, and 85.1% exhibit this same configuration in Lamas Quechua. So, according to the frequency criteria proposed by Dryer (2013a), both SVO and VO orders are dominant in Ulcumayo and Lamas Quechua.

Though Sánchez does not offer data regarding the proportion of marked and unmarked accusative objects across OV and VO orders, she ultimately concludes that both of these trends<sup>133</sup> provide “evidence of functional interference that leads to changes in the syntactic and morphosyntactic representations of bilinguals” in both Andean Spanish and Quechua (p. 155). Operating from a generative perspective, Sánchez (2003) takes the high incidence of SVO order in both dialects as “evidence of the projection of a ClP<sup>134</sup> in the bilingual Quechua grammar. A shift to dominant SVO order is a major syntactic change in the grammar of Quechua and is predicted by the Functional Interference Hypothesis. It is triggered by the constant activation of the Cl features that require verb movement to Cl in Spanish” (p. 101). To clarify, Sánchez notes

<sup>130</sup> Sánchez (2003) applies a story-telling task to elicit spoken data in this study.

<sup>131</sup> Sánchez (2003) provides data for Quechua DPs, Spanish DPs, and complex DPs separately.

<sup>132</sup> Sánchez reports raw frequencies and percentages for the the following word order configurations in transitive sentences: SVO, SOV, SV, VO, OV, V, OVS, OSV, VSO, VOS, VS. In orders that do not have an explicit constituent written, the pronominal subject or object is null.

<sup>133</sup> Along with the “the emergence of an indefinite determiner in bilingual Quechua as well as the gender-neutral specification of clitics and the emergence of null objects as continuing topics in bilingual Spanish” (p. 155).

<sup>134</sup> Clitic Phrase (Sánchez, 2003, p. 41)

earlier on that movement to a ClP is the primary syntactic difference between Quechua and Spanish. She writes, “verbs move outside the VP in Spanish to Cl, a functional category associated with topic, while in Quechua verbs stay inside the VP in sentences with canonical word order and move only for focus reasons” (p. 42). The *Functional Interference Hypothesis* that Sánchez advances proposes that syntactic change is the result of “the activation of functional features in one language triggered by input in the other language” (p. 13). Sánchez concludes that because the syntactic strategies from both languages (i.e., verb movement to Cl and the verb remaining in the VP) are available to the bilingual speaker, both SVO and SOV orders are grammatically acceptable in these the two varieties of Quechua under investigation. She anticipates that activation of Spanish features will ultimately result in convergence on SVO order, as predicted by her *Functional Convergence Hypothesis*.

Kalt and Geary (2021) also investigate the relationship between *ta*-marking and word order in two regional varieties of adolescent bilingual Quechua in Chuquisaca, Bolivia and Cusco, Peru. The authors find that, overall, about 19.6% of all tokens lack *ta*. Furthermore, there are regional differences regarding the rate of *ta*-omission such that Chuquisaca Quechua bilinguals omit *ta* more frequently than Cusco Quechua bilinguals. They posit that *ta*-omission may be more frequent for Chuquisaca bilinguals due to the availability of two linguistic strategies for marking accusative objects in Bolivian Quechua: *ta*-marking and final stress (Muntendam, 2015). Contrary to the observed association between SVO order and *ta*-omission reported by Sánchez (2003), Kalt and Geary (2021) find that *ta*-omission occurs more frequently with canonical OV-type orders. In fact, 98.2% of tokens lacking *ta* exhibit OV order. In general, OV-type orders are much more frequent in this study compared to Sánchez’s findings, and Chuquisaqueños employ canonical order at an even higher rate, 89.2%, than do Cusqueños, 78.0%. Kalt and Geary (2021) conclude that “the loss of case marking in Chuquisaca places a greater burden on word order to indicate grammatical roles, resulting in more consistent use of the canonical OV word order” (p. 18). Thus, a shift in stress from the penultimate to the final syllable does not appear to be a sufficient condition for distinguishing direct objects in this variety of Quechua. They liken this morphosyntactic adjustment to a case of diachronic linguistic change that transpired as Old English gave way to Modern English in which “case marking was lost and word order became a more rigid cue to grammatical relations” (p. 18).

The results reported by Sánchez (2003) and Kalt and Geary (2021) are nearly opposite regarding the proportions of OV/VO orders. One possible explanation for the discrepancy in ordering preferences may be the regional variety of

Quechua spoken by the adolescent participants. Moreover, the participants from Chuquisaca and Cusco reside in rural areas and were not exposed to Spanish prior to commencing formal education, and the majority come from monolingual Quechua-speaking homes (Kalt & Geary, 2021, p. 4). The participants from Lamas and Ulcumayo, on the other hand, reside in a more populated rural area, were often exposed to Spanish prior to receiving formal education,<sup>135</sup> and mostly come from Spanish-speaking or bilingual homes. One important methodological difference between the two studies is that, in the elicitation task employed by Kalt and Geary (2021), the participant is prompted with an SOV sentence, which could prime SOV in participant responses. The authors acknowledge, “it is possible that the use of SOV sentences in the comprehension task induced a syntactic priming effect among participants (J. K. Bock, 1986; Loebell & Bock, 2003), leading to an overall greater use of OV order during the production task.” Taking together the regional variety, linguistic profile of the participants, and methodological differences, it is possible to reconcile the near mirror-image proportions of OV/VO orders reported by Sánchez (2003) and Kalt and Geary (2021).

<sup>135</sup> Only 9 of 30 participants (30.0%) in the Lamas group and 10 of 28 participants in the Ulcumayo group (35.7%) claimed Quechua as their mother tongue.

Table 3.6 below provides the proportions of OV/VO orders and the rate of *ta*-omission in the studies discussed above. Recall that Sánchez (2003) and Kalt and Geary (2021) report an opposite conditioning effect of *ta*-omission on the syntax. In other words, *ta*-omission is associated with a higher rate of VO order according to Sánchez (2003); however, Kalt and Geary (2021) find that OV order is nearly categorical with direct objects lacking *-ta*.

Variety	OV	VO	<i>ta</i> -omission	Study
Chuquisaca	89.2%	10.8%	>19.6%	Kalt and Geary (2021)
Cusco	78.0%	22.0%	<19.6%	Kalt and Geary (2021)
Ulcumayo	35.2%	64.8%	21.3%	Sánchez (2003)
Lamas	14.9%	85.1%	60.2%	Sánchez (2003)

Table 3.6: OV/VO distribution in Quechua word order studies

### Information structure

Though word order in Quechua is notably flexible, it is not entirely free (Sánchez, 2010).<sup>136</sup> It has been argued that constituent order is restricted by the information structure of the sentence, i.e., the topical or focal status of one or more constituents. To clarify, Sánchez (2010, pp. 43–44) defines the **topic** of a sentence as the discourse-accessible information and the **focus** as the non-presupposed information, á la Zubizarreta (1998). In Quechua, information structure is con-

<sup>136</sup> The forthcoming description of the interaction between information status and word order in Southern Quechua (i.e., Quechua II-C) as described by Sánchez (2010) is taken from Hubbel (2023, pp. 119–120) with a few adjustments.

<sup>137</sup> The suffix *-mi* also functions an evidential particle, which communicates that the speaker has first-hand evidence of information conveyed by the utterance (Faller, 2007; Muysken, 1995). As summarized by Muntendam (2015), Muysken (1995) claims that *-mi* may either convey focused or first-hand evidential material when it appears on a sentence-initial constituent, but when *-mi* appears on later constituents, the *mi*-marked constituent takes on a focus interpretation. For more on the syncretic nature of this suffix, see Sánchez (2010, pp. 53–98).

<sup>138</sup> The focus morpheme has three primary allomorphs in Southern Peruvian Quechua: *-n*, *-m* *-mi*. The distribution of these allomorphs is dependent on the surrounding phonological context: the suffix *-mi* appears on words ending in a consonant, and the suffixes *-n* and *-m* appear on words ending in a vowel.

veyed both morphologically and syntactically. Regarding the morphological strategies, Quechua is equipped with a topic marker, *-qa*, and a focus marker, *-mi/-n*.<sup>137</sup> These suffixes are mutually exclusive in that they may not affix to the same constituent (Muntendam, 2015). The morphological strategies of marking topic and focus are illustrated in (64) and (65) respectively:

(64) *Atuq-qa kusi-ku-n.*  
 Fox-**TOP** happy-REFL-3 S G  
 ‘As for the fox, he is happy.’  
 (Sánchez, 2010, p. 31)

(65) *Pirdu wasi-ta-n ruwa-n.*  
 Pirdu house-ACC-**FOC** build-3 S G  
 ‘It is the house that Pirdu builds.’  
 (Muysken, 1995, p. 380)

In (64) the subject, *atuq*, ‘fox’, is also interpreted as the topic of the sentence, as it is accompanied by the topic suffix, *-qa*. Similarly, in (65), the focus suffix *-n*,<sup>138</sup> is affixed to the direct object, *wasi*, ‘house’, which highlights the focal, non-presupposed information of the sentence. Though both the topicalized and focalized constituents in examples (64-65) appear *in situ*, these constituents may also be subject to fronting. Consider (66) and (67) below:

(66) *Kawallu-ta-m allqu-qa kani-n.*  
 Horse-ACC-**FOC** dog-**TOP** bite-3 S G  
 ‘It is the horse that the dog bites.’  
 (Sánchez, 2010, p. 47)

(67) *Wasi-ta-qa Pirud-m ruwa-rqa-n.*  
 house-ACC-**TOP** Pirdu-**FOC** build-3 S  
 ‘The house, Pirdu built.’  
 (Sánchez, 2010, p. 71)

The focalized accusative object in (66) is fronted, appearing before the subject in a non-canonical position. Likewise, the object topic in (67) has been fronted.

Topicalized constituents in Quechua may also appear at the right periphery of a sentence, i.e., in post-verbal position, unlike focalized constituents, which

may only be fronted or remain *in situ* (Sánchez, 2010). Furthermore, Sánchez notes that occasionally, right-dislocated topics do not bear the topic marker, *-qa*. Consider (68) below:

- (68) *Hinaspa-n tariru-spa ka-rqa-n pajaru-cha-ta*  
 Then-FOC find-GER be-PST-3SG bird-DIM-ACC  
*qillu-cha-ta.*  
 yellow-DIM-ACC

‘Then, (she) had found a yellow bird.’

(Sánchez, 2010, p. 38)

Sánchez observes that the right-displaced object in (68), *pajaru-cha-ta qillu-cha-ta*, ‘a yellow bird’, is the topic of the sentence, which is why it appears in post-verbal position. In this example, the syntactic strategy of rightward displacement conveys topichood, in lieu of a morphological one. However, this particular example, which is uttered by a Quechua-Spanish bilingual child, is judged ungrammatical by an adult consultant, who requires that the displaced object be morphologically marked with *-qa*.

Another important difference between topic and focus in Quechua is that more than one constituent may bear the topic suffix within a single sentence, but only one constituent may bear the focus suffix. Muysken (1995, p. 385) offers the following description of topic and focus in Quechua:

- (69) {X-qa}o-2 {V/XP}-evi ... {Z-qa}o-3

This schema conveys that “between null and two *-qa* phrases may occur at the beginning of the sentence, followed by a constituent or the verb marked with an evidential,<sup>139</sup> eventually followed by between null and three other *-qa* phrases” (p. 385). This is illustrated by the following example:

<sup>139</sup> The term ‘evidential’ refers to the syncretic focus particle, *-mi*.

- (70) *Mama-y-qa Qusqu-ta-n ri-ra-n.*  
 mother-POSS.ISG-TOP Cusco-ACC-EVIDI to-PST-3SG

‘My mother went to Cusco.’

(Muysken, 1995, p. 385)

In (70), the subject of the sentence, *mamay*, is also the topic, as it bears the topic marker. The direct object, *Qusqu*, ‘Cusco’, is the focus, which is marked with the particle *-n*. No additional topic-marked constituents follow the focused element, which is a grammatically acceptable configuration according to Muysken’s schema (69).

Sánchez (2010) agrees that *qa*-marked topics may precede focalized elements; however, when an object is topic-marked *in situ* the preceding subject must be receive morphological focus-marking, as illustrated in (71) below:

- (71) *Mariya*-(*n*)                      *wasi-ta-qa*                      *riku-rqa-n*.  
 Mariya-\*(F O C / E V I D I) house-A C C - T O P see-P S T - 3 S G  
 ‘Mariya saw the house.’

(Sánchez, 2010, p. 32)

<sup>140</sup> Note that verbal topics are not included in the table as it is unclear from the previous literature if morphologically-marked verbal topics are grammatical.

<sup>141</sup> According to Sánchez (2003, p. 31), OVS order is also an acceptable order when the object topic is fronted.

<sup>142</sup> OV-type orders are OV, SOV, OVS, and OSV orders; VO-type orders are VO, SVO, VOS, and VSO orders. Note that VOS order is not listed as possible configuration in the above table. Sánchez (2003) reports only one instance of VOS order in Ulcumayo Quechua and one instance in Lamas Quechua, which accounts for 0.1% and 0.2% of all possible word orders respectively. In other words, though possible, VOS order is exceedingly rare in Quechua.

<sup>143</sup> Muntendam finds that other word orders are also exhibited, like OVS and OV, but she provides a more detailed comparison of only SOV versus SVO orders, since these are the most frequently used orders.

To review, both focalized and topicalized constituents may either remain *in situ* or be fronted, but only topics may also be right-dislocated. Table 3.7<sup>140</sup> below summarizes possible syntactic configurations when any constituent of a transitive predicate bears a topic or focus marker.<sup>141</sup>

	<i>in situ</i>	fronted	right-dislocated
<i>Focus-subject</i>	S-mi O V	S-mi O V	-
<i>Focus-object</i>	S O-mi V	O-mi S V	-
<i>Focus-verb</i>	S O V-mi	V-mi S O	-
<i>Topic-subject</i>	S-qa O V	S-qa O V	O V S-(qa)
<i>Topic-object</i>	S-mi O-qa V	O-qa S V	S V O-(qa)

Table 3.7: Order of topicalized and focalized constituents

The shaded cells in Table 3.7 are contexts that are expected to yield non-canonical VO-type orders.<sup>142</sup> That is, both focus-fronted verbs and right-dislocated object topics should exhibit VO order, according to Sánchez (2010).

Though, prescriptively, topicalized and focalized elements in Quechua must be marked with the respective topic and focus suffixes, there is evidence that these particles are lacking in some varieties. For instance, in both a picture-story task and question-answer pair elicitation task completed by Quechua-Spanish bilinguals from Tarata and Huayculí, Bolivia, Muntendam (2015) finds that the topic suffix, *-qa*, is used only sparingly by two of six participants, and focus suffixes are not employed at all. She concludes, “in many varieties of Quechua (e.g., Cusco Quechua), morphological markers are obligatory on a focused element when it is fronted, but not so in this [Bolivian] variety” (p. 222). Instead, she finds that a syntactic strategy of encoding information structure is employed, which involves the distinct distribution of SOV and SVO orders across three focus conditions— broad focus, contrastive focus on the subject, and contrastive focus on the object.<sup>143</sup> The questions that elicit these focus conditions are exemplified in (72) below:

- (72) a. **Broad focus**  
*Ima-taj pasa-sa-n?*  
 what-CONTR happen-PROG-3SG  
 ‘What is happening?’
- b. **Contrastive focus: subject**  
*Warmi-chu llama-ta q’ati-sa-n?*  
 woman-Q llama-ACC drive-PROG-3SG  
 ‘Is the woman driving the llama?’
- c. **Contrastive focus: object**  
*Runa kbuchi-ta-chu q’ati-sa-n?*  
 man pig-ACC-Q drive-PROG-3SG  
 ‘Is the man driving a pig?’

(Muntendam, 2015, p. 220)

In the broad focus condition, SOV order is exhibited in 70.9% of responses and SVO in 29.1%. This result indicates that SOV is the preferred order in pragmatically ‘neutral’ conditions, but non-canonical SVO order is still prominent. When there is contrastive focus on the object, SOV order is also preferred over SVO order (75% versus 25%). In this case, the object conveys new, contrastive information, and it has a tendency to appear *in situ*. However, post-verbal position is also available for focused objects, which should not be possible according to Sánchez (2010), who asserts that focused elements can either be fronted or appear *in situ*. Perhaps post-verbal position is also available for focalized elements in Bolivian Quechua due to the obsolescence of focus morphemes, or contrastive focus simply patterns differently than narrow focus. Lastly, when the subject receives contrastive focus, SOV order is only slightly preferred (56.7%) over SVO order (43.3%). It is important to note that, in these cases, the object is given via its explicit mention in the question. Though the object is likely not marked with *-qa*, considering its given information status, the object may appear either *in situ* or at the right periphery, which is predicted by Sánchez (2010) and borne out in Muntendam’s results for this particular focus condition.

Muntendam (2015) also provides evidence that suggests a correlation between contrastive focus and certain prosodic features, like “a higher Fo maximum, a higher intensity, and an increased duration of the the stressed syllable” (p. 241).<sup>144</sup> Overall, Muntendam concludes that, in this variety of Quechua, information structure is communicated through both syntactic and prosodic strategies, though somewhat inconsistently. She theorizes that the emergence and employment these non-prescriptive strategies are likely due to Spanish con-

<sup>144</sup> See Muntendam and Torreira (2016) for more detailed insights regarding the relationship between prosody and focus in Quechua and Spanish.

tact, which “may result in different linguistic outcomes in different regions” (p. 243).

### Variationist analysis

In earlier work, I conducted a variationist analysis of predicate constituent order in bilingual Cusco Quechua using spontaneous speech elicited via sociolinguistic interviews with a subset of 10 adult participants included in the present study (Hubbel, 2023). In general, OV order accounted for 63.7% of all tokens with an object and verb, and the remaining 36.4% exhibited VO order. Though OV order is more frequent than VO order in this sample, according to the statistical criteria for assessing dominant order advanced by Dryer (2013a), this sample of bilingual Quechua is characterized by a lack of dominant order among predicate constituents.

In this analysis, I also considered the conditioning effect of extralinguistic and linguistic variables on predicate constituent order, much like the present study. Regarding the analysis of extralinguistic predictors, I found that the rate of canonical OV order is higher for L1 Quechua bilinguals with higher BLP scores<sup>145</sup> (i.e., more Quechua-dominant) than for simultaneous Quechua-Spanish bilinguals with lower BLP scores (i.e., more Spanish-dominant). That the first language and linguistic dominance of the speaker is correlated with ordering preferences indicates a clear effect of language contact in this syntactic domain.

I also found that two linguistic variables— grammatical category of the object referent and subject expression— condition OV/VO variation in this variety of bilingual Quechua. Concerning the former, a direct object referent with multiple NPs (i.e., complex object constituent) exhibits a preference for post-verbal position (73a), and pronominal objects exhibit a preference for pre-verbal position (73b). Consider (73) below:

- (73) a. *tía-kuna-wan mikhu-sunchis buatia-ta queso-ta*  
aunt-PL-INST eat-FUT.IPL.INCL wathia-ACC cheese-ACC

‘We will eat wathia and cheese with the aunts.’

<sup>145</sup> The Bilingual Language Profile (BLP) score indicates the linguistic dominance of the speaker based on self-reported information (Birdsong et al., 2012).

b. *kan cebada-yku cb'uñu-yku*  
 there.is barley-POSS.IPL.INCL freeze.dried.potato-POSS.IPL.INCL  
*papa-yku oca lisas... chay-kuna-ta*  
 potato-POSS.IPL.INCL oca olluco that-PL-ACC  
*mikhu-yku*  
 eat-IPL.INCL

‘There is our barley, our freeze-dried potato, our potatoes, oca, oluco... we eat those [things].’

(Hubbel, 2023, p. 130)

In (73a) the object constituent contains two NPs, *huatiata*, ‘wathia’, and *quesota*, ‘cheese’, which are collocated in post-verbal position. In (73b), on the other hand, the pronominal object, *chaykunata*, ‘those’, appears in pre-verbal position. The above examples also illustrate that, in the operationalization of this variable, I effectively conflate two linguistic predictors that have been identified as independent predictors of ordering preferences in previous studies on word order variation— weight and definiteness. In other words, the difference between the object constituents in (73a) and (73b) is not just that one contains multiple NPs and the other a single pronoun, rather that the former represents a heavy, indefinite object constituent and the latter a light, definite object constituent. I rectify this conflation in the current study by evaluating the conditioning effect of definiteness and weight on predicate constituent order variation separately.

In Hubbel (2023), I also observed that subject expression exercises a significant effect on predicate constituent order variation such that overt subjects are associated with a higher rate of VO order than null subjects. I tentatively attribute the relationship between subject expression and ordering preferences to matters of iconicity (i.e., a grammatical manifestation of the subject acting on the object), though I myself am not convinced by this explanation. If it were the case that SVO order is ‘more iconic’ than SOV order, it may be reasonably expected that the world’s languages show a preference for SVO order in general. However, SVO is not universally preferred over SOV; in fact, SOV-dominant languages are slightly more statistically prominent than SVO-dominant languages (41.0% versus 35.5%). I refine my thoughts on the relationship between subject expression and predicate constituent order variation in Chapter 7.

Lastly, it is important to note two findings with respect to the interaction between information structure and word order reported in Hubbel (2023). First, I found no significant effect of information status (new versus given) on ordering preferences. That new object referents may occupy post-verbal position chal-

lenges Sanchez's (2010) claim that only topics (i.e., given, discourse-accessible) may be dislocated to the right. Second, there are very few instances of morphological expression of focus and topic on the subject, object, and verb constituents (i.e., the use of *-qa* and *-mi/-n*). This finding runs counter to Muntendam's (2015) assertion that focus-fronted elements are morphologically-marked in Cusco Quechua. The relationship between information structure (i.e., the information status of the object referent) and predicate constituent order variation will be examined further in the present study.

### 3.5 Summary

In broad strokes, this chapter has endeavored to both certify the canonical order of predicate constituents and present the factors that reportedly condition their variation in non-contact Spanish, Andean Spanish, and Quechua. Beginning with non-contact Spanish, in §3.2 I established that Spanish is an SVO-dominant language, though VSO order constitutes a frequent alternative (Muntendam, 2009; Zubizarreta, 1998). OV-type orders are reportedly possible in non-contact varieties; however, they are quite limited (Givón, 2001; Puerma Bonilla, 2019) and are motivated primarily by specific pragmatic circumstances, like the topical or focal status of the object constituent (F. A. Ocampo, 1992; F. A. Ocampo & Klee, 1995). Quantitatively speaking, the observed rate of OV-type orders in the extant literature ranges from <1% to nearly 11%; the lower end of this range reflects the syntactic patterns of literary and written works while the upper end of this range reflects those of natural speech.

The average rate of OV-type orders observed in Andean Spanish is substantially higher than that of non-contact varieties of Spanish, and the range of values is much larger: approximately 5% to 38%. The summarized literature in §3.3.1 reveals that variable predicate constituent order is constrained by social factors like sex, socioeconomic status, level of education, first language, and linguistic dominance. In general, L1 Quechua bilinguals (Klee et al., 2011), Quechua-dominant bilinguals (Muysken, 1981), individuals of a low socioeconomic status (Klee & Caravedo, 2006; F. A. Ocampo & Klee, 1995), and those with less formal education (Muntendam, 2009) exhibit the highest rates of non-canonical OV order. A cross-study comparison tentatively determines that the observed diametric effect of sex on OV/VO variation is contingent upon the locale in which the speaker resides (i.e., women in more rural regions employ OV order more frequently than their male counterparts in comparison to women in urban settings who exhibit a lower rate of OV order). Additionally, two linguistic factors— subject expression and pragmatic context— condition predicate

constituent order variation such that the presence of a subject inhibits non-canonical order, and pragmatic contexts that deviate from simply conveying information promote non-canonical order.

Additionally, I review the more generative explanations that account for OV/VO variation in §3.3.2. It has been generally asserted that indirect pragmatic transfer, rather than direct syntactic transfer, is responsible for the elevated rate of pre-verbal objects in Andean Spanish. Pragmatic transfer in this context has been construed as both the transfer of the Quechua focus-fronting mechanism and the loosening of pragmatic constraints on non-canonical word order that governs non-contact varieties of Spanish. (Camacho, 1999; Muntendam, 2008, 2009, 2010, 2013). Ultimately, I raise questions about the extent to which the surface-level replication of the Quechua focus fronting mechanism is responsible for frequent pre-verbal objects in Andean Spanish, especially given the finding that pre-verbal objects are accepted at the highest rates by Andean Spanish and Quechua speakers when the subject is the focalized element and the object is topical by way of its given status in the question that elicits this syntactic configuration in the response.

In the second half of this chapter, I turn my attention to Quechua, beginning with a description of its primary typological characteristics, which are most notably its (current) classification as an SOV-dominant language that encodes grammatical relationships primarily via the morphological mechanism of suffixation (Cerrón-Palomino, 1987a; Salas Cruz, 1993). The case-marking suffix most crucial to the present study is the accusative suffix, *-ta*. Though, prescriptively speaking, *ta*-marking is obligatory, *ta*-omission has been reported in several Bolivian and Peruvian Quechua varieties (Kalt & Geary, 2021; Muntendam, 2015; Sánchez, 2003). Overall, a higher incidence of *ta*-omission coincides with more rigid word order, though, in some varieties (Chuquisaca and Cusco Quechua), this means a stricter adherence to the canonical OV order and for others (Ulcumayo and Lamas Quechua) a concomitant shift to VO-dominant order. In either case, it appears that this agglutinative language is concurrently becoming less dependent on morphology and more dependent on syntax to successfully express grammatical relationships, much like the language with which it is in contact (Hintz, 2009, 2016).

As is the case in Spanish, word order in Quechua is determined in large part by information structure. Though the pragmatic notions of topic and focus are encoded via their corresponding suffixes, they also exhibit somewhat divergent syntactic patterns. Particularly, while both focus-marked and topic-marked constituents may appear in fronted positions and *in situ*, only topics may also be displaced to the right edge of a sentence (Muysken, 1995; Sánchez,

2010). However, as is the case with accusative case-marking, morphological marking of information structure appears to be attriting, and once again, syntactic structure is being called upon to communicate the information status of the object, at least in Bolivian Quechua varieties (Muntendam, 2015). In bilingual Cusqueño Quechua, on the other hand, a recent variationist analysis reports no correlation between the status of an object constituent as new or given and the syntactic configuration of predicate constituents. However, the linguistic dominance and first language of a speaker as well as the grammatical category of the object and subject expression do exert a significant effect on variation in this domain (Hubbel, 2023).

Barring the work of Luján et al. (1984) on Spanish second language acquisition among Quechua-speaking children, noticeably absent from past literature on Andean Spanish and Quechua is the consideration of age as a potential predictor of syntactic variation, despite the fact that age is reported alongside other demographic information in the tabulation of participant characteristics in the majority of the cited studies. As will be highlighted in Chapter 7, an age analysis is integral for documenting the trajectory of a linguistic feature and ultimately procuring suitable statistical evidence for confirming contact-induced language change. I address this perceived gap in the statistical analysis of the present study by evaluating the age of participants as a continuous, independent predictor. Furthermore, many of the linguistic predictors of variable word order in other syntactic domains and other languages addressed in Chapter 2 are overlooked in the Andean Spanish and Quechua literature, with the exception of information structure (or the information status of the object). Thus, I include animacy, definiteness, and weight of the object constituent as potential linguistic independent variables in the present study to determine their pertinence in accounting for variation in the syntactic domain of interest (i.e., the predicate) in both Andean Spanish and Quechua.

# CHAPTER 4

## METHODOLOGY

### 4.1 Introduction

The primary purpose of this chapter is to describe the methods applied to collect and code the data of the present study. In the section that follows, I outline the participant recruitment process, including recruitment locations and implemented sampling procedures. Furthermore, I provide select demographic information for each participant, which entails their reported *sex*, *age*, level of *education*, *residence*, and *language profile*. The general demographic characteristics of the participant sample are then compared to those of the Cusco population using census data with the objective of evaluating how well the sample represents the population from which it is drawn.

In §4.3, I outline the data collection procedure and detail the two data elicitation instruments of the study— the Language Background Questionnaire and the Sociolinguistic Interview. I highlight the purpose of the Language Background Questionnaire— to extract demographic and linguistic information about the participants— and describe the process by which the *dominant language* and *L1* of each participant was determined using this instrument. Next, I summarize the data codification process in §4.4 by first presenting the dependent variable of the present study— **predicate constituent order variation**. I then circumscribe the envelope of variation by identifying the contexts from which tokens of predicate constituent order are extracted, offering exclusionary and inclusionary criteria. Lastly, I present the linguistic independent variables— *animacy*, *definiteness*, *information status*, *polarity*, *subject expression*, *weight*, *accusative marking*, *argument type*, and *Spanish loanwords*— by listing their corresponding factor levels and providing examples from the Andean Spanish and Quechua datasets that exemplify the coding procedure applied to each independent variable.

Next, I restate the research questions that guide the present study and offer a null and alternative hypothesis for each question (§4.5). The final section, §4.6, is reserved for a summary of the chapter.

## 4.2 Participants

In the following sections, I provide the demographic characteristics of the participants of the present study and describe the process by which they were recruited. I begin by detailing the recruitment sites.

### 4.2.1 Location

The data of the present study were collected during grant-funded fieldwork<sup>146</sup> in Cusco, Peru in the summer of 2019. By Cusco, Peru<sup>147</sup> I refer not only to the city of Cusco, but also to the much larger department of Cusco, which is a geographic region comprised of thirteen provinces. Participants were recruited in four municipalities each pertaining to a distinct province within the department of Cusco: the **Cusco metropolitan** area in the province of Cusco, **Urcos** in the province of Quispicanchi, **Sicuani** in the province of Canchis, and **Chinchero** in the province of Urubamba. Though participant recruitment and data collection occurred only in these four locales, several participants did not reside in the municipality where they were recruited. These participants resided in Huaró (in the province of Quispicanchi), Anta (Anta), Raqch'i (Canchis), Calca (Calca), and the province of Canas. Figure 4.1 provides a map of the thirteen provinces of Cusco.

### 4.2.2 Sampling procedures

The participants of this study were randomly recruited in the four aforementioned locales by myself and my colleague, Bethany Bateman McDonald, by directly approaching the individual in a public space (e.g., the main plaza, vendor kiosks). Though participants were randomly recruited, particular individuals were approached on the basis of their perceived demographic characteristics with the goal of arriving at a representative, stratified sample. After providing the individual with the objective of the data collection project<sup>148</sup> and its associated tasks, they were politely asked for their participation. Most individuals who agreed to participate completed these tasks (to be described in §4.3.1) at the recruitment site; however, if an individual was interested but not immediately available to participate, the researchers and participant arranged to meet at a later date at an agreed upon location (e.g., el Museo del Café).

<sup>146</sup> The funds required for travel, lodging, and participant compensation were provided by the 2018 Willson Center Graduate Research Award and 2019 Tinker Graduate Field Research Award.

<sup>147</sup> The place name Cusco may refer to the city of Cusco, the province of Cusco, or the department of Cusco. Unless otherwise specified, I use the term Cusco to refer to the department of Cusco.

<sup>148</sup> The overarching objective of this project was to elicit and collect natural speech from speakers of Andean Spanish and Quechua in Cusco, Peru. The data collected for this project were also analyzed in Bateman McDonald's dissertation on past temporal reference in Andean Spanish (Bateman McDonald, 2022). The participants of this study also completed a forced-choice grammatical questionnaire task in which they were instructed to select a verb inflected in either the preterite or present perfect given a particular context. This questionnaire data was not used in the present study, as it is not relevant to its objective or research questions.



Figure 4.1: Map of the Department of Cusco (Instituto Nacional de Estadística e Informática, 2017, p. 15)

Moreover, a snowball sampling method was unintentionally employed with a few participants who kindly took it upon themselves to recruit additional participants due to their personal interest in the study. For example, one participant, a young woman who sold tourist packages in the main square of metropolitan Cusco, invited the researchers into her home to interview herself, her husband, mother-in-law, and two close friends. In another instance, a waiter at a restau-

rant we visited agreed to participate and subsequently arranged the participation of several friends and friends of friends on his own volition.

### 4.2.3 Demographic information

In all, 47 individuals participated in the present study. The inclusion criteria for participation was as follows: participants (i) must be at least 18 years of age, (ii) must be either a monolingual speaker of Andean Spanish or a bilingual speaker of Andean Spanish and Quechua, (iii) must be native to the department of Cusco and currently reside therein, and (iv) must not have resided in a region outside the department of Cusco for an extended period of time (i.e., more than a year). Ultimately, two of the 47 participants were not included in the analysis of the present study because, during the interview task, it was ascertained that they did not meet the aforementioned inclusion criteria. One participant was excluded because they had resided outside of Cusco for a considerable amount of time, and the other was excluded because they were a monolingual speaker of Quechua that was unable to complete the interview task in Andean Spanish.<sup>149</sup> Due to these exclusions, the participant group is comprised of 45 individuals.

<sup>149</sup> Both individuals were still monetarily compensated for their participation in the study.

The demographic information of each participant, including their **sex**, **age**, level of **education**, **residence**, and **language** profile, is presented in Table 4.1. The number corresponding to each participant has been assigned based on the chronological order of their participation (i.e., P1 was the first participant interviewed, and P45 was the the last participant interviewed.) Each of the demographic factors accounted for in Table 4.1 are detailed in the following subsections.

#### Sex

Of the 45 participants included in the present study, 27 (60.0%) identify as **female** and 18 (40.0%) identify as **male**. None of the participants identified as non-binary. Considering the monolingual and bilingual groups independently, the monolingual group is nearly balanced regarding sex (six males and five females); however, in the bilingual group, female participants outnumber male participants 22 to 12. According to the most recent Peruvian census data, in the department of Cusco, 49.5% of the population identify as male and 50.5% identify as female (Instituto Nacional de Estadística e Informática, 2017). Thus, the participant sample<sup>150</sup> of the present study slightly over-represents the female population of Cusco in general, and the female population is over-represented in the bilingual group to a greater degree.

<sup>150</sup> The term 'participant sample' refers to the 45 participants of the present study that were drawn from the target population.

Table 4.1: Demographic information of participants

	<b>Sex</b>	<b>Age</b>	<b>Residence</b>	<b>Education</b>	<b>Language</b>
<b>P1</b>	female	41	metro	secondary	bilingual
<b>P2</b>	male	24	metro	secondary	bilingual
<b>P3</b>	female	46	metro	university	monolingual
<b>P4</b>	male	23	metro	institute	monolingual
<b>P5</b>	male	19	metro	secondary	monolingual
<b>P6</b>	male	24	metro	university	monolingual
<b>P7</b>	male	24	metro	university	monolingual
<b>P8</b>	female	55	metro	superior	bilingual
<b>P9</b>	female	21	metro	superior	bilingual
<b>P10</b>	male	20	metro	university	bilingual
<b>P11</b>	female	19	metro	technical	monolingual
<b>P12</b>	female	22	metro	institute	monolingual
<b>P13</b>	male	20	metro	superior	bilingual
<b>P14</b>	female	19	metro	technical	monolingual
<b>P15</b>	female	53	department	superior	bilingual
<b>P16</b>	female	24	department	superior	bilingual
<b>P17</b>	male	59	department	superior	bilingual
<b>P18</b>	male	22	metro	secondary	bilingual
<b>P19</b>	male	20	department	superior	bilingual
<b>P20</b>	female	60	department	primary	bilingual
<b>P21</b>	male	34	metro	university	bilingual
<b>P22</b>	female	30	metro	technical	bilingual
<b>P23</b>	male	32	metro	secondary	bilingual
<b>P24</b>	female	30	department	superior	monolingual
<b>P25</b>	male	50	department	secondary	bilingual
<b>P26</b>	female	32	department	secondary	bilingual
<b>P27</b>	female	24	department	institute	bilingual
<b>P28</b>	female	34	department	secondary	bilingual
<b>P29</b>	female	49	department	primary	bilingual
<b>P30</b>	male	72	department	superior	monolingual
<b>P31</b>	female	44	department	primary	bilingual
<b>P32</b>	male	47	department	primary	bilingual
<b>P33</b>	female	52	department	primary	bilingual
<b>P34</b>	male	25	department	technical	bilingual
<b>P35</b>	female	32	metro	institute	bilingual
<b>P36</b>	female	22	metro	institute	bilingual
<b>P37</b>	female	29	metro	superior	bilingual

<b>P38</b>	male	19	metro	university	monolingual
<b>P39</b>	female	21	metro	superior	bilingual
<b>P40</b>	female	26	metro	superior	bilingual
<b>P41</b>	female	32	metro	university	bilingual
<b>P42</b>	female	26	metro	university	bilingual
<b>P43</b>	female	18	metro	secondary	bilingual
<b>P44</b>	male	68	metro	secondary	bilingual
<b>P45</b>	female	67	metro	none	bilingual

### Age

The age of the participants ranges from 18 to 72 years ( $\bar{x} = 34.0$  years). The density plot in Figure 4.2 illustrates the distribution of age across the participant sample, and the dashed line represents the mean age. It is clear that the age of participants is not normally distributed; the distribution curve is left-skewed, meaning there are more relatively younger participants than older participants in this sample.

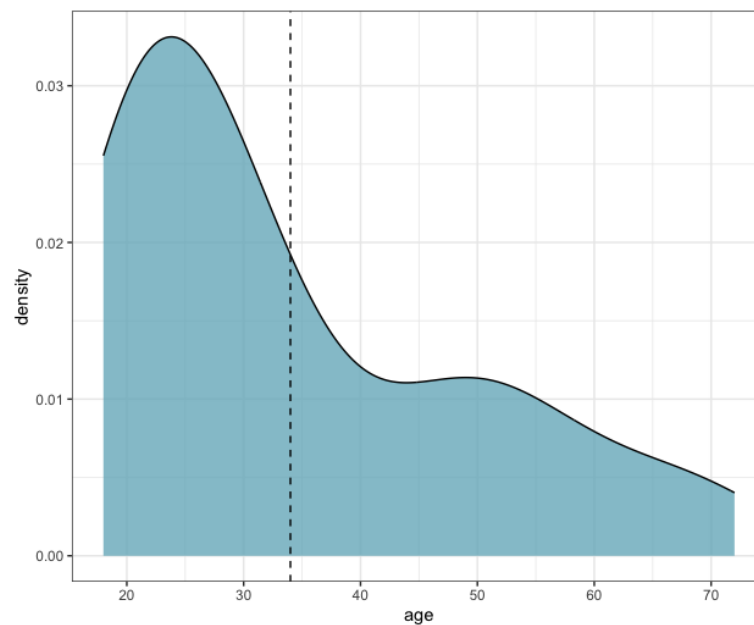


Figure 4.2: Density plot of participant age

Though age is treated as a **continuous** variable in the analysis of the present study, in Table 4.2, age is categorized by decade to compare the distribution of age in the sample to the Cusco population. Table 4.2 shows that half of

the participant sample is 18-29 years of age. It is also the case that half of the Cusco population is 29 years of age or younger (52.8%) (Instituto Nacional de Estadística e Informática, 2017); however, considering only those  $\geq 18$  years of age were permitted to participate in the present study, the participant sample over-represents the young adult population of Cusco.<sup>151</sup> Barring young adults (<30 years of age), overall, the distribution of the sample mostly reflects the age distribution of the Cusco population. In the Cusco population, after the age of 10, there is an inverse relationship between age and the population proportion. For example, 20-29 year-olds account for 16.4% of the Cusco population, 30-39 year-olds account for 14.6%, 40-49 year-olds account for 12.2%, and so on (see Table 4.2). Thus, despite an over-representation of young adults, the remaining participant sample generally reflects the overall age distribution of the Cusco population.

<sup>151</sup> In other words, 18-29 year-olds account for half of the participant sample, but 0-29 year-olds account for half of the Cusco population (Instituto Nacional de Estadística e Informática, 2017).

Age	Sample (n)	Sample (%)	Cusco %
18-19	5	10.9%	N/A
20-29	18	40.0%	16.4%
30-39	8	17.8%	14.6%
40-49	5	11.1%	12.2%
50-59	5	11.1%	9.4%
60-69	3	6.7%	5.9 %
70-79	1	2.2%	3.6%
<b>Total</b>	<b>46</b>	<b>100%</b>	—

Table 4.2: Age distribution by decade in participant sample and Cusco population

Finally, it is important to note that, in the present study, the monolingual participant group is considerably younger than the bilingual participant group ( $\bar{x}_{\text{monolingual}} = 28.8$  years,  $\bar{x}_{\text{bilingual}} = 35.7$  years). The density plot in Figure 4.3 demonstrates that, though both groups are represented by a left-skewed distribution curve, the monolingual group overlay shows a more extreme leftward skew than the bilingual group overlay. The possible effect of the disparate age distribution of monolingual and bilingual participant groups on the results of the study will be discussed in further detail in Chapter 7.

<sup>152</sup> Originally, participants were asked if they reside in an urban area or a rural area. Virtually all participants reported living in an urban area, even those from municipalities like Huaró, which has a population of less than 5,000 people and is located approximately 30 miles (one hour by car) from Cusco metro. It became clear through conversations with participants that the urban-rural construct is not viewed through the same lens in the Andes as it is in the United States. According to most participants, rural areas are those that fall between municipalities, which include mostly farmland. This is supported by the definition of the term ‘rural’ provided by the Instituto Nacional de Estadística e Informática (2017): “rural populations are those that have no more than 100 contiguously grouped dwellings and are not the district capital, or if there are more than 100 dwellings, they are dispersed or scattered without forming blocks” and those that “must not exceed 2,000 inhabitants” (my translation, 23). For this reason, participants were not classified as ‘urban’ or ‘rural’, rather as residing in metro Cusco or the department of Cusco.

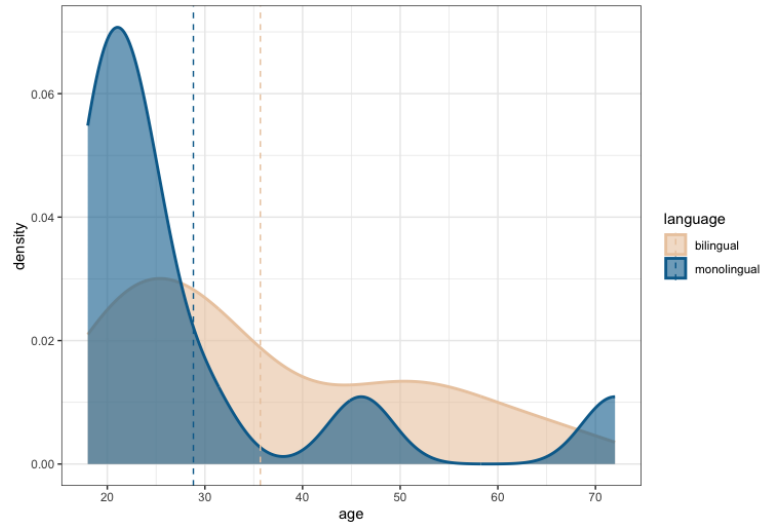


Figure 4.3: Density plot of monolingual and bilingual participant age

### Residence

The place of residence of the participants is partitioned into two categories: the metropolitan area of Cusco and the department of Cusco. Participants categorized as **metro** are those who live within the city limits of Cusco, and those labelled **department** reside outside of the city limits but within the department of Cusco in any of the 13 provinces depicted in Figure 4.1.<sup>152</sup> Participants in the *department* category reported residing in one of the following municipalities: Anta, Calca, Canas, Chinchero, Huaró, Raqch'i, Sicuani, and Urcos. It should be noted that the residence of the participants is determined by their dwelling place, not the recruitment location. Because it is common to commute to the metropolitan area for work, some participants recruited in the city of Cusco were categorized as residing in the department of Cusco, as that is where their home is located.

Of the total 45 participants, 29 (64.4%) reported residing in the **metropolitan** area of Cusco and 16 (35.6%) in the **department** of Cusco. In the monolingual group, 9 of the 11 participants were classified as metro, and 20 of the 34 bilingual participants received the same classification. Thus, monolinguals show an even stronger affiliation toward metro residence than bilinguals in the participant sample. To determine the metro-department proportions of the Cusco population, I first divided the urban population of the province of Cusco ( $n = 432,662$ ) by the total population of the department of Cusco ( $n = 1,205,527$ ) to find the percentage of metro residents (Instituto Nacional de Estadística e Informática, 2017). The proportion of the Cusco population residing in the department of Cusco was determined by first adding the number of

total rural residents within the province of Cusco ( $n = 14,926$ ) to the total number of residents residing in the other 12 provinces ( $n = 757,939$ ) then dividing the resulting sum ( $n = 772,865$ ) by the total Cusco population ( $n = 1,205,527$ ). In the Cusco population, approximately 35.9% of residents live in the city of Cusco, and the remaining 64.1% in the department of Cusco. Thus, metro participants are over-represented in the sample.<sup>153</sup>

## Education

Participants also reported their highest level of education attained at the time of their participation in the study. There were six catalogued responses: none, primary (*primaria*), secondary (*secundaria*), institute (*instituto*), superior (*superior*), technical (*técnico*), and university (*universitaria*). According to Monroy and Mackie (2022), these responses reflect the structure of the Peruvian education system, which is partitioned into two levels: basic education (initial, primary, and secondary) and higher education (superior, technical, and university). Only primary and secondary education are compulsory in Peru; however, only about 75% of secondary-aged children are enrolled in secondary education institutions, with the majority of unenrolled children residing in the Andean and Amazonian regions of Peru.

Prior to beginning primary education, Peruvian children may optionally complete 1-2 years of early childhood education, which concludes at the age of 5. Then, children complete six years of primary education followed by five years of secondary education. Secondary education is comprised of a 2-year lower-secondary period and 3-year upper-secondary period. In lower-secondary education, students follow a compulsory general education curriculum. Upon concluding the lower-secondary education period, students may elect to pursue either the academic or technical (i.e., vocational) tracks in the upper-secondary period, which is also compulsory (Monroy & Mackie, 2022).

Higher education beyond the secondary period is optional and is divided into two types— superior education and university education. Those pursuing a university degree may elect to complete 1-2 years of preparatory education (i.e., academy) prior to enrolling. Recipients of non-university, superior education report their highest level of education attained as *superior*, *technical*, or *institute*, depending on the type of certificate earned or program completed. A visual representation of the Peruvian education system based on Monroy and Mackie's (2022) description is provided in Figure 4.4.

<sup>153</sup> This sampling bias is the result of two circumstances. First, the interviewers resided in the city of Cusco through the duration of the data collection process. Second, individuals from the department of Cusco were generally less willing to be interviewed.

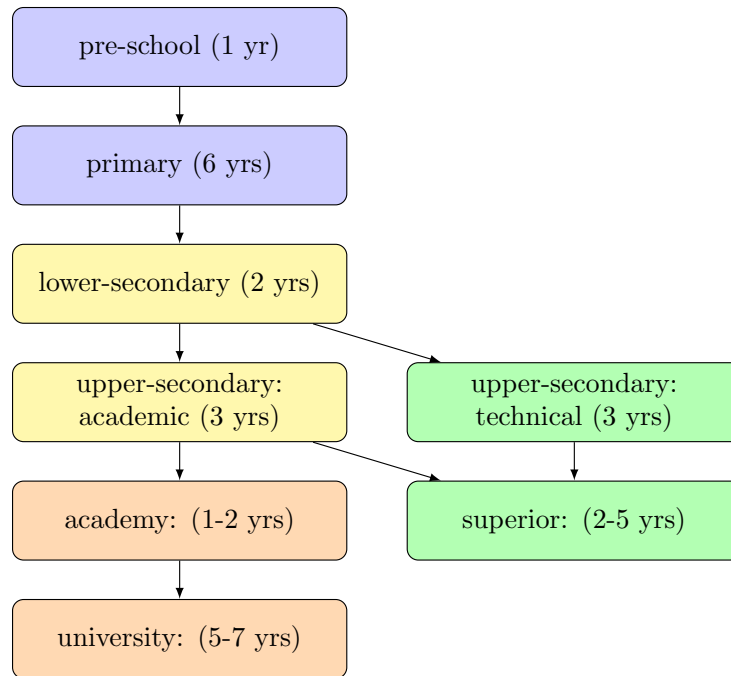


Figure 4.4: Educational system of Peru

<sup>154</sup> The percentages reported from the 2017 Peruvian census data reflect the highest level of education attained by all census takers 15 years of age or older. The report also provides percentages for the urban and rural populations separately; however, because the criteria for distinguishing urban and rural populations for the Peruvian census do not coincide with the criteria for distinguishing metro and department populations in the present study, only the total percentages that reflect the combined urban and rural populations are reported in Table 4.3.

<sup>155</sup> The Cusco data does not total to 100.0% because 0.3% of the population was categorized as completing *initial* education, which is synonymous with the optional early childhood or pre-school education mentioned above.

The various types of non-university, superior education reported by the participants— institute, superior, and technical— are combined into a single category, **post-secondary**, in the statistical analysis. Because the term ‘technical’ may refer to either secondary-level or superior-level education, other factors, such as the occupation and age of the participant, as well as details provided in their interview, were consulted to determine the highest level of education attained by the participant. Table 4.3 illustrates the distribution of the level of education across language profile (i.e., monolingual versus bilingual) and also compares the educational level of the participant sample to the overall Cusco population (Instituto Nacional de Estadística e Informática, 2017).<sup>154,155</sup>

Education	Language Profile		Total	Cusco (%)
	<i>monolingual</i>	<i>bilingual</i>		
<i>none</i>	0 (0.0%)	1 (2.9%)	1 (2.2%)	9.4%
<i>primary</i>	0 (0.0%)	5 (14.7%)	5 (11.1%)	20.9%
<i>secondary</i>	1 (9.1%)	9 (26.5%)	10 (22.2%)	38.2%
<i>post-secondary</i>	6 (54.5%)	15 (44.1%)	21 (46.7%)	13.1%
<i>university</i>	4 (36.4%)	4 (11.8%)	8 (17.8%)	18.1%
<b>Total</b>	<b>11 (100.0%)</b>	<b>34 (100.0%)</b>	<b>45 (100.0%)</b>	<b>99.7%</b>

Table 4.3: Level of education among participant sample and Cusco population

The above table illustrates that, overall, the participant sample has achieved a higher level of education on average than the Cusco population overall. For instance, 64.5% of the sample has attained post-secondary-level education (including university education), while only 31.2% of the Cusco population has reached the same academic level. This disparity is even more stark for monolingual participants, the majority of whom have completed some form of post-secondary education (90.9%).<sup>156</sup> The participant sample under-represents the proportion of the Cusco population that has completed primary and secondary education and over-represents the population that has attained higher education certificates and degrees, especially for monolinguals. It is important to note that, in the statistical analyses presented in the chapters that follow, *none* and *primary* were collapsed into a single category, as there is only one participant who reports receiving no formal education.

<sup>156</sup> This generalization does not apply to only one participant who had not yet completed post-secondary education due to his age, but was enrolled in a higher education institute at the time of the interview.

### Language profile

Both **bilingual** (i.e., native speakers of both Quechua and Andean Spanish) and **monolingual** (i.e., native speakers of Andean Spanish only) individuals were included in the present study to account for the possible effect of language profile on predicate constituent order variation. Of the 45 participants, 34 are bilingual (75.6%), and the remaining 11 are monolingual (24.4%).<sup>157</sup> The participant group is intentionally unbalanced, as the preeminent objective of the present study is to compare predicate constituent order variation in Andean Spanish and Quechua, and only bilingual participants were able to provide speech samples in Quechua.

<sup>157</sup> The language profile of the participant sample cannot accurately be compared to the overall population of the department of Cusco, because census-takers were asked to report their ‘mother tongue they learned to speak in childhood,’ and were not permitted to provide more than one language (Instituto Nacional de Estadística e Informática, 2017). Thus, the 2017 census results do not report the current proportion of monolinguals and bilinguals.

It is important to note that, though the monolingual participants do not consider themselves speakers of Quechua according to their responses on the Language Background Questionnaire (described below), these participants are indeed native speakers of the contact variety, Andean Spanish (Escobar, 2011). This is because all participants reside in the department of Cusco, where Quechua and Spanish are spoken in tandem in public spaces (e.g., marketplaces, town squares), even in the metropolitan area. Furthermore, the majority of monolingual participants (n=8, 72.2%) indicate having at least one parent who speaks Quechua, and, due to the pervasive presence of Quechua in this region, all monolingual participants likely have Quechua-speaking friends, relatives, colleagues, or acquaintances. Thus, the variety of Spanish that monolingual participants speak is not unaffected by a present and historical bilingualism in this region.

### 4.3 Data collection procedure

After recruited individuals signed a consent form confirming their participation in the present study, they completed two primary tasks— a Language Background Questionnaire and a Sociolinguistic Interview. As demonstrated in Figure 4.5, monolingual participants were interviewed in Andean Spanish only, and bilingual participants were interviewed in both Andean Spanish and Quechua. The order of interview languages is not specified in Figure 4.5 below, as the order alternated from one participant to the next to mitigate priming effects. Upon completion of the two tasks, participants were compensated for their participation in the study and provided a copy of the consent form. The two data-elicitation tasks are described in detail below, beginning with the Language Background Questionnaire.

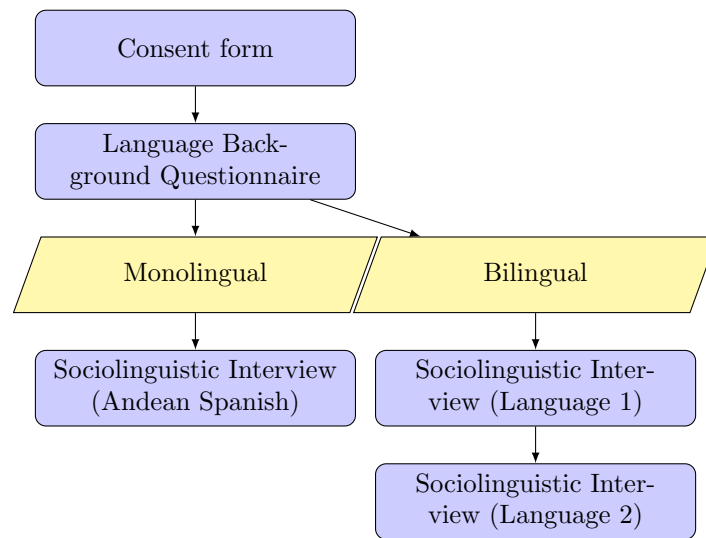


Figure 4.5: Data collection procedure flowchart

#### 4.3.1 Instruments

##### Language Background Questionnaire

The Language Background Questionnaire (henceforth, LBQ) was distributed to participants for the purpose of eliciting both demographic details and a comprehensive description of their linguistic profile.<sup>158</sup> The LBQ was adapted from the Bilingual Language Profile, an open-access survey tool that assesses the linguistic dominance of an individual using self-reports (Birdsong et al., 2012). The LBQ, like the Bilingual Language Profile, is divided into five thematic sections— (i) biographical information, (ii) language history, (iii) language use, (iv) lan-

<sup>158</sup> See Appendix C for the entire Language Background Questionnaire in Spanish.

guage proficiency, and (v) language attitudes. Using participant responses, the Bilingual Language Profile outputs a score on a continuous scale from -218 to +218. A positive score reflects dominance in one language (e.g., Quechua) and a negative score reflects dominance in the other language (e.g., Spanish). A score at either extreme (i.e., a score of -218 or +218) indicates monolingualism in the respective language, and a score of zero indicates that the individual is equally dominant in both languages. Each of the latter four sections of the Bilingual Language Profile— language history, language use, language proficiency, and language attitudes— are equally weighted.<sup>159</sup> Thus, for a monolingual individual, each section can account for a maximum of  $\pm 54.5$  points.

As previously mentioned, the LBQ is an adapted version of the Bilingual Language Profile. (See Appendix D for a list of the modifications made to the Bilingual Language Profile in designing the Language Background Questionnaire.) As a result of these modifications, the possible range of language dominance scores on the LBQ is -146 to +146, with the most extreme values reflecting monolingualism in either language. In the present study, a positive score indicates Quechua dominance, a negative score indicates Spanish dominance, and a score near zero indicates balanced bilingualism. A BLP score<sup>160</sup> is provided only for those participants who identify as bilinguals, as the output for monolinguals could not be properly computed due to non-applicable values.<sup>161</sup> Because the participants of this study reside in a region with widespread bilingualism, the BLP scores of the bilingual group tended to cluster near zero, i.e., the middle of the dominance spectrum. The actual dominance scores of the bilingual participants range from -68.39 to +63.12,  $\bar{x} = +0.82$ .<sup>162</sup> As visualized in Figure 4.6, the distribution of BLP scores approximates normal; that is, most scores cluster near the mean, and scores toward the minimum and maximum values are less frequent.

<sup>159</sup> The biographical information of the individual is not considered in the computation of the Bilingual Language Profile score.

<sup>160</sup> Though the LBQ was used to obtain demographic and linguistic information on participants, the resulting dominance score will be referred to using the term *BLP score*, as the BLP is the instrument that determined the score.

<sup>161</sup> Monolingual participants still completed the LBQ, primarily for the purpose of extracting demographic information.

<sup>162</sup> The range of possible responses is relatively small for this sample, in part due to the adaptations made to the questionnaire.

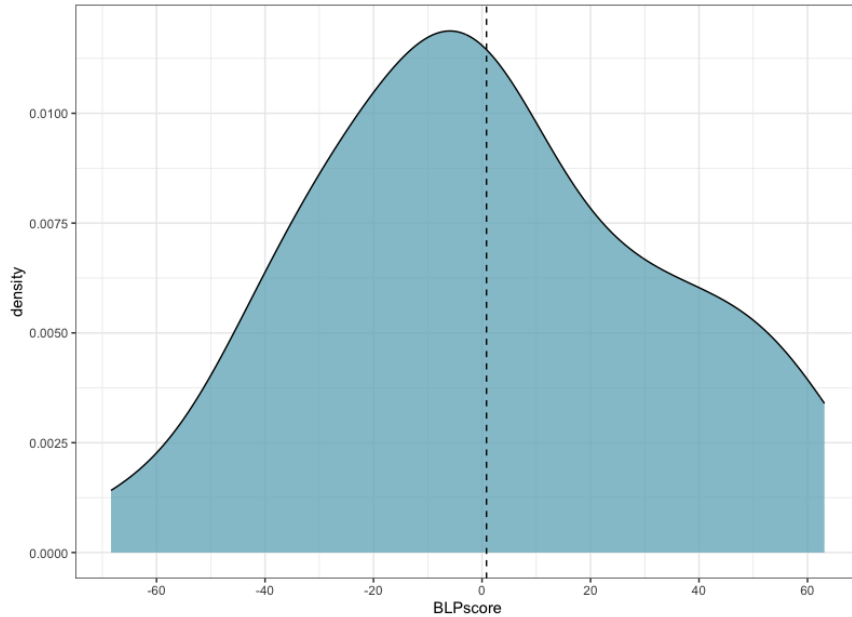


Figure 4.6: Distribution of BLP scores of bilingual participants

In addition to the demographic variables listed in Table 4.1, the **BLP score** of the participant and another linguistic characteristic extracted from LBQ responses— the **first language** (L1) of the participant— were also included as extralinguistic independent variables in the study. The L1 of a participant was determined by comparing the age of acquisition (AOA)<sup>163</sup> of Spanish and Quechua. For *simultaneous* (2L1) bilinguals  $AOA_{\text{Quechua}} = AOA_{\text{Spanish}}$ , for sequential, *L1 Quechua* bilinguals,  $AOA_{\text{Quechua}} < AOA_{\text{Spanish}}$ , and for sequential, *L1 Spanish* bilinguals  $AOA_{\text{Quechua}} > AOA_{\text{Spanish}}$ . The BLP scores, L1s, and AOAs of the bilingual participants are listed in Table 4.4.

The boxplots in Figure 4.7 illustrate the BLP score ranges for each L1 group. Each dot represents the BLP score of a single participant. This figure shows that L1 Spanish bilinguals have the smallest range of BLP scores, likely due to the low number of bilingual participants that report acquiring Spanish prior to Quechua. The BLP scores from this group are completely within the range of scores of the simultaneous bilinguals, which has a minimum value of -68.39 and a maximum value of +20.71. Additionally, the lower half of L1 Quechua scores overlap with the upper half of simultaneous bilingual scores. This indicates that not all L1 Quechua speakers are associated with higher BLP scores than simultaneous bilinguals, likely due to individual patterns of current language use. That is, some L1 Quechua speakers report using Quechua less frequently

<sup>163</sup> It appears that some participants interpreted this question to mean when they began speaking the language well or formally, which is why there are numerous participants who report acquiring their primary language after the age of 0.

and in fewer domains than simultaneous bilinguals. BLP scores of L1 Quechua bilinguals range from -20.62 to +63.12.

	<b>BLP score</b>	<b>L1</b>	<b>AOA(S)</b>	<b>AOA(Q)</b>
<b>P1</b>	4.36	simultaneous	0	0
<b>P2</b>	20.43	Quechua	4	2
<b>P8</b>	-68.39	simultaneous	0	0
<b>P9</b>	4.45	Quechua	8	0
<b>P10</b>	-31.24	simultaneous	4	4
<b>P13</b>	-13.62	Spanish	0	5
<b>P15</b>	-1.82	simultaneous	6.5	6.5
<b>P16</b>	-33.06	simultaneous	2	2
<b>P17</b>	0.45	simultaneous	1	1
<b>P18</b>	35.78	Quechua	10	6
<b>P19</b>	36.51	Quechua	13	1
<b>P20</b>	28.61	Quechua	10	0
<b>P21</b>	-1.19	Quechua	12	1
<b>P22</b>	-6.83	Quechua	11	1
<b>P23</b>	-20.62	Quechua	8	3
<b>P25</b>	-45.86	simultaneous	5	5
<b>P26</b>	20.71	simultaneous	1	1
<b>P27</b>	-2.72	Quechua	5	3
<b>P28</b>	52.68	Quechua	15	3
<b>P29</b>	41.41	Quechua	13	0
<b>P31</b>	55.94	Quechua	12	2
<b>P32</b>	51.49	Quechua	12	1
<b>P33</b>	63.12	Quechua	14	1
<b>P34</b>	1.36	Quechua	4	2
<b>P35</b>	-43.50	Spanish	2	5
<b>P36</b>	9.89	Quechua	12	3
<b>P37</b>	-15.08	simultaneous	1	1
<b>P39</b>	-25.52	simultaneous	1	1
<b>P40</b>	-40.05	simultaneous	8	8
<b>P41</b>	-18.43	simultaneous	1	1
<b>P42</b>	-29.61	Spanish	1	8
<b>P43</b>	-16.44	Spanish	3	5
<b>P44</b>	-5.29	Quechua	11	1
<b>P45</b>	19.98	Quechua	27	6

Table 4.4: Bilingual participants' linguistic characteristics

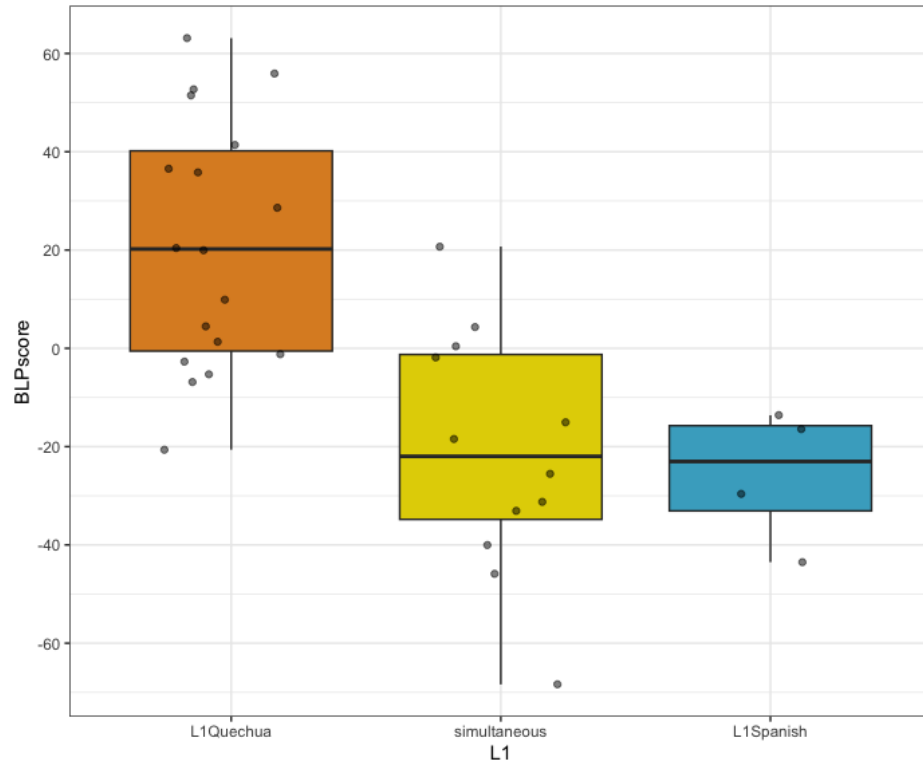


Figure 4.7: Boxplot of BLP score ranges across first language

After filling out the LBQ, the participants completed a semi-spontaneous speech task guided by a Sociolinguistic Interview.

### Sociolinguistic Interview

The primary data collection tool of the present study is the Sociolinguistic Interview, “[a] longstanding tool of sociolinguistic research” (Hoffman, 2014, p. 26). The goal of the Sociolinguistic Interview is to elicit semi-spontaneous speech that resembles the vernacular register of the participant, though, as Labov (1972, p. 209) suggests via the Observer’s Paradox, the presence of an observer (e.g., the interviewer) may somewhat inhibit completely naturalistic speech. A Sociolinguistic Interview is typically organized into tiers. In the Sociolinguistic Interview schedule of the present study,<sup>164</sup> the first tier included general, personal questions to encourage free conversation (e.g., *What were you like as a child?*; *Do you have a significant other? What are they like?*). The following two tiers included more specific questions about past circumstances (e.g., *Do you remember a time you were ill? In as much detail as possible, describe that experience to me.*) and cultural experiences (e.g., *Do you know of a traditional legend*

<sup>164</sup> See Appendix B for the entire interview schedule in English, Spanish, and Quechua.

or story that is important in your community? Could you tell me that story?). By asking more specific, personal questions participants think more about the experiences they are sharing with the interviewer and less about the mechanics of their speech, which ostensibly promotes spontaneous, unmonitored speech.

To imitate the dynamic nature of spontaneous discourse, the interviewers<sup>165</sup> used the Sociolinguistic Interview schedule only as a conversational guide. So, if a participant seemed eager to discuss a particular question, the interviewer would offer follow-up questions to extend the conversation on this topic. Similarly, not all questions were asked in each interview primarily to avoid asking redundant questions that may imply lack of attention on the interviewer's behalf (e.g., if a participant had already mentioned being single, they were not asked if they had a significant other).<sup>166</sup> Though the interview questions were not designed to probe highly sensitive or controversial topics, the participants were advised at the beginning of, and often throughout, the interview that they were not required to respond to any question that they did not wish to discuss. The interviewers were careful to move on from topics that the participant clearly found distressing.<sup>167</sup>

Sociolinguistic Interviews were audio-recorded using a Sony ICD-AX412 Stereo Digital Voice Recorder. Interviews were then transcribed using ELAN 6.3, "an annotation tool for audio and video recordings" ("ELAN", 2023). In total, 2,061 minutes (34 hours and 21 minutes) of audio were recorded. The interviews ranged from 18 minutes to 74 minutes in length, with an average length of approximately 46 minutes.

## 4.4 Coding the data

In this section, I begin by detailing the process of preparing the data to be coded for statistical analysis. I then define the envelope of variation, circumscribing the contexts in which the dependent variable, predicate constituent order, was considered. Next, I describe the linguistic variables of the present study, providing examples for each factor level. I conclude the section by listing all variables—the dependent variable and the extralinguistic and linguistic independent variables—and their corresponding levels.

### 4.4.1 Preparing the data

The data of the present study were prepared in accordance with the instrument used to elicit the data. All participant responses to the Language Background Questionnaire were stored in a folder accessible only to myself and my colleague.

<sup>165</sup> Myself and my colleague, Bethany Bateman McDonald, conducted all interviews.

<sup>166</sup> In a few instances, some questions were left unasked because the interview was cut short due a participant's previous engagement.

<sup>167</sup> For example, when one bilingual, female participant was asked to describe her first pregnancy, she became distraught by the memory of this experience. The interviewer comforted the participant and avoided this question in the subsequent interview language.

To calculate the BLP score of each participant, responses were entered into a Google Form that administers the Bilingual Language Profile online. The resulting score was imported into an Excel document with the accompanying extralinguistic information recorded for each participant— sex, age, education, residence, language profile, and LI— which were also codified on the basis of LBQ participant responses.

Regarding the Sociolinguistic Interviews, all participant interviews were first converted from MP3 to .wav files using Audacity (Audacity Team, 2022) then uploaded into ELAN, version 6.3 for transcription<sup>168</sup> (“ELAN”, 2023). The transcribed audio files were exported as a .txt files, separated by language, and compiled to construct two corpora— an **Andean Spanish corpus** and a **bilingual Cusco Quechua corpus**. Each corpus was then hand-searched for all instances of OV/VO word order that complied with the circumscribed envelope of variation detailed in §4.4.2. Each Andean Spanish instance (i.e., token) was entered into an Excel spreadsheet on a distinct row, coded for the linguistic variables presented in §4.4.3, and accompanied by the corresponding extralinguistic information of the participant who provided the token. The same process was repeated using the bilingual Cusco Quechua corpus to assemble a Quechua dataset.

#### 4.4.2 Defining the envelope of variation

The dependent variable of the present study is the **order of predicate constituents**, which refers to the syntactic configuration of the accusative (i.e., direct) object constituent and the verb constituent of a sentential unit. The dependent variable has two possible factor levels— *verb-object (VO)* and *object-verb (OV)* order. Recall that VO is the canonical order of predicate constituents in Spanish and OV the canonical order in Quechua. Below are examples of VO (a) and OV (b) orders in both Spanish (74) and Quechua (75) from the dataset. The accusative **object** is bolded and the verb is underlined.<sup>169</sup>

##### (74) Andean Spanish

- a. *sí, me regaló un par de poleras, esta polera, y, pues, yo en su cumpleaños, también le regalé **poleras** porque nos encantan las poleras*

‘yes, [he] gifted me a couple of polo shirts, this polo shirt, and, well, for his birthday, I also gifted him **polo shirts** because we love polo shirts’ [P14]

<sup>168</sup> Each interview was transcribed by the individual conducting that interview.

<sup>169</sup> Note that typical sentence mechanics, like periods and capital letters, are avoided in the forthcoming examples, as discourse does not typically parallel conventional sentence structure in writing. Ellipses are used to indicate pauses in the discourse, and commas are used conventionally to enhance the readability of the example. The participant who supplied the token is indicated in brackets on the translation line of the gloss.

- b. *todo estuvo organizado como mi mamá quería... me gustó porque **el vestido que utilicé** mi mamá lo bizo*  
 ‘everything was organized like my mom wanted... I liked [that] because my mom made **the dress that I used**’ [P3]

(75) **Quechua**

- a. *huq leyenda-ta-n cuenta-yki-man nuqa...*  
 a legend-ACC-EVIDI tell-2SG.OBJ.ISG-COND I  
*ka-ra-n-mi buk pastora... munaku-sqa buk*  
 be-PST-3SG-EVIDI a shepherdess love-PST2 a  
***ukuku-ta** binas chay ukuku unqu-chi-sqa, unquq*  
 bear-ACC thus that bear sick-CAUS-PST2, sick.person  
*tuku-sqa chay pastora*  
 finish-PST2 that shepherdess

‘I will tell you [about] a legend... there was a shepherdess... she loved **a bear**, so that bear impregnated her, that shepherdess became pregnant’ [P26]

- b. *nuqayku rima-yku anchayna... voltea-nki*  
 WE.EXCL speak-IPL.EXCL then turn.over-2SG  
*tierra-ta... anchay-pi **chakra-ta** sembra-nki, **papa-ta***  
 land-ACC that-LOC field-ACC sow-2SG potato-ACC  
*tarpu-nki*  
 plant-2SG

‘we talk then... you turn over the land... there you sow **the field**, you plant **the potatoes**’ [P33]

In accordance with the methodologies of previous variationist analyses on predicate constituent order in Andean Spanish (Klee, 1996; Klee et al., 2011; Muysken, 1984; F. A. Ocampo & Klee, 1995), the envelope of variation in the present study is delimited to predicates containing both an accusative object and verb constituent in main, declarative clauses. Only lexical NPs and pronouns<sup>170</sup> may constitute an accusative object, meaning clausal objects were excluded from the analysis (e.g., *hacia lo que podía*, ‘[I] did **what I could**’).<sup>171</sup> I depart from these methodologies in that I include predicates of both affirmative and negative

<sup>170</sup> Pronominal clitics were not included in the analysis.

<sup>171</sup> There is one exception to this. Infinitive verb objects were included in the Quechua dataset and will be discussed subsequently.

clauses, accounting for the possible effect of polarity on constituent order by treating polarity as an independent variable.

## Exclusions

Regarding the exclusion criteria, predicate constituents in subordinate (76), imperative (77), and interrogative clauses (78) were removed from the dataset. Predicates following conjunctions, like *y*, ‘and’, *pero*, *sino*, ‘but’, and *porque*, ‘because’, were included in the analysis provided that the conjunction joined together two independent clauses.<sup>172</sup> However, dependent clauses following subordinating conjunctions, like *aunque*, ‘(al)though’ and *cuando*, ‘when’, were not included in the statistical analysis. Also excluded from the analysis were predicates in subordinate clauses with an inflected verb in the subjunctive (76a) and relative clauses modifying a noun phrase (76b). In Quechua, predicates containing verbs marked with the subordinating morphemes *-spa* (76c) and *-qti* (76d) were excluded from the analysis.<sup>173</sup> Conversely, predicates following phrases like *dice que*, ‘it is said that’, *me acuerdo que*, ‘I remember that’, or *creo que*, ‘I think that’, were included in the analysis, as these phrases are often discursive in nature, and they precede independent clauses.<sup>174</sup>

<sup>172</sup> With the exception of *ichaqa*, ‘but’, these types of conjunctions do not exist as separate lexical items in Quechua, rather they are encoded through class-free morphemes. However, Spanish conjunctions like *pero*, ‘but’, are often borrowed into Quechua, especially among bilingual speakers.

<sup>173</sup> Though *-spa* and *-qti* are both subordinating morphemes, the former is used in cases where the subject of the main and subordinate clause are the same entity, and the latter is used in cases of a switch referent.

<sup>174</sup> One exception to this is when *dice que* triggers a subjunctive verb in the clause following the complementizer. Tokens of this nature were excluded, as the predicate was determined to be contained within a subordinate clause.

### (76) Subordinate Clause

- a. *yo trato de ponerle otras cosas más para que le dé **más sabor...**  
a la comida*

‘I try to put in other things more so that I give **more flavor...** to the food’ [P43]

- b. *entonces fue a un señor que cura eso < **esas fracturas, esos torcidos***

‘I went to a gentleman who cures that < **those fractures, those sprains**’ [P9]

- c. *tuta-ta-ña*            *hampu-ra-yku*            *siete ocho de*  
 night-ADV-PFV    return-PST-IPL.EXCL    seven eight of  
*la noche-ta-ña...*    *eh...*    *dur<*    *durmi-napaq, uh,*  
 the night-ADV-PFV    eh    sle<    sleep-PURP    uh  
*animal-kuna-ta...*    *wisq'a-spa*  
 animal-PL-ACC        close-SUB  
*puñu-na-yku-paq*  
 sleep-PURP-POSS.IPL.EXCL-PURP

‘at night we returned at seven, eight at night... eh.. sle< in order to sleep, uh, closing [up] the animals so that we [could] sleep’ [P1]

- d. *chay-ta*            *ni-wa-qti-n*            *nuqa*  
 that-ACC    say-OBJ.ISG-SUB.SR-3SG    I  
*senti-ri-ri-ku-ni*            *sumaq-ta, sunqu-y*  
 feel-INCP-INCP-REFL-ISG    good-ADV    heart-POSS.ISG  
*t'ika-ri-ru-n*  
 flower-INCP-EXH-3SG

‘when he told me **that**, I felt very good, my heart began to open (*lit.* flower)’ [P9]

The Andean Spanish and Quechua examples in (77) illustrate predicates with a verb inflected in the imperative form. Note that in Quechua, direct commands are formed using an uninflected verb (77b). Indirect commands (i.e., commands directed toward a third person) were also excluded.

### (77) Imperatives

- a. *y yo decía, le decía a mi esposo, al octavo mes le digo, “hazme cesárea,” le digo, “hazme cesárea, porque ya no aguanto”*  
 ‘and I told, I told my husband, in the eighth month I told him, “give me a C-section,” I told him, “give me a C-section because I can’t take it anymore” [P3]

- b. *abuelita-y ni-wa-n “phawa-y waka-ta...*  
 grandma-POSS.ISG say-OBJ.ISG-3SG run-INF cow-ACC  
***unu-ta toma-chi-y”***  
 water-ACC drink-CAUS-IMP  
 ‘my grandma said to me, “run [and] make the cows drink water”’  
 [P37]

Regarding interrogative clauses, predicate constituents in both open (78b) and closed questions were excluded, as were tokens in which the object constituent contained an interrogative pronoun (78a).

(78) **Interrogatives**

- a. *estaba mal, y creo que ese año fue el año que nos llevó a pasear... y nos habi<... ese d<... y no me recuerdo **qué fecha fue**, pero fuimos al cerro*

‘I was sick, and I think that that was the year that [he] took us to go for a walk... and he had us<... that d<... and I don’t remember what day it was, but we went to the mountain’ [P16]

- b. *allin-mi ka-sha-ni... hm... mm... **ima-ta-taq***  
 good-EVIDI be-PROG-ISG DM DM what-ACC-Q  
*willa-yki-man nuqa-manta?*  
 tell-OBJ.2SG-COND.ISG I-ABL

‘I’m doing well... hm... mm... **what should I tell** you about myself?’  
 [P15]

Additionally, only predicates with inflected verbs were included in the present study. Thus, predicates with non-finite verbs, like infinitives (79a) and gerunds (79b), were excluded.

(79) **Non-finite Verbs**

- a. *allí ya encontré mi vocación que era ayudar a los demás... o tratar ¿no? de... mínimo hacer algo ¿no?*

‘there I found my calling which was to help others... or try, right? to... at least to do something, right?’ [P41]

- b. *después eh... yo... he ido a la< al mercado a comprar para cocinar... ah qué mi esposo quiera pues, y después, regresamos del mercado y otra vuelta cocinando **almuerzo**... almorzamos... y a lavar ropa*

‘afterwards uh... I.. went to the market to go shopping in order to cook... ah whatever my husband wants, and afterwards, we return from the market and again cooking **lunch**... we eat... and [go] to wash clothes’ [P45]

The following instances were also excluded from the dataset— verbs like *gustar* (80), verbal complements requiring a prepositional phrase (81),<sup>175</sup> fixed expressions (82), and inaudible predicates (83).

<sup>175</sup> The exception to this is animate accusative objects with obligatory differential object marking, e.g., *voy a recoger a mi hijita*, ‘I’m going to pick up my daughter’.

(80) **Verbs like *gustar***

- a. *en un futuro me gustaría tener **mi empresa, restaurante, hotel, algo relacionado al turismo** porque aquí en Cusco es muy bueno*

‘in the future I would like to have **my business, restaurant, hotel, something related to tourism** because [it] is really good here in Cusco’ [P6]

- b. *buk kuti nuqa-qa musqhu-yku-sqa-ni buq uh*  
 one time I-TOP dream-INT-PST2-1SG a DM  
*wasi-ta-cha ranti-ri-ku-sqa-ni*  
 house-ACC-DIM buy-INC-REFL-PST2-1SG  
*munay-cha-ta chay animal-cha-kuna-wan...*  
 beautiful-DIM-ACC that animal-DIM-PL-INS  
***jardin-cha-ta** planta-cha-kuna-wan me*  
 garden-DIM-ACC plant-DIM-PL-INS CL.ACC.1SG  
*gusta-wa-n nuqa-man*  
 like-OBJ.1SG-3SG I-DAT

‘one time I dreamt I bought a, uh, little house, a beautiful one with those little animals... I want **a little garden** with little plants’ [P1]

- (81) *no salen de sus casas, o sea, no salen a divertirse... más se dedican al estudio*

‘they don’t go out from their house, I mean they don’t go out to have fun... they just dedicate themselves to **[their] studies**’ [P10]

- (82) *entonces mi amiga se dio cuenta, vino, y, ya pues, me sacó así superman junto a mi otra amiga*

‘then my other friend realized **[what was going on]**, she came, and then, well, she got me out like superman along with my other friend’ [P11]

(83) **Inaudible Predicates**

- a. *y las doctoras estaban ahí, “que no te va a doler no te va a doler,” y cuando fui, tuve que <inaudible> la vacuna a la fuerza, llorando*

‘and the doctors were there, “it’s not going to hurt you, it’s not going to hurt you,” and when I went, I had to <inaudible> the vaccine by force, crying’ [P12]

- b. *eh ima-ta muna-ra-ni unu-ta-chu muna-ra-ni*  
 eh what-ACC want-PST-1SG water-ACC-NEG want-PST-1SG  
*cerveza-ta-s muna-ra-ni ofrece-wa-n...*  
 beer-ACC-PL want-PST-1SG offer-OBJ.1SG-3SG  
 <inaudible> -cha-n-ta  
 <inaudible> -DIM-POSS.3SG-ACC

‘eh what did I want, I didn’t want water, I wanted beers, [he] offered me a little <inaudible>’ [P39]

Concerning the Quechua data only, constituents containing the placeholder *na* were also excluded. This particle may take the place of a constituent that the speaker is unable to think of during the speech act, comparable to *whatchamacallit* in English (Hayashi & Yoon, 2010). Syntactically, *na* may take the place of any part of speech, and it is often inflected for the part of speech that it is holding the place of. For instance, if a speaker is unable to think of the direct object of the sentence, they may say *nata*— the placeholder *na* with the accusative suffix *-ta*.

(84) Placeholder

*mana-n arroz fideo chay-kuna-ta na-yku-chu*  
NEG-EVIDI rice noodles that-PL-ACC PH-IPL.EXCL-NEG  
*wawa-yku-paq-qa... haba-s-ta-pas*  
child-IPL.EXCL-DAT-TOP... fava.bean-PL-ACC-COOR  
*bank'a-ru-yku, chayqa... na-ta... kuta-ra-yku,*  
toast-EXH-IPL.EXCL then PH-ACC pulverize-PST-IPL.EXCL  
*chay hak'u-ta ruwa-ra-pu-yku...*  
that flour-ACC make-PST-REG-IPL.EXCL

‘We don’t NA rice, noodles, those things for our kids... we toasted broad beans, then we pulverized NA, we made that flour...’ [P28]

In (84), the object of underlined verb, *kutarayku*, ‘we pulverized’, cannot be retrieved by the participant at utterance time, thus, *nata* is used instead. Tokens of this kind and tokens in which the inflected verb of the predicate constituent was replaced by *na* were not included in the Quechua dataset. Verbal placeholder *na* is exemplified earlier in the same example (84) in the predicate constituent *chaykunata naykuchu wawaykupaqqa*, ‘we don’t NA those for our children’. The placeholder *na* was excluded for two reasons. First, in some cases, when *na* is affixed with *-ta*, there is ambiguity surrounding the part of speech of the word it replaces, as *nata* may hold the place of either an accusative object or an adverb. Second, the placeholder is functionally a hesitation marker, and other hesitation markers, like *este* were not included in the analysis.

### Inclusions

Now I clarify certain instances in which a token of predicate constituent order was included in the dataset. First, there were a handful of tokens in which a single verb was both preceded and followed by an accusative object or a single object was both preceded and followed by a verb. These tokens could be construed as presenting either VO or OV order. In most cases, the interview audio was used to disambiguate the token by assessing the inflection of the utterance for inflectional breaks that would indicate which NP was intended to be the accusative object of the verb. However, in the few cases lacking inflectional breaks to disambiguate the token, the token was counted twice, once with OV order and once with VO order. For instance, in (85a), two tokens were extracted from the predicate with the verb, *sufrir*, ‘to suffer’— one in which *muchas cosas*, ‘a lot of things’, was considered the object constituent and one in which *un bul-*

*lying demasiado*, ‘a terrible bullying’, was considered the object. The former token exhibits OV order and the latter VO order. The Quechua example in (85b) illustrates the same phenomenon— the verb *ruway*, ‘to do’ is considered both part of an OV predicate, with the object *cesáreata*, ‘C-section’, and a VO predicate with the object *operacionta*, ‘operation’.

- (85) a. *por el cosa de mi boca con los ojos de mi rodilla, de mi ojo...*  
*muchas cosas sufri un bullying demasiado*  
 ‘because of the thing with my mouth with the eyes of my knee, of my eye... I suffered many things, a terrible bullying’ [P39]
- b. *uh huq-taq-qa allin-ta-n wacha-ku-ra-ni...*  
 uh one-CONTR-TOP good-ADV-EVIDI give.birth-REFL-PST-1SG  
*segundo-taq-qa... arí segundo-taq-qa cesárea-ta*  
 second-CONTR-TOP yes second-CONTR-TOP C.section-ACC  
*ruwa-wa-ra-nku operacion-ta*  
 do-OBJ.1SG-PST-3PL operation-ACC  
 ‘uh the first time, I birthed well... the second time... yes the second time they did a C-section on me, an operation’ [P1]

In Quechua, the accusative object is typically morphologically marked with the accusative case suffix, *-ta*, as illustrated in example (85b). However, it has been reported that in some Quechua varieties, accusative objects are not always accompanied by the corresponding case suffix (Kalt & Geary, 2021; Muntendam, 2015; Sánchez, 2003). In the present study, both morphologically marked (i.e., affixed with *-ta*) and unmarked objects are included in the dataset, but accusative case marking will be treated as an independent variable in the statistical analysis of the Quechua data.

As mentioned in Chapter 3, certain verbs (e.g., *munay*, ‘to want’, or *qallariy*, ‘to begin’) may take an infinitive verb object in Quechua. In these constructions, the infinitive verb functions as the accusative object of the inflected verb when marked with *-ta*, as illustrated in (86). Below, the infinitive verb *riy*, ‘to go’, is the object of the inflected verb, *munaq kani*, ‘[I] would want’. As infinitive verbs in this context exhibit nominal properties (Sánchez, 2010), they were included in the present analysis.

- (86) *hinaspa amiga-y ni-wa-n, mana riy-ta*  
 then friend-POSS.ISG say-OBJ.ISG-3SG NEG go-ACC  
*muna-q-chu ka-ni, 'baku tusu-q*  
 want-HAB.PST-NEG be-ISG come.on dance-PURP  
*ri-su-nchis'*  
 go-FUT-IPL.INCL

‘Then my friend says to me, I didn’t want to go, “come on, let’s go dance!”’ [P15]

Furthermore, there are a few tokens in which the accusative-marked, lexical verb of the complex verb phrase takes an accusative object of its own, as exemplified in (87). In the present study, these constructions are coded as two separate tokens. In one token, the inflected verb is considered the verb constituent and the lexical, infinitive verb is treated as the object (87a). In the other token, the complex VP containing both the inflected and uninflected verb is considered the verbal constituent, and the accusative object of the lexical verb is treated as the object constituent (87b). Thus, in (87a) the modal verb, *tukuy*, ‘to finish’, is the main verb of the sentence and the infinitive lexical verb marked with *-ta*, *estudiy*, ‘to study’, is the accusative object. On the other hand, in (87b), the entire complex VP, *estudiytaña tukusharaniña*, ‘I already finished studying’, is considered the verb constituent, and *enfermeriata*, ‘nursing’ is the accusative object. Regarding the dependent variable, the former configuration is coded as OV and the latter as VO.

- (87) a. *allin ka-sba-ra-ni, llank'a-sba-ra-ni... ah*  
 well be-PROG-PST-1SG work-PROG-PST-1SG ah  
*sapalla-y ka-sba-spa **estudia-y-ta-ña***  
 alone-POSS.ISG be-PROG-SUB study-INF-ACC-PFV  
*tuku-sba-ra-ni-ña enfermeria-ta*  
 finish-PROG-PST-1SG-PFV nursing-ACC
- b. *allin kasharani, llank'asharani... ah sapallay kashaspa*  
*estudiytaña tukusharaniña **enfermeriata***

‘I was doing well, I was working... ah, being alone, I finished studying nursing’ [P15]

Finally, the suffix *-ta* may also affix to the goal argument of a directional verb, like *riy*, ‘to go’, as mentioned in Chapter 2. Typically, to indicate movement

toward a destination in Quechua, the complement of a directional verb may be morphologically marked with either the allative case suffix *-man*, or accusative *-ta*. In the present study, goal complements, as exemplified in (88), were included in the analysis, and argument type is treated as an independent variable in the Quechua analysis to determine if the syntactic behavior of directional complements differs significantly from non-directional accusative complements. In (88), *yachay wasi*, ‘school’, the directional complement of the verb, *riy*, ‘to go’, is marked with the accusative suffix, *-ta*. Coincidentally, in this same example, the directional complement in the following predicate, *ch’usasunchis Urcosman*, ‘let’s travel to Urcos’, is marked with the allative case suffix, *-man*.

(88) *yacha-y masi-y... ni-wa-n.. ‘yaw... s< ah*  
 know-INF peer-POSS.ISG... say-OBJ.ISG-3SG ‘listen... s< DM  
*paqarin mana yacha-y wasi-ta ri-sunchis-chu...*  
 tomorrow NEG know-INF house-ACC go-FUT.IPL.INCL-NEG  
*mana ri-sunchis-chu... ch’usa-sunchis ahm*  
 NEG go-FUT.IPL.INCL-NEG travel-FUT.IPL.INCL ahm  
*Urcos-man’*  
 Urcos-DAT

‘my classmate... said to me... “hey... s< ah tomorrow, let’s not go to school... let’s not go... let’s travel to Urcos” [P41]

To summarize, the dependent variable of the present study is predicate constituent order, which may be either *object-verb (OV)* or *verb-object (VO)* order. Only predicates accommodating both a verb and an accusative object in a main, declarative clause with a finite verb were considered in the statistical analysis. The envelope of variation is delineated such that predicates in interrogative, imperative, and subordinate clauses were excluded from the analysis, as were predicates with clausal objects, non-finite verbs, verbs like *gustar*, verbal complements with a prepositional phrase, fixed expressions, inaudible components, and placeholders. In the Quechua dataset, both infinitive verb objects and goal complements were included in the analysis. In the next section, I present the independent variables considered in the present study.

#### 4.4.3 Independent variables

The **independent variables** of the present study are of two types– extralinguistic and linguistic. The **extralinguistic** independent variables considered for all participants include *sex, age, residence, education, and language profile*. The

variables *LI* and *BLP score* were also considered for bilingual participants. Refer back to §4.2.3 for a description of these variables.

The majority of the **linguistic** independent variables of the present study were evaluated for both the Andean Spanish and Quechua datasets. These include *animacy*, *definiteness*, *information status*, *polarity*, *subject*, *relative weight* of predicate constituents, and *absolute weight* of the object. In the Quechua analysis, three additional linguistic factors were included— the presence or absence of morphological *accusative marking*, *argument type*, and *Spanish loan-words*.<sup>176</sup> A description of each linguistic variable follows.<sup>177</sup>

### Animacy

It was discussed in Chapter 2 that animate referents tend to occupy prominent syntactic positions (J. K. Bock & Warren, 1985; Branigan et al., 2008; McDonald et al., 1993) or may move to earlier positions within a sentence (Levshina et al., 2023; Tanaka et al., 2011; van Bergen & de Swart, 2010) as a result of their high degree of conceptual accessibility and predicability. To account for the possible effect of the animacy of an object referent on its respective position within a predicate, animacy was included as an independent variable with three factor levels— human (89), animate (90), and inanimate (91). Each factor level is exemplified in both Spanish and Quechua below.

#### (89) Human

- a. *no estaba en su lado cuando en el momento que dio a luz, pero sus hermanas sí... y... cuando fui a verla mi< al día siguiente... en el día... a mi hija le vi y me asusté*

‘I wasn’t by her side the moment that she gave birth, but her sisters were... and... when I went to see her, my< the next day... in the day... I saw **my daughter** and I was startled’ [P17]

<sup>176</sup> Initially, I intended to include the presence or absence of morphological focus and topic marking as an independent factor as well; however, the presence of focus marker *-mi* and topic marker *-qa* were relatively scarce in the Quechua data. In the discussion chapter, I will provide a brief qualitative analysis of the relationship between the morphological expression of information structure and predicate constituent order.

<sup>177</sup> The terms *variable*, *predictor*, and *factor* interchangeably.

- b. *chay wañuy maki-lla-y-wan-mi...* *wayk'u-q*  
 this dead hand-LIM-POSS.ISG-INS-EVIDI cook-HAB.PST  
*ka-ni maki-lla-y-wan wawa-y-ta*  
 be-1SG hand-LIM-POSS.ISG-INS baby-POSS.ISG-ACC  
*cambia-q ka-ni, o maki-lla-y-wan*  
 change-HAB.PST be-1SG or hand-LIM-POSS.ISG-INS  
*t'aqsa-q ka-ni*  
 wash.clothes-HAB.PST be-1SG

‘with this little dead hand of mine... I would cook, with my little hand I would change my baby, or I would wash clothes with my little hand’ [P20]

(90) **Animate**

- a. *el niño tan feliz iba a pastear su ganado todo eso... un día*  
*un< el niño... perdió su vaca... pero su padre le dijo...*  
*“¿dónde está... un ganado?”*

‘the boy [was] so happy [and] he went to graze **his cattle** and everything... one day a< the boy... lost his cow... but his dad said to him... “where is the cow?”’ [P18]

- b. *irqi ka-spa-qa sapa p'unchay mamatayta*  
 child be-SUB-TOP every day parents  
*kamachi-wa-n chakra-ta ruwa-chi-wa-n...*  
 command-OBJ.1SG-3SG field-ACC do-CAUS-OBJ.1SG-3SG  
*uywa-ta michi-chi-wa-n*  
 animal-ACC shepherd-CAUS-OBJ.1SG-3SG

‘when I was a child, every day [my] parents [would] order me, [they would] make me prepare the field... [they would] make me shepherd the animal[s]’ [P31]

(91) **Inanimate**

- a. *porque esa chicha así naturalmente, pero te < te marea  
un poquito de lo que es < es como un alcohol, ¿no?  
la chicha esta tomábamos y cuando tomaba chicha, él se  
ponía ya mareadito y hablaba cualquier cosita*

‘because that chicha [is] natural, but it makes you < you a little tipsy because it’s like an alcohol, right? [we] drank **this chicha** and when [he] drank chicha, he became a little tipsy and [he] said any such thing’ [P19]

- b. *chay-pi buq amiga-ta riqsi-ra-ni también... eh  
that-LOC one friend-ACC meet-PST-1SG also DM  
rima-ra-n quechua-ta pay-pas... entonces...  
speak-PST-3SG quechua-ACC she-COOR then...  
nuqayku ri-ra-yku cafe palla-q  
we.EXCL go-PST-IPL.EXCL coffee gather-PURP*

‘there I met another friend as well... eh she also spoke **Quechua**... then... we went to gather coffee’ [P36]

Though these three categories are relatively straightforward, it bears mentioning that body parts were coded as inanimate referents, as body parts on their own do not constitute a human and do not have sentience apart from the living, inalienable possessor.

### Definiteness

Recall that the definiteness of an object refers to its degree of identifiability in the discourse. Previous word order studies have found that definite referents tend to appear in earlier, non-canonical positions when compared to less definite expressions (Faghiri & Samvelian, 2020; van Bergen & de Swart, 2009, 2010; Yao, 2018). Applying the methodology of van Bergen and de Swart (2009), I partition definiteness into four factor levels, listed in order from most definite to least definite: pronoun, proper noun, definite, and indefinite.

The factor level *pronoun* includes personal pronouns (e.g., *a mí* and *nuqa*, ‘me’), demonstrative pronouns (e.g., *eso* and *chay*, ‘that’), universally quantified pronouns (e.g., *todo* and *llapan*, ‘everything’), and existentially quantified pronouns (e.g., *nada* and *mana imatapas*, ‘nothing’).<sup>178</sup> It is important to note that demonstratives and quantitative expressions were not coded as pronouns when they functioned as determiners (e.g., *todos los países del mundo*, ‘all the

<sup>178</sup> In Quechua, negative, indefinite pronouns typically contain an interrogative expression. For instance, *mana imatapas* is literally ‘not what’. These types of objects were included in the present study because they do not serve an interrogative function.

countries in the world’, or *chay llaqtata*, ‘that town’). In the Spanish example (92a) below, the object of the predicate is the quantitative pronoun *todo*, ‘everything’, which is understood in this context as referring to the things that are sold in the kiosk owned by the participant. In (92b), the demonstrative pronoun *chay*, ‘that’, refers to the toasted wheat and mate that is given to the participant’s children.

(92) **Pronoun**

- a. *este negocio se llama mercachifle... como el mercachifle se vende de toda clase de cosas... pagos de tierra... vinos... cobetes zurriagos... para diferentes actividades... **todo** vendemos por acá... y me siento feliz con mi comercio*

‘this business is called “mercachifle”... like a merchant, things of every type are sold [here]... [supplies for] earth offerings, wines, fireworks, whips... for different activities... we sell everything here... and I am very happy with my business’ [P25]

- b. *habas-ta-pas                      mikhu-chi-yku*  
 fava.beans-ACC-COOR    eat-CAUS-IPL.EXCL  
*wawa-y-man...                      trigo-ta-pas    chay-ta    bank’a-spa*  
 child-POSS.ISG-DAT    wheat-ACC    that-ACC    toast-SUB  
*mate-ta    ruwa-yku                      **chay-ta**...    paykuna-man*  
 tea-ACC    make-IPL.EXCL    that-ACC    they-DAT  
*qu-yku*  
 give-IPL.EXCL

‘we make my kids eat fava beans... and wheat, toasting that, we make tea, **that**... we give to them’ [P28]

Next on the definiteness hierarchy are *proper* nouns. Some proper nouns that commonly appear in this dataset include language names (93a), toponyms (93b), celebrations (e.g., *Inti Raymi*<sup>179</sup>), and names of people (e.g., *Papa Noel*, ‘Santa Claus’).

<sup>179</sup> *Inti Raymi* is a native Quechua celebration held during the summer solstice to honor the sun god, Inti.

(93) **Proper Nouns**

- a. *bueno... este la verdad [es] que yo **el quechua** mastico... pero quiero aprender para llegar al pueblo*

‘well... so, the truth is I jumble up Quechua... but I want to learn so I can go to the community’ [P30]

- b. *Lima-ta* *kuti*< *ri-ni*, *chaymanta* *Arequipa-ta* *ri-ni*  
 Lima-ACC return< go-1SG then Arequipa-ACC go-1SG  
*chaymanta* ***Puno-ta*** *ri-ni* *chaymanta* ***Ocongate***  
 then Puno-ACC go-1SG then Ocongate  
*ri-ni* ***Quillabamba-ta*** *hayku-ni*  
 go-1SG Quillabamba-ACC enter-1SG

‘I return< went to Lima, then I went to Arequipa, then I went to Puno, then I went to Ocongate, [and I] entered Quillabamba’ [P29]

In the Spanish dataset, *definite* objects were differentiated from indefinite objects by the presence of a definite, demonstrative, or possessive determiner. The bolded object in (94a) is definite, as the possessive determiner certifies that the referent is identifiable. In Quechua, both demonstrative and possessive determiners are employed to identify a definite referent. In (94b), the demonstrative determiner *chay* means the referent *challwakunata* is identifiable, in this case, because it was introduced previously in the discourse as *challwa*. It is often the case that a referent is first introduced as an indefinite object, and subsequent mentions of the referent may be accompanied by a determiner or realized as a pronoun, making the referent ‘more’ definite.

(94) **Definite**

- a. *cuando cumplí quince años, me hice mi primera pro*<...  
*comunión que hacen normalmente la iglesia católica*  
 ‘when I turned fifteen years old, I did my first communion that the catholic church usually does’ [P35]
- b. *chakra-ta llank’a-spa uywa-ta michi-spa... ka-n*  
 field-ACC work-SUB animal-ACC graze-SUB be-3SG  
*kaq mayu-yku, chay mayu-pi challwa-ta*  
 one river-POSS.IPL.EXCL that river-LOC fish-ACC  
*hap’i-mu-lla-yku-taq chay challwa-kuna-ta*  
 grab-DIR-LIM-IPL.EXCL-CONTR that fish-PL-ACC  
*mikhu-yku*  
 eat-IPL.EXCL

‘working in the field, grazing animals... there is one our river, in that river, we just grab fish and we eat that fish’ [P32]

<sup>180</sup> Sánchez (2003) suggests that the accusative case-marking suffix *-ta* may also mark definiteness; however, Kalt and Geary (2021) provide an example of a *ta*-marked object with a generic interpretation, and there are numerous examples in the present dataset of indefinite objects marked with *-ta*, so I do not apply this criteria in determining definiteness.

<sup>181</sup> Mannheim et al. (2010) use the term ‘specific’ here instead of definiteness, but the same concept may be applied to definiteness.

It is important to note that definite determiners do not exist in Quechua as independent syntactic entities. Kalt and Geary (2021) assert that, in Quechua, definiteness may be “deduced from context and from the cumulative interpretation of other suffixes in a sentence” (42).<sup>180</sup> More specifically, Mannheim et al. (2010) claim that the default interpretation of referents lacking morphological marking is generic, and that evidential, temporal, and aspectual marking may indicate definiteness,<sup>181</sup> even if these suffixes do not appear on the referent itself, rather on other constituents within the sentential unit. Impressionistically speaking, it appears that topic marking may also alter the interpretation of definiteness. In some instances, the topic marker *-qa* seems to indicate that the object referent should be interpreted as definite. For instance, in (95), the object, *oveja-ta-qa*, ‘sheep’, refers to a particular, identifiable sheep within the group of sheep that was introduced earlier in the discourse. On the other hand, the topic-marked object, *irqitaqa*, ‘children’, in (96) refers to hypothetical children, not particular, identifiable children. In summary, definite referents in the Quechua data were included on the basis of the presence of 1) a demonstrative determiner (e.g., *kay*, this), 2) a possessive suffix, or 3) a topic or focus morpheme in conjunction with the context.

- (95) *huq sipas na-n-ta oveja-n-ta*  
 one young.woman PH-POSS.3SG-ACC sheep-POSS.3SG-ACC  
*michi-sqa ka-ra-n, hinaspa-s mana-s-yá*  
 graze-NMLZ be-PST-3SG then-EVID2 NEG-EVID2-EMO  
*oveja huñu-ra-ka-pu-n-si, o kaq<*  
 sheep gather-PST-REFL-REG-3SG-EVID2 or same  
*kaq-pi< oveja-kuna mana-s ayqi-ku-n-chu,*  
 same-LOC sheep-PL NEG-EVID2 escape-REFL-3SG-NEG  
*chawpi-ta-s atuq tiya-sqa ka-ra-n*  
 middle-ACC-EVID2 fox stand-NMLZ be-PST-3SG  
*chay-lla-man... alqu haqay-pi tiya-sha-n-qa riki*  
 that-LIM-ALL dog over.there-LOC stand-PROG-3SG-TOP DM  
*ni-spa... 'ima-taq ruwa-sha-n?' ni-spa... chay qunqaylla-s...*  
 say-SUB what-Q do-PROG-3SG say-SUB then quickly-3SG  
*atuq-qa q'ipi-ri-ku-n na-ta-qa*  
 fox-TOP carry.on.back-INCEP-REFL-EVID2 PH-ACC-TOP  
***oveja-ta-qa***  
 sheep-ACC-TOP

‘one young woman had been grazing her, whatchamacallit, sheep, then the sheep did not gather together, or the same [group of] sheep didn’t flee, there was a fox standing in the middle, then... a dog standing over there, right?, said ‘what is [the fox] doing?... then very quickly... the fox put the whatchamacallit, **the sheep**, on his back’ [P20]

(96) *nuqa pensa-ra-ni wañu-ku-lla-saq hospital-pi*  
 I think-PST-1SG die-REFL-LIM-FUT.1SG hospital-LOC  
*ni-spa... eh supay supay-ta manchaku-ra-ni chaymanta*  
 say-SUB... DM very very-ADV fear-PST-1SG then  
*nuqa-qa, mana irqi-yuq, uh mana, chay iskay*  
 I-TOP NEG child-ATT DM NEG that two  
*irqi-kuna-manta ka-sqa-n-manta mana nuqa-qa*  
 child-PL-ABL be-NMLZ-3SG-ABL NEG I-TOP  
***irqi-ta-qa** muna-ni-ña-chu*  
 child-ACC-TOP want-1SG-PFV-NEG

‘I thought, I will die in the hospital, I said... uh, I was very very afraid, then I without children, uh no, because there were those two children, I do not want **children** anymore’ [P22]

Finally, the label *indefinite* was reserved for non-identifiable object referents. These objects are either accompanied by an indefinite determiner<sup>182</sup> (e.g., *un* and *huk*, ‘a/an’), a quantitative adjective, (e.g., *algunos* and *wakin*, ‘some’, or *dos* and *iskay*, ‘two’), or are determinerless.<sup>183</sup> For example, the object *discriminación* in (97a) is indefinite because it lacks a determiner, and, as such, it does not refer to an identifiable, specific entity. Similarly, the bolded objects in (97b) refer to general belts and bracelets, not to any particular, identifiable set of belts or bracelets, so *chumpita* and *pulserata* are considered indefinite objects.

(97) **Indefinite**

a. *o sea, por lo, por tu apellido ya te discrimina... tus rasgos faciales por, por tu tamaño por tu color de piel... incluso también sufres **discriminación** aquí por, por tus creencias*

‘I mean, because of, because of your last name they discriminate [against] you... your facial features because, because of your size because of the color of your skin... you even endure **discrimination** here because of your beliefs’ [P24]

<sup>182</sup> For more on the emergence of *huk* as an indefinite determiner in Quechua, consult Sánchez (2003) and Kalt and Geary (2021).

<sup>183</sup> Refer to §8.4 of Chapter 8 for a discussion on evaluating the definiteness of determinerless tokens.

- b. *bueno manta-y-qa*                      *antes-qa*                      *mana-n*  
 well    mom-POSS.ISG-TOP    before-TOP    NEG-EVIDI  
*wasi-lla-y-pi,*                                      *chumpi-ta* *awa-ku-q*  
 house-LIM-POSS.ISG-LOC    belt-ACC    weave-REFL-HAB.PST  
*pulsera-ta*    *kipu-ku-q*  
 bracelet-ACC    tie-REFL-HAB.PST
- ‘well, before my mom was not just in my house, she would weave  
**belts** and would tie **bracelets**’ [P27]

### Information status

The term **information status** refers to the discursive status of a particular referent, in this case, the accusative object. If an object is given, this means it refers to a preceding entity that has previously been mentioned in the discourse. On the other hand, an object that does not refer to a previously introduced entity is new. Previous literature has attested that the information status of the object of a sentence may govern constituent order variation both in Andean Spanish and Quechua specifically (Camacho, 1999; Klee et al., 2011; Muntendam, 2009; F. A. Ocampo, 1995; Sánchez, 2010) and cross-linguistically (Gundel, 1988; Halliday, 1967; Heidinger, 2013; Levshina et al., 2023; Struik & Schoenmakers, 2023; Struik & Van Kemenade, 2018; van Bergen & de Swart, 2009, 2010; Yao, 2018). There are conflicting accounts about the relationship between information status and word order. For instance, according to the “Given before New Principle”, given referents tend to precede new referents (Gundel, 1988, p. 229). On the other hand, Sánchez (2010) proposes that non-canonical VO order is possible in Quechua if the accusative object is a right-dislocated topic, meaning given information may appear later.

To determine the effect of information status of the object on predicate constituent order in Andean Spanish and Quechua, the accusative object of the predicate was coded as either given or new. A *given* object is one that is discourse-accessible, meaning the referent to which the target object refers has been previously uttered in the discourse. A *new* object is one that is introduced into the discourse for the first time and conveys non-presupposed information. Recall that Struik and Van Kemenade (2018) conflate select Pentaset categories (Komen, 2013) to delineate given from new tokens in their work on OV/VO variation in Old English. In contrast to these authors, I take a more narrow interpretation of givenness, meaning only referents pertaining to the *identity* category— entities that have been previously introduced into the discourse—

<sup>184</sup> For Struik and Van Ke-menade (2018), given referents correspond to *identity*, *inferred*, and *assumed* categories.

are considered given in the present study. Referents that may be *inferred* from the previous discourse or *assumed* via shared cultural knowledge were considered *new* referents in the present study, in addition to those introduced into the discourse for the first time (i.e., *new* in Pentaset category terms).<sup>184</sup> Coding information status in this manner limited the often subjective endeavor of determining whether a referent may be inferred from previous discourse or belongs to a set of shared cultural knowledge.

Furthermore, objects were coded as either *new* or *given* with respect to each interview question or topic. For example, if a participant mentioned playing soccer in discussing their childhood, and then several questions later, mentioned watching soccer with their best friend, the referent *soccer* was considered *new* in both contexts. The Andean Spanish (98, 99) and Quechua (100, 101) examples below illustrate both new and given objects within a unit of discourse.

- (98) *ah bueno tengo una experiencia, era cuando estaba, sí creo que era en primaria o jardín, no recuerdo, era pequeña, y pues, recuerdo que me habían... comprado una manzana acaramelada, y pues... recuerdo que no la quería comer yo porque, no sé qué pensaba, no recuerdo bien, pero solo mi mamá dice que no no no quería comerla, y estaba caminando mucho y no, no comía la manzana, y como que mi tío agarró, y me agarró la manzana un rato, y él mordió*

‘ah okay I have an experience, it was when I was, yes, I think it was [when I was in] primary school or kindergarten, I don’t remember, I was little, and well, I remember that they had bought me **a caramel apple**, and well... I remember that I didn’t want to eat it because, I don’t know what I was thinking, I don’t remember well, just that my mom says that I didn’t want to eat it, and I was walking a lot and I didn’t eat the apple, and so my uncle grabbed it, and he grabbed the apple from me a moment, and he bit it’ [P11]

In example (98) above, the accusative object, *una manzana acaramelada*, ‘a caramel apple’, is mentioned for the first time in the predicate, *me habían comprado una manzana acaramelada*, ‘they had bought me a caramel apple’, and is thus coded as a *new* referent. However, in subsequent predicates (e.g., *no comía la manzana*, ‘I didn’t eat the apple’ and *me agarró la manzana*, ‘[he] grabbed the apple from me’), the direct object is coded as *given*, as it has already been introduced into the discourse. Even though the lexical material of the ob-

ject is not precisely the same each time the object is mentioned (i.e., the object is introduced as *una manzana acaramelada* and is subsequently referred to as *la manzana*), because the referent is the same, the accusative object is considered a *given* referent in the second and third predicates emphasized in example (98).

- (99) *me contaron que hacían traer... leche del campo buevo del campo y eso me lo daron*

‘they told me that they made [them] bring milk from the countryside [and] eggs from the countryside and they gave me that’ [P5]

In (99),<sup>185</sup> the object, *leche del campo, buevo del campo*, ‘milk from the countryside, eggs from the countryside’, is new in the first emphasized predicate, as prior to this utterance, this referent had not yet been introduced into the discourse. In the following predicate, the object, *eso*, ‘that’, is considered given because the demonstrative pronoun is co-referential with the object of the previous predicate. Example (99) illustrates that demonstrative pronouns are often considered given, as they are generally discourse deictic.<sup>186</sup>

- (100) *Haqaypi mikbuna gustawaran, ah buq... mikbuna papayuy ch'uñuyuy caldoyuy, ah, chaymanta, ahm, challwayuy, chaypa sutinmi “challwa t'impu” chay mikbunaq sutin, y sumaqllaña, ah, chaytan kuraq ñañay pusawaq mercadota rantinaykupaq... binaspa niwaq, “baku, mikburamusunchis challwa t'imputa,” y sumaqta mikhumuq kayku challwa t'imputa*

‘Over there I liked a food, ah, one... food with potatoes, with dehydrated potato, with broth, ah, then, ahm with fish. Its name is “**boiled fish**”, [it] is that food’s name, and it’s delicious. Ah, my older sister would take me to the market to buy that... then, she would say to me, “Come on, let’s eat boiled fish,” and we would eat the boiled fish well.’ [P15]

In the above Quechua example (100), the entity *challwa t'impu*, ‘boiled fish’ is mentioned for the first time in the sentence, *chaypa sutinmi “challwa t'impu”*, ‘Its name is “boiled fish”’. Subsequently, this referent appears as the accusative object of two predicates that follow in the discourse.<sup>187</sup> The accusative object *challwa t'impu* is considered given in both of these predicates as it was previously mentioned in the discourse. This particular referent is never coded as new in the dataset because the first mention of the referent is as the predicate nominative of the null copulative verb *kay*, ‘to be’, and does not appear in a transitive predicate.

<sup>185</sup> The non-standard use of *daron* (in lieu of *dieron*) in this example illustrates another linguistic feature of Andean Spanish—the “morphological regularization” of certain verbal inflectional paradigms (Escobar, 2011). Escobar contends that this is a second-language feature of Andean Spanish, but the speaker employing this verb form in (99) is a monolingual Andean Spanish speaker.

<sup>186</sup> In some cases, the demonstrative pronoun was cataphoric, meaning it was co-referential with a linguistic entity that appeared later in the discourse. In these cases, the demonstrative pronoun received the label *new*. For example, in the sentence *I will tell you this: as a child, I was very happy*, the demonstrative pronoun, *this*, is new because it is co-referential with the statement *as a child, I was very happy*, which comes after the demonstrative pronoun.

<sup>187</sup> The object *chaytan* in *chaytan kuraq ñañay pusawaq mercadota rantinaykupaq*, ‘my sister would take me to the market to buy that’, is excluded as a token because it is part of a subordinate clause predicate, with the subordinate verb *rantinaykupaq*, ‘in order that we buy’.

(101) *hinaspa papa-y-ta* *waqya-ni...* “*allin-chu ka-sba-n*”  
 then father-POSS.ISG-ACC call-1SG well-Q be-PROG-3SG  
*chay-ta* *tapu-ra-ni* *ni-spa* *ni-wa-n* “*ari*”  
 that-ACC ask-PST-1SG say-SUB say-OBJ.ISG-3SG “yes  
*allin-mi* *ka-sba-ni*”  
 good-EVIDI be-PROG-1SG”

‘then I call my dad... “Are you doing well?” I ask him **that**, he says to me, “Yes, I am well.” [P43]

The Quechua token in (101) mirrors the Andean Spanish token in (99) in that the demonstrative pronoun, *chayta*, ‘that’, is co-referential with an element of the previous discourse. In this case, *chayta* refers to the question that the speaker asks her father, *allin chu kashan?*, ‘are you doing well?’, and is therefore considered a given referent.

### Polarity

In previous analyses on constituent order in Andean Spanish, the corresponding methodologies have either excluded negated main clauses (Klee et al., 2011; F. A. Ocampo & Klee, 1995) or have not specified whether negative polarity contexts were included or excluded (Muntendam, 2009; Muysken, 1984). In the present study, polarity is considered as an independent variable to account for the methodological incongruities of the extant literature and determine its potential conditioning effect on predicate constituent order variation. Only cases in which the negative particle held scope over the entire predicate constituent were coded as *negative*;<sup>188</sup> conversely, tokens lacking a negator were coded as *affirmative*.<sup>189</sup> In Spanish, the presence of a negative adverb (e.g., *no*, *nunca*, ‘never’), noun (e.g., *nada* ‘nothing’, *nadie* ‘no one’), and/or adjective (e.g., *ninguna*, ‘not one’) may result in an overall negative interpretation at the sentential level. In Quechua, a combination of the adverbial negator *mana*, ‘no’, and the negative suffix *-chu*, is the standard means by which negation is expressed, but predicates with a negative pronoun object constituent, like *mana imatapas*, ‘nothing’, also engender a negative interpretation. The examples below illustrate sentential negation in Andean Spanish (102a) and Quechua (102b) due to the presence of adverbial negators *no* and *mana... -chu* respectively.

<sup>188</sup> In Quechua, the negator *mana* may precede an adjective to indicate that the adjective conveys the opposite quality expressed by the non-negated adjective. For example, *allin* may be translated as ‘good’ and *mana allin* may translated as ‘bad’. Where the negator *mana* modifies an adjective, the token is not considered negative, unless there is an additional negator that holds scope over the predicate.

<sup>189</sup> I use the term *affirmative* to enhance the contrast between negative and non-negative polarity; however, ‘affirmative’ may be better interpreted as ‘neutral’.

(102) **Negative Polarity**

- a. *eso también me ch< me afectó... una etapa traumática se puede decir, pero ahora normal... no le guardo rencor, tal vez en ese momento, sí*

‘that also affected me... one could say [it was] a traumatic time, but [everything is] normal now... I don’t hold a grudge against him, maybe in that moment I did’ [P38]

- b. *isqun wata-cha-yuq ka-sba-qti-y...*  
nine year-DIM-ATT be-PROG-SUB.SR-POSS.ISG  
*nishu-ta unquku-ra-ni... mm estomago-y-manta*  
very-ADV get.sick-PST-1SG mm stomach-POSS.ISG-ABL  
*wiksa-y-manta mal ka-ra-ni anchay...*  
stomach-POSS.ISG-ABL ill be-PST-1SG very  
*mana puñu-y-ta ati-ra-ni-chu*  
NEG sleep-INF-ACC able.to-PST-1SG-NEG

‘when I was just nine years old... I became very sick... mm in my stomach, in my stomach I was very ill... I could not **sleep**’ [P43]

**Subject expression**

Because Andean Spanish and Quechua permit both null and overt subjects (Cerrón-Palomino, 1987a; Zagona, 2002), the presence (*overt*) or absence (*null*) of a subject was treated as an independent variable. Subjects are verbal arguments, like objects, and their presence may affect the syntactic structure of the predicate (Hubbel, 2023; Muysken, 1984). The tokens below provide examples of *overt* subjects (103) and *null* subjects (104) in each language.

(103) **Overt Subject**

- a. *dice mi abuelita que ahí hay duende... y que, que ahí vienen los, los, los duendes y que bueno... mi abuelita había visto un duende ahí... ajá... un hombre chiquitito*

‘my grandma says that there are elves there... and that, that they come there the, the, the elves and that, well... my grandma had seen an elf there... uh huh... a very little man’ [P11]

- b. *mana nuqa-qa yacha-ni-chu maneja-y-ta... pero*  
 NEG I-TOP know-1SG-NEG drive-1INF-ACC but  
*musquy-ni-y-pi-qa... supay-ta maneja-sba-ni*  
 dream-CONN-POSS.1SG-LOC-TOP devil-ADV drive-PROG-1SG  
*nuqa... carro-ta-qa... allin-ta maneja-ni carro-ta-qa*  
 I car-ACC-TOP good-ADV drive-1SG car-ACC-TOP
- ‘I don’t know how to drive... but in my dream... like crazy I drive...  
**the car...** I drive the car well’ [P8]

(104) **Null Subject**

- a. *y mi mamá estaba gestando... Ø tenía como sus ocho meses*  
*y mi papá nos dejó*
- ‘and my mom was pregnant... [she] was eight months along (*lit.*  
had eight months) and my dad left us’ [P11]
- b. *nuqa-qa llank’a-ni... Plaza de Armas-pi Ø vende-ku-ni*  
 I-TOP work-1SG Plaza de Armas-LOC Øsell-REFL-1SG  
*artesanía-ta... eh Ø ch’ullu-kuna-ta vendi-ku-ni*  
 handicrafts-ACC eh Øknitted.cap-PL-ACC sell-REFL-1SG
- ‘I work... in the Plaza de Armas Ø [I] sell handicrafts... Ø [I] sell  
**knitted caps [with earflaps]**’ [P1]

Note that in both the Andean Spanish and Quechua examples of null subjects in (104), there are overt subjects in the sentential units immediately preceding each token, e.g., *mi mamá estaba gestando*, ‘my mom was pregnant’, and *nuqa llank’ani Plaza de Armaspi*, ‘I work in the Plaza de Armas’. Semantically speaking, the subjects of the predicates with an expressed subject and null subject are the same; that is, they are co-referential. However, because there is no overt subject licensed by the verb of the target predicate, the emphasized tokens are coded as having a *null* subject.

**Weight**

As discussed in Chapter 2, the weight of constituents may determine their syntactic configuration. The evaluation of weight as a relative or absolute measurement is dependent on the locus of variation. In studies that examine the order of two constituents contained to either the pre-verbal or post-verbal space

(e.g., NP-PP-V versus PP-NP-V), relative weight is the standardly applied metric (Arnold et al., 2000; Heidinger, 2013; Wasow, 1997, 2002). Relative weight refers to the weight of both constituents with respect to one another (e.g., short-before-long or long-before-short). However, in studies on cross domain phrasal shift (i.e., variation in the respective order of object and verb), only the absolute weight of the direct object is evaluated (Struik & Van Kemenade, 2018; Yao, 2018). In the present study, both the **relative** weight of the verb and object constituents and the **absolute** weight of the object constituent were included in the analysis as *continuous* variables. The relative weight is calculated by subtracting the weight of the verb from the weight of the object. Thus, a negative value indicates that the verb constituent is relatively heavier than the object constituent, and a positive value indicates that the object is heavier than the verb. A value of zero indicates that the verb and object are equal in terms of weight. The absolute weight metric considers only the heaviness of the object constituent.

Due to typological differences between the two languages, weight is operationalized distinctly in the Andean Spanish and Quechua datasets. In Andean Spanish, the weight of the verb and object are reported as the number of words of each constituent. However, because Quechua is an agglutinative language that encodes grammatical relationships in morphemes instead of function words, like prepositions and articles, weight in Quechua determined by the number of syllables comprising each constituent.<sup>190</sup>

The examples that follow illustrate instances in which the object of the sentence is relatively heavier than the verb (105), the verb is relatively heavier than the object (106), and the verb and object have similar weight measurements (107).

(105) **Heavy Object**

- a. *este no nos, no nos bañamos como dice con jabón y agua no... solo nos, nos echamos de cabeza para los pies... para que se nos vaya todo, y **el agua que se queda de lo que usamos lo botamos a la calle***

‘uhm we don’t, we don’t bathe ourselves, as they say, with soap and water, no... we just pour [it] from our head to our feet... so that everything leaves us, and we throw [away] **the water that remains from what we use** to the street’ [P37]

<sup>190</sup> Recall that there are many effective ways of measuring weight (Szmrecsanyi, 2004). Wasow (1997, 2002) finds, for example, that the number of words, nodes, and phrasal nodes dominated by the NP predict variation in constituent ordering similarly. Both Yao (2018) and van Bergen and de Swart (2010) operationalize weight as the number of characters of the object. In Struik and Van Kemenade (2018), weight is the “log base 2 of the number of letters” of the object (p. 10).

- b. *eh, taytamama-y-mi llank'a-n eh chakra-ta,*  
 eh parents-POSS.ISG-EVIDI work-3SG eh field-ACC  
*kaq eh, na-ta-n ganaderia-ta uywa-kuna-ta*  
 DM eh PH-ACC-EVIDI livestock-ACC animal-PL-ACC  
*uywa-n... papa-y mama-y eh*  
 care.for-3SG father-POSS.ISG mother-POSS.ISG DM  
*llank'a-n sara-ta papa-ta habas-ta*  
 work-3SG corn-ACC potato-ACC broad.beans-ACC

‘eh, my parents work, eh, the field, uh eh, they take care of whatchamacal-  
 lit, **livestock**, and **animals**... my parents work [in] **corn, potatoes,**  
**and beans**’ [P9]

The examples in (105) illustrate relatively heavy objects with respect to the verb constituent, as the number of words or syllables in the object constituents outweigh the number of words or syllables of the verb constituent. In the Andean Spanish example (105a), the accusative object, *el agua que se queda de lo que usamos*, ‘the water that remains from what we use’, contains a relative clause and is comprised of eight words. This accusative object is relatively heavier than the simple verb constituent, *botamos*, ‘we throw [away]’, which is only one word. The value reported for the relative weight of this predicate is -7 and the value reported for absolute weight of the object is 8. Likewise, the accusative object of the emphasized Quechua tokens in (105b), *ganaderiata, uywakunata*, ‘livestock, animals’ and *sarata, papata, habasta*, ‘corn, potatoes, and broad beans’, contain 10 and 9 syllables respectively, and are thus heavier, than their respective disyllabic verbs, *uywan*, ‘cares for’, and *llank’an* ‘works’.

(106) **Heavy Verb**

- a. *y, o sea, cuando pinto paisajes o algo referente a la  
 naturaleza, trato de darle **perfección**, ¿no?... en cambio las  
 personas, no*

‘and, let’s see, when I paint landscapes or something related to nature, I try to give it **perfection**, right?... however, [with] people, I don’t’ [P13]

- b. *allin ka-sha-ra-ni llank'a-sha-ra-ni... ah*  
 good be-PROG-PST-IS work-PROG-PST-ISG DM  
*sapa-lla-y ka-sha-spa estudia-y-ta-ña*  
 only-LIM-POSS.ISG be-PROG-SUB study-INF-ACC-PFV  
*tuku-sha-ra-ni-ña enfermeria-ta...*  
 finish-PROG-PST-ISG-PFV nursing-ACC  
 ‘I was doing good, I was working... ah being alone, I was finishing  
studying nursing...’ [P15]

On the contrary, in examples (106a) and (106b), the verb constituent is relatively heavier than the object constituent. In the former example, the Andean Spanish verb constituent is a complex VP made up of three words, *tratar de dar*, ‘try to give’, which is heavier than the single-word object constituent, *perfección*, ‘perfection’. Likewise, the five-syllable Quechua object, *enfermeriata*, ‘nursing’, is outweighed by the eleven-syllable, complex VP, *estudiytaña tukusharaniña*, ‘I was finishing studying’.

(107) **Balanced Constituents**

- a. *pero era travieso también en el fondo y normalmente*  
*solía hacer muchas travesuras con mis compañeros de*  
*colegio*  
 ‘but ultimately, I was also mischievous, and normally, I tended to make  
a lot of mischief with my high school classmates’ [P4]
- b. *chaymanta-ña mm suegro-y-pa*  
 then-PFV DM father.in.law-POSS.ISG-GEN  
*wasi-n-pi tiya-ra-ni...*  
 house-POSS.3SG-LOC live-PST-ISG  
*tiya-ra-yku-n, cuarto-cha-ta*  
 live-PST-IPL.EXCL-EVIDI room-DIM-ACC  
*qu-wa-ra-n-ku*  
 give-OBJ.I-PST-3SG-OBJ.PL  
 ‘then, mm, I lived in my father-in-law’s house... we lived, he gave  
 us **a little room**’ [P23]

Finally, in (107a), the object and verb constituents are equal in weight, as both the object, *muchas travesuras*, ‘a lot of mischief’, and verb, *solía hacer*,

‘used to do’, are two-word constituents. In (107b), the object, *cuartochata*, ‘a little room’, and verb, *quwaranku*, ‘he gave [us]’, are comprised of four syllables each, and are thus equal in weight.

### Accusative marking (Quechua only)

All examples of Quechua accusative objects provided in this chapter to this point have been marked with the accusative suffix, *-ta*. However, it has been reported that, in some Quechua varieties, accusative case-marking is not categorical, and that the presence or absence of case-marking may correlate with constituent order variation (Kalt & Geary, 2021; Muntendam, 2015; Sánchez, 2003). To account for the possible effect of accusative case-marking on predicate constituent order, morphological marking is accounted for as an independent variable with two factor levels— *present* and *absent*— in the statistical analysis of the Quechua data. Consider example (108) below:

- (108) *paykuna*, “*baku-chu wasi-y-ta*,” *ah o na*  
 they go-NEG house-POSS.ISG-ACC ah or DM  
 “*wayk’u-ka-mu-sunchis*” *mikhuna*, *tarea-ta*  
 cook-REFL-DIR-FUT.IPL.INCL food homework-ACC  
*ruwa-mu-sunchis*”  
 do-DIR-FUT.IPL.INCL

‘they [said], “why don’t we go to my house?” ah or um, “let’s cook **food**, let’s do homework”’ [P15]

Example (108) illustrates an instance in which the object, *mikhuna*, ‘food’, is not case-marked, but is clearly the accusative complement of the verb *wayk’uy*, ‘to cook’. In the predicate that follows, the object, *tarea*, ‘homework’ precedes the inflected verb, *ruway*, ‘to do’, and is marked with *-ta*. Consequently, example (108) contains two tokens, one in which accusative marking is present and one in which it is absent.

### Argument type (Quechua only)

Recall that in Quechua, the goal argument of a directional verb (e.g., *riy*, ‘to go’, or *kutiy*, ‘to return’) is often marked with the allative suffix, *-man*, but may alternatively be accompanied by accusative *-ta*, as illustrated in (88) in §4.4.2. In this example, the argument of the directional verb, *riy*, ‘to go’, is the destination, *yachay wasi*, ‘school’. This object may be described as a goal because it is the

place toward which an action is directed (Longacre, 1983). Because goal arguments are typically encoded in a prepositional phrase in Spanish (e.g., *vamos a la playa*, ‘let’s go **to** the beach’), they are not considered accusative objects, and were thus not included in the Andean Spanish dataset. Directional complements are accounted for in the Quechua analysis as an independent variable with two factor levels— *goal* and *non-goal*.

It is important to note that goal arguments that were accompanied by neither *-ta* nor *-man* were also included in the dataset, just as accusative object tokens lacking *-ta* on non-goal arguments were included. In (109) below, the goal argument, *chay llaqtay*, ‘my [that] town’, is not marked with either *-ta* or *-man*; nevertheless, it functions as the goal argument of the inflected verb *riy*, ‘to go’.

- (109) *mana irqi-cha-kuna-pas ka-n-ña-chu*  
 NEG child-DIM-PL-COOR be-3SG-PFV-NEG  
*llaqta-y-pi... machu-cha kunallan-mi*  
 town-POSS.ISG-LOC old.person-DIM now.LIM-EVIDI  
*ka-sba-n... mana tanto ri-pu-nku-ña-chu*  
 be-PROG-3SG NEG so.much go-REG-3PL-PFV-NEG  
*llaqta-y-man... ajá llaqta-y-pi ruwa-nku*  
 town-POSS.ISG-ALL uh.huh town-POSS.ISG-LOC make-3PL  
*sara-ta papa-ta haba-s-ta ruwa-nku... pero*  
 corn-ACC potato-ACC fava.bean-PL-ACC do-3PL but  
*kunan-qa, mana-ña tanto runa-kuna ri-nku-ña-chu*  
 now-TOP NEG-PFV so.much person-PL go-3PL-PFV-NEG  
*chay llaqta-y ajá*  
 that town-POSS.ISG uh.huh

‘there are no longer any children in my town... now there just old people... they don’t really go to my town so much anymore... uh huh, in my town they make corn, potatoes, and fava beans... but now, people don’t go to my town so much anymore’ [P22]

Note that in example (109), prior to the emphasized token, the speaker says *ripunkuñachu llaqtayman*, ‘[they] no longer go to my town’. In this case, the allative case marker *-man* is affixed to the goal argument of the inflected verb, *llaqtayman*. Though I will not explore *-man* versus *-ta* versus zero-marking on goal arguments, this example and (88) above suggest that there is some degree of intra-speaker morphological variation in this domain.

### Spanish loanwords (Quechua only)

As a result of prolonged Spanish-Quechua contact, a substantial proportion of the modern Quechua lexicon is comprised of Spanish loanwords, with estimates for Imbabura Quechua around 25% (Gómez Rendón & Adelaar, 2009, p. 955). Recall that it has been suggested in the contact literature that, in situations of language maintenance, structural borrowing is mediated by lexical borrowing (Thomason & Kaufman, 1988; Winford, 2003). To determine if the realization of a Spanish lexical borrowing primes canonical Spanish VO order, each token was evaluated for its *presence* or *absence* of a Spanish loanword in the accusative object and/or verb. All examples in (110) are coded as having a Spanish loanword *present* in the predicate constituent.

- (110) a. *yaqa yaqa wañu-pu-n wawa-y ajá yaqa*  
 almost almost die-REG-3SG baby-POSS.IS uh.huh almost  
*entonces... entonces eh hospital-pi... indu< inducción-ta*  
 then then eh hospital-LOC indu< induction-ACC  
*ruwa-wa-rqa-nku*  
 do-OBJ.ISG-PST-3PL

‘my baby almost, almost died, uh huh, almost, then... then eh in the hospital... induc< they did **an induction** [on] me’ [P8]

- b. *eh... wayna-pura buñuka-ru-ku-spa chay*  
 eh young.man-INTER gather-EXH-REFL-SUB that  
*puklla-q ka-yku... taki-yu-q*  
 play-HAB.PST be-IPL.EXCL sing-INT-HAB.PST  
*ka-yku <inaudible>... eh quena-cha-ta*  
 be-IPL.EXCL <inaudible> eh Andean.flute-DIM-ACC  
*toca-q ka-yku*  
 play-HAB.PST be-IPL.EXCL

‘eh... gathering just us young men, we would play that... we would sing... eh we would play **the Andean flute**’ [P44]

- c. *y ri-ra-yku... nn na-n... paseo< pasea-q*  
 and go-PST-IPL.EXCL DM PH-EVIDI paseo< stroll-PURP  
*chay Quillabamba llaqta-ta... y foto-kuna-ta*  
 that Quillabamba town-ACC and picture-PL-ACC  
*toma-ka-mu-ra-yku hermanu-y-kuna-wan*  
 take-REFL-DIR-PST-IPL.EXCL brother-POSS.ISG-PL-INS  
 ‘and we went... mm uh... to stro< to stroll to that town of Quillabamba... and with my brother we took **pictures** of ourselves’ [P43]

In the above examples, the object, *inducciónta*, ‘induction’ in (110a); the verb, *tocay*, ‘to play’ in (110b); and both predicate constituents, *fotokunata*, ‘photos’ and *tomay*, ‘to take’ in (110c) are Spanish loanwords.

All independent variables and their respective factor levels are listed in Table 4.5 below.<sup>191,192,193</sup>

<sup>191</sup> Absolute weight is operationalized as number of words of the object constituent in Spanish and number of syllables in Quechua.

<sup>192</sup> Relative weight is operationalized as the relative number of words of predicate constituents (relative weight = weight of object - weight of verb) in Spanish and number of syllables in Quechua.

<sup>193</sup> The *first language* and *BLP score* are evaluated only for bilingual participants. The presence or absence of *accusative marking*, *argument type*, and *Spanish loanwords* are considered only in the Quechua analysis.

<b>Factor</b>	<b>Levels</b>
<b><i>Extralinguistic</i></b>	
<b>Sex</b>	<i>male</i> <i>female</i>
<b>Age</b>	<i>continuous</i> (range: 18-72 years of age)
<b>Education</b>	<i>primary</i> <i>secondary</i> <i>post-secondary</i> <i>university</i>
<b>Residence</b>	<i>metro</i> <i>department</i>
<b>First language</b>	<i>L1 Quechua</i> <i>simultaneous</i> <i>L1 Spanish</i>
<b>BLP score</b>	<i>continuous</i> (range: -68.39 — +63.12)
<b><i>Linguistic</i></b>	
<b>Animacy</b>	<i>human</i> <i>animate</i> <i>inanimate</i>
<b>Definiteness</b>	<i>pronoun</i> <i>proper noun</i> <i>definite</i> <i>indefinite</i>
<b>Information status</b>	<i>given</i> <i>new</i>
<b>Polarity</b>	<i>affirmative</i> <i>negative</i>
<b>Subject</b>	<i>null</i> <i>overt</i>
<b>Weight (absolute of object)</b>	<i>continuous</i>
<b>Weight (relative)</b>	<i>continuous</i>
<b>Accusative marking</b>	<i>present</i> <i>absent</i>
<b>Argument type</b>	<i>goal</i> <i>non-goal</i>
<b>Spanish loanwords</b>	<i>present</i> <i>absent</i>

Table 4.5: Independent variables of the present study

## 4.5 Research questions

Recall that the general objective of the present study is to examine variation in the order of predicate constituents (i.e., OV versus VO order) in Andean Spanish and bilingual Cusqueño Quechua. The specific research questions intended to guide the present study are restated below. A null hypothesis and an alternative hypothesis accompanies each of the research questions. I espouse the alternative hypotheses.

1. What are the overall proportions of OV/VO orders in Andean Spanish? Do monolingual and bilingual speakers differ in their distribution of predicate constituent orders?

**Null hypothesis:** The rate of non-canonical OV order in Andean Spanish will resemble the rates reported for non-contact varieties of Spanish ( $\approx 1-11\%$ ). There will be no difference between monolingual and bilingual speakers regarding predicate constituent order variation.

**Alternative hypothesis:** The rate of non-canonical OV order in Andean Spanish will resemble the average rate of OV order reported in the extant literature on word order variation in Andean Spanish,  $\approx 17-28\%$  (Klee, 1996; Klee et al., 2011; Muntendam, 2009; Muysken, 1984; F. A. Ocampo & Klee, 1995).<sup>194</sup> Based on the correlation between OV/VO variation and language profile of the speaker reported in previous studies (Klee et al., 2011; Muysken, 1984), I hypothesize that bilinguals will exhibit higher rates of non-canonical OV order than monolinguals.

<sup>194</sup> See Table 7.3

2. Which extralinguistic and linguistic factors condition OV/VO variation in the varieties of Andean Spanish spoken by monolingual and bilingual speakers?

**Null hypothesis:** Extralinguistic and linguistic factors will not be shown to significantly condition predicate constituent order in Andean Spanish.

**Alternative hypothesis:** In line with the findings of previous studies, non-canonical OV order rates will be higher among Andean Spanish speakers who are female,<sup>195</sup> older, less educated and reside in the department of Cusco. Furthermore, Quechua-dominant bilingual participants who identify Quechua as their first language

<sup>195</sup> Muntendam (2009) and Klee et al. (2011) report an opposite effect of sex on OV/VO variation. I anticipate that Muntendam's findings will be more applicable to the present study given that the data of her study was collected in an Andean region, contrary to the data of Klee et al., which was collected in Lima.

will also exhibit relatively higher OV order rates (Muysken, 1984). Though the age of the speaker has not been evaluated as an independent predictor of OV/VO variation in the extant literature, I hypothesize that there will be a positive correlation between OV order rate and age, as language change is often reflected in generational differences across speakers (Tagliamonte, 2012).

Regarding the linguistic variables, I hypothesize that non-canonical OV order will be more frequent with given, animate, definite, and light objects accompanied by null subjects, given the findings summarized in Chapters 3 and 2 (Faghiri and Samvelian, 2020; Gundel, 1988; Heidinger, 2013; Struik and Schoenmakers, 2023; Struik and Van Kemenade, 2018; Tanaka et al., 2011; van Bergen and de Swart, 2009, 2010; Yao, 2018, to name a few). Additionally, I anticipate that polarity will govern OV/VO variation such that more variation (i.e., higher OV order rates) will be permitted in affirmative contexts compared to negative contexts, as additional constituents may affect syntactic variation (Klee, 1996; Klee et al., 2011; F. A. Ocampo & Klee, 1995).

3. What is the overall proportion of OV/VO orders in bilingual Cusco Quechua?

**Null hypothesis:** The proportions of OV/VO orders in the bilingual Cusco Quechua will coincide with OV-dominant order in accordance with its typological classification.

**Alternative hypothesis:** The rate of OV order in bilingual Cusco Quechua will be consistent with a language lacking a dominant order of predicate constituents due to the prevalence of non-canonical VO order in the corpus (Sánchez, 2003).

4. Which extralinguistic and linguistic factors condition OV/VO variation in bilingual Cusco Quechua?

**Null hypothesis:** Extralinguistic and linguistic factors will not be shown to significantly condition predicate constituent order variation in Quechua.

**Alternative hypothesis:** Given the findings of Hubbel (2023), I hypothesize that significantly higher rates of canonical OV order will coincide with Quechua-dominant L1 Quechua speakers, and that the sex, age, residence, and level of education of the speaker will not exert a significant effect on OV/VO variation.

Concerning the linguistic predictors, I hypothesize that canonical OV order will be motivated by light, definite, new, accusatively-marked<sup>196</sup> objects with a null subject based on the findings of Hubbel (2023), Kalt and Geary (2021), and Sánchez (2003, 2010). Based on the cross-linguistic tendency for ‘more’ animate entities to appear earlier (J. K. Bock & Warren, 1985; McDonald et al., 1993; Tanaka et al., 2011), I also anticipate that objects arranged in OV order will tend to be more animate. I speculate that predicates containing a Spanish loanword will exhibit VO order more frequently than those lacking a Spanish loanword on the basis of the claim that structural borrowing is mediated by lexical borrowing (Winford, 2003). Finally, I posit that goal arguments will exhibit higher rates of VO order as a mechanism of structural differentiation from non-goal, NP arguments.

<sup>196</sup> Recall that Kalt and Geary (2021) and Sánchez (2003) report opposite findings regarding the relationship between *ta*-marking and OV/VO variation in Quechua. My hypotheses align with Sánchez’s findings because the methodology of her study (in particular, the manner in which she elicited data) aligns more closely with that of the present study.

5. How do the factors selected as predictors of predicate constituent order variation in Andean Spanish compare to those selected for Quechua? What might this comparison suggest about contact-induced language change in this syntactic domain?

**Null hypothesis:** Since predicate constituent order variation will not be found to be governed by the examined extralinguistic or linguistic variables in neither Andean Spanish nor Quechua, there will be no indication of contact-induced language change.

**Alternative hypothesis:** Syntactic variation will be conditioned similarly in Andean Spanish and bilingual Cusco Quechua, which will suggest bidirectional contact-induced change such that contact with Quechua structurally influences Andean Spanish and vice versa (Poplack & Levey, 2010). The linguistic outcome of the contact situation will resemble convergence, i.e., Andean Spanish and Quechua are becoming structurally more similar to one another (Bullock & Toribio, 2004; Winford, 2003).

## 4.6 Summary

This chapter has provided the methodological architecture of the present study, including a description of the participants, data collection procedure, and data coding process. The 45 participants of this study were recruited using a combination of random sampling and snowball sampling methods. Overall, the participant sample slightly over-represents females, younger generations, metropolitan

inhabitants, and formally-educated individuals when compared to the general population of the department of Cusco. Approximately three-quarters of the participant sample is bilingual, which was a deliberate choice to adequately compare predicate constituent order variation in bilingual Andean Spanish and Cusco Quechua, which is required for the consideration of this variable structure as an example of contact-induced change.

Regarding the data collection instruments, all participants completed the Language Background Questionnaire (LBQ) and a Sociolinguistic Interview. The LBQ included questions that extracted the participants' biographical information as well as their linguistic history, current use, proficiency, and attitudes. For bilingual participants, this instrument was used to determine their linguistic dominance (i.e., BLP score) and first language (L1), in addition to general demographic information— sex, age, residence, and level of education— collected for both monolingual and bilingual participants. In essence, LBQ responses inform the extralinguistic variables that are associated with each participant. The Sociolinguistic Interview was completed in Andean Spanish by all participants<sup>197</sup> and in Quechua by bilingual participants.

<sup>197</sup> With the exception of P15, who completed only an interview in Quechua

The spontaneous speech extracted via the Sociolinguistic Interview was transcribed and subsequently hand-searched for all predicates containing an explicit object and verb that fell within the circumscribed envelope of variation. Only predicates with both an object and verb in main, declarative sentences were included, effectively excluding predicates in subordinate, imperative, and interrogative clauses. Also excluded from the analysis were predicates comprised of non-finite verbs, verbs like *gustar*, verbal complements requiring a preposition, fixed expressions, inaudible components, the Quechua placeholder *na*, and clausal objects. Both affirmative and negative predicates were included in the analysis, as were lexical NP objects and pronominal objects, though pronominal clitics (e.g., *lo, la, los, las*) were excluded. Tokens exhibiting OVO or VOV order that could not be disambiguated using the intonation of the utterance were counted twice— once as OV and once as VO order. In the Quechua dataset, infinitive verb objects and goal arguments are included, though the latter are regarded as an independent variable. The independent variables, particularly the linguistic variables, were presented in §4.4.3 and summarized in Table 4.5

The hypotheses that accompany each research questions in §4.5 will be put to the test in the following two chapters, which provide the findings of the statistical analyses of the Andean Spanish and bilingual Cusco Quechua datasets respectively.

## CHAPTER 5

# ANALYSIS: ANDEAN SPANISH

### 5.1 Introduction

The purpose of this chapter is to provide a comprehensive statistical analysis of predicate constituent order variation in the Andean Spanish corpus. I reserve a discussion of these results for Chapter 7, which will incorporate the findings of the statistical analyses conducted on both languages under investigation. The results provided here pertain directly to the first two research questions, which are repeated below.

**RQ1:** What are the overall proportions of OV/VO orders in Andean Spanish? Do monolingual and bilingual speakers differ in their distribution of predicate constituent orders?

**RQ2:** Which extralinguistic and linguistic factors condition OV/VO variation in the varieties of Andean Spanish spoken by monolingual and bilingual speakers?

In §5.2, I supply the overall proportions of predicate and clausal constituent orders for the entire Andean Spanish corpus. Subsequently, I compare the distribution of OV/VO orders among the monolingual and bilingual groups. From here, I move to a separate treatment of the monolingual and bilingual groups.

In the monolingual analysis (§5.3), I first present the general OV/VO proportions of the group as a whole followed by the proportions corresponding to each individual monolingual participant to determine the degree of interspeaker variation within this group. Then, I supply the descriptive statistics of the extralinguistic (*sex*, *residence*, level of *education*, *age*) and linguistic factors (*animacy*, *definiteness*, *information status*, *polarity*, *subject expression*, *weight*)

by reporting both the distribution of OV/VO order across the levels of each categorical variable and the measures of central tendency and spread (e.g., mean, median and quantile values) associated with each continuous variable. Finally, I conduct a binomial logistic regression analysis to identify the predictors that significantly condition predicate constituent order variation for the monolingual group. I close this section by discussing significant interactions among the linguistic predictors.

In §5.4, I observe a similar process to report the results of the bilingual group. Structurally, this section differs slightly from the preceding section in that I offer a separate binomial logistic regression analysis of the extralinguistic and linguistic variables for expository purposes. Incidentally, the extralinguistic variables considered in the bilingual analysis also include *first language (L1)* and *BLP score*, which do not apply to the monolingual group. Lastly, I conclude the chapter in §5.5 with a summary of the results presented for the monolingual and bilingual groups, offering a brief comparison of the respective trends uncovered in the data.

## 5.2 General results

Of the 45 total participants of the study, 44 provided interviews in Andean Spanish.<sup>198</sup> In the data collected from these 44 interviews, there were **3,687** instances in which the predicate constituent, comprised of a verb and object, adhered to the exclusion and inclusion criteria described in §4.4.2. Of the total 3,687 tokens, **405** (10.98%) exhibited non-canonical OV order and **3,282** (89.02%) exhibited VO order.

When an explicit subject constituent is considered, all possible word orders are accounted for in the Andean Spanish dataset. Consider Table 5.1 below:

	VO Type				OV Type			
	<i>VO</i>	<i>SVO</i>	<i>VSO</i>	<i>VOS</i>	<i>OV</i>	<i>SOV</i>	<i>OSV</i>	<i>OVS</i>
<i>n</i>	2,456	780	13	33	329	17	39	20
%	66.61%	21.16%	0.35%	0.90%	8.92%	0.46%	1.06%	0.54%

Table 5.1: Frequency of clausal constituent orders in Andean Spanish dataset

Table 5.1 illustrates that null subjects are far more frequent than overt subjects in this dataset, as tokens lacking overt subject expression account for approximately three-fourths of the data (75.53%). Moreover, proportionally speaking, overt subjects coincide more frequently with VO-type orders than OV-type

<sup>198</sup> P15 was the sole participant who completed a sociolinguistic interview in Quechua only.

orders. Regarding the former, 25.17% of VO-type orders have an expressed subject, while overt subjects appear in only 18.77% of instances of OV-type orders. It is interesting to note that when subjects are overt, subject-initial order is the second most frequent order for VO-type tokens, but there is no clear preference for subject position when expressed in OV-type orders. The possible conditioning effect of subject expression on predicate constituent order will be inspected in §5.3 and §5.4.2, where it is evaluated as a linguistic predictor.

Recall that the objectives of this chapter are to calculate the general OV/VO order proportions and to determine the variables that condition predicate constituent order variation in two groups— monolingual and bilingual Andean Spanish— with the intention of comparing their results in Chapter 7. Monolingual participants account for 36.05% of the total tokens in the Andean Spanish corpus ( $n = 1,329$ ) and bilingual participants the remaining 63.95% ( $n = 2,358$ ). Regarding the general OV/VO proportions of each group, OV order is more frequent among the bilingual group (13.19%) compared to the monolingual group (7.07%). According to a Pearson’s Chi-squared test with Yates’ continuity correction, there is a significant difference,  $p = 1.63 \times 10^{-08}$ , between monolingual and bilingual speakers of Andean Spanish with respect to predicate constituent order variation. Consider Table 5.2 below.

PCO	Monolingual		Bilingual	
	<i>n</i>	%	<i>n</i>	%
OV	94	7.07%	311	13.19%
VO	1,235	92.93%	2,047	86.81%
<b>Total</b>	<b>1,329</b>	<b>100%</b>	<b>2,358</b>	<b>100%</b>

Table 5.2: OV/VO order across language profile of Andean Spanish speakers

To further illustrate the difference in predicate constituent order variation between monolingual and bilingual groups, the categorical, binary dependent variable was transformed into a continuous variable—rate of OV order— by creating a data frame composed of the OV/VO rates, in percentages, of each participant. The descriptive statistics of both groups are visually represented in the boxplots in Figure 5.1.

In Figure 5.1, the blue boxplot on the left summarizes the range OV order rates of the bilingual group, and the yellow boxplot on the right summarizes the corresponding data of the monolingual group. The middle, horizontal line of each boxplot represents the *median* OV order rate, with the exact numerical value in the yellow box; the value in the blue box is the *mean* OV order rate. The colored boxes for each group represent the interquartile range of OV order rates; that is, the middle 50% of the data falling between Q<sub>1</sub> (the first quartile)

and  $Q_3$  (the third quartile). The upper vertical line represents the values that fall between the maximum value and  $Q_3$ , and the lower vertical line represents the values that fall between the minimum value and  $Q_1$ . The large black dot above the bilingual boxplot identifies a single outlier. Each small, black dot corresponds to the OV order rate of a single participant.

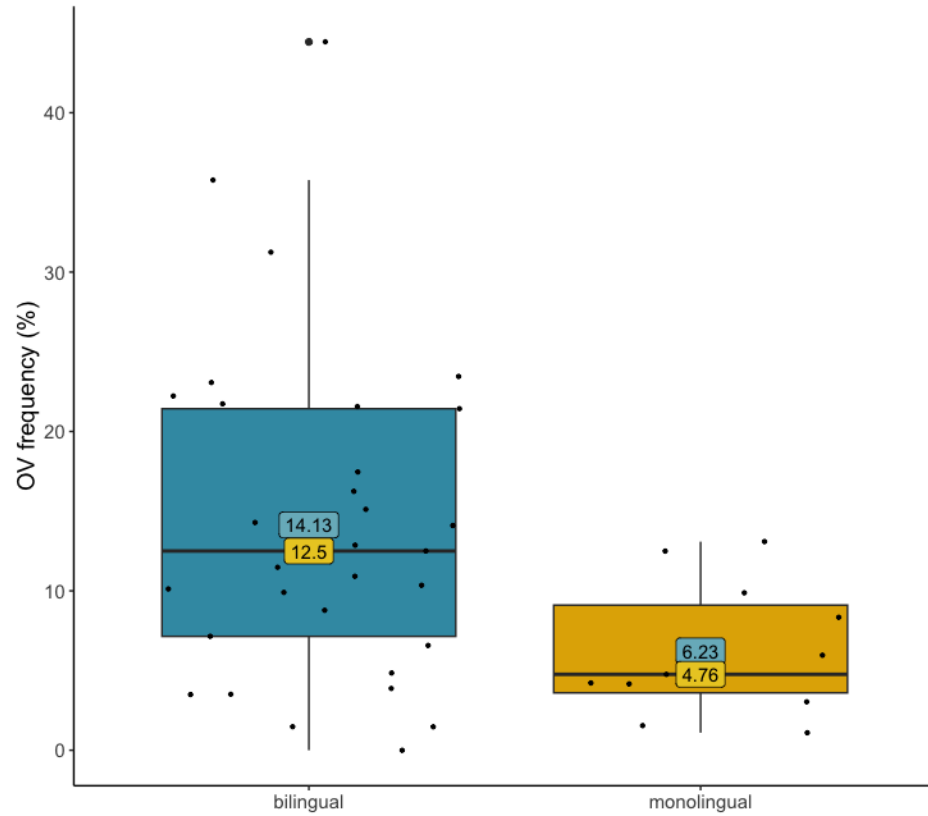


Figure 5.1: Rate of OV order (%) across language profile of Andean Spanish speakers

Figure 5.1 not only illustrates that OV order rates are higher overall for bilingual participants compared to monolingual participants, but also that there is considerable inter-speaker variation among bilingual participants. Specifically, OV order rates range from 0.0% to 44.0% for the bilingual group, and from 1.10% to 13.09% for the monolingual group. So, there is much less inter-speaker variation among monolingual participants, as evidenced by a compressed boxplot that represents a smaller range of values.

Further comparing the monolingual and bilingual results necessarily requires a separate analysis for each group. In the next two sections, I provide a statistical analysis of the distribution of predicate constituent order across all

extralinguistic and linguistic variables evaluated in the present study for monolinguals and bilinguals. I begin with the monolingual Andean Spanish results.

## 5.3 Monolingual Andean Spanish results

### 5.3.1 Descriptive statistics

Based on self-reported information provided in the Language Background Questionnaire, 11 of the 45 participants are monolingual speakers of Andean Spanish. As reported in Table 5.2, of the total **1,329** tokens supplied by the monolingual group, **1,235** (92.93%) predicate constituents were arranged in VO order, and the remaining **94** (7.07%) in OV order. The raw (*n*) and relative (%) frequencies of OV/VO orders for each participant are presented in Table 5.3 and illustrated graphically in Figure 5.2.<sup>199</sup>

<sup>199</sup> PCO is an abbreviation for *predicate constituent order*.

	OV	OV(%)	VO	VO(%)	Total( <i>n</i> )
P3	10	5.95%	158	94.05%	168
P4	1	1.10%	90	98.90%	91
P5	5	4.76%	100	95.24%	105
P6	5	12.50%	35	87.5%	40
P7	3	4.17%	69	95.83%	72
P11	7	4.22%	159	95.78%	166
P12	3	3.03%	96	96.97%	99
P14	1	1.54%	64	98.46%	65
P24	39	13.09%	259	86.91%	298
P30	8	9.88%	73	90.12%	81
P38	12	8.33%	132	91.67	144
<b>Total</b>	<b>94</b>	<b>7.07%</b>	<b>1,235</b>	<b>92.93%</b>	<b>1,329</b>

Table 5.3: OV/VO order across monolingual participants

Though monolingual speakers do vary with respect to their ordering preferences, the degree of inter-participant variation is small, especially in comparison to the bilingual group. As mentioned earlier, OV order rates range from 1.10% to 13.09% across monolingual participants, with an average value of 6.23%.<sup>200</sup> What appears to be more variable among this group is the number of tokens provided by each participant. For instance, P24 accounts for 298 (22.42%) of the monolingual tokens, but P6 accounts for only 60 (4.51%). The disproportionate number of tokens offered by some participants is simply due to the duration of their respective interviews.

<sup>200</sup> To clarify, OV orders account for 7.07% of monolingual tokens overall; the 6.23% figure is the average OV order rate across monolingual participants (see Figure 5.1).

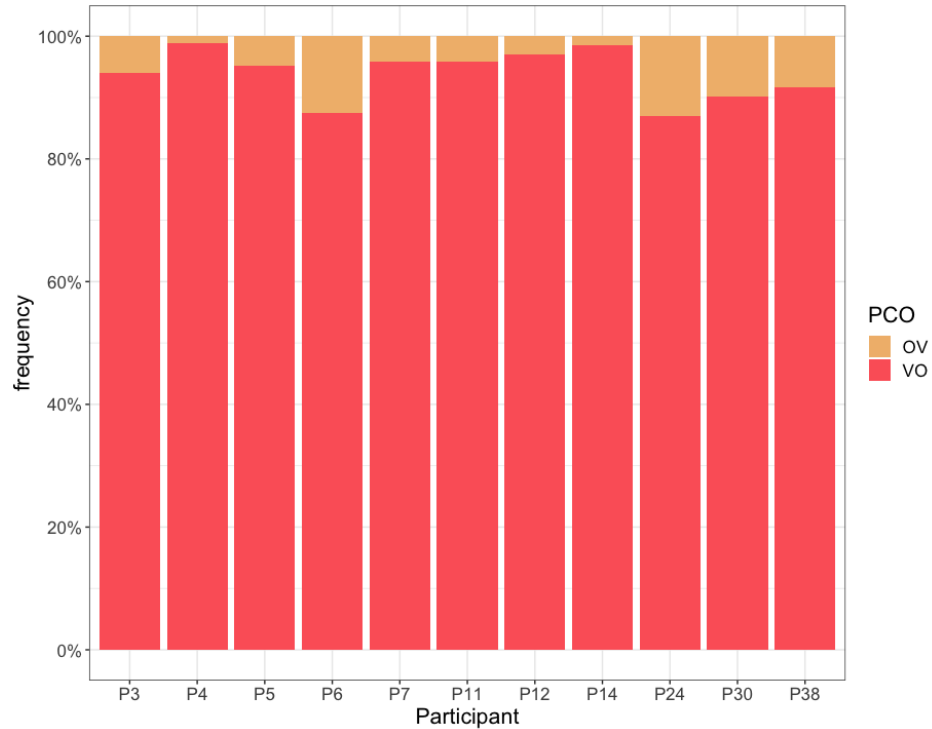


Figure 5.2: Barplot of OV/VO order across monolingual participants

### Extralinguistic variables

The extralinguistic variables evaluated for the monolingual group include *sex*, *residence*, level of *education*, and *age*. Table 5.4 presents the raw and relative frequencies of OV and VO orders across the categorical extralinguistic variables—*sex*, *residence*, *education*— and Table 5.5 provides the descriptive statistics for the continuous variable, *age*.

To summarize Table 5.4, regarding **sex**, OV order is slightly higher for *women* (7.53%) than for *men* (6.38%). Additionally, female participants supply a larger proportion of tokens in general, despite the fact that the monolingual group is comprised of six males and five females.<sup>201</sup> The contrast between OV and VO proportions is more apparent for the variable **residence**. Participants from the *department* of Cusco realize OV order at a higher rate (12.40%) than participants from the *city* of Cusco (4.95%). These factor levels are also unbalanced, with metro participants contributing nearly three times as many tokens as department participants.

<sup>201</sup> It is possible that female participants felt more comfortable and, thus, willing to converse about personal topics due to the fact that both interviewers also identify as female.

	OV	OV (%)	VO	VO (%)	Total (n)
<b>Sex</b>					
Male	34	6.38%	499	93.62%	533
Female	60	7.53%	736	92.46%	796
<b>Residence</b>					
Metro	47	4.95%	903	95.05%	950
Department	47	12.40%	332	87.60%	379
<b>Education</b>					
Secondary	5	4.76%	100	95.24%	105
Post-secondary	59	7.38%	741	92.63%	800
University	30	7.08%	394	92.92%	424

Table 5.4: OV/VO order across extralinguistic variables for monolingual group

It is important to note that all monolingual participants completed secondary education at a minimum. Thus, there are only three levels of the variable **education** that apply to the monolingual group— *secondary*, *post-secondary*, and *university*. Unexpectedly, OV proportions are lower for monolingual participants with a secondary-level education (4.79%) in comparison to those that are more highly educated (7.38% and 7.08%). However, because only one of the eleven monolingual participants had attained a secondary-level education, the lower proportion of OV tokens is perhaps the result of individual ordering preferences. Regardless, descriptively speaking, there is little difference across education factor levels concerning the distribution of predicate constituent orders.

The descriptive statistics, including the minimum, maximum, quantile, median, and mean values, for the continuous variable, **age**, are available in Table 5.5.

	min	Q1	median	mean	Q3	max
<b>OV</b>	19	19	30	31.5	30	72
<b>VO</b>	18	19	23	28.8	30	72

Table 5.5: Descriptive statistics of age for monolingual group

Overall, the mean age of OV order is **31.5 years**, which 2.7 years higher than the mean of VO order, **28.8 years**. This indicates that OV order is employed at higher rates among relatively older participants. It is also the case that the lower quartile represents a larger range of values for OV than for VO, meaning the lower 25% of the data is spread out across a larger range of ages for OV order

(19-30 years) compared to VO order (18-23 years). The minimum and maximum values are the same for OV and VO orders because all participants employed both orders at least once.

### Linguistic variables

The linguistic independent variables considered in the present study are *animacy*, *definiteness*, *information status*, *polarity*, *subject expression*, and *weight* (*absolute* and *relative*). Table 5.6 provides the raw and relative frequencies of OV and VO orders across the levels of each categorical linguistic variable.

	OV	OV(%)	VO	VO(%)	Total( <i>n</i> )
<b>Animacy</b>					
animate	22	11.11%	235	88.89%	257
inanimate	72	6.72%	1,000	93.28%	1,072
<b>Definiteness</b>					
pronoun	34	20.24%	134	79.76%	168
definite	26	5.21%	473	94.79%	499
indefinite	34	5.14%	628	94.86%	662
<b>Information status</b>					
given	61	14.15%	370	85.85%	431
new	33	3.67%	865	96.33%	898
<b>Polarity</b>					
affirmative	88	7.38%	1,105	92.62%	1,193
negative	6	4.41%	130	95.59%	136
<b>Subject expression</b>					
null	79	7.57%	965	92.43%	1,044
overt	15	5.26%	270	94.74%	285

Table 5.6: OV/VO order across linguistic variables for monolingual group

Beginning with **animacy**, it is important to note that the factor levels *human* and *animate* were ultimately conflated due to the small number of non-human, animate tokens ( $n = 54$ ). Still, the factor levels are unbalanced, with *inanimate* referents accounting for 80.67% of all tokens. There is a slight preference for pre-verbal position among *animate* referents (11.11%) compared to inanimate referents (6.72%). On the contrary, the relationship between **definiteness** and predicate constituent order is more transparent. While *definite* and *indefinite* referents exhibit virtually the same rate of OV order (5.21% and 5.14%), *pronouns* appear in pre-verbal position more frequently (20.24%). To clarify, *proper* and *definite* NPs were conflated into a single category— *defi-*

*nite*— as *proper* nouns were exceedingly infrequent in the Andean Spanish and Quechua datasets.

Regarding **information status**, the Table 5.6 shows that, overall, there are more new referents in the dataset than given referents; in fact, new referents account for two-thirds of the data. That new referents outnumber given referents is not surprising considering the lack of common ground between interlocutors at the outset of the conversation. Furthermore, participants were asked several questions on distinct topics that would necessarily require the frequent introduction of new referents into the discourse. Concerning the distribution of predicate constituent orders across information status, it is clear that there is a near categorical preference for VO order when the referent is new (3.67%). The proportion of OV order, on the other hand, is higher for given referents (14.15%), meaning predicate constituent order is more variable when the object referent has already been introduced into the discourse.

For the remaining variables— **polarity** and **subject expression**— there is only a marginal difference in OV/VO distribution across their accompanying factor levels. In general, *affirmative* polarity and *null* subjects exhibit a slightly higher proportion of OV order in comparison to their counterparts (*negative* polarity and *overt* subjects, respectively). Notice also that the overall distribution of tokens across the factor levels of each variable is quite unbalanced; the factor level with a slightly higher OV preference (*affirmative* and *null*) is that which accounts for at least three-fourths of the data within the respective factor.

The *continuous* linguistic variables evaluated in the present study are **absolute weight** and **relative weight**. Recall that in the Andean Spanish dataset, absolute weight is operationalized as the number of words comprising the object constituent. On the contrary, relative weight considers both the weight of the object and verb constituents; this value is calculated by subtracting the number of words of the verb constituent from the number of words of the object constituent. Thus, a negative value denotes a relatively heavy verb constituent, a positive value denotes a relatively heavy object constituent, and a value of zero denotes predicate constituents of equal weight. The descriptive statistics for absolute and relative weight are given in Table 5.7.<sup>202</sup>

<sup>202</sup> The (+) symbol accompanies positive values for relative weight because these values may be either positive or negative.

	min	Q1	median	mean	Q3	max
<b>Absolute weight</b>						
OV	1	1	2	2.02	2	9
VO	1	1	2	2.47	3	13
<b>Relative weight</b>						
OV	-2	0	0	+0.72	+1	+8
VO	-3	0	+1	+1.13	+2	+11

Table 5.7: Descriptive statistics of weight for monolingual group

Regarding **absolute weight**, the range of values is slightly larger for VO order compared to OV order. The minimum values are the same for both orders, but the maximum object length for VO order is 13 words and the maximum for OV order is only 9 words. The mean value for VO order, 2.47 words, is also larger than the mean value for OV order, 2.02 words. In the density plot<sup>203</sup> in Figure 5.3, the OV and VO order overlays are skewed to the left, meaning the majority of object constituents are only a few words in length. However, the yellow OV overlay is skewed to the left comparably more than the red VO overlay, meaning that OV order is preferred with lighter objects.

<sup>203</sup> The yellow dashed line represents the mean value for OV order, and the red dashed line represents the mean value for VO order.

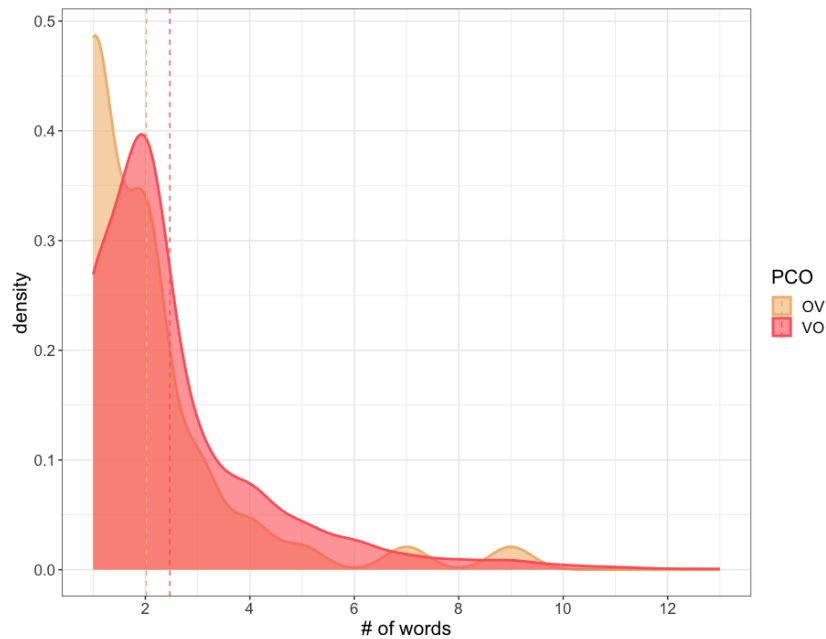


Figure 5.3: Density plot of OV/VO order across absolute weight for monolingual group

The **relative weight** statistics follow a similar pattern: VO order encompasses a larger range of values than OV order, especially considering the respective maximum values. The mean values corresponding to both OV and VO orders are positive, meaning objects tend to outweigh verbs in general. This is unsurprising given that nouns tend to be accompanied by determiners and other modifiers (including heavy modifiers, like relative clauses) while verb phrases are typically comprised of only one word, or two to three words in complex verbal phrases. It is also the case for relative weight that the mean value of VO order is slightly higher than the mean value of OV order. From the density plot in Figure 5.4 it is also clear that the distribution of VO order is more normal<sup>204</sup> than OV order, as the yellow OV overlay both skews more to the left and has a sharper peak, yielding less of a bell-curve shape than its VO counterpart.

<sup>204</sup> A normal distribution is represented by a line in the shape of a bell curve that is centered in the middle of the x-axis.

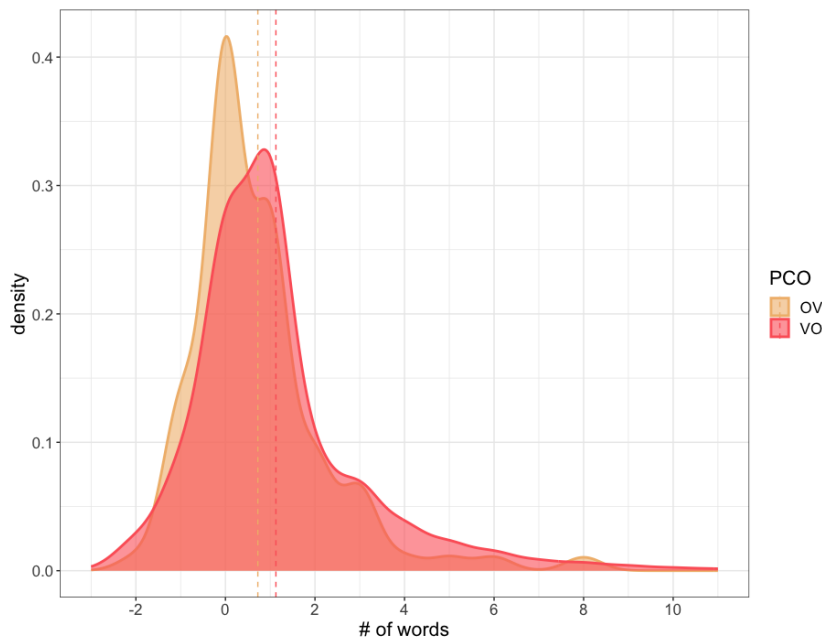


Figure 5.4: Density plot of OV/VO order across relative weight for monolingual group

### 5.3.2 Binomial logistic regression analysis

In this section, I discuss the results of a binomial logistical regression analysis. Binomial logistic regression is a type of statistical analysis that estimates the probability of a categorical binary response variable<sup>205</sup> given predictor variables using a maximum likelihood estimation technique (Baayen, 2008). A logistic regression analysis enhances the descriptive statistics by determining which predictor variables significantly effect the probability that a particular predicate

<sup>205</sup> The term ‘response variable’ is synonymous to ‘dependent variable’, and the term ‘predictor variable’ is synonymous to ‘independent variable’.

constituent order will be employed. I conducted a logistic regression analysis in R (R Core Team, 2023) using the function *glm()*, which is typically used to fit generalized linear models but may also be used for logistic regression analysis when it is specified that the response variable is binary.<sup>206</sup>

<sup>206</sup> The website <https://www.statology.org> was an indispensable resource in the interpretation of these results.

The purpose of conducting a logistic regression analysis is to ascertain the model that best explains the variation associated with the response variable (i.e., best-fit model). To find the best-fit model, I first constructed a model containing the response variable, *predicate constituent order*, and all extralinguistic and linguistic predictor variables listed above. For the monolingual group, I considered only fixed effects, as inter-speaker variation is negligible.<sup>207</sup> Then, I used the *step()* function to apply an automatic stepwise algorithm to determine the model with the lowest AIC value. The AIC value, *Akaike Information Criterion*, does not provide any information on its own; however, comparing across models, the lowest AIC value accompanies the model that best fits the data. The model with the lowest AIC value for the monolingual group includes the predictors *education*, *residence*, *definiteness*, *information status*, *polarity*, and *subject expression*. A summary of the best-fit model is presented in Table 5.8.

<sup>207</sup> I attempted to use *glmer()* to identify a mixed effects model that evaluated *participant* as a random variable; however, I received error messages because *participant* was not an appropriate random variable due to lack of variation.

	Coefficients	Std. Error	z value	Pr(> z )
(Intercept)	1.73	0.25	6.86	6.87e-12***
education:secondary	-0.67	0.56	-1.20	0.23
education:university	-1.09	0.36	-3.05	0.00232**
residence:metro	1.42	0.34	4.17	3.09e-05***
definiteness:indefinite	-0.25	0.28	-0.91	0.36
definiteness:pronoun	-1.32	0.30	-4.47	7.92e-06***
info.status:new	1.20	0.25	4.84	1.30e-06***
polarity:negative	0.68	0.45	1.51	0.13
subject:overt	0.42	0.30	1.38	0.17
Null deviance:	679.18 on 1328 degrees of freedom			
Residual deviance:	585.39 on 1320 degrees of freedom			
AIC:	603.39			

Table 5.8: Results of best-fit mixed-effects model for monolingual group

The coefficient estimate corresponding to the intercept indicates the log odds of VO order when each predictor is equal to zero. For categorical predictors, ‘zero’ refers to the log odds of VO order at the reference level of each variable, which is determined by alphabetical order. The reference levels of the selected categorical predictors are as follows: *post-secondary* (education), *department* (residence), *given* (anaphoricity), *definite* (definiteness), *negative* (polar-

ity), and *null* (subject). To enhance comprehensibility of the results, the log odds coefficient estimate may be transformed into an odds ratio— the average change in odds— by configuring the logit value as the exponent of the constant  $e$ . Log odds may also be converted into a probability expression through the application of the formula  $e^{\text{logit}}/(1+e^{\text{logit}})$ . According to this formula, given the log odds coefficient estimate of 1.73, the probability of VO order is 84.94% at the intercept.

The coefficient estimates provided to the right of each factor level represent the average change in log odds from the reference level at the intercept to the factor level expressed in the leftmost column, assuming all other variables are held constant. For instance, when the highest level of education attained is *secondary*, there is a decrease of -0.67 in the log odds of VO order from the reference level, *post-secondary*. Since  $e^{-0.67} = 0.51$ , this means the odds of VO order for participants with a secondary-level education is 0.51 times lower than participants with post-secondary-level education. That is, those with a secondary-level education have a 49% (1-0.51) lower odds of employing VO order than those with post-secondary-level education.

The  $p$ -value in the rightmost column ( $p = 0.23$ ) indicates that the odds of VO order are not significantly different between participants with a secondary-level and post-secondary-level education, as only a  $p$ -value  $< 0.05$  is considered significant in the present study. Technically speaking, the  $p$ -value provides the probability of a given  $z$ -value, which is determined by dividing the coefficient estimate by the standard error. The null and residual deviance values may be used to calculate a Chi-square ( $X^2$ ) statistic by subtracting the latter from the former. For this model  $X^2 = 93.79$  with 8 degrees of freedom, and is associated with a significant  $p$ -value  $< 0.05$ . This means that the model summarized in Table 5.8 accounts for the variation in the response variable well.

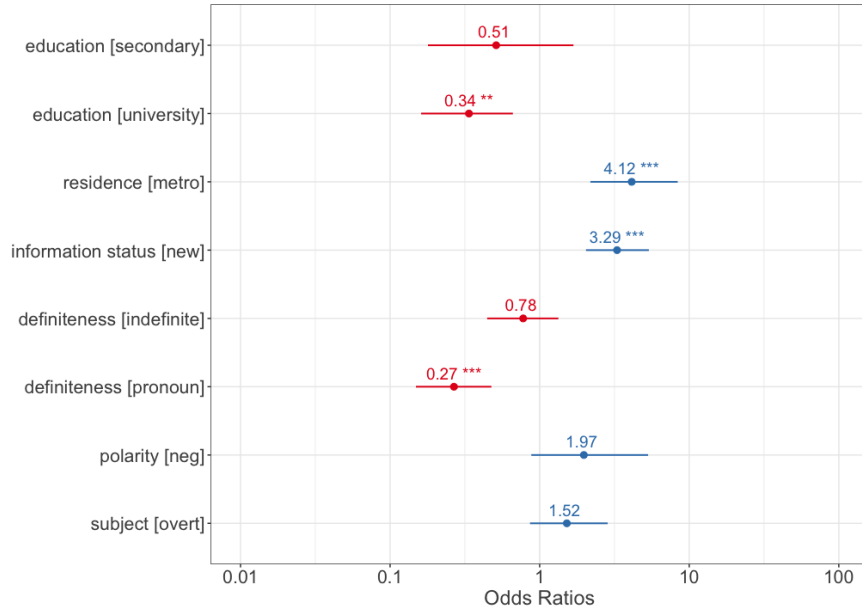


Figure 5.5: Visualization of best-fit logistic regression model for monolingual group

The fixed-effects logistic regression model detailed in Table 5.8 is visualized in Figure 5.5. The values in Figure 5.5 are odds ratios, which have been transformed from the log odds values provided in Table 5.8 using the exponentiation procedure described above to facilitate interpretation of the results. The red values represent a negative change in odds and the blue values represent a positive change in odds. A value accompanied by asterisks means that there is a significant difference in the odds of VO order between the reference level and the listed level of the predictor variable. Thus, the difference in the odds of VO order between *secondary* and *post-secondary* is not significantly different, but, the post-secondary group and *university* group do differ significantly from one another. Specifically, those with a university-level education have 66% (1-0.34) lower odds of employing VO order than those with post-secondary education. The effect of **residence** on predicate constituent order variation is also significant such that participants residing in *metro* Cusco are associated with a 312% (4.12-1) increase in the odds of VO order compared to those residing in the *department* of Cusco.

The linguistic variables, **definiteness** and **information status** also significantly condition OV/VO variation in the monolingual group. Particularly, a *new* referent increases the odds of VO order by 3.29, or 229%, considering given referents as the reference level and assuming all other variables are held constant. Regarding definiteness, though there is no significant difference in the odds of

VO order between *definite* and *indefinite* objects, when the object constituent is *pronominal*, the odds of VO order decrease by 73%.

The last two selected linguistic predictors, **polarity** and **subject**, are not significant predictors of predicate constituent order variation. However, eliminating these predictors from the model increases the AIC value, meaning that a model that does not include these variables does not offer the best fit. Both negative polarity and overt subjects increase the odds of VO order by 97% and 52% respectively.

Lastly, I considered two potential interactions in the data— definiteness-animacy and definiteness-polarity. Concerning the first interaction, I entertained the possibility that pronouns referring to human entities (e.g., *a mí*, ‘me’) had a unique effect on predicate constituent order compared to pronouns referring to non-human entities (e.g., *eso*, ‘that’ or *todo*, ‘everything’). Similarly, I considered a possible interaction between definiteness and polarity to determine whether negative pronouns, like *nada*, ‘nothing’, behave differently than affirmative pronouns, like *todo*, ‘everything’.<sup>208</sup>

I consulted conditional inference trees both to identify the existence of a significant interaction between these two sets of variables and to understand the relationship between this interaction and predicate constituent order variation. The first interaction, definiteness-animacy, is not associated with a significant *p*-value. The interaction between definiteness and polarity, on the other hand, is significant,  $p < 0.05$ . Figure 5.6 visualizes this interaction.

<sup>208</sup> The variable *polarity* does not refer just to the polarity of the object rather to the sentence as a whole. However, the majority of instances of pronominal objects in a negative context do contain the negative object, *nada* ( $n = 26, 78.78\%$ ).

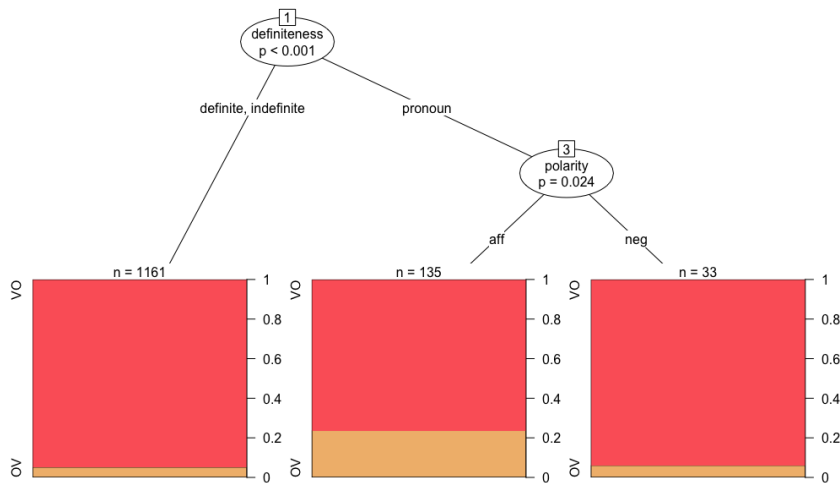


Figure 5.6: Definiteness-polarity interaction in monolingual group

In Figure 5.6, VO proportions are represented in red, and OV proportions are represented in yellow. This conditional inference tree shows that the pro-

portion of OV order is substantially higher for pronouns in affirmative contexts (23.70%) versus pronouns in negative contexts (6.06%). This means that pronouns like *a mi*, ‘me’ and *eso*, ‘that’ show a higher preference for OV order than negative pronouns like *nada*, ‘nothing’. The distribution of OV/VO orders is unaffected by polarity for definite and indefinite objects.

Upon adding the definiteness-polarity interaction to the model, the AIC value dropped from 603.39 to 602.96, and the Chi-squared statistic increased from  $X^2 = 93.79$  to  $X^2 = 98.22$ . This means that the model including the interaction between definiteness and polarity better accounts for variation in predicate constituent order than the model sans this interaction.

To summarize, in this section I provided both descriptive statistics and a binomial logistic regression analysis. Taken together, the data presented here elucidate the statistical relationship between predicate constituent order variation and all extralinguistic and linguistic predictors evaluated for the monolingual Andean Spanish group. Ultimately, the logistic regression analysis revealed that *education*, *residence*, *definiteness*, and *information status* are significant predictors of OV/VO variation in the monolingual group. The linguistic predictors *polarity* and *subject expression* were also selected by the best-fit model, though the *p*-value associated with these factors was not significant,  $p > 0.05$ . Finally, there is a significant interaction between polarity and definiteness such that negative polarity restrict the pre-verbal collocation of pronominal objects. I turn now to a similar analysis of the bilingual Andean Spanish group.

## 5.4 Bilingual Andean Spanish results

The bilingual Andean Spanish corpus is comprised of the interviews of 33 bilingual participants.<sup>209</sup> The bilingual group realized **2,358** tokens of OV/VO orders, which were recorded in the corresponding bilingual Andean Spanish dataset. As presented in Table 5.2, **311** (13.19%) instances were arranged in non-canonical OV order and the remaining **2,047** (86.81%) in VO order. The descriptive statistics corresponding to the distribution of OV and VO order rates per bilingual individual, including the minimum, maximum, quantile, median, and mean values, appears in Table 5.9.

	min	Q1	median	mean	Q3	max
<b>OV (%)</b>	0.0	7.1	12.5	14.1	21.4	44.4
<b>VO (%)</b>	55.6	78.6	87.5	85.9	92.9	100.00

Table 5.9: Distribution of OV/VO order rates per bilingual participant

<sup>209</sup> Though 34 bilingual participants completed an interview, one participant, P15, completed a sociolinguistic interview only in Quechua.

In opposition to the monolingual results, there is considerable inter-speaker variation in ordering preferences, which is visually represented in Figure 5.7. At one extreme, P<sub>41</sub> did not employ non-canonical OV order in a single predicate; that is, she used VO order categorically. At the other extreme, P<sub>28</sub> appears to lack a strong preference for one order over the other, as OV order accounts for 44.44% of her total tokens, and VO order accounts for the remaining 55.56%.

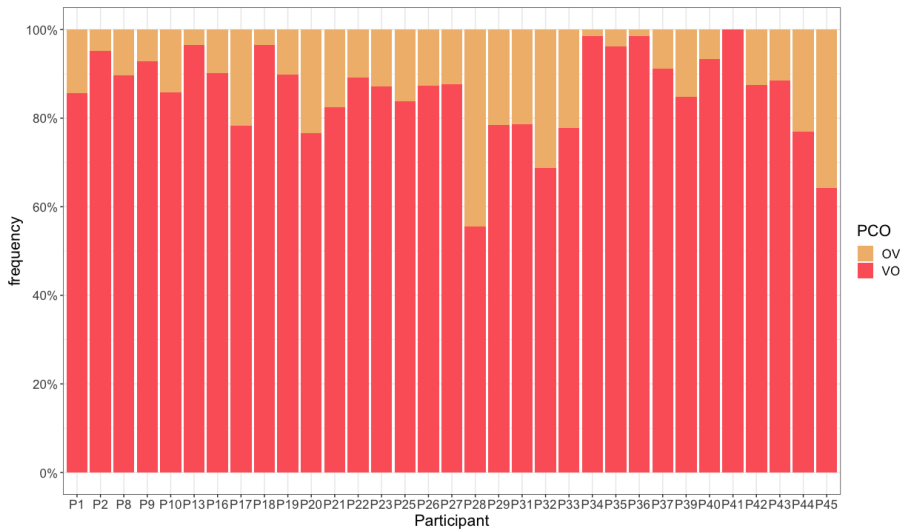


Figure 5.7: OV/VO proportions across bilingual group participants

Bilingual participants also differ from one another with respect to their demographic and linguistic profiles, which have been codified in multiple extralinguistic variables. I turn now to an analysis of these extralinguistic predictors.

### 5.4.1 Analysis of extralinguistic predictors

#### Descriptive statistics

To review, two additional extralinguistic predictors were considered for the bilingual group that were not evaluated for the monolingual group simply because these predictors did not apply to the latter group. In addition to the extralinguistic variables *sex*, *age*, *residence*, and level of *education*, the variables *first language* (*L1*) and *BLP score* were also included. Recall that *L1* is a categorical variable with three levels—*L1* Quechua, simultaneous, and *L1* Spanish. *BLP score* is a continuous variable reflecting the linguistic dominance of the participant. The raw and relative frequencies of OV/VO orders for each categorical predictor—*sex*, *residence*, *education*, and *L1*— and their corresponding levels are listed in Table 5.10. A visualization of the distribution of OV/VO orders across the cat-

egorical extralinguistic predictors is provided in Table 5.11. Note that the red partition represents VO order frequency, and the yellow partition represents OV order frequency.

	VO	OV (%)	OV	VO (%)	Total (n)
<b>Sex</b>					
Female	216	14.26%	1299	85.74%	1,515
Male	95	11.27%	748	88.73%	843
<b>Residence</b>					
Department	141	16.13%	733	83.87%	874
Metro	170	11.46%	1314	88.54%	1,484
<b>Education</b>					
Primary	84	28.57%	210	71.42%	294
Secondary	97	15.50%	529	84.50%	626
Post-secondary	130	9.04%	1,308	90.96%	1,438
<b>First language</b>					
LI Quechua	118	16.05%	983	83.95%	1,101
simultaneous	99	12.06%	722	87.94%	821
LI Spanish	24	6.56%	342	93.44%	366

Table 5.10: OV/VO order across extralinguistic variables for bilingual group

Regarding the first categorical variable presented in the table, **sex**, OV/VO order proportions are similar for *males* and *females*, though female participants employ non-canonical OV order slightly more frequently than male participants, 14.26% versus 11.27%. It is important to mention the discrepancy between men and women regarding the total number of tokens provided by each. Female participants account for nearly two-thirds of the data (64.25%) because they also accounted for two-thirds of the total participants ( $n = 22$ ) in this group. Thus, the bilingual group over-represents the female population. Consequently, OV order frequency in this sample may be marginally higher than a more balanced sample for sex, as female participants employ non-canonical order slightly more frequently than their male counterparts in this sample.

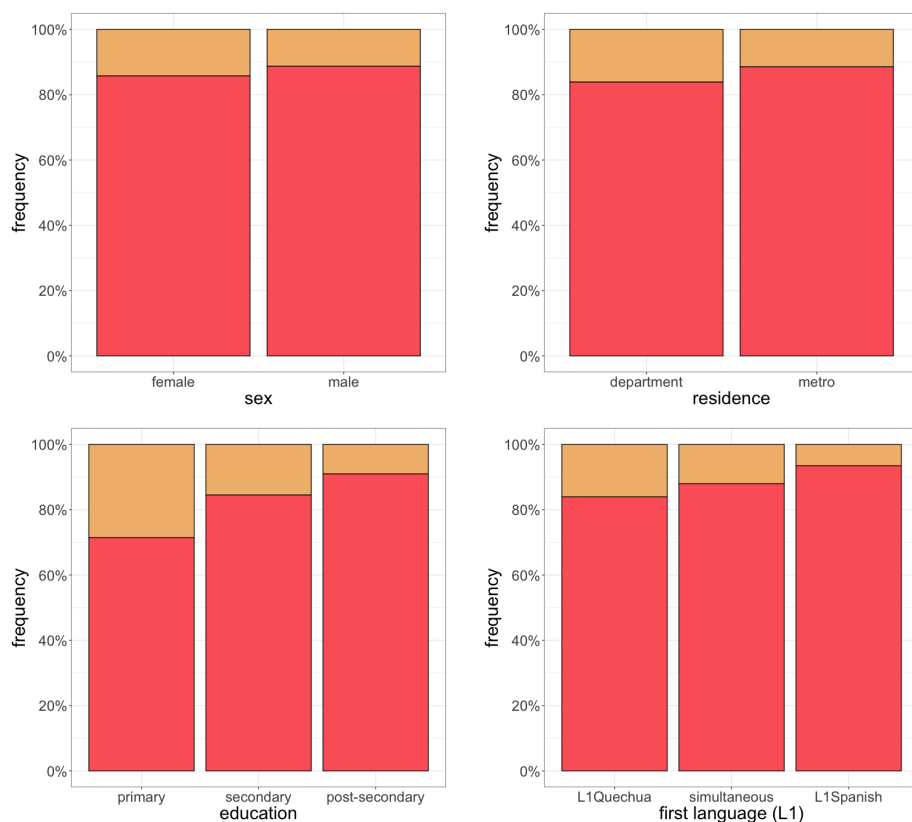


Table 5.11: Barplots of OV/VO order across extralinguistic variables for bilingual group

Similarly, the factor **residence** is unbalanced in that metropolitan residents are over-represented. Nearly two-thirds of the data (62.93%) is provided by metropolitan participants, who also account for approximately two-thirds of the bilingual participant group ( $n = 20$ ).<sup>210</sup> Irrespective of the sampling bias, there is tendency for *department* residents to employ non-canonical OV order at a slightly higher rate, 16.13%, than *metropolitan* residents, 11.46%.

Prior to discussing the descriptive results corresponding to the variable **education**, I must first justify the conflation of the levels *university* and *post-secondary* for the bilingual group. First, *university* tokens accounted for a small proportion of the data ( $n = 250$ ). Second, there are even fewer participants with a university-level education who provided an interview in Quechua, and, in order to effectively compare the bilingual Andean Spanish and Quechua results, the levels of a given predictor variable ought to be identical. Regarding the relationship between education and predicate constituent order after this adjustment, there appears to be a robust negative correlation between level of

<sup>210</sup> As determined in the previous chapter, this is opposite the proportion of metro to department residents in the Cusco population as reported by the Instituto Nacional de Estadística e Informática (2017)— 35.9% of Cusco inhabitants reside in the city, and the remaining 64.1% live outside the city.

education and OV order frequency. That is, the less educated a participant, the more likely they are to employ OV order. The relevant barplot in Table 5.11 highlights this trend; the proportion of OV order is the highest for participants with a primary-level education and the lowest for those with a post-secondary-level education.

Lastly, the **first language** of the speaker is associated with disparate proportions of OV/VO order. *L1 Quechua* bilinguals exhibit the highest rates of non-canonical OV order, which is the dominant order of Quechua. *Simultaneous* bilinguals— those who reported identical ages of acquisition for Andean Spanish and Quechua— employ OV order slightly less frequently than L1 Quechua bilinguals, but twice as frequently as *L1 Spanish* bilinguals, proportionally speaking. In fact, the OV order rate corresponding to L1 Spanish bilinguals mirror that of monolingual participants, 6.56% and 7.07% respectively. Regarding the overall distribution of tokens across first language, L1 Quechua and simultaneous bilinguals provide a substantial proportion of the data in comparison to L1 Spanish bilinguals. This skew in the data could mean that a more balanced sample would exhibit even lower rates of OV overall.

For the continuous variables, *age* and *BLP score*, I supply descriptive statistics, which include the minimum, maximum, quantile, median, and mean values in years and dominance scores respectively for both predicate constituent orders. Table 5.12 below provides the statistics associated with age.

	min	Q1	median	mean	Q3	max
<b>OV</b>	18	24.5	34	40.3	59	68
<b>VO</b>	18	22	29	32.4	34	68

Table 5.12: Descriptive statistics of age (in years) for bilingual group

The age range of the speaker, 18-68 years, is equal across both predicate constituent orders because all speakers<sup>211</sup> produce both OV and VO orders. The mean and median age values are higher for OV order than for VO order. The average age of the participant for OV order is **40.3 years**, and the average age for VO order is **32.4 years**, a difference of 7.9 years. According to the provided quantile values, it is also the case that, for VO order, the data falling between Q3 and the maximum corresponds to a range of 34-68 years ( $\Delta = 34$  years), meaning the upper 25% of the data is spread out over a large span of ages, as VO order is more concentrated in the lower quantiles. The opposite is true for OV order— the upper 25% encompasses a much smaller range of ages, 59-68 ( $\Delta 9$  years), meaning OV order is more concentrated among relatively older ages. The density plot in Figure 5.8 supplies a visual representation of the distribution

<sup>211</sup> With the exception of P41, who is 32 years old.

of age for both orders. The red, dashed line represents the mean value of VO order, and the yellow, dashed line represents the mean value of OV order.

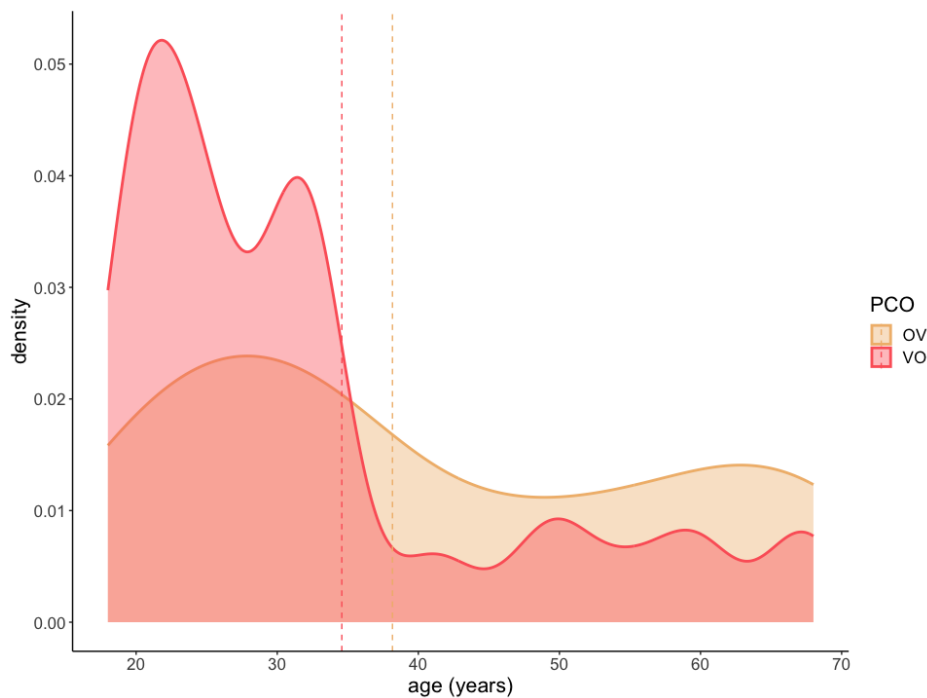


Figure 5.8: Density plot of OV/VO order across age for bilingual group

The density plot illustrates that age is not normally distributed, rather it is left-skewed. Given that the x-axis increases from left to right, this means that relatively younger participants (<40 years old) constitute a comparatively larger proportion of the data overall. It is also clear from this graph that the red overlay corresponding to VO order is left-skewed, meaning that VO order is employed more frequently among younger speakers. The yellow overlay that represents OV order, also skews to the left, but less dramatically. There is a second peak in the data around  $\approx 65$  years of age, meaning OV proportions are higher for older speakers, even in this population that over-represents young speakers. The OV density plot overlay exceeds the VO overlay after  $\approx 35$  years of age, which demonstrates that there is a positive correlation between OV order frequency and age— as age increases, so does OV order frequency.

In Table 5.13, I present the descriptive statistics of the BLP scores for both predicate constituent orders. Recall that negative BLP scores reflect Spanish dominance, positive BLP scores reflect Quechua dominance, and values near zero indicate balanced dominance (i.e., a participant is equally balanced in both languages).

	min	Q1	median	mean	Q3	max
<b>OV</b>	-68.4	-20.6	+4.45	+6.11	+28.6	+63.1
<b>VO</b>	-68.4	-29.6	-2.72	-4.13	+20.4	+63.1

Table 5.13: Descriptive statistics of BLP score for bilingual group

Like age, the range of BLP scores is the same for both orders because all participants produce both OV and VO order, with the exception of P<sub>41</sub>, whose BLP score falls within this range. However, the mean BLP score values for OV and VO orders differ by an index of **10.24**. The higher average BLP score associated with OV order indicates that OV order frequency and BLP scores are positively correlated. That is, Quechua-dominant bilinguals employ OV order more frequently than Spanish-dominant bilinguals. The opposite notion may be applied to VO order— VO order is associated with a lower BLP score, which is indicative of Spanish dominance. This is supported by the distribution of the data indicated by the quantile values. A smaller range of BLP scores corresponds to the data that falls between the median and maximum values for OV order, meaning that OV order is concentrated among higher BLP score values. Again, the opposite is true of VO order— a comparatively smaller range of BLP scores accounts for the data between the minimum and median values for VO order, meaning VO order is concentrated among lower BLP score values.

Figure 5.9 shows that, in general, the data corresponding to the BLP scores of bilingual participants, are more normally distributed than age. This means that the BLP scores of participants in general cluster around 0, and taper off toward each of the extremes. This is especially true of the red VO overlay despite the secondary peak corresponding to a score of approximately -40. The yellow OV overlay in the density plot skews more to the right, meaning OV order occurs more frequently with positive BLP scores.

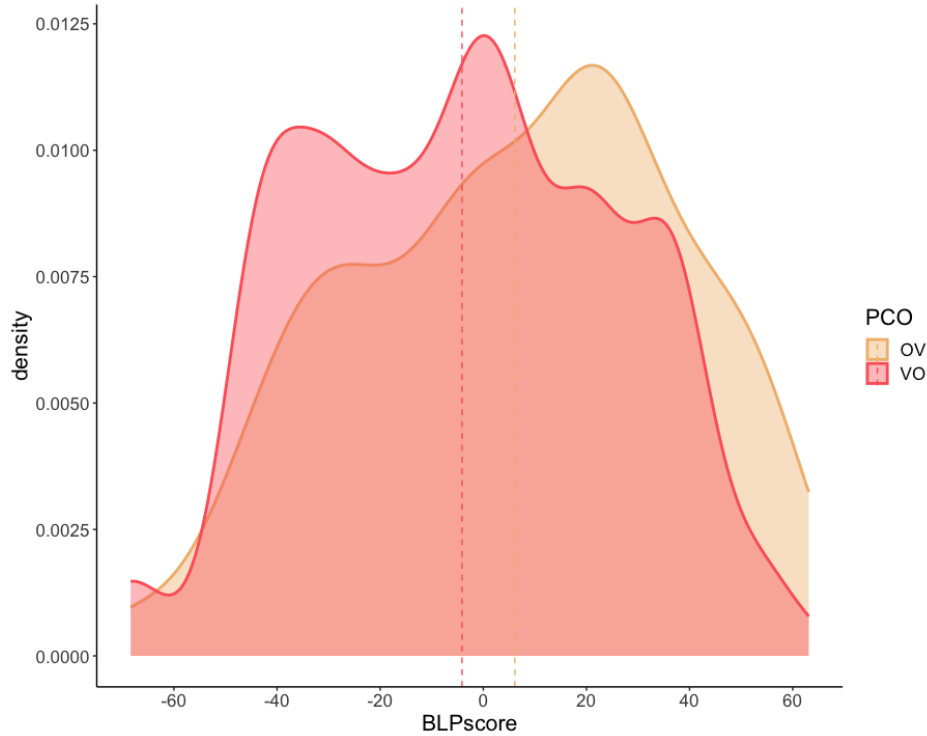


Figure 5.9: Density plot of OV/VO order across BLP score for bilingual group

To summarize, the descriptive statistics of the extralinguistic variables demonstrate that, for some variables, there is little difference in the proportion of OV/VO orders across factor levels (e.g., *sex* and *residence*). For other variables (e.g., *education* and *L1*), there is a clear proportional difference across levels. Regarding education, participants who have achieved a comparatively higher level of education employ canonical VO order more frequently than their less educated counterparts. The L1 of the participant appears to govern OV/VO order proportions such that L1 Quechua bilinguals exhibit the highest rate of OV order, followed by simultaneous bilinguals, then by L1 Spanish bilinguals. Concerning the continuous variables, *age* and *BLP score*, the mean values associated with OV order are relatively higher than those associated with VO order for both variables (i.e., older participants and participants with positive BLP scores employ OV order more frequently). In the next section, I determine which of these extralinguistic predictors significantly condition OV/VO variation in the bilingual group using a more sophisticated statistical method.

### Binomial logistic regression analysis

To enhance the descriptive statistics offered above, I provide a binomial logistic regression analysis, similar to that conducted for the monolingual group in §5.3.

For the bilingual group, I offer a separate logistic regression analysis for extralinguistic and linguistic variables for expository purposes. Furthermore, due to considerable inter-speaker variation in the bilingual group (see Figure 5.7), I conduct a mixed-effects logistic regression analysis in which participant is included as a random effect to account for inter-participant variation. All extralinguistic predictors listed above are treated as fixed effects. To identify the best-fit model, I conducted a manual step-up step-wise analysis, beginning with a model including predicate constituent order as the response variable and only participant as a random variable. From there, I built up the model one variable at a time and compared the resulting AIC values. If a variable did not lower the AIC value, it was subsequently removed from the next instantiation of the model. Recall that a lower AIC value indicates a better-fitting model. A summary of the best-fit model, which selected *age* and *BLP score* as significant predictors of predicate constituent order variation, is presented in Table 5.14.

	<b>Coefficient</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>
(Intercept)	3.09	0.29	10.85	2e-16***
age	-0.031	0.0072	-4.29	1.78e-05***
BLPscore	-0.081	0.0035	-2.31	0.021*
<b>AIC: 1733.4</b>				

Table 5.14: Results of best-fit mixed-effects model of extralinguistic predictors for bilingual group

In the best-fit model, both selected predictor variables are continuous. This means that the coefficient estimate to the right of the intercept provides the log odds of the reference level of the response variable, VO order, when age and BLP score are zero. Recall that log odds may be transformed into a probability by applying the formula  $e^{\text{logit}}/(1+e^{\text{logit}})$ , which computes a probability of VO order of 95.65% when age and BLP score are null. The coefficient estimate of a continuous variable indicates the change in log odds in the response variable for each unit increase in the predictor variable. Note that the negative sign accompanying the coefficients of each predictor variable denotes a negative correlation between VO likelihood and age and BLP score. So, as age and BLP score increase, VO likelihood decreases. Regarding age, the log odds of VO decreases by -0.031 for each unit (i.e., year) increase in age when BLP score is held constant. In terms of simple odds (i.e., odds ratios), the odds of VO order decrease by 3% (1 - 0.97) for each unit increase in age. Similarly, the odds of VO order decrease by 8% (1-0.92) for each unit increase in BLP score when age is held constant. Both predictors significantly effect predicate constituent order variation according to

their associated  $p$ -values in Table 5.14. The relationships between the probability of VO order and age and BLP score are visually illustrated in the side-by-side graphs in Table 5.15.

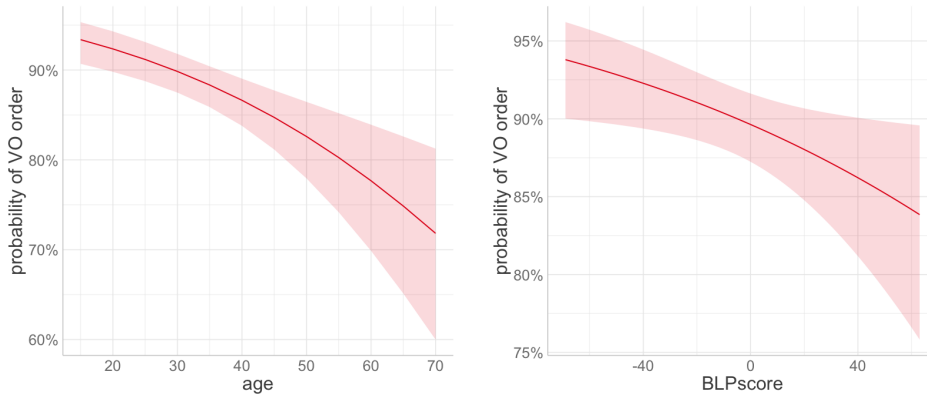


Table 5.15: Probability of VO order across age (left) and BLP score (right) for bilingual group

It is curious that **education** is not selected as a significant predictor of predicate constituent order variation, especially given the notable difference in OV/VO proportions between the primary and post-secondary participant groups. A Chi-square test between predicate constituent order and education is indeed significant,  $p < 0.05$ , but this variable is not selected as a significant predictor of variation in the dependent variable by the best-fit mixed-effects logistic regression model. Upon constructing a new model that includes education as a fixed effect, the AIC value increases, meaning this model less effectively accounts for variation in the response variable. Though *education* is not selected by the best-fit model, the trends revealed by the descriptive statistics will be further discussed in Chapter 7.

### 5.4.2 Analysis of linguistic predictors

In this section, I offer an analysis of the linguistic predictors considered in the current study: *animacy*, *definiteness*, *information status*, *polarity*, *subject expression*, and *weight* (*relative* and *absolute*). I begin by presenting the descriptive statistics of the categorical variables followed by the continuous variables. Then, I report the linguistic variables that significantly condition the distribution of predicate constituent orders in the data as determined by a binomial logistic regression model with mixed effects.

## Descriptive statistics

The raw and relative frequencies of OV and VO orders across the levels of each categorical linguistic predictor are supplied in Table 5.16 and visually represented in Table 5.17.

	OV	OV (%)	VO	VO (%)	Total ( <i>n</i> )
<b>Animacy</b>					
animate	79	16.16%	410	83.84%	489
inanimate	232	12.41%	1,637	87.59%	1,869
<b>Definiteness</b>					
pronoun	92	33.58%	182	66.42%	274
definite	113	12.70%	776	87.30%	889
indefinite	106	8.87%	1,089	91.13%	1,195
<b>Information status</b>					
given	180	21.05%	675	78.95%	855
new	131	8.72%	1,372	91.28%	1,503
<b>Polarity</b>					
affirmative	293	14.10%	1,785	85.90%	2,078
negative	18	6.43%	262	93.57%	280
<b>Subject expression</b>					
null	250	14.34%	1,493	85.66%	1,743
overt	61	9.92%	554	90.08%	615

Table 5.16: OV/VO order across linguistic variables for bilingual group

Recall that for the predictor **animacy**, the levels *human* and *animate*, were collapsed into a single category due to the small number of animate tokens ( $n = 59$ ). Note that there is little difference in the distribution of OV/VO across animate and inanimate referents according to the raw frequencies presented by Table 5.16. When the object referent is *animate*, OV is realized at only a slightly higher rate (16.16%) than when the object referent is *inanimate* (12.41%).

By contrast, there is a clear difference in the distribution of OV/VO orders across the levels of the predictor **definiteness**. Non-canonical OV order is most frequent with *pronominal* objects (33.58%), followed by *definite* objects (12.70%), then *indefinite* objects (8.87%). The large proportion of indefinite objects may be due to the following reasoning— when a referent is first introduced into the discourse, it typically appears in its indefinite form (e.g., *a woman*). After the initial introduction, a definite noun phrase (e.g., *the woman*) or pronoun (e.g., *she* or *her*), which usually contain discourse deictic particles like definite

or demonstrative articles, may be used to refer back to the discursively anchored referent; however, this is not necessary if the referent is not applicable in later discourse.

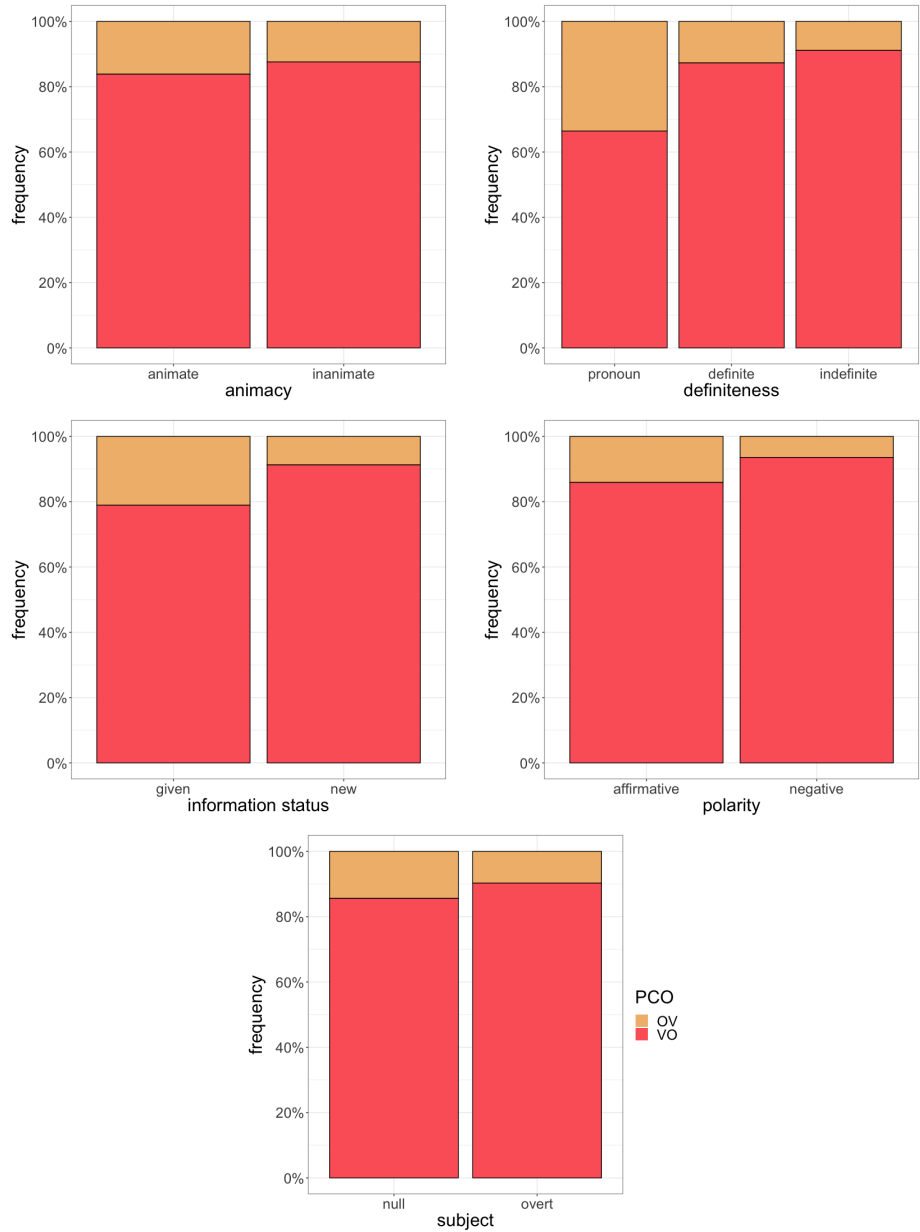


Table 5.17: Barplots of OV/VO order across linguistic variables for bilingual group

Similarly, the table demonstrates that referents with a *new information status* exceed *given* referents in general, which, again, is likely due to the lack

of common ground between interlocutors, as mentioned in the monolingual group analysis. When the object referent is given, predicate constituent order is more flexible, as 21.05% of given objects are collocated in pre-verbal position, but only 8.72% of *new* objects appear in this position.

Concerning the **polarity** of the token, the proportion of *affirmative* contexts far exceed *negative* contexts. This is simply because, impressionistically speaking, affirmative (i.e., neutral) statements are far more common in language than negative statements. Nevertheless, word order tends to be more rigid in negative contexts; that is, VO order is employed nearly categorically when the polarity of the sentence is negative (93.57%). Conversely, affirmative polarity permits a higher degree of variability in predicate constituent order.

The final predictor, **subject expression**, considers the possible syntactic effect of the presence of an additional argument. Overall, the majority of the tokens contain only one expressed argument—the accusative object. This is supported by the sizable proportion of tokens with a *null* subject (73.92%). As observed in Table 5.16, canonical VO order occurs at a high rate with *overt* subjects (90.08%), while null subjects appear to permit slightly more syntactic variation.

The descriptive statistics of the continuous linguistic predictors— **relative weight** of the predicate constituents and **absolute weight** of the object constituent— are supplied in Table 5.18. To review, for both relative weight and absolute weight, number of words is the relevant unit of measurement.

	min	Q1	median	mean	Q3	max
<b>Relative weight</b>						
OV	-2	0	0	+0.51	+1	+6
VO	-4	0	+1	+0.87	+1	+14
<b>Absolute weight of object</b>						
OV	1	1	2	1.95	2	9
VO	1	1	2	2.31	2	15

Table 5.18: Descriptive statistics of relative and absolute weight for bilingual group

Table 5.18 illustrates that, for **relative weight**, VO order is associated with a larger range of values than OV order, especially considering the maximum values. Based on relatively lower mean and median (0.51 and 0.0, respectively) values, it appears that OV order is preferred for predicate constituents of approximately equal weight. The mean and median values corresponding to VO order are slightly higher, 0.87 and 1 respectively, which may suggest that VO or-

der is slightly more frequent with relatively heavier objects. The density plot in Figure 5.10 reveals that the distribution of VO and OV order across the number of words corresponding to the relative weight of predicate constituents is nearly identical. There are small differences in the PCO overlays (e.g., the OV overlay has a slightly higher peak around 0, and the VO overlay exceeds the OV overlay after  $\approx 3$  words), but in general, the distribution is similar for both orders.

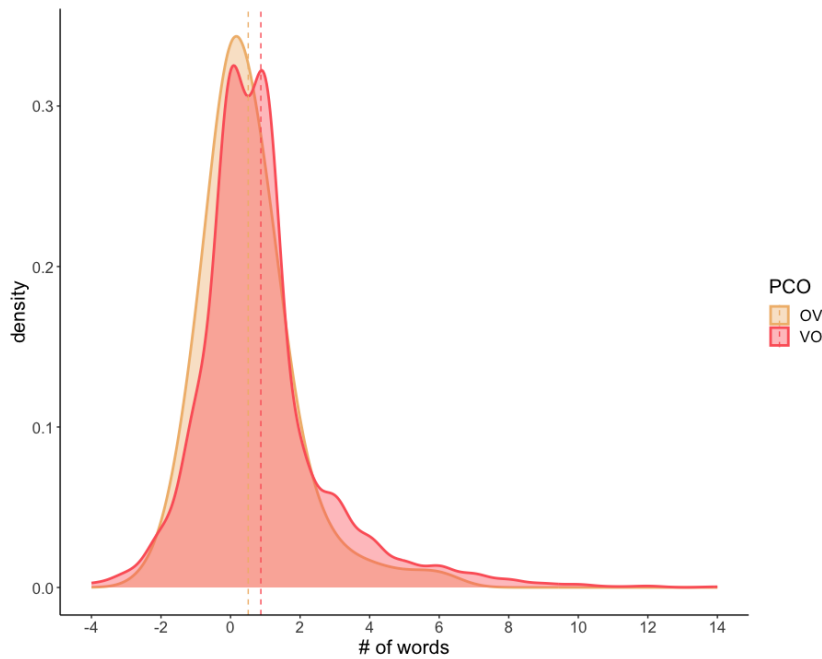


Figure 5.10: Density plot of OV/VO order across relative weight for bilingual group

The absolute weight of the object appears to follow a comparable trend. Specifically, the range of values corresponding to VO order is higher because the maximum value of VO order exceeds that of OV order. Furthermore, the mean value of OV order ( $\bar{x} = 1.95$ ) is only slightly lower than that of VO order ( $\bar{x} = 2.31$ ). The density plot in Figure 5.11 again shows that both orders have a similar distribution with respect to the number of words of the object constituent. Particularly, both orders are left-skewed, meaning, in general, the length of an object is typically four words or fewer. There are two small differences in OV/VO order distribution according to the density plot: first, the OV overlay peak is situated slightly further to the left, and, second, VO is marginally preferred with objects that exceed approximately four words.

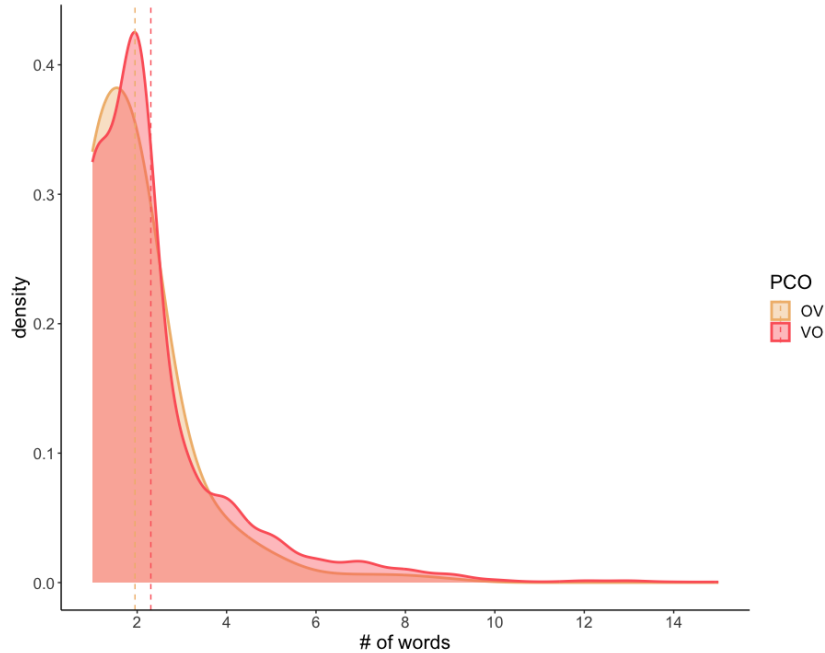


Figure 5.11: Density plot of OV/VO order across absolute weight for bilingual group

To summarize, in this section, I provided the descriptive statistics of the categorical and continuous linguistic variables considered for bilingual Andean Spanish. Several of these predictor variables do not have an observable effect on predicate constituent order variation, e.g., *animacy*, *relative weight*, and *absolute weight*. Based on the raw and relative frequencies of OV/VO orders presented in Table 5.16, there is reason to suspect that the remaining predictors—*definiteness*, *information status*, *polarity*, and *subject* expression— may affect the distribution of predicate constituent orders. Once again, in the following section, I turn to the results of a binomial logistic regression analysis of the linguistic variables, which will conclusively identify the predictors that significantly condition predicate constituent order variation in the bilingual group.

### Binomial logistic regression analysis

The process involved in determining the model the best accounts for variation in the order of predicate constituents with respect to the linguistic predictors mirrors that which was applied to the extralinguistic analysis in the previous section. I conduct a mixed-effects binomial logistic regression analysis that includes participant as a random intercept. The best-fitting model with the lowest AIC value selected the variables *definiteness*, *information status*, *polarity*, and

*subject expression* as significant predictors of OV/VO variation for the bilingual group. Consider Table 5.19:

	Coefficient	Std. Error	z value	Pr(> z )
(Intercept)	1.41	0.19	7.31	2.55e-13***
definiteness:indefinite	0.36	0.16	2.31	0.021*
definiteness:pronoun	-1.28	0.18	-6.94	3.88e-12***
info.status:new	0.81	0.14	5.79	7.02e-09***
polarity:negative	1.22	0.27	4.54	5.65e-06***
subject:overt	0.59	0.16	3.60	0.00032***

**AIC:** 1589.3

Table 5.19: Results of best-fit mixed-effects model of linguistic predictors for bilingual group

Recall that the intercept selects a reference level for each variable on the basis of alphabetical order. Thus, when the object is *given* and *definite*, the polarity of the sentence is *affirmative*, and there is no expressed subject (i.e., the subject is *null*), the log odds of VO order are **1.41**, which may be calculated as a probability value of **80.44%** using the formula presented earlier. The low *p*-values associated with each predictor variable indicates that all of the selected linguistic variables are significant predictors of variation in the response variable. The mixed-effects model is visually represented in Figure 5.12, which provides odds ratios in lieu of log odds to enhance comprehensibility of the results.

Beginning with **definiteness**, both *indefinite* and *pronominal* objects are significantly different from *definite* objects in terms of their effect on word order. Considering definite objects as the reference level, when the object is indefinite, the odds of VO order increase by 43% (1.43 - 1); however, when the object is pronominal, the odds of VO order decrease significantly by 72% (1 - 0.28). Thus, there is a negative correlation between VO likelihood and definiteness such that as definiteness increases,<sup>212</sup> VO likelihood increases. Furthermore, according to Figure 5.12 when the **information status** of the object is *new*, and all other predictor variables are held constant, the odds of VO order increases by 125% (2.25 - 1) considering *given* referents as the reference level.

When the **polarity** of the sentence is *negative*, the odds of VO order increase by 241% (3.41 - 1). An *overt subject* also increases the likelihood of VO order compared to *null* subjects, though to a smaller degree (81%). Note that these two predictor variables were also selected by the best-fit model for the monolingual group, but did not reach a significant *p*-value. For the bilingual

<sup>212</sup> Pronouns are considered ‘more definite’ than definite nouns, and definite nouns ‘more definite’ than indefinite nouns (van Bergen & de Swart, 2009, 2010).

group, however, polarity and subject expression are significant predictors of OV/VO variation.

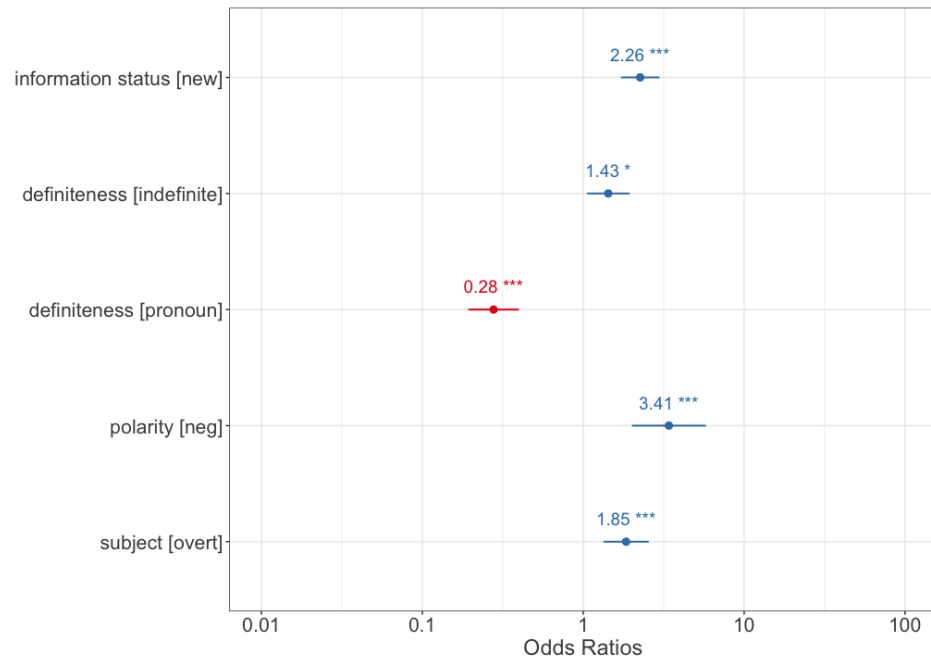


Figure 5.12: Visualization of best-fit mixed-effects model of linguistic predictors for bilingual group

Again, the interactions definiteness-polarity and definiteness-animacy were evaluated for the bilingual group. Though the interaction between the former pair of variables is associated with a significant *p-value*, the nature of the interaction is slightly different for bilinguals compared to monolinguals. As illustrated in Figure 5.13, it is still the case that *pronominal* objects in *affirmative* contexts are associated with a higher rate of OV order (40.85%) compared to pronominal objects in *negative* contexts (8.20%). However, for the bilingual group, the polarity of the token also conditions the distribution of OV/VO orders for *indefinite* objects. Specifically, OV order is more frequent among indefinite objects in affirmative contexts (9.82%) compared to negative contexts (0.79%). In fact, VO order is nearly categorical for indefinite objects in a sentence with negative polarity.<sup>213</sup> Ordering preferences among *definite* objects, on the other hand, appear to be unaffected by polarity.

<sup>213</sup> This unexpected piece of the interaction will be discussed further in Chapter 7.

Contrary to the monolingual analysis, the definiteness-animacy interaction is significant for the bilingual group. However, it is not *pronominal* objects that interact with the animacy of the referent, rather *definite* objects. Specifically, *animate* definite objects exhibit higher rates of OV order (16.93%) than *inanimate* definite objects (8.77%), as illustrated in Figure 5.14. This figure also

displays a lack of significant interactions between animacy and definiteness for *pronouns* and *indefinite* referents.

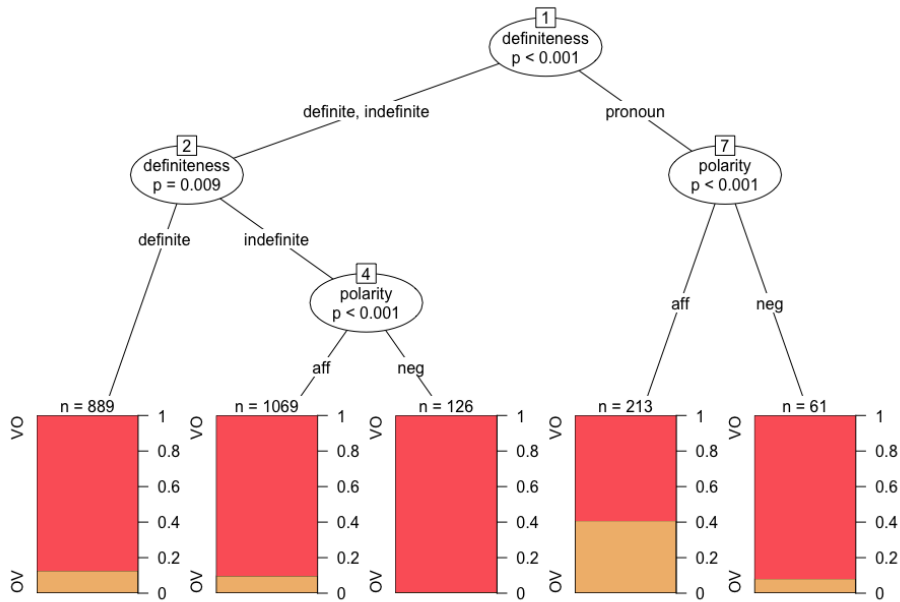


Figure 5.13: Definiteness-polarity interaction in bilingual group

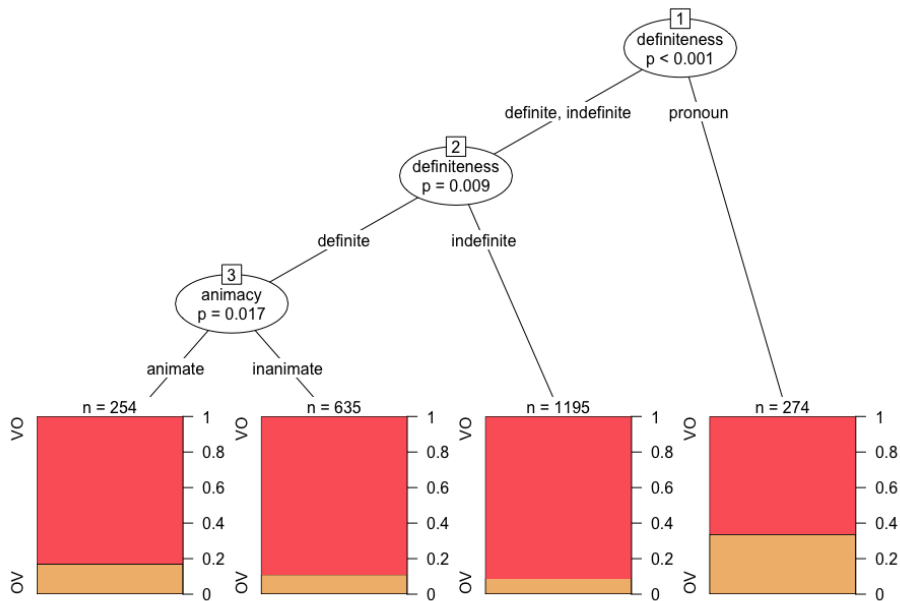


Figure 5.14: Definiteness-animacy interaction in bilingual group

To determine whether these interactions improved the fitness of the mixed-effects model, I created three new models— one including definiteness-polarity, one including definiteness-animacy, and one including both interactions— and compared their resulting AIC values. The model that considers only the interaction between definiteness and animacy does not improve the fitness of the model, but the remaining two models do yield a lower AIC value than the mixed-effects model without interactions. The model with the lowest AIC value overall, 1574.7, includes only the interaction between definiteness and polarity. Thus, in addition to the fixed effects listed above, the definiteness-polarity interaction also significantly conditions the distribution of OV/VO orders in the bilingual group.

## 5.5 Summary

The statistical analysis detailed in this chapter provides the necessary quantitative information to successfully respond to the first two research questions concerning OV/VO variation in monolingual and bilingual Andean Spanish, though a discussion of these results is reserved for Chapter 7. Overall, the rate of non-canonical OV order in the Andean Spanish corpus is 10.98%; however, bilinguals realize OV order at a significantly higher rate than monolinguals, 13.19% and 7.07% respectively. Bilingual Andean Spanish also differs from the monolingual variety in that inter-participant variation is substantially higher, with OV rates ranging from 0.0% to 44.4%

Interestingly, the same linguistic predictors condition predicate constituent order variation in monolingual and bilingual Andean Spanish, notwithstanding a few minor differences. For both varieties, definiteness, information status of the object, polarity, and subject expression are selected by the best-fit logistic regression model. Though, for the monolingual group, there is no significant difference between the factor levels corresponding to polarity and subject expression, their inclusion in their model enhances its fitness (i.e., lowers the AIC value). Furthermore, these factors condition the dependent variable in the same way— for both varieties, pronouns, given referents, affirmative polarity, and overt subjects are associated with a higher likelihood of non-canonical order. That predicate constituent order variation is conditioned in the same way by the same set of linguistic predictors in monolingual and bilingual Andean Spanish indicates that non-canonical OV order is promoted by the same linguistic circumstances in these two varieties, though the non-canonical configuration is triggered by these linguistic contexts more frequently in bilingual Andean Spanish. This observation will be discussed at length in Chapter 7.

In contrast with the linguistic factors, the extralinguistic predictors selected by the logistic regression models are not equivalent for both groups. Residence and education significantly condition OV/VO variation in the monolingual group, though I suggested that the selection of education by the model may be due to individual differences given that one factor level (*secondary*) reflects only the speech of a single participant. For the monolingual group, participants residing in metro Cusco realize lower rates of OV order than those who inhabit the areas surrounding the departmental capital. For bilingual participants, age and BLP score share an inverse correlation with VO order likelihood, meaning as the age of the participant and their level of Quechua-dominance increases, the likelihood of VO order decreases.<sup>214</sup> In Chapter 7, I argue that these findings suggest both a change in progress and contact-induced language change.

<sup>214</sup> The trend described here is not meant to be interpreted as applying at the level of the individual. In other words, I do not mean to suggest that a speaker will employ OV order more frequently as they age; rather, older speakers employ OV order more frequently than younger speakers on the aggregate. Similarly, I do not mean to suggest that an individual who decides to acquire Quechua in adulthood will subsequently employ OV order more frequently in their Spanish speech; rather, Quechua-dominant speakers as a group employ OV order more frequently than Spanish-dominant speakers.

# CHAPTER 6

## ANALYSIS: BILINGUAL CUSCO QUECHUA

### 6.1 Introduction

<sup>215</sup> For the sake of brevity, I will refer to this variety simply as ‘Quechua’ throughout the remainder of the chapter.

In this chapter, I present the results of the statistical analyses conducted on predicate constituent order variation in the bilingual Cusco Quechua<sup>215</sup> corpus. The results I offer here relate directly to the third and fourth research questions advanced in Chapter 1, which are restated below:

**RQ3:** What is the general proportion of OV/VO orders in bilingual Cusco Quechua?

**RQ4:** Which extralinguistic and linguistic factors condition OV/VO variation in bilingual Cusco Quechua?

In §6.2, I first identify the participants whose interviews were excluded from the dataset. Then, I discuss a pair of caveats concerning the envelope of variation delimited in Chapter 4. Specifically, this section provides a justification for the decision to ultimately exclude both purposive constructions and infinitive verb objects from the Quechua dataset.

The three sections that follow supply the results of various statistical analyses motivated by the above research questions. Accordingly, in §6.3 I provide the raw and relative frequencies of OV/VO orders in the Quechua dataset. Additionally, I present the individual proportions of OV/VO orders for each participant to highlight the prominence of intra- and inter-participant variation in the Quechua sample.

The following section, §6.4, is reserved for an analysis of the extralinguistic variables explored in this study. I first tabulate the proportions of OV/VO

orders across the categorical extralinguistic predictors— *sex*, *residence*, level of *education*, and *first language (L1)*. Then, I compute and report the measures of central tendency (i.e., mean and median values) and measures of spread (i.e., range and quantile values) for the continuous variables, *age* and *BLP score*. To enhance the presented descriptive statistics, I submit all extralinguistic predictor variables to a binomial logistic regression analysis with mixed effects, accounting for inter-participant variation by treating participant as a random variable. The output of this model reveals the extralinguistic predictors that significantly condition predicate constituent order variation in Quechua.

In §6.5 a parallel process is applied to the linguistic predictors— *animacy*, *definiteness*, *information status*, *polarity*, *subject expression*, *weight*, *accusative marking*, *argument type*, and *Spanish loanwords*. Note that the last three predictors listed are investigated only in the Quechua analysis, as these possible predictors of constituent order are language-specific and, thus, do not apply to the Andean Spanish analysis. I do not provide a discussion of the findings here, as the forthcoming chapter offers a comprehensive comparison of the Andean Spanish and Quechua results in accordance with the research questions.

In the final section, §6.6, I synthesize the results of the variationist statistical analyses and prepare the reader for a discussion of these results in the subsequent chapter.

## 6.2 Housekeeping

In the present study, 34 of the total 45 participants identified as bilingual speakers of Andean Spanish and Quechua. Of these 34 participants, 30 completed an interview in Quechua. Four participants— P13, P16, P40, P42— chose to complete an interview only in Andean Spanish due to either time constraints on the participant's behalf or their hesitancy to record a conversation in their non-dominant language. Additionally, two participants— P10 and P21— completed an interview in Quechua but were not ultimately included in the analysis because the former participant contributed only one token, and the audio quality of the interview with the latter impeded transcription. Thus, the Quechua interviews of **28** bilingual participants were included in the Quechua corpus.

In the Quechua dataset, there were **1,370** total tokens with an object and verb that adhered to the circumscribed envelope of variation detailed in §4.4.2. Canonical **OV** order accounted for **68.98%** ( $n = 945$ ) of the tokens, and non-canonical **VO** order accounted for the remaining **31.02%** ( $n = 425$ ).

### 6.2.1 A caveat: Purposive constructions

Whilst coding the data, I noticed a pattern concerning the relationship between morphological accusative marking (*-ta*) and a particular Quechua verbal construction—the purposive verb phrase. A purposive verb phrase conveys the reason or purpose for an action directed toward a certain destination. Structurally speaking, it is a periphrastic verbal construction comprised of a lexical verb and a directional verb. The lexical verb root is affixed with the purposive suffix, *-q*, and the directional verb, like *riy*, ‘to go’, is inflected with a subject-agreement particle, as exemplified in (III). The verb is underlined and the object is bolded.

- (III) *chay San Sebastian-pi sobrina-y tiya-rqa-n alquila-pi*  
 that San Sebastian-LOC niece-POSS.ISG live-PST-3SG rent-LOC  
*huq... wasi alquila-pi... entonces ñuqa-qa **sobrina-y***  
 one... house rent-LOC then I-TOP niece-POSS.ISG  
*visita-q* *ri-rqa-ni*  
 visit-PURP go-PST-1SG

‘my niece lived there in San Sebastian in a rented, in a rented house... so, I went to visit my niece’ [P8]

As illustrated by the above example, for the majority of tokens with a purposive verb phrase in this dataset, the object of the sentence is left unmarked for case. This is true of (III) in which the object, *sobrinay*, ‘my niece’, of the purposive verb constituent, *visitaq rirqani*, ‘went to visit’, is not accompanied by the accusative suffix, *-ta*. Upon further investigation, I found that the direct object is not marked with *-ta* in **62.07%** ( $n = 18$ ) of all instances with a purposive construction, while only **7.98%** ( $n = 107$ ) of non-purposive tokens are not marked for case. According to a Pearson’s Chi-squared test, there is a significant difference between purposive and non-purposive verb phrases regarding the presence or absence of accusative marking,  $p < 0.05$ .

The presence or absence of accusative morphological marking is included as a linguistic independent variable in the Quechua analysis to determine its possible conditioning effect on OV/VO variation. However, the relationship between accusative marking and predicate constituent order variation may be obscured by the apparent coincidence of zero *-ta* marking and purposive constructions, as evidenced by the proportions offered in the previous paragraph. That is, it may be the case that *-ta* expression correlates with predicate constituent order variation, but it is certainly the case that the presence or absence

of *-ta* is reliant on the type of construction comprising the verbal constituent (i.e., purposive versus non-purposive). Furthermore, in many purposive constructions, the direct object intervenes between the lexical verb and the inflected, directional verb, as illustrated in (112).

- (112) *huq amiga-ta riqsi-ra-ni también... eh rima-ra-n*  
 another friend-ACC meet-PST-1SG also uh speak-PST-3SG  
*quechua-ta pay-pas... entonces... nuqayku ri-ra-yku*  
 quechua-ACC she-COOR then we.EXCL go-PST-1PL.EXCL  
*café palla-q*  
 coffee gather-PURP

‘I met another friend as well... uh, she also spoke Quechua... then, we went to gather **coffee**’ [P36]

In (112), the direct object, *café*, ‘coffee’, is collocated after the inflected, directional verb, *ri-rayku* ‘[we] went’, but before the purposive verb, *pallaq*, ‘to gather’. This type of construction presents a challenge for coding the dependent variable. Should the inflected verb or the lexical verb, which licenses the object, be considered the main verb when determining the order of predicate constituents? For these two reasons, the 29 tokens containing a purposive verb phrase were eliminated from the dataset. After removing these tokens from the dataset, the proportion of OV/VO orders remained about the same— **68.68%** ( $n = 921$ ) and **31.32%** ( $n = 420$ ), respectively.

### 6.2.2 Another caveat: Infinitive verb objects

Recall that in Quechua, an infinitive verb marked with *-ta* functions as the accusative object of the verb (Sánchez, 2010). Tokens of this variety were originally included in the Quechua dataset; however, because several of the linguistic predictors evaluated in the present study did not apply to infinitive verb objects, they were ultimately removed from the dataset. For instance, tokens with an infinitive verb object could not be coded for the independent factors *animacy* and *definiteness* because none of the available factor levels applied to an infinitive verb object (i.e., it is not appropriate to classify a verb as either animate or inanimate, or definite or indefinite). Additionally, all infinitive verb objects were marked with *-ta*<sup>216</sup> and were not considered goals, so the inclusion of tokens of this type would cause an over-representation of case-marked objects and an under-representation of goal arguments.

<sup>216</sup> There is one token that presents an exception to this generalization: *estudi-aytaq qallaripuni*, ‘I began to study.’ The suffix on the infinitive verb object, *estudi-aytaq*, ‘to study’, is *-taq*, a contrastive morpheme that coordinates predicates. It could be the case that the speaker did not employ *-ta* in this token for phonological reasons, i.e., duplication of the syllable *-ta*.

Before removing infinitive verb objects from the dataset, I used a Pearson’s Chi-square test to determine whether there is a significant difference between infinitive verb objects and NP objects regarding their corresponding distributions of predicate constituent orders. As illustrated in Table 6.1, though the frequency of canonical OV order is slightly higher for infinitive verb objects compared to NP objects, this difference does not reach statistical significance,  $p = 0.26$ .

	OV	%	VO	%	Total (n)
infinitive verb	94	73.44%	34	26.56%	<b>128</b>
noun phrase (NP)	827	68.18%	386	31.82%	<b>1,213</b>

Table 6.1: OV/VO order across object type in Quechua dataset

Like purposive constructions, the non-canonical syntax of some predicate constituents with infinitive verb objects present codification challenges. As exemplified in (113), occasionally, the infinitive verb object licences an additional object, which may intervene between the infinitive verb object and inflected verb.

- (113) *Wawa-ta-taq apa-ra-n-taq wiksa-n-pi,*  
 baby-ACC-CONTR carry-PST-3SG-CONTR belly-POSS.3SG-LOC  
*paqari-ra-n binaspa anchay-pi mana ati-pu-n-chu*  
 birth-PST-3SG then that-LOC NEG able.to-REG-NEG  
*wawa-ta burqumu-y-ta binaspa... pana-y*  
 baby-ACC take.out-INF-ACC then sister-POSS.ISG  
*yaqa-lla wañu-ra-pu-pu-n-pas binaspa*  
 almost-LIM die-PST-REG-REG-3SG-COOR then

‘she carried the baby in her belly, then [she] gave birth, there she couldn’t get **the baby** out, so... then my sister almost died’ [P36]

In (113), the infinitive verb object, *burqumyta*, ‘take out’, licenses an NP object, *wawata*, ‘the baby’, which also carries an accusative suffix. This NP object intervenes between the inflected verb, *atipunchu*, ‘couldn’t’, and the infinitive verb object. Again, the question arises, should the position of the direct object be considered with respect to the inflected verb or the lexical verb? Though infinitive verb objects were removed from the dataset, the token in (113) is included in the dataset, in which *wawata* is considered the object constituent and *atipunchu burqumyta* the complex verbal constituent. The order of this token

and similar tokens is considered OV, as the lexical verb, *burqumuy*, licenses the object.

### 6.3 General results

After removing both purposive constructions and infinitive verb objects, the resulting dataset contained **1,212** instances of OV/VO orders. Canonical OV order accounted for **68.23%** ( $n = 827$ ) of the data and non-canonical VO order **31.77%** ( $n = 385$ ). All clausal constituent orders including an explicit subject are also attested in the Quechua dataset, as shown in Table 6.2.

	OV Type				VO Type			
	<i>OV</i>	<i>SOV</i>	<i>OSV</i>	<i>OVS</i>	<i>VO</i>	<i>SVO</i>	<i>VSO</i>	<i>VOS</i>
<i>n</i>	659	108	35	25	257	113	5	10
%	54.4%	8.9%	2.9%	2.1%	21.2%	9.3%	0.4%	0.8%

Table 6.2: Frequency of clausal constituent orders in Quechua dataset

It is clear from the data provided by the table that, regardless of the configuration of object and verb, null subjects are preferred over overt subjects. Overall, three-fourths (75.58%) of the tokens do not have an expressed subject. Additionally, when subjects are overtly expressed, subject-initial order is preferred for both OV- and VO-type orders over subject-internal or subject-final orders. It is also apparent that subjects are expressed more frequently in VO-type orders than OV-type orders. The subject is overtly expressed in only 20.31% of OV-type tokens, but in 33.25% of VO-type tokens. The correlation between subject expression and predicate constituent order will be further explored in the analysis of linguistic variables in §6.5.

Regarding predicate constituent order variation at the individual level, there was considerable inter-speaker variation in this sample, as illustrated in Figure 6.1. In this figure, each dot represents the rate of OV order for a single participant. The size of the dot represents the number of tokens that pertain to the corresponding participant, and the color of the dot represents the sex of the participant. For instance, the large pink dot corresponding to P19 indicates that P19 is a male participant who produced a relatively large amount of tokens ( $n = 101$ ). The rate of OV order corresponding to P19 is 61.39%. The small purple dot that represents P8, on the other hand, indicates that P8, a female participant, contributed a relatively small number of tokens ( $n = 26$ ) to the dataset. Concerning inter-speaker variation, on one end of the spectrum, P44 does not use non-canonical VO order at all in this speech sample, and on the other end

of the spectrum, P26 employs non-canonical VO order more frequently than canonical OV order, 60.00% and 40.00% respectively.

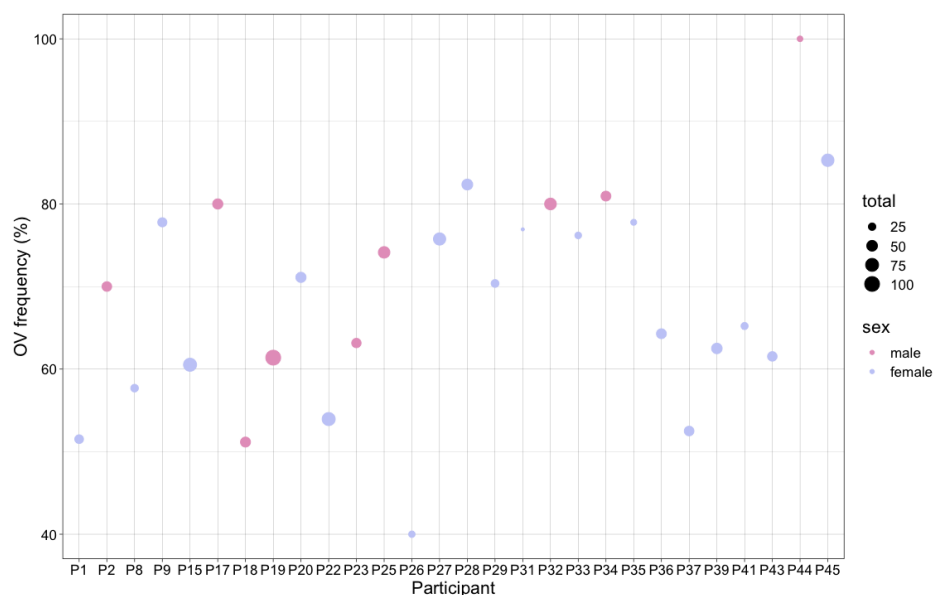


Figure 6.1: OV order rate per participants in Quechua dataset

The objective of the next two sections is to ascertain the extralinguistic and linguistic variables that condition this observed variation in predicate constituent order in the Quechua dataset.

## 6.4 Analysis of extralinguistic variables

The extralinguistic independent variables considered for the Quechua sample are identical to those evaluated for the bilingual Andean Spanish group: *sex*, *age*, *residence*, level of *education*, *first language (L1)* and *BLP score*. I begin by presenting the descriptive statistics of each variable.

### 6.4.1 Descriptive statistics

The raw and relative frequencies of OV/VO orders for each level of the categorical factors— *sex*, *residence*, *education*, and *L1*— are presented in Table 6.3 and visualized in barplots in Table 6.4. The yellow partition represents OV order and the red partition represents VO order.

	OV	OV(%)	VO	VO(%)	Total(n)
<b>Sex</b>					
Female	513	66.80%	255	33.20%	768
Male	314	70.72%	130	29.28%	444
<b>Residence</b>					
Department	446	71.36%	179	28.64%	625
Metro	381	64.91%	206	35.09%	587
<b>Education</b>					
Primary	183	78.21%	51	21.79%	234
Secondary	225	66.37%	114	33.63%	339
Post-secondary	419	65.57%	220	34.43%	639
<b>First Language (L1)</b>					
L1 Quechua	558	70.99%	228	29.01%	786
simultaneous	231	62.60%	138	37.40%	369
L1 Spanish	38	66.67%	19	33.33%	57

Table 6.3: OV/VO order across extralinguistic variables in Quechua dataset

Note that in Table 6.3, the sum of the tokens ( $n$ ) across all levels within a particular factor equal the total number of tokens in the dataset, 1,212. Beginning with **sex**, Table 6.3 reports that there is little difference in the distribution of predicate constituent orders between *male* and *female* participants. Women employ non-canonical VO order slightly more frequently than men, 33.20% and 29.28% respectively. The table also illustrates that these groups are somewhat unbalanced. Because the majority of the participants who completed a Quechua interview were women ( $n = 19$ ), women provided nearly two-thirds (63.37%) of the data. There is a greater discrepancy in the proportion of OV/VO tokens between factor levels regarding the **residence** of the participant. Participants who reside in the city of Cusco (i.e., *metro* Cusco), employ non-canonical VO order more frequently than do those who live in the *department* of Cusco, 35.09% and 28.64% respectively. Unlike the variable **sex**, residence is more balanced in terms of the overall proportions of tokens provided by metro and department participants.

It is important to note that, for the variable **education**, the factor levels *university* and *post-secondary* were conflated, as is the case for the bilingual Andean Spanish group.<sup>217</sup> Regarding the distribution of OV/VO orders across education level more generally, the factor level *post-secondary* accounts for a little more than half the data, 52.72%. Concerning the distribution of the dependent variable, participants with *primary*-level education (or lower) produce canonical

<sup>217</sup> Prior to combining these two factor levels, I observed that the distribution of predicate constituent orders across levels *university* and *post-secondary* was nearly identical— the OV rate of among participants who had attained a *university*-level education was 65.22%.

OV order more frequently, 78.21%, than their counterparts with more formal education (66.37% for participants who completed secondary education and 65.57% for participants who completed post-secondary education). Considering only the relative frequencies, there appears to be a negligible difference in predicate constituent order variation between participants with a secondary-level education and participants with a post-secondary-level education.

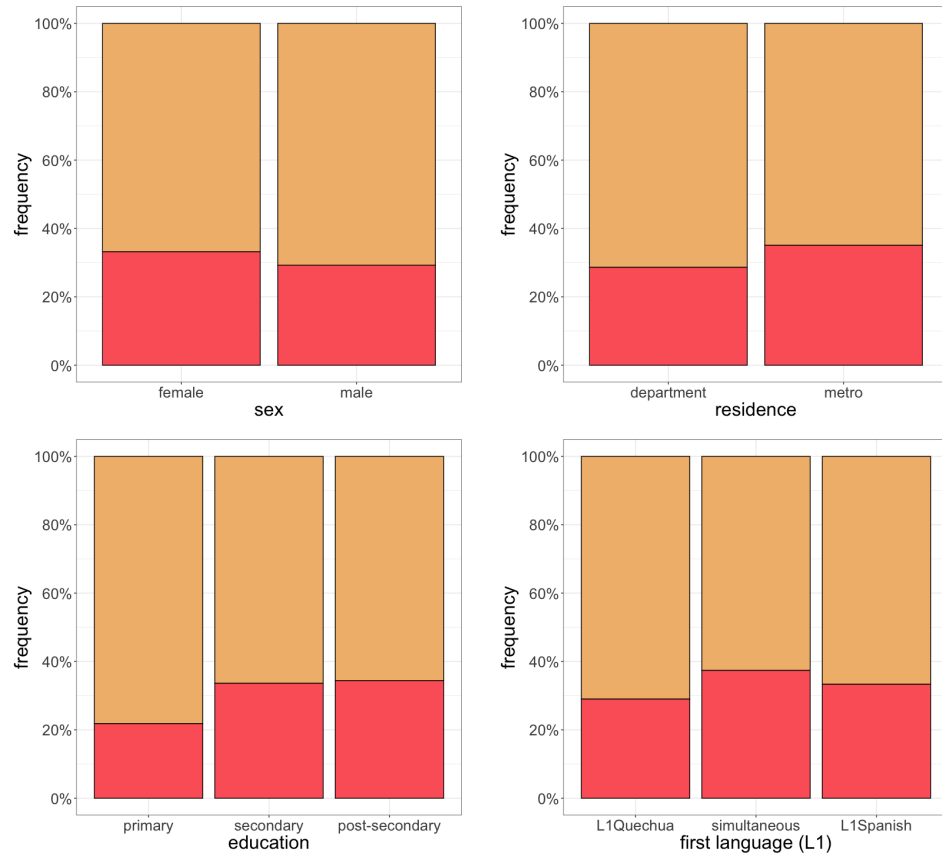


Table 6.4: Barplots of OV/VO proportions across extralinguistic variables in Quechua dataset

Recall that the **first language (L<sub>1</sub>)** of the participant was determined by comparing their respective ages of acquisition. Simultaneous bilinguals acquired both languages in tandem, L<sub>1</sub> Quechua participants acquired Quechua at a comparatively earlier age, and L<sub>1</sub> Spanish participants acquired Spanish at a comparatively earlier age. Though four bilingual participants identified as sequential L<sub>1</sub> Spanish bilinguals, two of these participants did not complete a Quechua interview. Thus, only two sequential L<sub>1</sub> Spanish bilinguals account for the data of the corresponding factor level. The majority of participants who

completed a Quechua interview identified as sequential L1 Quechua bilinguals ( $n = 18$ ), and thus, provide the majority of the tokens (64.9%). Table 6.3 illustrates that L1 Quechua bilinguals employ canonical OV order most frequently (70.99%), followed by L1 Spanish bilinguals (66.67%), then simultaneous bilinguals (62.60%). Though it is surprising that L1 Spanish bilinguals do not produce the highest proportion of non-canonical VO order, this result may be due to individual differences given the low number of participants belonging to this category.

Table 6.5 reports the descriptive statistics, including measures of central tendency and spread, for the continuous variables— age and BLP score. The values associated with age represent number of years and the values associated with BLP represent dominance scores.

	min	Q1	median	mean	Q3	max
<b>Age</b>						
OV	18	24	32	38.2	52	68
VO	18	22	30	34.6	49	67
<b>BLP score</b>						
OV	-68.4	-5.29	+4.45	+9.08	+35.8	+63.1
VO	-68.4	-6.83	+4.36	+6.52	+35.8	+63.1

Table 6.5: Descriptive statistics for continuous extralinguistic variables in Quechua dataset

Regarding, **age**, the ranges of ages for OV and VO orders are nearly identical. The maximum age is slightly higher for OV than VO order because P<sub>44</sub>, the oldest participant in the Quechua dataset, realized only OV order. Otherwise, all participants use both OV and VO orders. The mean age corresponding to OV order is 38.2 years, and the mean age of VO order is 34.6 years, a difference of 3.6 years. The density plot in Figure 6.2 visually depicts the distribution of predicate constituent orders across age. The red overlay represents the distribution of VO order across age, and the yellow overlay the distribution of OV order across age. The mean age for OV order is represented by the yellow dashed line, and the mean age for VO order is represented by the red dashed line.

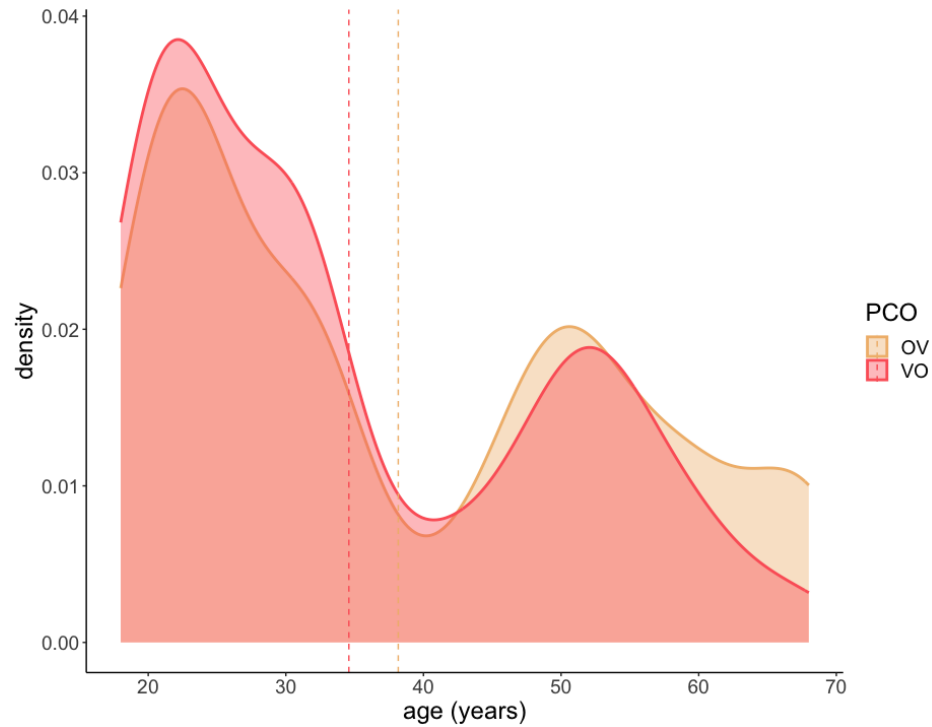


Figure 6.2: Density plot of OV/VO order across age

As visible in Figure 6.2, from 18 years of age to about 42 years of age, VO order is favored over OV order, which is illustrated by the disparity in vertical distance between the red and yellow overlays. From approximately 43 years of age to 68 years of age, OV is the preferred order, which is evidenced by the taller yellow overlay. In general, Figure 6.2 also shows that age is not normally distributed— there is a cluster of relatively younger participants (i.e., 20s to mid 30s) and a cluster of relatively older participants (i.e., mid 40s through 50s) demonstrated by the two peaks in both overlays. Comparatively, there are fairly few middle-aged (i.e., mid 30s to mid 40s) and older (i.e., 60+) participants.

Like age, the **BLP score** ranges are the same for OV and VO orders. Again, this means that both orders are used by all participants, with the exception of P44, who does not affect the range in this case because his BLP score is neither the lowest nor the highest of the participants. Recall that a negative BLP score corresponds to Spanish dominance, a positive BLP score to Quechua dominance, and a score around 0 to balanced linguistic dominance. According to the data presented in Table 6.5, the average BLP score is only **2.56** points higher for OV order than for VO order. The corresponding density plot in Figure 6.3 shows that there is no consistent pattern regarding a correlation between the dependent variable and BLP score. In the age density plot, there is a clear set of values for which VO order is favored and for which OV order is favored. Con-

versely, in the BLP score density plot, VO is preferred for the most extreme negative BLP score values, then OV is preferred for negative values between approximately -50 to -40, then VO is more frequently realized for values -40 to -5, then OV is favored for BLP scores around 0, and so on. Ostensibly, this density plot represents a lack of a linear correlation between BLP score and predicate constituent order variation, though a more sophisticated statistical analysis will confirm this observation.

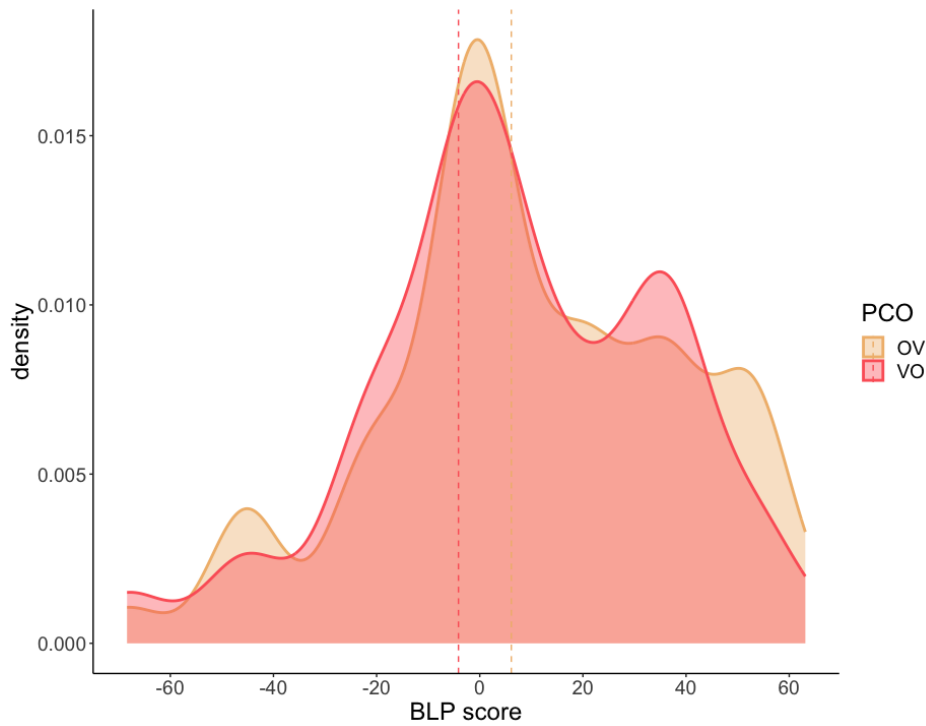


Figure 6.3: Density plot of OV/VO order across BLP score in Quechua dataset

To conclude, in this section, the descriptive statistics of the extralinguistic predictors revealed that some variables appear to condition predicate constituent order distribution, like age, and others appear to lack an effect on OV/VO variation, like sex. In the following section, I provide a binomial logistic regression analysis to enhance the above-presented descriptive statistics.

### 6.4.2 Binomial logistic regression analysis

Due to the evident inter-participant variation in the Quechua dataset (see Figure 6.1), I conducted a mixed-effects binomial logistic regression analysis that accounts for participant as a random effect. Recall from Chapter 5 that logistic regression is a type of statistical analysis that estimates the probability of a binary

response variable given various predictor variables using a maximum likelihood estimation technique (Baayen, 2008). A logistic regression analysis enhances the descriptive statistics, as it has the capacity to determine which predictor variables significantly affect variation in the response variable. I again conducted the mixed-effects logistic regression analysis in R (R Core Team, 2023) using the function *glmer()*— a function that is used to fit generalized linear models, but may also be used for logistic regression analysis when it is specified that the response variable is binary.

To arrive at the best-fit model, I first constructed a model with only the response variable and participant as a random intercept. From there, I conducted a manual step-up step-wise analysis to identify the model with the lowest AIC value. Recall that the AIC value, or *Akaike Information Criterion*, does not mean anything its own, but comparatively, the lowest AIC value accompanies the model that best accounts for variation in the data. Regarding the extralinguistic predictors, the best-fit model for the Quechua sample includes only the predictors *age* and *first language (L1)*. A summary of this model is provided in Table 6.6 below.<sup>218</sup>

<sup>218</sup> For a more detailed description of how to interpret glm and glmer output, see §5.3.2.

	Coefficient	Std. Error	z value	Pr(> z )
(Intercept)	-0.21	0.24	-0.88	0.38
age	-0.021	0.0061	-3.50	0.000466***
L1:Spanish	-0.050	0.38	-0.13	0.90
L1:simultaneous	0.61	0.19	3.18	0.0015**
<b>AIC: 1489.2</b>				

Table 6.6: Results of best-fit mixed-effects model of extralinguistic predictors in Quechua dataset

At the intercept, the selected reference level for the categorical variable first language is *L1 Quechua*,<sup>219</sup> and the reference level of the continuous variable age is null. Thus, the coefficient estimate corresponding to the intercept, -0.21, gives the log odds of VO order when the L1 is Quechua and age is zero. As stated in the previous chapter, log odds can be transformed into a probability by applying the formula  $e^{\text{logit}} / (1 + e^{\text{logit}})$ . Thus, a coefficient of -0.21 means that the probability of VO order at the intercept is 44.77%.

<sup>219</sup> Reference levels are determined by alphabetical order.

The model summary in Table 6.6 indicates that age is a significant predictor of predicate constituent order variation ( $p < 0.05$ ). The negative coefficient corresponding to age indicates that for each unit increase in age, the log odds of VO order decrease by -0.021, assuming the L1 of the participant is held constant. In simple odds terms, there is a 2% ( $1 - 0.98$ ) decrease in the odds of VO order

for each unit increase in age. Figure 6.4 offers a visual representation of the correlation between age and the probability of VO order.

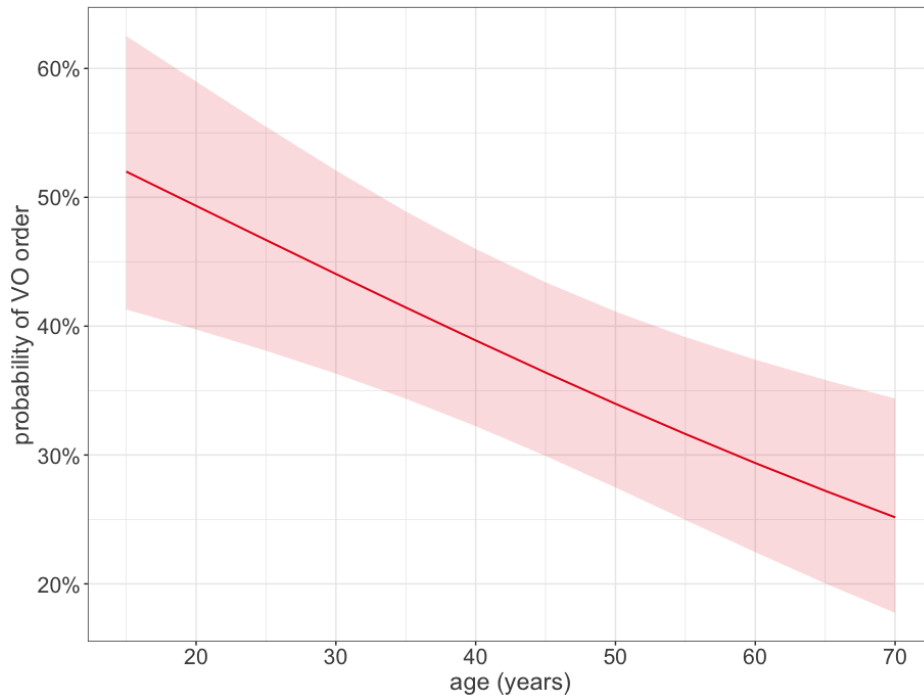


Figure 6.4: Probability of VO order across age in Quechua dataset

The *L1* of the speaker is also a significant predictor of VO order with an important caveat: there is a significant difference between *L1* Quechua and simultaneous (*2L1*) bilinguals, but not *L1* Quechua and *L1* Spanish bilinguals. When the participant is a simultaneous bilingual, the odds of VO order increase by 84% ( $1.84 - 1$ ), considering *L1* Quechua bilinguals as the reference level. The odds of VO order decrease by 5% ( $1 - 0.95$ ) for *L1* Spanish bilinguals, but, again, the difference between these two groups is not significant,  $p = 0.90$ . A visualization of the effect of first language on the probability of VO order is provided in Figure 6.4.

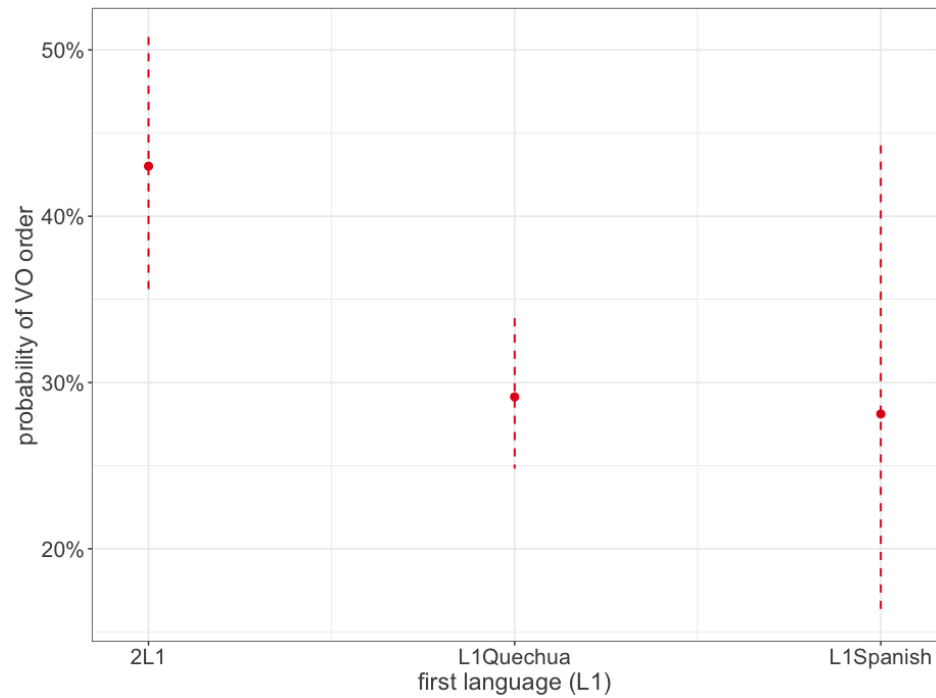


Figure 6.5: Probability of VO order across first language (L1) in Quechua dataset

In the following section, I investigate the effect of the linguistic predictors on predicate constituent order variation in Quechua.

## 6.5 Analysis of linguistic variables

The linguistic independent variables considered for the Quechua analysis include *animacy*, *definiteness*, *information status*, *polarity*, *subject*, *weight*, *accusative marking*, *argument type*, and *Spanish loanwords*. The last three variables listed are language-specific, meaning they apply only to the Quechua dataset. I begin with the descriptive statistics of the categorical variables.

### 6.5.1 Descriptive statistics

Table 6.7 presents the raw and relative frequencies of OV/VO orders across each factor level of the categorical predictors (i.e., all linguistic predictors save weight). Table 6.8 visualizes the data in the table using barplots. Once again, yellow partitions correspond to OV order and red partitions correspond to VO order.

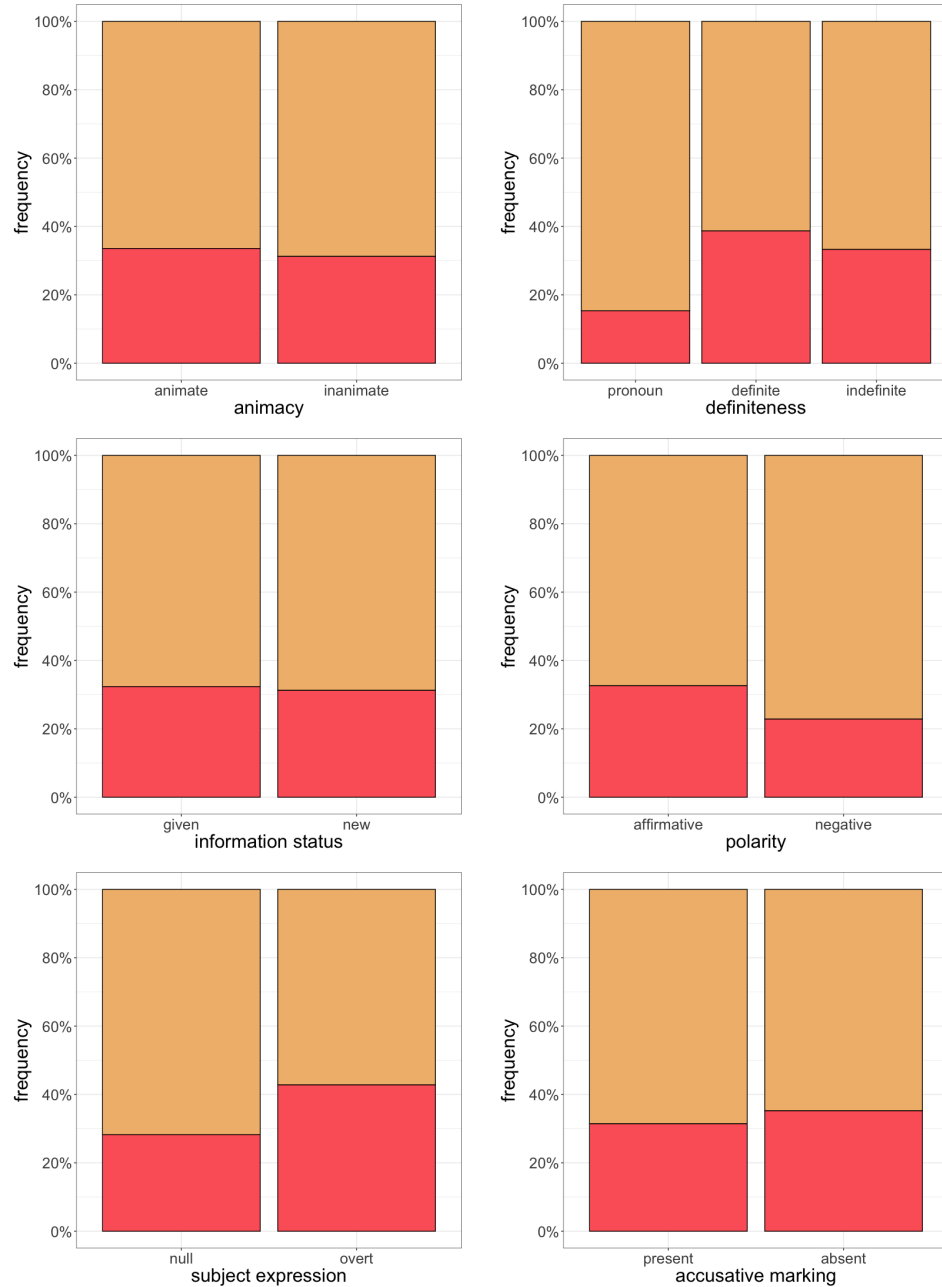
	OV	OV(%)	VO	VO(%)	Total( <i>n</i> )
<b>Animacy</b>					
animate	179	66.30%	91	33.70%	270
inanimate	648	68.79%	294	31.21%	942
<b>Definiteness</b>					
pronoun	182	84.65%	33	15.35%	215
definite	228	61.29%	144	38.71%	372
indefinite	417	66.72%	208	33.28%	625
<b>Information status</b>					
given	374	67.63%	179	32.37%	553
new	453	68.74%	206	31.26%	659
<b>Polarity</b>					
affirmative	746	67.39%	361	32.61%	1107
negative	81	77.14%	24	22.86%	105
<b>Subject</b>					
overt	167	57.19%	125	42.81%	292
null	660	71.74%	260	28.26%	920
<b>Accusative marking</b>					
present	759	68.56%	348	31.44%	1107
absent	68	64.76%	37	35.24%	105
<b>Argument type</b>					
goal	64	60.95%	41	39.05%	105
non-goal	763	68.93%	344	31.07%	1107
<b>Spanish loanwords</b>					
present	329	63.51%	189	36.49%	518
absent	498	71.76%	196	28.24%	694

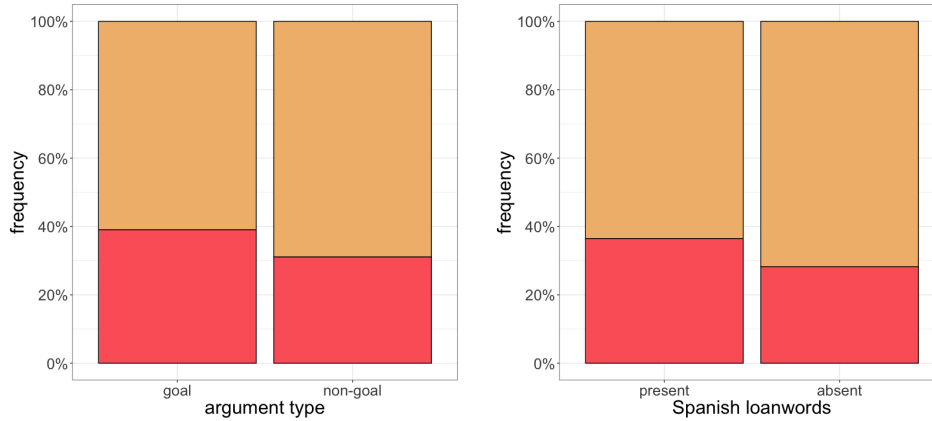
Table 6.7: OV/VO order across linguistic variables in Quechua dataset

Beginning with **animacy**, recall that the factor levels *animate* and *human* were combined due to the small number of animate tokens ( $n = 63$ ). In spite of this effort to balance the groups, the levels of this variable remain unbalanced, as inanimate entities account for over three-fourths (77.72%) of the data. On the surface, it appears that there is very little difference in OV/VO proportions for animate and inanimate objects. This pattern is repeated for the **information status** of the object, as there is very little difference in predicate constituent order distribution across factor levels. In fact, the rate of OV order is virtually identical for *new* and *given* referents, 68.74% and 67.63% respectively. Because the

levels are nearly balanced, it is unlikely that the similar distribution of OV/VO orders across information status is due to unequal groups.

Table 6.8: Barplots of OV/VO order across linguistic variables in Quechua dataset





Contrarily, there is an observable relationship between **definiteness** and predicate constituent order variation in the descriptive statistics. Canonical OV order is highly favored with pronominal objects (84.65%) compared to definite and indefinite nominal objects. Definite objects exhibit the highest proportion of VO order (38.71%), followed closely by indefinite objects (33.28%). Analogously, the data in Table 6.7 suggest a correlation between **polarity** and constituent ordering. Sentences containing a negator adhere more rigidly to canonical OV order (77.14%) than do sentences lacking a negator (67.39%). In other words, non-canonical VO order is exhibited more frequently in affirmative sentences than in negative sentences. **Subject** expression also appears to govern word order. As mentioned earlier in discussing Table 6.2, the proportion of OV order is lower when there is an *overt* subject compared to a *null* subject. Like polarity, subject expression factor levels are also unbalanced; null subjects are observed three times more frequently than overt subjects, 75.91% and 24.09% respectively.

Concerning the factors relevant only for the Quechua analysis, **accusative marking** is relatively robust in this sample, as only 8.66% of all accusative objects lack case marking. According to the raw data presented in Table 6.7, there is very little difference in OV/VO proportions between tokens with objects that exhibit accusative marking and those lacking accusative marking. Specifically, objects affixed with *-ta* appear in pre-verbal position in 68.56% of the tokens, and objects lacking *-ta* appear in the same position in 64.76% of the tokens. Regarding **argument type**, a fairly small proportion of the tokens, 8.66%, are considered goals. Nevertheless, it appears that goal arguments are realized in non-canonical VO order slightly more frequently than non-goal objects, 39.05% and 31.07% respectively, which exhibit a higher rate of OV order. Likewise, the presence of a Spanish loanword seems to govern predicate constituent order variation. When a Spanish loanword is present, the proportion of VO order—

the canonical order of predicate constituents in Spanish— is higher (36.49% versus 28.24%).

The continuous variable, **weight**, is evaluated as two separate predictors. To review, in the Quechua dataset, the *relative weight* of predicate constituents is computed by subtracting the number of syllables of the verb constituent from the number of syllables of the object constituent. Hence, positive values represent predicate constituents containing an object that is relatively heavier than the verb, and negative values represent predicate constituents for which the verb outweighs the object. The *absolute weight* of the object is calculated by simply tallying the total number of syllables of the object constituent. Table 6.9 presents the measures of central tendencies and spread— mean, median, range, and quantile values— for the relative weight of predicate constituents and the absolute weight of the object constituent.

	min	Q1	median	mean	Q3	max
<b>Relative weight</b>						
OV	-6	-1	0	+0.31	+1	+17
VO	-6	-1	+1	+1.12	+2	+20
<b>Absolute weight of object</b>						
OV	1	3	4	4.29	5	20
VO	2	3	4	5.09	6	23

Table 6.9: Descriptive statistics of weight for Quechua dataset

Concerning **relative weight**, the ranges corresponding to each order are only marginally different. Though the minimum value for both orders is -6 syllables (i.e., the verb is six syllables heavier than the object), the maximum value for OV order is +17 syllables (i.e., the object is 17 syllables heavier than the verb) and the maximum value for VO order is slightly higher, +20 syllables. On average, the mean relative weight for OV order, 0.31 syllables, is slightly lower than the VO order mean, 1.12.

The descriptive statistics for the **absolute weight** of the object closely mirror those corresponding to the relative weight. The primary difference is that the minimum weight is not equivalent for both orders— the minimum weight is 2 syllables for VO order and 1 syllable for OV order.<sup>220</sup> Likewise, the maximum value for VO order is higher than the maximum value for OV order, 23 and 20 syllables respectively. Thus, the range of absolute weight values for VO order is slightly larger than the range of values corresponding to OV order. Additionally, the mean number of syllables for VO order is 0.8 syllables heavier than the mean corresponding to OV order, 5.09 and 4.29 respectively.

<sup>220</sup> An absolute weight of 1 syllable occurs in only 7 tokens, all of which are necessarily unmarked for case, as the suffix *-ta* is included in the syllable count. Six of these objects are *chay*, ‘that’, and one is *huq*, ‘another’.

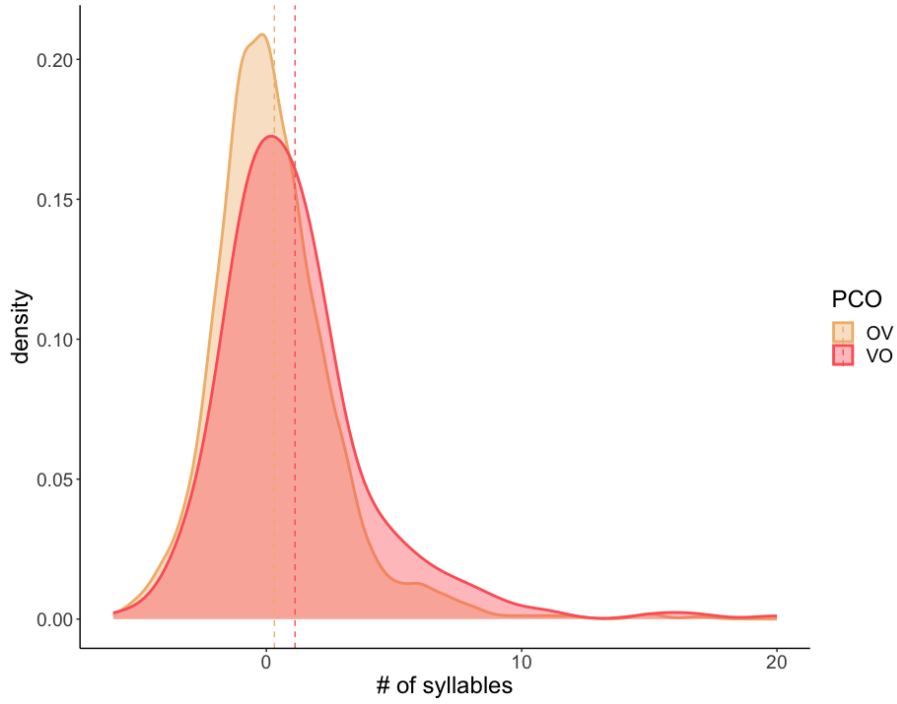


Figure 6.6: Density plot of OV/VO order across **relative** weight for Quechua dataset

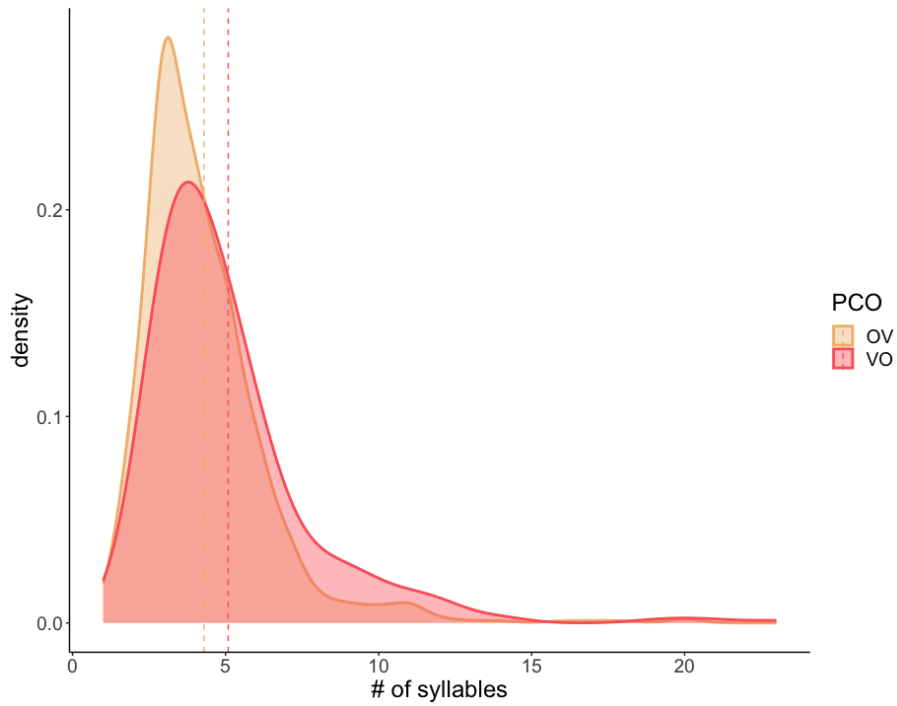


Figure 6.7: Density plot of OV/VO order across **absolute** weight for Quechua dataset

The above density plots in Figure 6.6 and Figure 6.7 provide visual evidence of a similar distribution of OV/VO orders for both the relative and absolute weights. Both plots are skewed to the left, meaning the data is concentrated around relatively lower syllable counts. In both plots, OV predominates among smaller values. (These values correspond to approximately -2 to 1 syllables for relative weight and 2 to 4 syllables for absolute weight.) In the relative weight density plot, VO order is preferred for relative weights between 1 and 12 syllables. For relative weights above 12 syllables, there is no clear ordering preference due to the relative infrequency of these types of tokens. In the absolute weight density plot, between approximately 4 and 5 syllables, OV and VO orders are favored almost equally. Figure 6.7 shows that, when the syllable count of an object surpasses 5 syllables, a preference for VO order emerges. As is the case with relative weight, objects with the highest number of syllables (>15) do not have a clear ordering preference.

In sum, the incidence of VO order increases as the relative weight of predicate constituents and absolute weight of the object increases. Based on Table 6.9, Figure 6.6, and Figure 6.7, there is a positive, direct relationship between VO likelihood and weight.

### 6.5.2 Binomial logistic regression analysis

The linguistic predictors summarized in Table 6.7 and Table 6.9 were submitted to a binomial logistic regression analysis in a similar fashion as the extralinguistic predictors. Due to the high degree of inter-speaker variation, I used a mixed-effects model that accounts for *participant* as a random effect. The summary of the best-fit model in Table 6.10 shows that the variables *definiteness*, *information status*, *polarity*, *subject expression*, *argument type*, and *absolute weight* were selected as predictors of predicate constituent order variation. Again, the AIC values were compared to determine fitness throughout the process of applying a manual step-wise analysis.

The intercept of the model selected the following reference levels: definiteness=definite, information status=given, subject=null, polarity=affirmative, argument=non goal, and absolute weight=null. At the intercept, the log odds of VO order is -0.67, which converts to a probability of 33.85% using the formula  $e^{\text{logit}}/(1+e^{\text{logit}})$ . In Figure 6.8, the coefficient estimates for each factor level are transformed into an odds ratio to facilitate interpretation of the results. Red values indicate a negative change in the odds of VO order (i.e., a decrease in odds) and blue values represent a positive change in odds (i.e., an increase in odds) with respect to the intercept. Significant odds ratio values are accompanied by asterisks.

	Coefficients	Std. Error	z value	Pr(> z )
(Intercept)	-0.67	0.28	-2.37	0.0178*
definiteness:indefinite	-0.18	0.15	-1.21	0.22
definiteness:pronoun	-1.05	0.23	-4.53	6.04e-06***
info.status:new	-0.38	0.14	-2.67	0.00757**
polarity:negative	-0.59	0.26	-2.26	0.0240*
subject:overt	0.67	0.15	4.47	7.85e-06***
argument:non-goal	-0.41	0.23	-1.82	0.069
absolute.weight	0.12	0.029	4.21	2.57e-05***

AIC: 1430.1

Table 6.10: Results of best-fit mixed-effects model of linguistic predictors for Quechua dataset

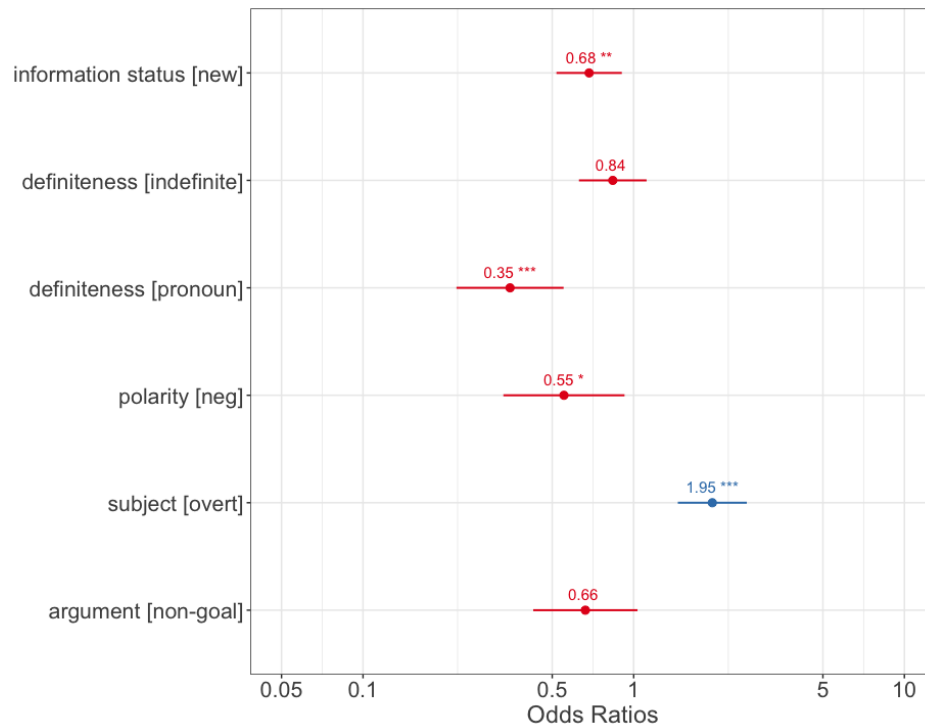


Figure 6.8: Visualization of best-fit mixed-effects model of linguistic predictors in Quechua dataset

Beginning with **definiteness**, Figure 6.8 clarifies that there is no significant difference in the odds of VO order between *definite* and *indefinite* objects ( $p = 0.22$ ); however, there is a significant difference between definite objects and *pronominal* objects. The odds of VO order decrease by **65%** ( $1 - 0.35$ ) when

the object is pronominal. Thus, there is a notable preference for pronominal objects to be collocated in pre-verbal position.

Though in Table 6.7 there is very little difference between new and given objects regarding their respective distribution of OV/VO order, when participant is considered as a random variable, **information status** is selected as a significant predictor of predicate constituent order variation. The negative coefficient accompanying the level *new* indicates that, when all other predictor variables are held constant, the odds of VO order decrease by 32% (1- 0.68) for *new* referents. Based on the *p-value* associated with this trend ( $p = 0.0076$ ), there is a significant difference between *given* and *new* referents. If one were to evaluate the effect of information status on predicate constituent order solely on basis of the raw and relative frequencies provided in Table 6.7, it would be reasonable to conclude that a statistical relationship between the dependent variable and information status is lacking. This finding highlights the importance of employing statistical modeling to understand the conditioning effect of predictor variables on the response variable.

When the **polarity** of a sentence is *negative*, the odds of VO order decrease by 45%. Based on the *p-value* ( $p = 0.024$ ), there is a significant difference between affirmative and negative contexts. According to the model, **subject** expression is also a significant predictor of VO order, though the change in odds represented by the coefficient is positive. That is, when the subject is *overt*, the odds of VO order increase by 95%. So, the likelihood of non-canonical order is higher for predicates with an expressed subject than for predicates with a *null* subject.

Lastly, the predictor **argument type** was selected by the model because its inclusion enhances the fitness (i.e., lowers the AIC value). In fact, when this predictor is removed from the model, the AIC value increases by more than a point to 1431.2. However, the presence or absence of a goal argument is not considered a significant predictor of predicate constituent order variation,  $p = 0.075$ . Nevertheless, the negative coefficient accompanying this predictor indicates that the odds of VO order decreases by 34% (1 - 0.66) when the object of the predicate is a *non-goal* argument as opposed to a *goal* argument.

Because the **absolute weight** of the object is a continuous variable, a visual depiction of the relationship between absolute weight and predicate constituent order variation is offered separately in Figure 6.9. Note that the y-axis measures the probability of VO order in lieu of the odds. The positive coefficient corresponding to absolute weight may be interpreted as follows: for each unit (i.e., syllable) increase in the absolute weight of the object constituent, the odds of VO order also increase by 12% (1-1.12). In other words, heavier objects have a higher likelihood of post-verbal collocation than do lighter objects. Figure 6.9

illustrates the positive correlation between VO order and number of syllables of the object.

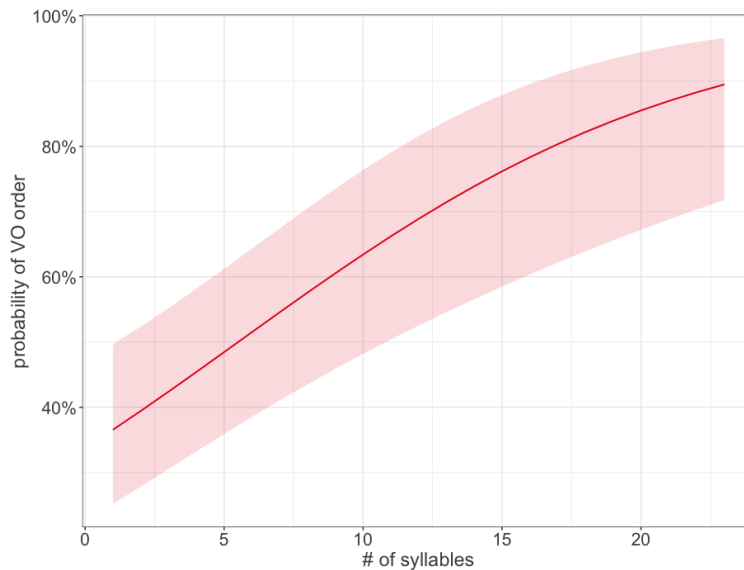


Figure 6.9: Probability of VO order across absolute weight in Quechua dataset

Finally, to ensure that the Andean Spanish and Quechua analyses are comparable, I added two interactions to the model to determine their effect on predicate constituent order variation in Quechua: definiteness-animacy and definiteness-polarity. Contrary to the Andean Spanish analysis, neither of these interactions improve the best-fit model, either individually or jointly.

## 6.6 Summary

This chapter has supplied a quantitative analysis of predicate constituent order variation in bilingual Cusco Quechua based on the speech of 28 bilinguals. Prior to conducting a statistical analysis, I removed purposive constructions from the dataset due to their correlation with *ta*-omission. Furthermore, I removed infinitive verb objects due to the codification challenges accompanying these tokens, particularly the lack of relevancy of many linguistic predictors considered in analysis (i.e., a verbal object cannot be classified as animate or inanimate). After eliminating purposive constructions and infinitive verb objects from the Quechua dataset, I reported that canonical OV order is present in 68.23% of the evaluated predicates and non-canonical VO order in 31.77%. As is the case for the bilingual Andean Spanish group, inter-participant variation

in Quechua is quite high, with the rate of canonical OV order ranging from 40.0% to 100.0%.

The extralinguistic variables selected as significant predictors of OV/VO variation by the logistic regression model include age and first language. Parallel to the relationship between age and OV/VO orders reported for Andean Spanish bilinguals, there is an inverse correlation between age and VO order likelihood— as the age of the speaker increases, the likelihood of VO order decreases. Again, I use this finding in the forthcoming chapter to argue that bilingual Cusco Quechua is currently experiencing a syntactic change in progress. The effect of the first language of the speaker is such that simultaneous bilinguals are more likely to employ VO order than L1 Quechua bilinguals. The lack of significant difference between L1 Quechua and L1 Spanish participants is attributed to the unbalancedness of the groups, considering there are only two participants who identify as L1 Spanish speakers.

Regarding the linguistic predictors, definiteness, information status, polarity, subject expression, and absolute weight significantly condition predicate constituent order variation in bilingual Cusco Quechua, and the inclusion of the argument type enhances the fitness of the model. In general, canonical OV order is associated with new referents, pronominal objects, negative polarity, null subjects, non-goal arguments, and light objects. Two of these predictors— weight and goal arguments— exercise an effect on the dependent variable in only the bilingual Cusco Quechua dataset. The remaining predictors (definiteness, information status, polarity, and subject expression) also condition OV/VO variation in monolingual and bilingual Andean Spanish; however, polarity and information status affect predicate constituent order variation oppositely in Quechua. In discussing the fifth research question in the following chapter, I propose that polarity and information status of the object exert language-internal effects on syntactic variation. Furthermore, I identify convergent structures associated with the two predictors that condition OV/VO variation in the same way in Quechua and Andean Spanish— definiteness and subject expression.

# CHAPTER 7

## DISCUSSION

### 7.1 Introduction

The objective of Chapter 7 is to discuss the results of the statistical analyses on predicate constituent order variation in Andean Spanish (Chapter 5) and bilingual Cusco Quechua (Chapter 5) in light of the research questions and hypotheses posed in Chapter 4. This chapter is partitioned into five sections that correspond to each research question in chronological order. Each research question is repeated at the beginning of the section to facilitate a discussion of the results. I conclude the chapter in §7.7 by summarizing the discussed results.

### 7.2 Research Question 1

The first two research questions inquire about predicate constituent order variation in the Andean Spanish dataset. The first research question is repeated below:

**RQ1:** What is the overall proportion of OV/VO orders in Andean Spanish? Do monolingual and bilingual speakers differ in their distribution of predicate constituent orders?

Considering all 44 participants who completed a sociolinguistic interview in Andean Spanish, OV order accounted for **11.0%** ( $n=405$ ) of all predicates and VO order the remaining **89.0%** ( $n=3,282$ ). As predicted, Andean Spanish bilinguals employ non-canonical OV order more frequently than monolinguals. The distribution of OV and VO orders across bilingual and monolingual Andean Spanish groups originally reported in Chapter 5 is repeated below in Table 7.1 for convenience.

PCO	Monolingual		Bilingual		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
OV	94	7.1%	311	13.2%	405	11.0%
VO	1,235	92.9%	2,047	86.8%	3,282	89.0%
<b>Total</b>	<b>1,329</b>	<b>100%</b>	<b>2,358</b>	<b>100%</b>	<b>3,687</b>	<b>100%</b>

Table 7.1: OV/VO order across Andean Spanish groups

Table 7.1 shows that bilingual speakers of Andean Spanish employ non-canonical OV order nearly twice as frequently as monolingual speakers, 13.2% and 7.1% respectively. Incidentally, these disparate OV/VO proportions across the language profile of the speaker are significantly different. This may suggest that the linguistic outcomes of language contact are more pervasive the level of the bilingual individual or within the bilingual speech community. To elaborate, bilingual individuals have both clausal word orders available to them as a result of their being native speakers of two languages with mirror-image word order typologies. Additionally, the social networks to which these individuals belong likely include other bilingual individuals, which may have the linguistic effect of maintaining features of Andean Spanish, like pre-verbal objects.<sup>221</sup>

<sup>221</sup> This reasoning is used by Klee et al. (2011) to account for the maintenance of Andean features in the speech of male Andean migrants after relocating to Lima.

Recall that monolingual and bilingual groups also differ regarding the degree of inter-participant variation within them. Though not every monolingual participant employs OV order at the same rate, the range of proportions only fluctuates between 1% and 13% across all speakers. For bilingual participants, however, this range is substantially larger— 0% to 44%. Thus, monolinguals are more homogeneous in terms of their linguistic behavior than bilinguals, at least regarding the relative order of predicate constituents. It is reasonable to presume that the differing OV/VO order proportions offered by monolinguals and bilinguals is partially a reflection of their separate speech communities, which reinforce the respective linguistic norms of each group. For bilinguals, the target vernacular has likely been established through a process of L2 group acquisition in which learner features (e.g., increased OV order) are adopted as non-standard features that characterize the regional contact variety (Muysken, 1984). In other words, the vernacular target for bilinguals, especially those who have strong social associations with other bilingual speakers, is not a non-contact variety exhibiting more standard features (e.g., Limeño Spanish), rather a regional variety that has been established within their speech communities and reflects the widespread bilingualism therein. Crucially, the vernacular target will vary more widely both within and across bilingual speech communities as a function of the types of speakers that comprise them. This is supported by the various responses provided by bilinguals on the Language Background Questionnaire

regarding their current language use. For instance, P33 indicates that she uses only Quechua in the workplace and with family and friends; contrarily, P16 uses primarily Spanish in these domains, though she also reports communicating in Quechua slightly more frequently with her family members.

To situate findings presented in Table 7.1 within the context of the previous literature on this topic, the OV/VO proportions reported for **bilingual** participants in the present study are compared to those reported by previous quantitative studies on word order in bilingual Andean Spanish<sup>222</sup> in Table 7.2 below.<sup>223</sup> The table is organized by descending order of OV proportions. The results of the present study are supplied in the shaded row.

<sup>222</sup> The findings presented in Luján et al. (1984) are not included in the table because the participants of this study are children in the process of acquiring Spanish.

<sup>223</sup> Recall that Klee (1996) presents results for two sets of studies— one in which she applies the methodology provided by Muysken (1984) and the other in which she applies the methodology of F. A. Ocampo and Klee (1995). The results of these studies are referred to in the table as Klee (1996)–M and Klee (1996)–O&K respectively.

Demographics	OV (%)	VO (%)	Study
lower group	37.6%	62.4%	F. A. Ocampo and Klee (1995)
lower group	36.3%	63.7%	Klee (1996)–O&K
Quechua-dominant	34.9%	65.1%	Muysken (1984)
incipient bilinguals	32.7%	67.3%	Muysken (1984)
lower group	22.6%	77.4%	Klee (1996)–M
sequential bilinguals	19.9%	80.1%	Muntendam (2009)
indigenous L1	18.9%	81.1%	Klee et al. (2011)
middle group	17.9%	82.1%	Klee (1996)–M
simultaneous bilinguals	17.4%	82.6%	Muntendam (2009)
professionals	15.2%	84.8%	Klee (1996)–M
<b>bilinguals</b>	<b>13.2%</b>	<b>86.8%</b>	<b>present study</b>
professionals	12.9%	87.1%	F. A. Ocampo and Klee (1995)
Spanish-dominant	8.3%	91.7%	Muysken (1984)
professionals	8.1%	91.9%	Klee (1996)–O&K
middle group	4.8%	95.2%	Klee (1996)–O&K

Table 7.2: OV/VO order in Andean Spanish word order studies

The average OV/VO proportions for each study are provided in Table 7.3 to better facilitate cross-study comparison, especially considering that the demographic groups listed in the second column of Table 7.2 are defined by different clusters of characteristics across studies (e.g., Muysken (1984) forms groups based on linguistic dominance and Muntendam (2009) separates groups based on region and bilingual type).

<b>Region</b>	<b>OV (%)</b>	<b>VO (%)</b>	<b>Study</b>
Calca, Peru	28.3%	71.7%	F. A. Ocampo and Klee (1995)
Ecuador	19.6%	80.4%	Muysken (1984)
Lima, Peru	18.9%	81.1%	Klee et al. (2011)
Ecuador/Bolivia	18.5%	81.5%	Muntendam (2009)
Calca, Peru	18.3%	81.7%	Klee (1996)–M
Calca, Peru	17.3%	82.7%	Klee (1996)–O&K
<b>Cusco, Peru</b>	<b>13.2%</b>	<b>86.8%</b>	<b>present study</b>

Table 7.3: Average OV/VO order in Andean Spanish word order studies

According to both Table 7.2 and Table 7.3, in comparison to previous studies on predicate constituent order variation, the rate of OV order among bilingual speakers of the present study is relatively low, contrary to the advanced hypothesis in Chapter 4. Table 7.2 illustrates that the highest rates of OV order correspond to individuals that belong to a low socioeconomic stratum (Klee, 1996; F. A. Ocampo & Klee, 1995) and to incipient/Quechua-dominant bilinguals (Muysken, 1984). The lowest rates of OV order are associated with individuals belonging to a relatively high socioeconomic stratum and to Spanish-dominant bilinguals. It does not appear that region exerts an effect on OV/VO variation as both the highest and lower rates of OV order are associated with groups from Peruvian and Ecuadorian regions. The overall rate of OV order reported for the bilingual group of the present study most closely resembles the proportion of OV order reported for professionals in Calca, Peru nearly three decades ago (Klee, 1996; F. A. Ocampo & Klee, 1995). Though the bilingual speakers of the present study may be slightly younger, more urban, and more educated on average than the Cusco population from which they are drawn, this heterogeneous group of speakers would certainly not be characterized as ‘professionals’.

In general, methodological differences across studies on OV/VO variation may be responsible for the variable results. For example, one notable difference between the methodologies employed by Muysken (1984) and F. A. Ocampo and Klee (1995) is the manner in which the object is defined. For Muysken (1984), the object constituent includes direct and indirect object NPs, but, for F. A. Ocampo and Klee (1995), objects are “NPs that can be replaced by the clitics *lo, la, los, las*” (p. 74). Additionally, F. A. Ocampo and Klee (1995) discard tokens that contain additional constituents (e.g., and adverbial phrase), whereas Muysken (1984) includes tokens of this kind. The impact of these methodological differences is illustrated both by the disparate overall rates of OV order reported in Muysken (1984) and F. A. Ocampo and Klee (1995)— 19.6% versus

28.3%— and the two unequal sets of results presented by Klee (1996) generated through the application of each of the aforementioned methodologies (see Table 3.5 in Chapter 3).

The inclusion criteria of the present study converges with past studies in that object clitics are not considered, and only predicates in main clauses are included. I diverge from the methodology advanced by Klee et al. (2011) and F. A. Ocampo and Klee (1995) in that I include clauses that contain additional constituents, negative polarity, and pronominal objects.<sup>224</sup> Given that Muntendam (2009) includes tokens with additional constituents and personal pronouns, and the rate of OV order in Muntendam's sample is lower than that reported by F. A. Ocampo and Klee (1995), perhaps it is the inclusion of these tokens that contribute to a relatively lower incidence of OV order. This possibility will be explored in the forthcoming section given that polarity and definiteness (which includes pronominal objects as a factor level) are considered as independent variables in the present study.<sup>225</sup>

It is also possible that the relatively low proportion of OV order reported in the present study reflects a change in ordering preferences over time, especially given the young average age of the bilingual group in the present study ( $\bar{x} = \approx 33$  years). Though there is still considerable predicate constituent order variation in the present study compared to non-contact varieties of Spanish (Givón, 2001; F. A. Ocampo, 1995; Puerma Bonilla, 2019), overall, the rate of OV order is lower than what has been observed in studies conducted approximately one to three decades ago. This shift toward a preference for canonical structures coincides with a demographic shift that has taken place in Peru over the past several decades from Quechua monolingualism to Quechua-Spanish bilingualism and Spanish monolingualism in the region (Table 1.1 in Chapter 1). Thus, it is possible that the bilingual vernacular target mentioned above is shifting as a consequence of the changing sociohistorical context. That is, as the Cusco speech community becomes more Spanish-dominant, and social process like globalization and urbanization increase exposure to more prestigious, non-contact varieties, the linguistic target within the bilingual the speech community also shifts to reflect features of a non-contact Spanish (Muysken, 1984). Of course these hypotheses must be confirmed by longitudinal data or an apparent time analysis, which I provide in the next section.

In Table 7.4, I list the OV/VO order proportions that correspond to the **monolingual** participants of the present study alongside the proportions supplied by previous quantitative studies on monolingual and non-contact Spanish varieties.

<sup>224</sup> It is unclear if Muysken (1984) includes tokens of this type.

<sup>225</sup> Since I do not consider the presence of additional constituents as an independent variable, with the exception of subject expression, it will remain undetermined if this methodological difference is somewhat responsible for the relatively low rate of OV order reported in the present study.

Region/Register	OV (%)	VO (%)	Study
La Plata, Argentina	10.5%	89.5%	F. A. Ocampo (1992)
<b>Cusco, Peru</b>	<b>7.1%</b>	<b>92.9%</b>	<b>present study</b>
La Plata, Argentina	6.9%	93.1%	F. A. Ocampo (1995)
Lima, Peru	5.6%	94.4%	Klee et al. (2011)
<i>Don Quijote</i> analysis	2.2%	97.8%	Givón (2001)
text corpus analysis	0.4%	99.6%	Puerma Bonilla (2019)

Table 7.4: Average OV/VO distribution in monolingual Spanish word order studies

It bears mentioning that the results provided by Givón (2001) and Puerma Bonilla (2019) are quite different than the other studies listed. This is likely a consequence of methodological differences. Specifically, Givón (2001) and Puerma Bonilla (2019) consult written works, while the remaining studies elicit spontaneous speech data via sociolinguistic interviews. The proportion of OV/VO orders produced by monolinguals in the present study mirrors those that analyze the natural speech of non-contact Rio Platense Spanish (F. A. Ocampo, 1992, 1995) and the variety of monolingual Spanish spoken by individuals residing in “shanty towns inhabited by Andean migrants” in Lima<sup>226</sup> (Klee et al., 2011, p. 11). This confirms that monolingual Andean Spanish is more similar to non-contact varieties of Spanish than bilingual Andean Spanish, as regards predicate constituent order variation. In the next section, I discuss the extralinguistic and linguistic predictors of predicate constituent order variation in Andean Spanish to determine if monolinguals and bilinguals differ only in the frequency with which OV order is employed or also the independent factors that condition the use of the non-canonical structure.

### 7.3 Research Question 2

The second research question involves the possible conditioning effect of the evaluated extralinguistic and linguistic variables on predicate constituent order in Andean Spanish:

**RQ2:** Which extralinguistic and linguistic factors condition OV/VO variation in the varieties of Andean Spanish spoken by monolingual and bilingual speakers?

I discuss the results of the extralinguistic and linguistic predictors separately in the forthcoming subsections.

<sup>226</sup> This group of speakers is comprised of individuals that were either born in the Andean region and moved to Lima before the age of 12 or have parents (second generation) or grandparents (third generation) that had migrated to Lima from the Andean region. The Spanish variety spoken by this group of individuals may be considered the result of dialect contact (e.g., contact between Limeño Spanish and Andean Spanish), not language contact.

### 7.3.1 Extralinguistic predictors

The extralinguistic variables that condition predicate constituent order variation in Andean Spanish differ across bilingual and monolingual groups. Thus, I will discuss the results of each group separately, beginning with the former. The mixed-effects logistic regression model pertaining to the **bilingual** group selected two significant extralinguistic predictors— age and BLP score. The negative coefficient accompanying the **age** of the bilingual speaker indicates an inverse correlation such that, as age increases, the probability of VO order decreases. In other words, higher rates of OV order characterize the discourse of relatively older speakers, while the opposite is true of relatively younger speakers. Figure 7.1 below illustrates this correlation.

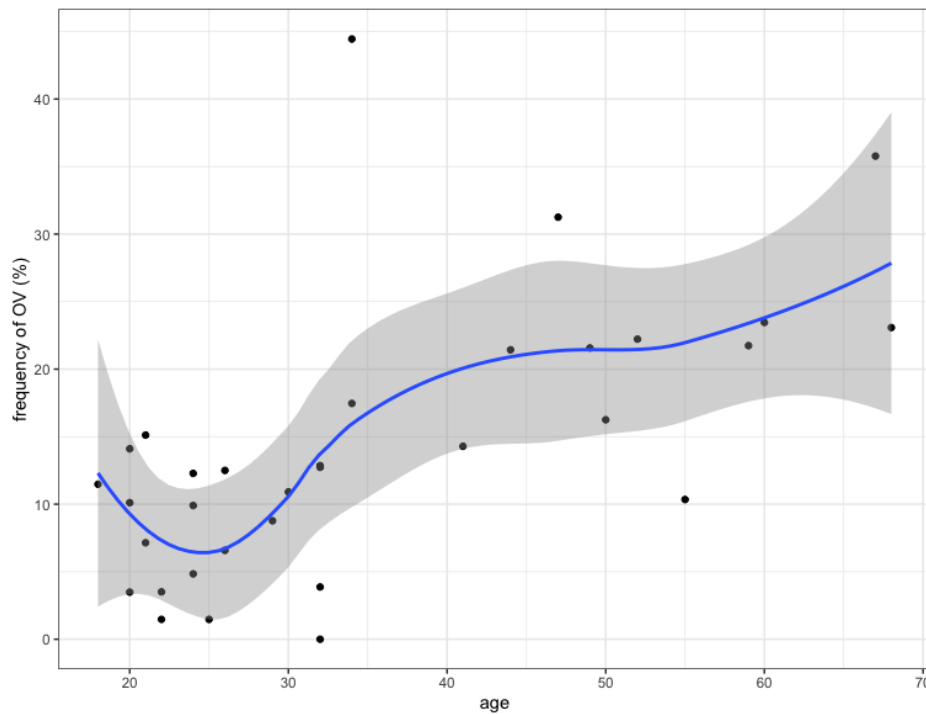


Figure 7.1: Scatterplot of age and OV order frequency in bilingual Andean Spanish group

The scatterplot above clearly shows that, overall, as the age of the speaker increases, so does the frequency of OV order in their speech. In the absence of longitudinal data, an examination of the relationship between age of the speaker and the behavior of a particular linguistic feature may provide insight into its diachronic trajectory. Tagliamonte (2012) writes, “[i]n an apparent time study, generational differences are compared at a single point and are used to make inferences about how a change may have taken place in the (recent) past. Age

differences are assumed to be temporal analogues, reflecting historical stages in the progress of the change” (p. 43). Two types of linguistic change that may be revealed through an apparent time analysis include age-graded change and generational change. Age-grading refers to the stratification of a particular linguistic feature whereby speakers of different ages use particular features as a result of their being that age. Graphically, an age-graded change presents as a u- or v-shaped curve, at least when the feature is non-prestigious. This is because middle aged people (30-55 years) “experience maximum social pressure to conform to the norms of the standard language,” likely due to their position in the workforce, but adolescents and older generations are not as affected by this societal pressure. Generational change, on the other hand, refers to the process whereby one structure or feature of a language replaces another over time. This brand of linguistic change is typically represented by an S-curve or a monotonic slope (Tagliamonte, 2012, p. 44). Though the loess line in Figure 7.1 is not a perfect S-curve, it represents a general increase in y-values (i.e., OV frequency) from left to right, which represents a statistical trend corresponding to generational change. Given this apparent time analysis, it appears that there is a change in progress in which non-canonical OV order is falling out of favor over time among bilingual speakers.

This finding informs the observation presented in §7.2 that the rate of OV order reported for the bilingual group is low in comparison to the OV order rates reported in previous studies conducted in the past one to three decades. Because the most recent study was conducted approximately a decade ago, I removed the youngest decade of speakers from the bilingual Andean Spanish dataset to enhance the comparability of the results. Upon removing all participants  $\leq 29$  years of age, the mean rate of OV order<sup>227</sup> increased from 14.1% to 19.6%. This figure is nearly identical to the proportion of OV order corresponding to Andean Spanish bilinguals from Juncal, Ecuador (19.9%) reported in 2009— a decade prior to the year that data was collected for the present study (Muntendam, 2009). Taken together, there is strong evidence to suggest that a decrease in the use of non-canonical structures, particularly OV order, constitutes a change in progress among bilingual speakers of Andean Spanish.

The **linguistic dominance** of the speaker, determined via BLP scores, also exercised a significant effect on predicate constituent order variation among bilingual speakers. Recall that, in this study, negative BLP score values indicate Spanish dominance, scores near zero represent balanced bilingualism, and positive values correspond to Quechua dominance. As the BLP score increases from negative to positive values, so does the rate and probability of OV order. This trend mirrors Muysken’s (1984) finding that the rate of OV order associated

<sup>227</sup> The mean rate of OV order is determined by calculating the average rate of OV order across all bilingual participants.

with Quechua-dominant bilinguals far surpasses that associated with Spanish-dominant bilinguals. That Quechua-dominant speakers employ non-canonical, Quechua-like structures more frequently than their Spanish-dominant counterparts provides clear evidence that linguistic contact with Quechua is partially responsible for an elevated incidence of OV-type structures, as argued in Hubbel (2023).

Despite my hypothesis that all extralinguistic factors would condition syntactic variation, the remaining extralinguistic predictors— sex, residence, education, and first language— were not selected as significant predictors by the statistical model. However, the descriptive statistics corresponding to the level of **education** reveals a trend in which participants with less formal education employ OV order more frequently than those with more formal education. In fact, participants who completed primary education employ OV order nearly twice as often as those who completed secondary education (28.6% versus 15.5%) and more than three times as frequently as those who completed post-secondary education (9.0%). This observation coincides with Muntendam’s (2009) finding that non-professional participants exhibit higher rates of OV order than professional participants in both Bolivia and Ecuador.<sup>228</sup> Because there is a diglossic relationship between Spanish and Quechua (i.e. Spanish is the prestige language that dominates in more formal domains, like education, and Quechua is the language of informal social domains, like the marketplace), bilinguals who have attained a higher level of education have likely been more consistently exposed to standardized varieties of Spanish, or at least, less consistently exposed to Quechua. One possible reason that the statistical model did not select education as a significant predictor is the lack of balance across groups. Specifically, participants with a primary-level education provide only 12.5% of the total tokens, whereas participants with post-secondary-level education provide 61.0%.

Though the **first language** of the speaker was also not selected as a significant predictor of variable predicate constituent order, the descriptive statistics show that L1 Quechua bilinguals employ OV order more frequently than simultaneous and L1 Spanish bilinguals. Again, it is likely the case that this variable was not selected as a significant predictor because these groups are unbalanced. To elaborate, there are three times as many L1 Quechua bilinguals as L1 Spanish bilinguals in the dataset. I presume that this discrepancy is representative of the Cusco population, as, impressionistically speaking, in this region, it is more common for an individual to first grow up speaking Quechua in the home then acquire Spanish later through formal education or exposure to urban life than to grow up speaking Spanish in the home to later learn Quechua. It bears mentioning that the four participants who reported acquiring Spanish prior to

<sup>228</sup> According to Muntendam (2009), “professionals received some form of higher education after secondary school (for instance, they attended a teacher training college or university), whereas non-professionals received no more than secondary education” (p. 56).

Quechua are much younger on average than the participant group as a whole ( $\bar{x} = 24$  years) and may have been exposed to Quechua more consistently beginning in primary school as a result of the recent Intercultural Bilingual Education policies.

Contrary to the findings supplied by Klee et al. (2011) and Muntendam (2009), **sex** is not a significant predictor of OV/VO variation in this sample of Andean Spanish bilinguals. Recall that Klee et al. (2011) determine that OV order is more frequent among males than females (16.3% versus 5.6%), while Muntendam (2009) reports the opposite effect of sex among Bolivian bilinguals—women employ OV order at a higher frequency than men. Upon closer inspection, the variable sex in these studies seems to be a proxy for occupation and/or the social network of the speaker. Klee et al. (2011) observe that, in Lima, women tend to find employment in positions in which they work closely with speakers of the regional, standard variety (e.g., live-in nannies), and men tend to work in labor-intensive industries with other Andean migrants. In more rural locales, like Tarata, Bolivia, for example, women tend to remain in the Andean community where their regional variety is reinforced through daily communication with other women in similar social positions. Men residing in these communities may be more inclined to travel to urban areas for work, which exposes them to more standard varieties. Sex may not significantly condition predicate constituent order variation in the present study because these female and male participants reside in both urban and rural areas, have attained varying levels of education, and are employed in many different types of occupations.

Turning now to the **monolingual** group, two extralinguistic variables are selected as significant predictors of predicate constituent order variation—residence and level of education. Concerning **residence**, participants that reside in the department of Cusco employ OV order 2.5 times more frequently than those residing within the city of Cusco (12.4% versus 5.0%). This finding may be explained by Peruvian Census data, which reports that a higher percentage of inhabitants claim Quechua as their maternal language within more rural provinces compared to the province of Cusco, which primarily contains the city of Cusco and its suburbs (Instituto Nacional de Estadística e Informática, 2017). For example, 76.4% of Quispicanchi inhabitants claim Quechua as their maternal language, while only 31.7% of provincial Cusco residents self-identify as L1 Quechua speakers. Presumably, participants residing outside the city of Cusco would interact more frequently and intimately with bilingual Quechua-Spanish speakers than those in the city. That OV order is more frequent among department monolinguals suggest that these social relationships shape the linguistic features of their speech.

The effect of level of **education** on OV/VO variation among monolinguals is quite unexpected. According to the statistical model, participants who have attained a university education exhibit a lower probability of employing canonical VO order than those who have completed post-secondary education (e.g., technical certification). However, there is no significant difference between secondary and post-secondary participants. Descriptively speaking, secondary participants use OV order the least, which is the opposite tendency displayed in the bilingual dataset. A closer look reveals that all secondary-level tokens correspond to a single participant, P5. This participant is only 19 years of age; thus, it is nearly impossible that he would have been able to complete post-secondary-level education prior to his participation in the present study. In fact, he was enrolled in a post-secondary institution at the time of his interview. Ultimately, it is difficult to ascertain the actual effect of education on predicate constituent order variation among monolingual speakers because all monolingual participants are highly educated in comparison to the overall population of Cusco—less than a third of the Cusco population (30.4%) has attained a degree or certification beyond secondary education in comparison to 90.9% of monolingual group participants (Instituto Nacional de Estadística e Informática, 2017).

### 7.3.2 Linguistic predictors

In opposition to the findings reported for the extralinguistic variables, the conditioning effect of the linguistic predictors on predicate constituent order variation is nearly identical for both monolingual and bilingual participants. Specifically, information status, definiteness, polarity, and subject expression exercise a statistically significant effect on syntactic variation, while animacy, and weight do not exercise a significant effect.

#### Information status

Beginning with **information status** of the object, for both monolinguals and bilinguals, the odds of canonical VO order increase when the direct object referent is new, considering given objects as the reference level. For monolingual participants, even though new referents are twice as frequent in the dataset overall, given objects account for 64.9% of all objects configured in OV-type orders. Among bilinguals, the odds of VO order also increases when the direct object referent is new, but the preference for given objects to collocate in pre-verbal position is not quite as strong. This observation is substantiated through a comparison of the odds of VO order associated with information status in each group. For monolingual speakers, a new referent increases the odds of VO order

by 229%, and in the bilingual group, a new referent increases the odds of VO order by only 125%.

These results may lend quantitative support to the claim made in previous literature that a higher incidence of OV order in Andean Spanish is a syntactic consequence of the loosening of the pragmatic restrictions that govern pre-verbal objects in non-contact varieties of Spanish (Camacho, 1999; Muntendam, 2009; Muysken, 1984). Though monolingual Andean Spanish cannot be considered a non-contact variety, it is ostensibly less affected by contact with Quechua than bilingual Andean Spanish. So, it may be the case that pre-verbal objects in monolingual Andean Spanish are more restricted by pragmatic factors than bilingual Andean Spanish. This is supported by the disparate odds ratios mentioned above, which demonstrate that pre-verbal objects are better predicted by the information status of the object in monolingual Andean Spanish. Though information status of the object also conditions predicate constituent order variation among bilinguals, the syntactic-pragmatic relationship is not as strong in this variety.

That predicate constituent order variation in both monolingual and bilingual Andean Spanish is conditioned by the information status of the referent such that given referents occupy earlier positions aligns with the Given Before New Principle (Gundel, 1988; Halliday, 1967). The effect of information status reported in the present study for Andean Spanish also coincides with past studies on cross-domain phrasal shift,<sup>229</sup> which ascertain that given objects are associated with an increased probability of pre-verbal collocation both in SVO-dominant languages, like Mandarin Chinese (Yao, 2018) and Old English (Struik & Van Kemenade, 2018) and SOV-dominant languages, like historical Dutch (Struik & Schoenmakers, 2023).

<sup>229</sup> Recall that cross-domain phrasal shift refers to the “shift of a phrase from one domain to another,” e.g., the shift of a direct object across the verb in either direction (Yao, 2018, p. 5).

### **Definiteness**

Another linguistic variable that significantly conditions syntactic variation in Andean Spanish is **definiteness**. For both bilingual and monolingual groups, there is a significant decrease in the odds of canonical VO order when the direct object is a pronoun versus a definite referent. In fact, in approximately a third of OV tokens<sup>230</sup> the object is pronominal, despite the fact that pronouns account for little more than 10% of all object referents in the Andean Spanish dataset overall. A pronoun is comparatively more definite than a definite referent, according to the Givenness Hierarchy proposed by Gundel et al. (1993), and previous studies that have assessed the impact of definiteness on word order variation share the general finding that relatively more definite expressions appear in non-canonical positions to the left (Faghiri & Samvelian, 2020; Nac-

<sup>230</sup> This proportion is slightly more than a third for the monolingual group and slightly less than a third for the bilingual group.

carato et al., 2021; van Bergen & de Swart, 2009, 2010; Yao, 2018). Based on the finding that a higher frequency of OV order coincides with pronominal objects compared to definite and indefinite expressions, definiteness appears to condition predicate constituent order in Andean Spanish in a similar fashion.

The monolingual and bilingual groups diverge regarding the syntactic behavior of indefinite object referents compared to their definite counterparts. For the bilingual group, there is a significant difference between definites and indefinites such that indefinites have a higher likelihood of exhibiting VO order. Thus, for bilinguals, there is a consistent inverse relationship between definiteness and canonical predicate constituent order frequency across all factor levels—the likelihood of VO order increases along the definiteness hierarchy (e.g., pronouns > definites > indefinites). In the monolingual Andean Spanish group, there is no significant difference in word order preferences between definites and indefinites. Descriptively speaking, both types of referents are collocated in pre-verbal position at nearly the same rate (5.2% versus 5.1%). This pattern fits with the findings of van Bergen and de Swart (2010) that pronouns scramble (i.e., move to the left of the adverb) nearly obligatorily and that definite and indefinite referents remain unscrambled nearly categorically. The present study is the first, to my knowledge, that has established a correlation between the definiteness of the object and predicate constituent order variation in Andean Spanish.

## **Polarity**

According to the logistic regression analyses, the effect of sentential **polarity** is a significant predictor of OV/VO variation in bilingual Andean Spanish, but not in monolingual Andean Spanish. However, the inclusion of polarity as a predictor in the monolingual statistical model increases its fitness. Thus, it may be tentatively assumed that, in both groups, negative polarity increases the likelihood of VO order. Recall that polarity was included as an independent variable in the present study because past studies on variable predicate constituent order in both Andean Spanish and Quechua have either excluded negative sentences from the dataset (Klee, 1996; Klee et al., 2011; F. A. Ocampo & Klee, 1995) or have not specified whether sentential polarity was considered in the process of delimiting the variable context (Kalt & Geary, 2021; Munten-dam, 2009; Muysken, 1984; Sánchez, 2003). Because negation does significantly affect syntactic ordering, its exclusion from the envelope of variation in the former set of studies is justified. It follows that if negation inhibits non-canonical structures, the overall proportion of OV in the present study should be higher, and thus, more akin to that reported by previous studies if negative tokens were

eliminated. For the bilingual Andean Spanish group, the proportion of OV order increases from 13.2% to 14.1% when only tokens in affirmative contexts are considered. This figure remains lower than the proportion of OV order attested in previous studies, but this methodological adjustment does cause a slight increase in the OV order proportion. The inclusion of polarity as an independent predictor demonstrates that establishing a consistent methodology is crucial for cross-study comparison.

### Subject expression

Like negation, **subject** expression also significantly conditions syntactic variation in the bilingual group, but not the monolingual group, though it is still selected as a predictor that enhances the fitness of the monolingual statistical model. Considering null subjects as the reference level, the likelihood of VO order increases when the subject is overt. In other words, explicit subjects exercise the effect of maintaining canonical constituent order in Andean Spanish. This finding aligns with the observation by Muysken (1984) that “the presence of a subject inhibits to some extent verb final order” (p. 114). The absence of SOV structures in Muysken’s data leads him to conclude that verb-final order “is only very indirectly due to Quechua influence.” Instead, a higher incidence of OV order is due to an over-extension of the “preposing rule” that is already operational in non-contact varieties of Spanish (p. 114). This conclusion is somewhat substantiated by the distribution of clausal constituent orders with three explicit constituents— subject, object, and verb— which are provided in Table 7.5 below.

	OV Type			VO Type			Total
	SOV	OSV	OVS	SVO	VSO	VOS	
<b>Monolingual</b>							
<i>n</i>	4	5	6	256	5	13	<b>289</b>
%	1.4%	1.7%	2.1%	88.6%	1.7%	4.5%	<b>100.0%</b>
<b>Bilingual</b>							
<i>n</i>	12	34	14	524	9	20	<b>613</b>
%	1.9%	5.5%	2.3%	85.5%	1.5%	3.3%	<b>100.0%</b>

Table 7.5: Distribution of clausal constituent orders in Andean Spanish

Contrary to Muysken’s results, Table 7.5 shows that SOV order is not completely absent from the Andean Spanish dataset, though it is the most infrequent order for monolinguals and the second most infrequent order for bilin-

guals. The most notable difference between the monolingual and bilingual groups is that, proportionally speaking, OSV order is over three times as frequent for bilinguals. In fact, OSV order accounts for over half (56.7%) of all OV-type tokens with an explicit subject for bilinguals. OSV order results when the object is fronted, which according to F. A. Ocampo (1992) is only possible when the object is the topic (or given) and the subject is the focus of contrast. However, of the 34 OSV tokens in the bilingual dataset, 13 (38.2%) contain new objects. The example below illustrates that OSV order may occur with a new object constituent and a subject that is not the focus of contrast.

(114) *entonces tres de la mañana yo he levantado... **comidita** yo he  
hecho para mi suegra desayuno... cafecito*

‘So at three in the morning, I woke up... I made my mother-in-law a little food, breakfast... a little coffee’ [P45]

In (114), the object referent, *comidita*, ‘food’, has not been mentioned prior to this utterance, so it is new. Additionally, the subject, *yo*, ‘I’, is the same as the subject in the clause that immediately precedes the exemplified OSV sentence, so it cannot be considered the focus of contrast. This example, and others like it, demonstrates that non-canonical clausal orders do not correspond in a one-to-one fashion with particular pragmatic situations, at least not in Andean Spanish.

It bears mentioning that the most obvious trend provided by Table 7.5 is that SVO is by far the most frequent order when there is an explicit subject. According to Dryer’s criteria of establishing dominant word order, the dominant order should be at least twice as frequent as the next most frequent alternative, which is certainly the case for SVO order. For monolinguals, SVO order is nearly 20 times more frequent than the next most frequent alternative, VOS, and for bilinguals, SVO order is more than 15 times more frequent than OSV order. Thus, though Andean Spanish exhibits a higher incidence of OV-type orders, the basic order typology of both clausal constituents and predicate constituents does not differ from non-contact varieties of Spanish. In fact, given the findings regarding the relationship between age and OV order frequency, it appears that Andean Spanish is approximating the prescriptive Spanish norm over time, at least concerning the order of predicate constituents.

### **Interaction: Definiteness and polarity**

The final significant result noted in Chapter 5 involves an **interaction** between two variables— definiteness and polarity. The inclusion of this interaction improves the fitness of both the monolingual and bilingual statistical models.

Recall that the definiteness-polarity interaction was considered to determine whether negative pronouns, like *nada*, ‘nothing’, behave differently than affirmative or neutral pronouns, like *todo*, ‘everything’, and *a mí*, ‘me’. Within this interaction, polarity exercises a significant effect on pronouns such that pronouns in affirmative contexts are associated with higher rates of OV order than pronouns in negative contexts. For the bilingual group, this difference is evident— pronouns in affirmative contexts appear in OV order at a rate of 40.8% and pronouns in negative contexts are configured in OV order at a rate of 8.2%. In the subset of data corresponding to pronouns in negative contexts, there are a handful of cases in which the pronoun is not negative, rather the sentence is negative, but the pronominal object is affirmative or neutral. Isolating only the negative pronouns that appear in the dataset, *nada*, ‘nothing’, and *nadie*, ‘noone’, only one token of 40 total (2.5%) is configured in OV order. Incidentally, this one instance of OV order occurs with the only token of *nadie* in the dataset. An example of the negative pronominal object *nada* is provided in (115) below.

(115) *me dediqué a vender por necesidad bebidas alcohólicas para poder tener ingreso... sí, y ese día no nos cocinamos **nada** porque... no teníamos creo economía porque mi papá tampoco no no estaba trabajando, no ha habido mucha producción en las papas, así es*

‘I was dedicated to selling alcoholic beverages out of necessity in order to make money... yeah, and that day we didn’t cook **anything** because... we didn’t have, I believe, money, because my dad wasn’t working either, there hadn’t been a lot of potato production, that’s how it is’ [P27]

To account for the rigidity of VO order with negative pronominal objects, I posit that the negative adverbial *no* syntactically blocks the pre-verbal collocation of an additional negative constituent, like *nada*.<sup>231</sup> That the non-canonical order, *no **nada** nos cocinamos* is judged infelicitous by a native speaker consultant of a non-Andean Spanish dialect further supports this possibility. Additional investigation is needed to better understand the relationship between negative pronominals and predicate constituent order variation in Andean Spanish.

For the bilingual group only, polarity also significantly conditions the syntactic configuration of indefinite referents, again with affirmative sentences exhibiting a preference for pre-verbal objects when compared to negative sentences. In fact, only one of 126 total indefinite referents in negative contexts is arranged in OV order (0.8%). Thus, the preference for canonical order in these contexts is nearly categorical. Perhaps an example of indefinite referents in a negative context can clarify this result. Consider (116)<sup>232</sup> below:

<sup>231</sup> This syntactic tendency may be overridden when the object is animate (e.g., *nadie*).

<sup>232</sup> The word *nada* in this example is not included as part of the object constituent, because it may be interpreted as a general extender and translated as, ‘or anything like that’ (see Chapter 4).

(116) *mi mamá sabe mirar coca.. ¿cuánto? dice cinco años he sufrido y ni siquiera no hacía mirar ni **coca** nada, no no creía pues yo*

‘my mom knows how to read coca... how long, I mean, I have suffered for five years and I haven’t even made [her] read **coca** or anything, I didn’t believe [in it]’ [P20]

The above example illustrates that negative sentences contain one or more negative adverbial expression, in this case *no* (‘no’), *ni siquiera* (‘not even’), and *ni* (‘not even, nor’). The presence of extra constituents may complicate the operation of preposing an object constituent, as is the case with explicit subjects (Muysken, 1984). For instance, should the object be preposed prior to the first negative adverbial, *ni siquiera*, or the second negative adverbial, *no*? Are these options grammatical for speakers of Andean Spanish? However, this observation does not explain why indefinite expressions are affected more by negation than definite expressions. One possibility is that indefinite expressions are more likely to be new than definite referents or pronouns, as shown in Table 7.6.

	new		given		Total( <i>n</i> )
	<i>n</i>	%	<i>n</i>	%	
<b>pronoun</b>	98	35.8%	176	64.2%	274
<b>definite</b>	484	54.4%	405	45.6%	889
<b>indefinite</b>	921	77.1%	274	22.9%	1,195

Table 7.6: Distribution of information status of the object across definiteness

Thus, it may be the case that the combination of features that indefinites in negative sentences exhibit increases the likelihood of VO order. That is, indefinites are often new, and both new and indefinite referents are associated with increased odds of VO order in comparison to their counterparts. Furthermore, negation as an independent predictor tends to inhibit non-canonical order. Thus, indefinites in negative sentences largely exhibit three linguistic features— indefinite, negative, and new— that, on their own, are associated with relatively higher rates of canonical word order. The same cannot be said for definite referents because they are evenly distributed across the information statuses, new and given. To further illustrate this point, when I extract all tokens that are characterized by only the factor levels that exhibit a preference for VO order— new, indefinite, negative, and overt subjects— the proportion of VO order is nearly categorical, 95.7%. On the other hand, when I extract all tokens that are characterized by the factor levels that exhibit a preference for

OV order— given, pronouns, affirmative, and null subjects— the proportion of OV order exceeds the canonical order, 54.4%.

To summarize, information status, definiteness, polarity, and subject expression are significant predictors of predicate constituent order variation in bilingual Andean Spanish. In monolingual Andean Spanish, the latter two predictors are not statistically significant, however, they do contribute to the fitness of the statistical model, which accounts for variation in the data. The remaining predictors— animacy and weight— were not found to exert a significant effect on OV/VO variation for either group. This suggests that these cross-linguistic predictors either do not apply to the languages of this contact situation or the investigated syntactic domain. That predicate constituent order variation in monolingual and bilingual Andean Spanish is conditioned by the same linguistic predictors suggests that bilinguals may employ OV order more frequently, but not differently than monolinguals. In the words of Levshina et al. (2023), these varieties differ in “degree, not kind” (p. 865). This hypothesis will be explored further in §7.6.

## 7.4 Research Question 3

The next two research questions inquire about predicate constituent order in the Quechua dataset. The third research question is repeated below:

**RQ3:** What is the overall proportion of OV/VO orders in bilingual Cusco Quechua?

Recall that, according to prescriptive grammar sources, canonical word order in Quechua is SOV, though other word orders are permitted (Cerrón-Palomino, 1987a; Cusihamán, 1976, 2001; Salas Cruz, 1993). In this sample of adult bilingual Cusco Quechua, slightly more than two-thirds of the included tokens contain predicates that exhibit OV order (**68.2%**) and slightly less than one-third of the tokens are configured in VO order (**31.8%**). Contrary to my hypothesis that this sample of bilingual Cusco Quechua would lack a dominant order of predicate constituents, OV order may be considered the dominant order of this sample because it is twice as frequent as the alternative, VO order (Dryer, 2013a). However, it is important to note that there is a great deal of inter-participant variation. In fact, for 13 of the 28 total participants,<sup>233</sup> OV order does not account for at least two-thirds of their total tokens; that is, the percentage of tokens with OV order provided by each of these participants is less than 66.7%. That nearly half of the participants do not employ OV order

<sup>233</sup> P1, P8, P15, P18, P19, P22, P23, P26, P36, P37, P39, P41, P43

twice as frequently as VO order suggests that OV is not always the dominant order at the idiolectal level.

In Table 7.7 below, I provide the general distribution of OV/VO orders reported in the present study alongside two previous studies that provide quantitative data on predicate constituent order variation in Quechua— Sánchez (2003) and Kalt and Geary (2021). The results of the present study are highlighted.

Variety	OV (%)	VO (%)	Study
Chuquisaca	89.2%	10.8%	Kalt and Geary (2021)
Cusco	78.0%	22.0%	Kalt and Geary (2021)
Cusco	68.2%	31.8%	present study
Ulcumayo	35.2%	64.8%	Sánchez (2003)
Lamas	14.9%	85.1%	Sánchez (2003)

Table 7.7: OV/VO distribution in Quechua word order studies

A comparison of the data reveals that the distribution of OV/VO order associated with the bilingual Cusco Quechua speakers of the present study falls between those reported in Kalt and Geary (2021) and Sánchez (2003). It is important to note that the participants consulted in each of these studies are child bilinguals. I will not attempt to discern how the disparate average age of the participant groups may effect the statistical results, especially given the fact that age is not the only variable that distinguishes these investigations from one another. For instance, the Quechua varieties investigated also differ regarding the region in which they are spoken and their consequent classification within the Quechua language family; Chuquisaca and Cusco Quechua pertain to the Quechua II-C classification, Lamas Quechua to the II-B classification, and Ulcumayo Quechua to the Quechua I classification (Kalt & Geary, 2021; Sánchez, 2003; Torero, 1964). Most critically, the methodological approach to data collection varies greatly across studies. To be more precise, a sociolinguistic interview is used to elicit spontaneous speech in the present study, Kalt and Geary (2021) employ a picture-based elicitation task to extract structured speech, and Sánchez (2003) applies a story-telling task to elicit semi-spontaneous speech.

The reported OV/VO proportions in Table 7.7 for both Chuquisaca and Cusco varieties of Quechua by child bilinguals indicate that OV is the dominant order in these varieties. It is important to note that the bilingual children of this study are L<sub>1</sub> Quechua speakers, which may partially explain the difference between the rates of OV order cited in Kalt and Geary’s (2021) study versus the present study. If I were to consider only the data provided by L<sub>1</sub> Quechua

bilinguals, the rate of OV order of the present study would be 71.0%, which is slightly closer to the 78.0% figure reported by Kalt and Geary (2021). Additionally, the authors employ a picture-based elicitation task in which participants are prompted with an SOV stimulus sentence. As previously mentioned, it is possible that this stimulus sentence primed SOV order, resulting in a higher proportion of canonical word order than what would be found in naturalistic speech.

Using a less-structured task, Sánchez (2003) reports much lower OV proportions for the bilingual children of Ulcumayo and Lamas. In fact, the figures reported for Lamas in particular suggest that VO is the dominant order in this variety of Quechua, at least among bilingual children. Like the participants of the present study, the Ulcumayo and Lamas participant groups include sequential (both L1 Quechua and L1 Spanish) and simultaneous bilinguals. Perhaps consistent contact with Spanish in educational and social domains is partially responsible for the high incidence of Spanish-like structures within the speech of these children. Sánchez’s findings may also foreshadow a change in progress in which the basic order of predicate constituents is shifting from OV to VO. I look further into this possibility in the next section by examining the effect of age on predicate constituent order variation in Quechua.

Though the dependent variable of the present study is the order of predicate constituents, when considering the order of clausal constituents, it is clear that there is no dominant order in bilingual Cusco Quechua. The frequencies of all OV-type and VO-type orders with an explicit subject are supplied in Table 7.8 below.

	OV Type			VO Type			Total
	SOV	OSV	OVS	SVO	VSO	VOS	
<i>n</i>	108	35	25	113	5	10	<b>296</b>
%	36.5%	11.8%	8.4%	38.2%	1.7%	3.4%	<b>100.0%</b>

Table 7.8: Distribution of clausal constituent orders in Quechua data

Unexpectedly, the most frequent configuration of clausal constituents is SVO, which accounts for 38.2% of predicates with an explicit subject. Thus, clausal constituent order preferences in Cusco Quechua parallel those of Lamas and Ulcumayo Quechua, as SVO is the most frequent order in these varieties as well. However, SVO cannot be considered the dominant order of clausal constituents based on the criteria advanced by Dryer (2013a) because SVO is not twice as frequent as the next most frequent alternative, SOV. The above data suggest that, though OV is the dominant order of predicate constituents,

contemporary bilingual Cusco Quechua should be typologically classified as a language lacking dominant order with respect to the syntactic arrangement of subject, object, and verb constituents. This pattern aligns with Dryer's observation that some languages that lack a dominant clausal order maintain a clear ordering preference for two elements, like object and verb. The potential typological implications of the prevalence of SVO order in a language traditionally classified as SOV-dominant will be discussed in the concluding chapter.

## 7.5 Research Question 4

The fourth research question is restated below:

**RQ4:** Which extralinguistic and linguistic factors condition OV/VO variation in bilingual Cusco Quechua?

The results of a logistic regression analysis presented in the previous chapter confirmed that two extralinguistic (age and first language) and five linguistic predictors (information status, definiteness, subject expression, polarity and absolute weight) significantly condition predicate constituent order variation in this sample of bilingual Cusco Quechua. I begin with a discussion of the significant extralinguistic variables.

### 7.5.1 Extralinguistic predictors

As mentioned above, both the age and first language of the speaker are significant predictors of OV/VO variation in the Quechua dataset, contrary to my hypothesis that only first language and linguistic dominance (i.e., BLP score) would be selected as significant predictors in line with the findings of Hubbel (2023). Beginning with the **age** of the speaker, it was determined that there is an inverse correlation between age and the likelihood of non-canonical VO order. Put simply, older speakers tend to employ canonical OV order more frequently than younger speakers, who tend to prefer non-canonical VO order. Figure 7.2 provides a graphical representation of the relationship between age and of OV order frequency.

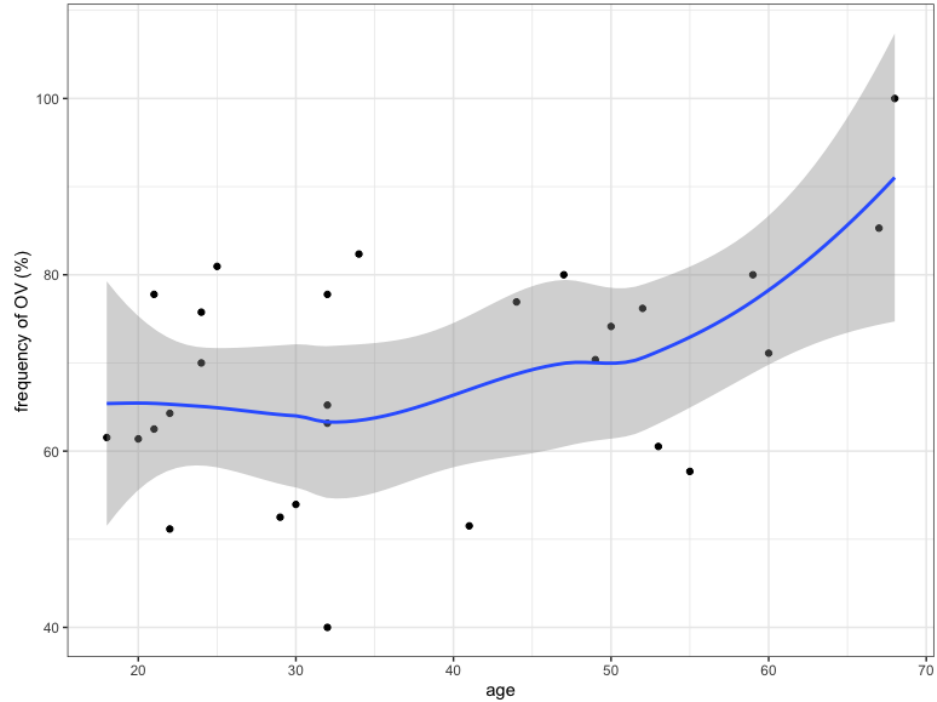


Figure 7.2: Scatterplot of age and OV order frequency in bilingual Cusco Quechua

Recall from §7.3 that the line representing the relationship between age and a particular linguistic feature may indicate either age-grading or generational change; the former is represented by a u-shaped curve and the latter by an S-curve or monotonic slope (Tagliamonte, 2012). Though the loess line in the above graph is not completely sigmoidal, it is clearly not a u-shaped curve. The line in Figure 7.2 represents a relationship between the two variables such that as age increases, so does the frequency of OV order. From 18 to  $\approx 35$  years of age, the line is relatively flat, which is characteristic of an S-curve. From there, the line takes on a positive slope, which increases further around 52 years of age. Though the loess line does not flatten out as it approaches the right-most edge of the graph, as would a sigmoid function, the trend illustrated here clearly corresponds to a change in progress. It remains unclear at this point in time whether this change will ultimately result in the typological classification of Quechua as a language with two primary alternating orders (Dryer & Haspelmath, 2013) or a syntactic shift from OV to VO order. Longitudinal data is required to determine the final destination of this change in progress.

The **first language** of the speaker is the other extralinguistic factor that significantly conditions word order variation in this sample of bilingual Cusco Quechua. First, recall that there is no significant difference between L1 Quechua

and L1 Spanish bilinguals. The lack of difference between these groups is likely due to the fact that only two participants identify as L1 Spanish bilinguals. The statistical analysis does reveal, however, that the odds of VO order increase by 84% for simultaneous bilinguals considering L1 Quechua as the reference level. This means that bilinguals who acquired Quechua prior to Spanish employ the canonical Quechua structure more frequently than those who acquired Quechua and Spanish in tandem. That simultaneous bilinguals exhibit higher rates of Spanish-like VO order than L1 Quechua bilinguals confirms that language contact is one driving force behind the increased word order flexibility documented in contemporary bilingual Quechua (Hubbel, 2023). In the absence of historical spoken data on word order variation in Quechua, it is unclear how frequent VO-type orders were in spoken Quechua prior to initial contact with Spanish or even prior to the acceleration in the shift from Quechua to Spanish witnessed only a few decades ago (Escobar, 2011); however, it appears that language contact in this region has permitted more word order flexibility in Quechua. This is substantiated by the relationship between age and OV order frequency (Figure 7.2), which suggests that canonical order is falling out of favor over time.

Though **education** was not selected as a significant predictor of predicate constituent order variation, it is worth mentioning that the proportion of OV order is a great deal higher for participants with a primary-level education (78.2%) versus those with a secondary- or post-secondary-level education (66.4% and 65.6%, respectively). Perhaps one of the reasons that this variable was not selected as a significant predictor is because the factor levels are not completely balanced. Namely, there are more tokens provided by participants with a post-secondary-level education than those with primary-level and secondary-level education combined. Another possible reason this predictor was not selected by the logistic regression model is, in this sample, level of education may actually be a proxy for age of the speaker. It appears that participants who have attained a primary-level education are, on the aggregate, older, than those associated with higher levels of education in this sample. Figure 7.3 illustrates the relationship between age and level of education.

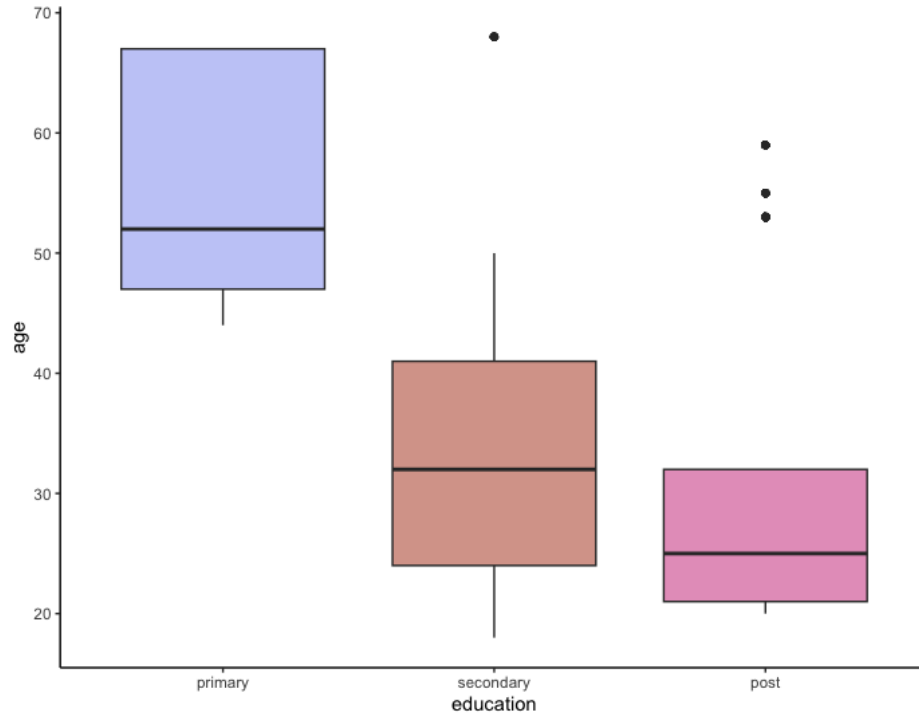


Figure 7.3: Age of Quechua participants across educational level

It is not surprising that participants with a primary-level education are generally older than more highly educated participants, considering the current demographic trend characterized by an increase in the proportion of highly educated individuals in Peru, which is corroborated by recent census data (Instituto Nacional de Estadística e Informática, 2017). For instance, in 2007, 12.9% of the Peruvian population had attained a university-level education, and in 2017, 18.1% had attained this same level of education, a 40.3% increase. Because the proportion of the population with a higher level of education is increasing, and consequently, the proportion of the population with no education or primary-level education is decreasing, it is conceivable that older participants would report lower levels of education than younger participants.

As a point of comparison, the two significant variables, age and first language, do not show this same relationship in which one variable could be simply considered epiphenomenal of the other. As illustrated in Figure 7.4, the age of each group does not increase from L1 Spanish bilingual, to simultaneous bilingual, to L1 Quechua bilingual, as would be expected if first language was simply a proxy for age.

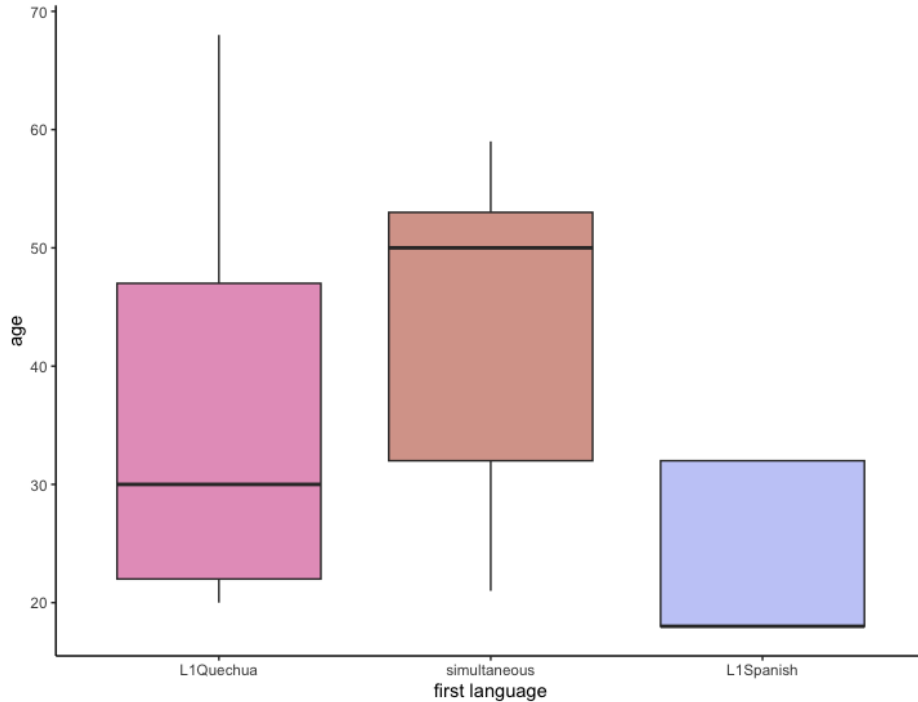


Figure 7.4: Age of Quechua participants across first language

In sum, age and first language are the only extralinguistic predictors that significantly condition predicate constituent order variation in Quechua. Though there is a tendency for less formally educated participants to employ canonical OV order more frequently than more formally educated participants, education is not selected as significant predictor by the statistical model and may simply be epiphenomenal of age. Now I turn to a discussion of the linguistic predictors.

### 7.5.2 Linguistic predictors

The five linguistic variables that were selected as significant predictors of predicate constituent order variation in Quechua were information status, definiteness, subject expression, polarity, and absolute weight of the object constituent. I will also briefly discuss the effect of goal arguments, as this predictor improved the fitness of the statistical model, even though it was not significant ( $p=0.075$ ). The variables animacy, accusative marking, and Spanish loanwords did not exercise a significant effect on predicate constituent order variation in this sample of bilingual Cusco Quechua.

## Information status

Beginning with **information status** of the object, on the surface it appears that there is an even distribution of OV/VO orders across new and given referents. According to the descriptive statistics, 67.6% of given referents and 68.7% of new referents are configured in OV order. However, when the participant is considered as a random effect, the statistical analysis reports a significant difference across information status of the object referent such that the odds of non-canonical VO order decrease for new referents with respect to given referents, meaning that new referents are more likely appear in pre-verbal position than given referents.

This finding contradicts the cross-linguistic observation that familiar, accessible referents tend to appear earlier in the utterance, as captured in the Given Before New Principle (Gundel, 1988; Halliday, 1967).<sup>234</sup> Recall from Chapter 2 that two studies investigating the relationship between the information status of the object and OV/VO variation were conducted on head-initial languages, which exhibit SVO-dominant order— Old English (Struik & Van Kemenade, 2018) and Mandarin Chinese (Yao, 2018). These studies coincide on the finding that there is an increased probability of OV order when the object constituent is given. In this sample of bilingual Cusco Quechua, the opposite tendency is observed. Note that the description of the relationship between syntax and morphological topic- and focus-marking in Quechua offered by Sánchez (2010) coincides with the finding that given referents appear in post-verbal position. Specifically, she asserts that focused-marked (new) elements may appear in fronted positions, and topic-marked (given) elements may appear in either fronted positions or be dislocated to the right. The findings of the present study regarding information status constitutes preliminary evidence that head directionality may impact the interaction between syntax and information structure.

Recall that, in the Conceptual-Accessibility Hypothesis advanced by Yamashita and Chang (2001), it is proposed that head-final languages are more sensitive to conceptual accessibility than form accessibility,<sup>235</sup> as is the case with head-initial languages. Consequently, constituents that are semantically richer are more accessible in head-final languages, and thus appear in earlier positions. Though this hypothesis has been applied to explain weight based phenomena in head-final languages (e.g., a long-before-short preference), it is reasonable to assume that it may also apply to the information status of a referent. New referents have not been mentioned earlier in the discourse so they typically require more lexical material to be successfully identified, whereas given referents have been mentioned earlier so their label need not be as descriptive. This is substantiated by the fact that, in the bilingual Quechua sample, new referents

<sup>234</sup> Seemingly, in Quechua, the *First Things First* principle applies, which Gundel (1988) posits may be active in particular languages.

<sup>235</sup> Conceptual accessibility is a semantics-based metric, and form accessibility is a weight-based metric.

are, on average, a syllable longer than given referents, 5.0 and 4.0 syllables respectively.<sup>236</sup> Consider (117) below:

<sup>236</sup> According to a one-way ANOVA test with no assumption of equal variances, this difference is significant,  $p = 4.26 \times 10^{-12}$ .

- (117) a. *kay-pi-qa*            ***Chinchero-q***        ***llaqta-q***  
 this-LOC-TOP    Chinchero-GEN    town-GEN  
                   ***aniversario-n-ta***                    *celebra-yku*  
                   anniversary-POSS.3SG-ACC    celebrate-IPL.EXCL  
 ‘Here we celebrate **the town of Chinchero’s anniversary**’ [P27]
- b. *ñuqa-qa ni-ra-ni,*            “*mana ruwa-ni-chu tarea*        *mana*  
 I-TOP    say-PST-ISG NEG    do-ISG-NEG homework NEG  
                   *ruwa-mu-ni-chu*        ***tarea-ta***”  
                   do-DIR-ISG-NEG homework-ACC  
 ‘I said, “I didn’t do **[my] homework**, I didn’t do **[my] homework**”’  
 [P36]

The predicate constituents in (117a) are configured in OV order, and the object constituent is a new referent. Perhaps because it is a new referent, 11 syllables are required to identify this entity, so it is conceptually accessible as regards semantic richness and salience. In later reference to this entity, it would be acceptable to use a lighter referent, like *aniversario*, ‘it’s anniversary’, or even a pronoun, which would be both given and less semantically rich. In (117b), the given object referent is collocated in post-verbal position. Because *tarea* (‘the homework’) has been mentioned prior to the first bolded object, conceptual accessibility (i.e., semantic richness) is low, and it may be referred to without other words or morphemes that would enhance its salience. This is evidenced by the lack of accusative marking on the first bolded object referent in (117b).

As established in chapter 3, topic and focus, which roughly correspond to given and new information status respectively, are prescriptively expressed via suffixation in Quechua. However, in the bilingual Quechua dataset of the present study, overt topic and focus suffixes are relatively infrequent. Table 7.9 below presents the distribution of topic and focus markers on all clausal constituents across OV and VO word orders.

	OV		VO		Total
	<i>n</i>	%	<i>n</i>	%	
Focus-subject	1	25.0%	3	75.0%	4
Focus-object	16	94.1	1	5.9%	17
Focus-verb	15	78.9%	4	21.1%	19
Topic-subject	18	58.1%	13	41.9%	31
Topic-object	33	57.9%	24	42.1%	57
Topic-verb	0	0.0%	1	100.0%	1

Table 7.9: OV/VO order across constituents marked for topic and focus

<sup>237</sup> Six of these tokens bear a topic marker on both the subject and the object. They are counted twice in the table— once as topic-subject and once as object-topic.

Table 7.9 illustrates that, overall, only 123<sup>237</sup> tokens (10.6%) exhibit morphological topic or focus marking on the subject, object or verb. Though topic-marking is more frequent than focus-marking, topic suffixes still appear on a relatively small proportion of utterances (7.3%) in the dataset.

Recall from Table 3.7 in chapter 3 that Sánchez (2010) predicts that only focus-marked verbs and topic-marked objects are expected to activate non-canonical VO-type orders. That is, VO order results when focus-marked verbs are fronted or object topics are displaced to the right. These information structure configurations may also exhibit OV order when a focus-marked verb remains *in situ* and an object topic is either fronted or remains in *in situ*. The table confirms that both focus-marked verbs and topic-marked objects exhibit both orders: OV and VO. However, according to Sánchez (2010), the remaining information structure configurations are predicted to result only in OV-type orders. Focus-marked objects do appear to adhere to this tendency, with the exception of one token of VO order. However, topic-marked subjects are not always arranged in an OV-type order as predicted. After removing 6 of the 31 tokens of topic-marked subjects that are accompanied by topic-marked objects, which may permit VO order, 40% ( $n=10$ ) of the remaining tokens with topic-marked subjects are configured in VO order, as exemplified below:

- (118) *kuraq ñaña-y-wan ri-rqa-ni campu-ta...*  
 older sister-POSS.ISG-INS go-PST-ISG field-ACC...  
*ñaña-y-qa cuida-rqa-n papa-y-paq*  
 sister-POSS.ISG-TOP care.for-PST-3SG father-POSS.ISG-DAT  
*uywa-n-kuna-ta*  
 animal-POSS.3SG-PL-ACC

‘I went to the field with my older sister... my sister took care of my dad’s animals for him’ [P8]

In this example, the topic-marked subject appears *in situ*, which is permitted by Sánchez’s predictions. However, the object of the sentence, *uywankunata*, ‘his animals’, is dislocated to the right. According to Sánchez (2003), this configuration would be acceptable if the object were also topic-marked, but this is not the case.<sup>238</sup> In fact, the object here cannot be construed as the topic because it is a new referent that is being introduced into the discourse for the first time in the exemplified utterance.

In summation, because topic-marked objects and focus-marked verbs (i.e., the two information structure configurations that permit VO order, according to Sánchez) account for only 28 of the 385 tokens exhibiting VO order (7.3%), it appears that the relationship between Quechua morphosyntax and information structure is not as robust as Sánchez (2010) proposes. This finding also directly contradicts the idea that these morphological markers are obligatory in Cusco Quechua (Muntendam, 2015). As is the case in Bolivian Quechua,<sup>239</sup> morphological topic and focus suffixes are used only sparingly in the bilingual Quechua dataset of the present study.

## Definiteness

Next, **definiteness** is the strongest categorical linguistic predictor<sup>240</sup> of predicate constituent order variation in Quechua. Though definite and indefinite referents are not significantly different from one another, pronouns are significantly more likely to appear in canonical pre-verbal position with respect to definite referents. The studies discussed in Chapter 2 report similar findings—that more definite referents appear in earlier positions—in a variety of syntactic domains (e.g., the genitive phrase) across several languages (Faghiri & Samvelian, 2020; Naccarato et al., 2021; van Bergen & de Swart, 2009, 2010; Yao, 2018). Additionally, for each of the languages and loci of variation investigated in these studies, the earlier position was the non-canonical position. This is not the case in Quechua—though pronouns are more likely to appear in an earlier position (i.e., pre-verbal) than definites and indefinites, pre-verbal position in Quechua is the canonical position of an object.

It bears mentioning that pronouns, which are the most definite type of referent, are also typically given<sup>241</sup> and light. It is reasonable to speculate that pronouns are more likely to appear earlier in a syntactic sequence not just as a consequence of their definiteness, but also the cluster of linguistic properties (e.g., givenness and lightness) that characterize them. However, it has already been determined that, in Quechua, given referents are less likely to be configured in OV order than new referents, possibly as a consequence of its head directionality. In fact, the inclusion of pronouns in this dataset may be some-

<sup>238</sup> Recall that Sánchez (2010) does encounter instances of right-dislocated topics lacking *-qa*, but maintains that this is an ungrammatical structure, according to the grammatical intuitions offered by an adult Quechua speaker.

<sup>239</sup> Specifically the dialect of Quechua spoken by Spanish-Quechua bilinguals in Tarata and Huayculí, Bolivia.

<sup>240</sup> This is confirmed by the magnitude of the coefficient that accompanies the factor level *pronoun* in the statistical model, which is greater than all other coefficients.

<sup>241</sup> Cataphoric pronouns, however, are not considered given because they are co-referential with an entity mentioned later in the discourse.

<sup>242</sup> In the Quechua dataset, 74.9% of all pronominal referents are given, versus 52.7% of definite referents and 31.4% of indefinite referents.

<sup>243</sup> According to a one-way ANOVA test with no assumption of equal variances, this difference is significant,  $p < 0.05$ .

what obscuring the effect of information status, as when they are removed, the proportion of given referents in pre-verbal position drops from 67.3% to 61.0%, while the proportion of pre-verbal new referents remains nearly constant. Thus, the fact that pronouns are often given<sup>242</sup> does not help promote their incidence of OV order, as pronouns and given referents are associated with opposite syntactic tendencies. Regarding weight, it is also the case that pronouns are lighter on average than indefinite and definite referents. In fact, the mean number of syllables of pronominal referents is 3.2, while the mean number of syllables of definite and indefinite referents is 4.9 and 4.8 respectively.<sup>243</sup> So, it may be the case that the syntactic effect of definiteness is compounded by the low weight of the constituent with respect to pronominal constituents. The syntactic behavior of pronouns will be explored in greater depth in the next section in which the Quechua and Andean Spanish results are compared.

### Subject expression

**Subject** expression is the next strongest categorical predictor of predicate constituent order in Quechua. When the subject is overt, the odds of non-canonical VO order increase by 95% compared to when the subject is null. So, when there is an explicit subject, SVO order is preferred, but when the subject is null, OV order is preferred. In fact, when the order of explicit clausal constituents is considered, SVO is the most frequent order in this dataset, followed closely by SOV (see Table 7.8). The same finding was reported in Hubbel (2023) in which a subset of this data was analyzed. In this work, I suggest that perhaps SVO order is preferred when the subject is explicit because “this order iconically conveys the concept of a subject acting on an object” (p. 130). However, this explanation does not account for the fact that 41.0% of the world’s languages exhibit SOV order (Dryer & Haspelmath, 2013), therefore iconicity may not be a determining factor of ordering preferences. An alternative explanation may be that the presence of an explicit subject is more likely to yield SVO order because, canonically, both arguments exist in the same syntactic space (pre-verbal), and by dislocating the object to the right, there is a partition between the two arguments of the verb, which may facilitate processing. When there is no explicit subject, there is no other constituent that occupies the same syntactic space as the object, and thus no need to separate the object from the subject.

However, this is not the most convincing justification either because accusative marking is fairly robust in this variety of bilingual Quechua.<sup>244</sup> Ostensibly, morphological marking, not syntactic structure, should bear the weight of differentiating the object from the subject. Regardless, when the subject is null, OV order may be preferred because, even in the absence of morphological

<sup>244</sup> Omission of *-ta* occurs at a rate of only 8.7% in this dataset.

accusative case marking, interpretation would likely remain unaffected because no other constituent could be construed as the object. Recall that Sánchez (2003) finds that a higher rate of SVO order coincides with high rates of *ta*-omission in Ulcumayo and Lamas Quechua. Thus, in these varieties, speakers rely on the syntax to disambiguate the subject and object constituents because accusative marking is inconsistent. Perhaps accusative marking remains a reliable mechanism of identifying the object of the sentence in this sample of bilingual Cusco Quechua because a shift to SVO order must precede a reduction in *ta*-marking. This way, when morphological case-marking declines, the language is already equipped with a linguistic mechanism for grammatically differentiating arguments (e.g., word order).<sup>245</sup> Upon closer inspection of the relationship between age and accusative case-marking, I find that the mean age of tokens with *ta*-marked objects is higher compared to *ta*-less objects, as illustrated in Figure 7.5.<sup>246</sup> That younger speakers omit *ta* more frequently than older speakers suggests that perhaps in the coming generations, *ta*-marking will become more variable and less obligatory, mirroring the linguistic change from OV-dominant to more variable predicate constituent order already observed in bilingual Cusco Quechua (see Figure 7.2).

<sup>245</sup> Because Sanchez's data is synchronic and representative of child bilingual speech, it is not possible to determine whether a syntactic shift preceded *ta*-omission (or vice versa) or if they developed in tandem.

<sup>246</sup> According to a one-way ANOVA test with no assumption of equal variances, this difference is significant,  $p = 0.029$ .

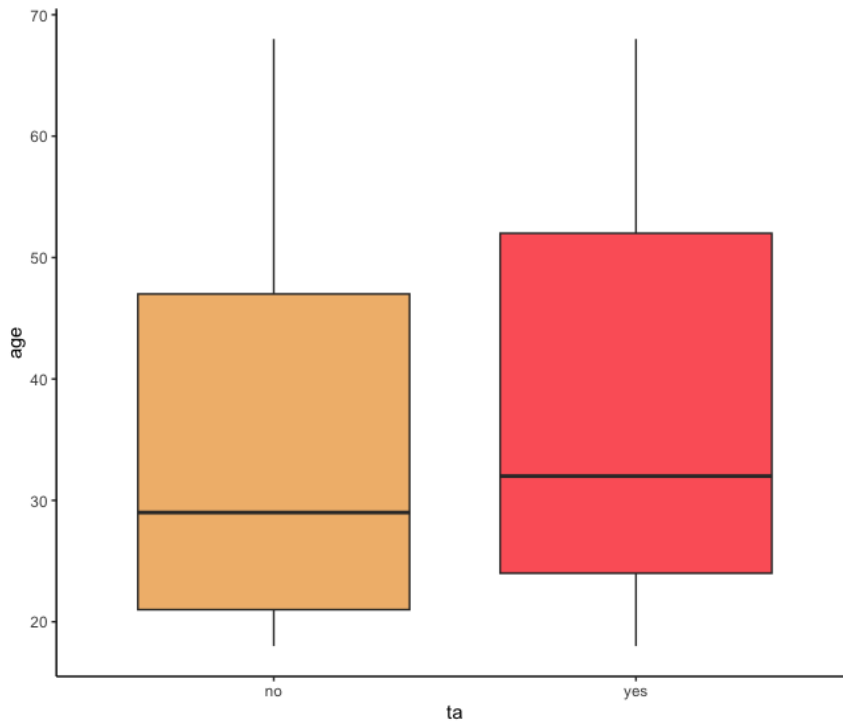


Figure 7.5: Distribution of age across *ta*-marking in Quechua

Another possible approach to deciphering the correlation between explicit subjects and an increased frequency of VO order is to compare other linguistic features that are associated with SOV and SVO orders. Surprisingly, there is very little difference in the proportion of objects that lack *ta*-marking for SOV (9.3%) and SVO (9.6%) structures. However, non-*ta*-marked object constituents in sOV and SVO constructions do differ in another way. In SOV constructions, there is a disproportionate amount of inanimate objects lacking *ta*-marking (80.0%), compared to human objects (20.0%). However, in SVO constructions, the contrast is even more stark— out of the 11 objects lacking *ta*-marking, 10 (90.9%) are inanimate.<sup>247</sup> So, only one object lacking case-marking is animate in SVO constructions, as exemplified in (119) below. Perhaps, then, SVO is a syntactic tool used to disambiguate subject-like human objects from the explicit subject, which is often higher on the animacy hierarchy than the object (J. K. Bock & Warren, 1985; McDonald et al., 1993). This is supported by near-categorical *ta*-marking on human objects in this construction. If it were the case that explicit subjects trigger VO order to differentiate a human object from the subject, a higher proportion of human object referents would be expected in SVO order compared to SOV order. However, this is not the case, as the proportion of human objects in SOV order is slightly higher than the proportion of human objects in SVO order, 23.4% and 18.6% respectively.

<sup>247</sup> These percentages are based on a small number of tokens, and should thus be regarded with caution.

(119) *nuqa mana riqsi-ni-chu buk persona cercana...*  
 I NEG know-1SG-NEG one person close  
*riqsi-ni-chu manan pero ni-nku runa ninku*  
 know-1SG-NEG NEG-EVIDI but say-3PL person say-3PL  
*chiqaq*  
 true

‘I don’t know a close person, I don’t know [them] but they say, people say it’s true’ [P37]

It is certainly not the case that SVO order occurs only with human object referents, so the effect of an explicit subject on predicate constituent order cannot be reduced to animacy alone. That overt subjects increase the likelihood of VO-type orders remains a somewhat puzzling finding. As suggested, perhaps this the high incidence of SVO order is merely the first step in a typological shift from SOV to SVO order, but longitudinal studies are required to arrive at this conclusion. I return to subject expression in §7.6.

## Polarity

Recall that **polarity** was evaluated as an independent predictor in the present study because past studies on word order variation in Quechua have not clarified whether negative tokens were included in the envelope of variation. The statistical model reveals that polarity is a significant predictor of predicate constituent order variation in bilingual Cusco Quechua. Specifically, sentences with a negator are associated with a lower likelihood of VO order. In other words, negative expressions restrict non-canonical word order. In Quechua, negation is expressed through both the a negative adverbial, *mana*, ‘no’, and the class-free negative suffix, *-chu*, which is typically affixed to the verb in instances of sentential negation. Consider (120) below:

- (120) *papa-y-taq*                      *eh nace-ra-n*                      *Urubamba llaqta-pi,*  
dad-POSS.ISG-COOR eh born-PST-3SG Urubamba town-LOC  
*paykuna quechua-ta*      *rima-nku...* *manta-y-taq*  
they quechua-ACC speak-3PL mom-POSS.ISG-COOR  
*mana...* *mana rima-ta-pas*      *rima-y-ta*      *yacha-n-chu*  
NEG NEG word-ACC-COOR say-INF-ACC know-3SG-NEG
- ‘And my dad was born in the town of Urubamba, they spoke Quechua...  
but my mom didn’t... didn’t know how to speak **the words**’ [P35]

In this example, the negative adverbial *mana* and negative particle *-chu* bookend the sentence. Because *-chu* appears on the verb, the scope of negation in this example is sentential.<sup>248</sup> Upon closer inspection, I find that not all negative tokens in the dataset include both negators that are prescriptively required in Quechua. In fact, of the 105 negative tokens, 13.3% ( $n = 14$ ) lack the class-free negative suffix *-chu*. When comparing the syntactic patterns in negative sentences with both negators versus *chu*-less negative sentences, canonical OV order is more frequent without *-chu*, 92.9%, versus sentences with both negators, 74.2%. According to a chi-square test, this difference is not significant ( $p = 0.25$ ), likely due to the small number of *chu*-less tokens. However, the trend here suggests that perhaps it is the absence of *-chu* in negative sentences that restricts non-canonical order in lieu of negation itself. In fact, when the 14 tokens without *-chu* are removed from the dataset, polarity is no longer selected as a significant predictor of word order variation in the logistic regression model ( $p = 0.092$ ).<sup>249</sup>

It has been established in other circumstances that the loss of morphological marking coincides with a syntax that adheres more strictly to the dominant order. For example, Kalt and Geary (2021) find that higher *-ta* omission correlates

<sup>248</sup> There are a few instances of adverbial negation in the Quechua dataset, in which *-chu* is affixed to the adverb. However, the majority of negative tokens exhibit sentential negation.

<sup>249</sup> However, the inclusion of this independent variable does still improve the fitness of the statistical model.

with stricter OV order in Chuquisaca Quechua. Perhaps one reason that OV order is higher in negative sentences, particularly in those lacking *-chu*, is because this structure is also experiencing attrition of morphological marking. Thus, the syntax compensates for the loss of morphological expression of negation by becoming more rigid.

Furthermore, I observe that the negative pronominal, *mana imatapas*, which accounts for nearly one-fifth of all negative tokens (19.0%), is always collocated in canonical, pre-verbal position. This is exemplified in (121) below:

- (121) *entonces chay p'unchay nishu mancha-sqa ka-sha-ra-ni...*  
 then that day very scare-NMLZ be-PROG-PST-1SG  
*mana... ima-ta-pas ruwa-y-ta ati-ra-ni-chu*  
 NEG what-ACC-COOR do-INF-ACC can-PST-1SG-NEG  
*porque huch'uy-cha-raq ka-ra-ni*  
 because small-DIM-IMPV be-PST-1SG

'So, that day I was very scared... I couldn't do anything because I was still very small.' [P43]

In the next section, I explore the behavior of pronominal expressions in greater detail, including existentially quantified pronouns like *mana imatapas*, in both Andean Spanish and Quechua.

### Weight

Regarding syntactic **weight**, it was determined that there is a positive correlation between the number of syllables of the object constituent and the probability of VO order. In broad strokes, heavier object constituents tend to appear in post-verbal position (VO) and lighter constituents tend to appear in pre-verbal position (OV). Only the absolute weight of the object constituent, not the relative weight of predicate constituents, was selected as a significant predictor of syntactic variation. This finding contradicts the assertion of Wasow (2002) that weight is a relative measurement. Like Yao (2018), I find that the absolute weight of a single constituent may be a more appropriate conceptualization of this variable in the context of cross-domain phrasal shift.

Consequently, it would be misleading to conclude that the short-before-long ordering preference observed in other languages (Arnold et al., 2000; Heindinger, 2013; Stallings et al., 1998; van Bergen & de Swart, 2009, 2010; Wasow, 1997, 2002) also applies to Quechua because only the absolute weight of the constituent, not relative weight, affects syntactic ordering. This means that

objects are considered lighter with respect to other objects but not necessarily with respect to the verb that appears within the same predicate. Thus, the term **end-weight** (Quirk et al., 1972) may be a more appropriate label to apply to weight-based phenomena in Quechua, at least in this syntactic domain. Recall that it has also been predicted that head-final languages exhibit a long-before-short preference (Hawkins, 1994, 2004), which has been substantiated by word order investigations on Japanese (Yamashita & Chang, 2001) and Persian (Faghiri & Samvelian, 2020). This does not seem to be the case for Quechua, at least when the predicate is the locus of variation. Therefore, the findings on weight reported for bilingual Cusco Quechua tentatively contradict the MiD Principle advanced by Hawkins.

The interaction between weight and the order of predicate constituents in Quechua may differ from other head-final languages for several reasons. First, it may be the case that there is a long-before-short ordering preference in other syntactic domains, but, in situations of cross domain phrasal shift, the same weight-based principles do not apply. Second, it is possible Quechua is simply more sensitive to form accessibility (i.e., number of words) than conceptual accessibility (i.e., saliency), despite what is predicted for head-final languages (Yamashita & Chang, 2001). Though it is reasonable to hypothesize that the end-weight pattern exhibited in Quechua may be compounded by contact with a head-initial language, weight was not selected as a significant predictor of OV/VO variation in Andean Spanish.

It is worth mentioning that the weight-based metric applied in the Quechua analysis was number of syllables because Quechua is an agglutinative language that communicates grammatical relationships through the morphological strategy of suffixation rather than separate lexical items (i.e., prepositions), as is the case in Spanish. When weight is operationalized as the number of words of an object constituent in Quechua, a one-way ANOVA test indicates that there is still a significant effect of weight on predicate constituent order such that pre-verbal objects are shorter than post-verbal objects, 1.3 words and 1.5 words respectively. When both weight measurements are evaluated in a logistic regression analysis, only weight defined as number of syllables significantly conditions predicate constituent order variation. This finding tentatively contradicts the assertion that all weight-based measurements exercise a similar effect on syntactic variation (Wasow, 1997). Instead, it appears that considering the morphological type (e.g., synthetic, polysynthetic) of a language in deciding which weight-based metric to apply is a crucial methodological step.

## Goal arguments

Lastly, the status of an accusative-marked object as a **goal** argument was not selected as a significant predictor of predicate constituent order variation, but its inclusion in the statistical model improves the fitness. Generally speaking, goal arguments are associated with an increase in the odds of VO order. To review, in Quechua, goal arguments are marked with the accusative suffix, *-ta*, just like thematic arguments, so, grammatically speaking, they may be considered direct objects. This is illustrated in (122):

- (122) *mama-y*                      *siempre*    *hamu-n-puni*                      ***Cusco-ta***,  
 mom-POSS.ISG    always    come-3SG-EMPH    Cusco-ACC  
*papa-ta*            *apa-wa-ku-spa*,                      *chuño-ta*  
 potato-ACC    carry-OBJ.ISG-DIR-SUB    freeze.dried.potato-ACC  
*apa-wa-ku-spa*,                      *a*            *veces*    *uqa-ta*  
 carry-OBJ.ISG-DIR-SUB,    some    times    oca-ACC  
*apa-wa-ku-spa*,                      *hamu-n*  
 carry-OBJ.ISG-DIR-SUB    come-3SG

‘My mom always comes to **Cusco**, bringing me potatoes, bringing me freeze dried potatoes, sometimes bringing me oca, she comes.’ [P19]

Recall that goal arguments were included as an independent variable in the Quechua analysis to enhance cross-corpora comparability given that these arguments are not considered accusative objects in Spanish. Consequently, goal arguments are not included in the Andean Spanish corpus. The results confirm that the syntactic behavior of goal arguments differ from non-goal arguments. It could be the case that goal arguments appear in post-verbal position more frequently than non-goal arguments due to language contact under the assumption that goal arguments in Andean Spanish exhibit less syntactic flexibility than non-goal arguments.<sup>250</sup> This possibility is substantiated by the distribution of goal and non-goal arguments across OV and VO orders for L1 Quechua bilinguals versus simultaneous bilinguals.<sup>251</sup> Consider Table 7.10 below:

<sup>250</sup> To my knowledge, this assumption has not been corroborated by quantitative data.

<sup>251</sup> L1 Spanish bilinguals are not included in the table due to the small number of tokens pertaining to this group.

	<b>L1 Quechua</b>		<b>simultaneous</b>	
	OV (%)	VO (%)	OV (%)	VO (%)
goal	71.0%	29.0%	32.1%	67.9%
non-goal	71.2%	28.8%	65.1%	34.9%

Table 7.10: OV/VO order across argument type for L1 Quechua and simultaneous bilinguals

The above table shows that though the rate of OV/VO orders is identical across goal and non-goal arguments for L1 Quechua bilinguals, for simultaneous bilinguals, there is a clear preference for OV order with non-goal arguments and VO order with goal arguments. Because the status of the object as either a goal or non-goal argument affects predicate constituent order variation only for simultaneous bilinguals, I hypothesize that simultaneous bilinguals differentiate these arguments syntactically, possibly because goal arguments are not grammatically accusative in Spanish but are marked with *-ta* in Quechua, like other accusative arguments. Additional investigations that assess the grammaticality of pre-verbal goal arguments versus pre-verbal non-goal arguments in Andean Spanish may offer important insights into this matter.

To summarize, age, first language, information status, definiteness, subject expression, polarity, and absolute weight significantly condition predicate constituent order variation in bilingual Cusco Quechua. It is important to recognize that there is a somewhat contradictory explanation given for the effect of information status and weight in this section. Regarding information status, I theorize that new referents may appear in earlier positions in Quechua because these referents are typically longer, as they require more semantic material to successfully identify them in comparison to given referents, which have already been activated in the discourse and may thus be referred to using less material (e.g., pronouns). This would suggest that Quechua is more sensitive to conceptual accessibility (i.e., saliency of a referent) than form accessibility (i.e., weight defined as number of syllables). However, in discussing the effect of weight, I take the opposite stance— Quechua is sensitive to form accessibility based on the finding that lighter constituents tend to appear in earlier positions and heavier constituents in later positions. Though these results seem difficult to reconcile, they align with the assertion of Levshina et al. (2023) that factors affecting word order may exercise contradictory effects.

## 7.6 Research Question 5

The fifth and final research question motivates a comparison between the two languages investigated in the present study:

**RQ5:** How do the factors selected as predictors of predicate constituent order variation in Andean Spanish compare to those selected for Quechua? What might this comparison suggest about contact-induced language change and/or variation in this syntactic domain?

Before comparing the significant predictors of OV/VO variation in Andean Spanish and Quechua, I briefly discuss the general findings regarding the distribution of OV/VO orders in the three varieties investigated: monolingual Andean Spanish, bilingual Andean Spanish, and bilingual Cusco Quechua. To facilitate a cross-variety comparison, the overall rates of OV and VO orders for each group are listed in Table 7.11 below.

	OV	OV(%)	VO	VO(%)	Total( <i>n</i> )
<b>Monolingual AS</b>	94	7.1%	1,235	92.9%	1,329
<b>Bilingual AS</b>	311	13.2%	2,047	86.8%	2,358
<b>Quechua</b>	827	68.2%	385	31.8%	1,212

Table 7.11: OV/VO order for Andean Spanish and Quechua

Based on the frequency criteria of determining dominant order (Dryer, 2013a), both varieties of Andean Spanish exhibit dominant VO order, while OV is the dominant order in bilingual Cusco Quechua. It is also clear that the degree of predicate constituent order variation is higher in Quechua than Andean Spanish given the relatively high rate of non-canonical VO order. Recall that one-third of the Quechua participants exhibit a distribution of OV/VO orders that suggests a lack of dominant order (i.e., OV order is not at least twice as frequent as VO order). On the other hand, VO is the dominant order in Andean Spanish for all participants, with the exception of two bilingual speakers, P28 and P45.<sup>252</sup>

Of the 34 bilingual participants, 27 completed an interview in both Andean Spanish and Quechua that were ultimately included in the statistical analyses of the present study. For these participants, it is possible to directly compare their proportions of OV/VO orders in each language to determine whether syntactic preferences in one language correlate with those of the other language. Consequently, I compiled a new dataset containing the rates of OV/VO orders in

<sup>252</sup> P28 employs VO order at a rate of 55.6%, and P45 employs VO order at a rate of 64.2%

Andean Spanish and Quechua for each of these 25 participants. Figure 7.6 displays the correlation between OV order rates in Andean Spanish and Quechua in yellow and VO order rates in red. Each dot represents a single participant.

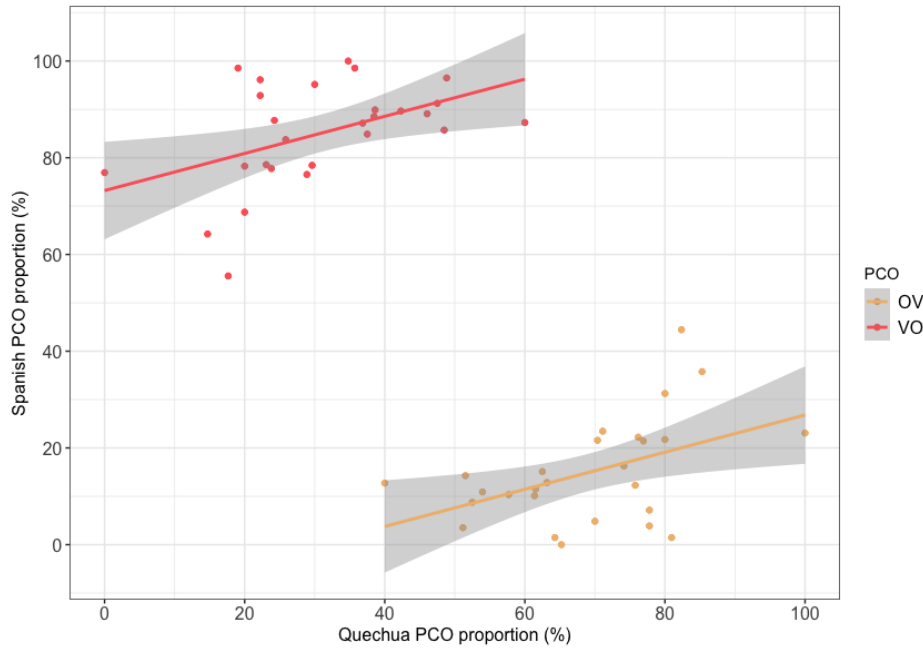


Figure 7.6: Correlation between OV/VO order rates of Quechua and Andean Spanish corpora for bilinguals

Figure 7.6 shows that there is a positive correlation<sup>253</sup> associated with both word orders; that is, a higher frequency of OV order in Quechua is correlated with a higher frequency of OV order in Spanish, and the same is true of VO order. This finding suggests that, at the idiolectal level, the two syntactic systems available to the bilingual speaker impact one another as regards the frequency of these structures. In other words, bilinguals show structural convergence in their speech concerning ordering preferences within the predicate. It is important to note that, though the correlation between word order in bilingual Quechua and Andean Spanish is significant, the corresponding R-squared value<sup>254</sup> is not objectively high, 0.22. This means that some of the variance in predicate constituent order can be explained by the syntactic patterns of the other language, but there are also other factors at play, like the extralinguistic and linguistic variables discussed in §7.3 and §7.5.

According to Poplack and Levey (2010), comparing the effect of these independent predictors across varieties reveals whether a linguistic feature of a contact language can indeed be classified as a contact-induced change, which

<sup>253</sup> According to a linear regression analysis, this correlation is significant,  $p < 0.05$ .

<sup>254</sup> The R-squared value calculates the proportion of variance in the dependent variable that can be explained by a given independent variable. An R-squared statistic of 1 means that all variation can be explained by the independent predictor, and an R-squared statistic of 0 means that none of the variation can be explained by the independent predictor.

is the main inquiry of the fifth research question. The criteria that the authors advance for establishing contact-induced language change are repeated below:

A candidate for contact-induced change in a contact variety is present in the presumed source variety and either 1) absent in the pre-contact or non-contact variety, or if present (e.g., through interlingual coincidence), is not conditioned in the same way as the source, *and* 3) can also be shown to parallel in some non-trivial way the behavior of a counterpart feature in the source (Poplack & Levey, 2010, p. 398).

Evaluating these criteria necessarily requires both horizontal and vertical comparisons in both languages (Poplack & Levey, 2010). A vertical comparison involves assessing the differences between current and earlier versions of the contact variety, which has already been carried out through an apparent-time age analysis for both Quechua and Andean Spanish. The results of these analyses show that the frequency of VO order is increasing over time in both languages. For Andean Spanish, this trend means that, structurally, Andean Spanish is approximating more standard varieties of Spanish in this particular syntactic domain. For Quechua, this trend means that non-canonical structures (i.e., more Spanish-like structures) are favored among younger generations. In the concluding chapter I discuss how these trends may inform an assessment of the mechanisms by which contact-induced change occurs in the Quechua-Spanish contact situation.

A horizontal comparison, on the other hand, involves a three-way comparison across the contact variety, non-contact variety, and source language. Considering Andean Spanish, this involves a comparison of OV/VO variation across Andean Spanish, non-contact Spanish, and Quechua. Though monolingual Andean Spanish is not a non-contact or pre-contact variety, this variety will act as the non-contact variety in the forthcoming comparison to assess the possibility of contact-induced change because, 1) there are no quantitative analyses currently available on non-contact varieties of Spanish that consider the same predictors evaluated in the present study, and 2) monolingual Andean Spanish may be considered 'less' of a contact language than bilingual Andean Spanish, which is substantiated by the fact that the rate of OV frequency in monolingual Andean Spanish is nearly identical to that reported for non-contact varieties, like Río Platense Spanish (F. A. Ocampo, 1992, 1995). Bilingual Andean Spanish is considered the contact variety in this comparison. Regarding Quechua, a horizontal comparison involves comparing OV/VO variation in bilingual Quechua, monolingual Quechua, and Andean Spanish. Unfortunately, spoken monolingual Quechua data is lacking from the present study and elsewhere, so I

must consult Quechua grammars as a proxy for monolingual Quechua (Cerrón-Palomino, 1987a; Cusihuamán, 1976, 2001). According to Poplack and Levey (2010), a horizontal comparison revealing that 1) predicate constituent order variation is conditioned differently in the contact versus non-contact varieties and 2) structural parallels exist between the contact variety and source language is sufficient grounds for confirming contact-induced change.

Applying these criteria to the present study, I first compare the findings of OV/VO variation in monolingual and bilingual Andean Spanish. I provide Table 7.12 below to facilitate a cross-variety comparison of the conditioning effect of extralinguistic and linguistic variables across these groups. The symbol **X** indicates that the corresponding variable is not a significant predictor of syntactic variation and the symbol  $\checkmark$  indicates that the corresponding variable is a significant predictor.<sup>255</sup> A dash (—) appears where the information is not applicable (e.g., L1 of the speaker was not considered for monolinguals because they have only one L1). The fourth column reports whether the corresponding predictor conditioned variation of the dependent variable in the same way in both varieties.<sup>256</sup> The right-most column is reserved for a description of the effect of the significant predictor variable.

	AS-mono	AS-bi	same?	effect
<b>Extralinguistic</b>				
sex	X	X	—	—
residence	$\checkmark$	X	—	—
education	$\checkmark$	X	—	—
L1	—	X	—	—
age	X	$\checkmark$	—	—
BLP score	—	$\checkmark$	—	—
<b>Linguistic</b>				
animacy	X	X	—	—
definiteness	$\checkmark$	$\checkmark$	yes*	pronoun: VO ↓
information status	$\checkmark$	$\checkmark$	yes	new: VO ↑
polarity	$\checkmark^*$	$\checkmark$	yes	negative: VO ↑
subject	$\checkmark^*$	$\checkmark$	yes	overt: VO ↑
weight	X	X	—	—

Table 7.12: Comparison of significant predictors of predicate constituent order variation across monolingual and bilingual Andean Spanish groups

Though there is no coincidence of extralinguistic predictors across monolingual and bilingual varieties, all significant linguistic predictors condition pred-

<sup>255</sup> The asterisks that accompany the  $\checkmark$  in the polarity and subject rows simply indicate that the predictor was not associated with a significant *p*-value, but is still considered a predictor of variation because its inclusion in the logistic regression analysis enhanced the fitness of the model.

<sup>256</sup> The rows that correspond to those predictors that condition variation identically in both varieties are also highlighted.

icate constituent order variation in the same way in both varieties. There is only one exception to this generalization: for bilingual Andean Spanish, there is a significant difference between definite and indefinite referents such that indefinites are associated with a higher likelihood of VO order, but there is no significant difference between these two types of referents for monolinguals. For both groups, however, the likelihood of VO order with pronominal objects is significantly lower than that of definite referents.

Though Table 7.12 clearly illustrates that OV/VO variation in monolingual and bilingual Andean Spanish is conditioned in the same way by the same linguistic predictors, there is still a significant difference regarding the proportions of OV/VO orders across bilingual and monolingual groups.<sup>257</sup> One notable difference between the two groups is the coefficient associated with the information status of the direct object. For monolingual speakers, twice as many given referents appear in OV order as new referents, despite the fact that new referents outweigh given referents two to one overall. Though given referents also exhibit a higher likelihood of OV order in bilingual Andean Spanish, the difference in syntactic behavior between the two factor levels is not as evident. This finding lends credence to the idea that OV order in Andean Spanish is less restricted by pragmatic constraints compared to other varieties of Spanish (Camacho, 1999; Muntendam, 2009). It is also the case that polarity and subject expression exert a stronger conditioning effect in the bilingual group. Specifically, both affirmative sentences and null subjects are associated with a higher likelihood of OV order. That these factor levels account for a higher proportion of the total tokens than their counterpart factor levels (negative polarity and overt subjects) may also partly explain the higher proportion of OV order in the bilingual group. Though the effect of some independent predictors may be stronger than others in each respective group, all linguistic predictors selected by the statistical models condition predicate constituent order variation in the same way for both groups. Thus, as previously stated, monolingual and bilingual Andean Spanish varieties differ in “degree but not kind” (Levshina et al., 2023, p. 826).

In Table 7.13 below, I compare the effect of the extralinguistic and linguistic variables on OV/VO variation in bilingual Andean Spanish and bilingual Cusco Quechua.

<sup>257</sup> According to Pearson’s Chi-square test with Yate’s continuity correction:  $\chi^2 = 31.9, p = 1.63 \times 10^{-8}$ .

	Quechua	AS-bi	same?	effect
<b>Extralinguistic</b>				
sex	X	X	—	—
residence	X	X	—	—
education	X	X	—	—
L1	✓	X	—	—
age	✓	✓	yes	age ↑ VO ↓
BLP score	X	✓	—	—
<b>Linguistic</b>				
animacy	X	X	—	—
definiteness	✓	✓	yes*	pronoun: VO ↓
information status	✓	✓	no	<b>Quechua</b> new: VO ↓ <b>AS</b> new: VO ↑
polarity	✓	✓	no	<b>Quechua</b> neg: VO ↓ <b>AS</b> neg: VO ↑
subject	✓	✓	yes	overt: VO ↑
weight	✓	X	—	—

Table 7.13: Comparison of significant predictors of OV/VO variation across Quechua and Andean Spanish bilinguals

The one extralinguistic predictor that is significant among both the Quechua and Andean Spanish bilingual groups is **age**. For both groups, there is an inverse relationship between age and VO order likelihood. In §7.3 and §7.5 I argued that this correlation suggests a change in progress (i.e., generational change). In Andean Spanish, the distribution of OV/VO orders is moving toward the canonical, prescriptive order, and in Quechua, the distribution of OV/VO orders is moving toward a higher degree of syntactic variation. For Quechua, it is unclear whether this variability will ultimately culminate in word order shift, which has been documented in other situations of intense language contact (Heine, 2008; Thomason & Kaufman, 1988), or if syntactic variation will persist indefinitely, ultimately resulting in its typological classification as a language with no dominant order for object and verb constituents, like Nez Perce (Dryer & Haspelmath, 2013).

Though the two other significant extralinguistic predictors for bilingual Andean Spanish and Quechua groups are not equivalent— **BLP score** and **first language**— it is important to note that these variables are moderately interrelated. Figure 4.7 in Chapter 4 demonstrates that on average, the reported BLP scores are higher (i.e., more Quechua-dominant) for L1 Quechua bilinguals and lower (i.e., more Spanish-dominant) for simultaneous and L1 Spanish

bilinguals. Of course, these variables are not synonymous because they are not both selected as significant predictors in each language. Regardless, it is the case that OV order is more frequent in both languages when the level of Quechua activation is higher, whether this be through the acquisition of Quechua as a first language or a relatively higher level of Quechua dominance.

Four of the linguistic variables are selected as significant predictors for both bilingual Andean Spanish and Quechua, but only two of them— **definiteness** and **subject expression**— condition predicate constituent order variation in the same way. In particular, in both varieties, pronouns are associated with lower odds of VO order<sup>258</sup> and overt subjects are associated with higher odds of VO. The information status of the object and polarity exercise language-specific effects, which is perhaps a consequence of their distinct head directionalities. In this way, each significant variable exercises the same effect regarding adherence to canonical order. In other words, new objects and negative polarity promote a stronger adherence to the canonical order in Andean Spanish (VO) and Quechua (OV). In the following subsections, I take a closer look at only those linguistic predictors for which there is evidence of syntactic convergence between the two languages— definiteness and subject expression.

<sup>258</sup> Indefinite referents are not significantly different from definite referents in Quechua, but there is a significant difference between indefinites and definites in Andean Spanish.

### 7.6.1 A closer look at pronouns

To reiterate, pronominal objects are correlated with high rates of OV order in all three investigated varieties— Quechua (84.7%), bilingual Andean Spanish (33.6%), and monolingual Andean Spanish (20.2%). For Andean Spanish varieties, this means pronouns have a higher likelihood of appearing in non-canonical pre-verbal position, and thus have a greater degree of syntactic freedom; in Quechua, on the other hand, a greater likelihood of OV order means pronouns adhere more strictly to canonical OV order.

In the initial operationalization of the variable **definiteness**, the factor level *pronoun* conflates several pronominal categories for the sake of balancing factor levels. However, research on adverb-object scrambling in Dutch suggests that not all pronouns interact with word order in the same way. Partitioning pronouns into three categories— personal pronouns, universally quantified pronouns, and existentially quantified pronouns— van Bergen and de Swart (2010) find that personal pronouns scramble obligatorily, universal pronouns scramble some of the time, and existential pronouns hardly ever scramble. Universally quantified pronouns include lexical items like *everything* and *everyone*, and existentially quantified pronouns consist of both negative pronouns (e.g., *nothing*) and affirmative pronouns (e.g., *somebody*). With the objective of examining pronouns in a more fine-grained manner, I divide the pronominal factor

level into four categories based on both van Bergen and de Swart’s analysis and the types of pronouns that appear most frequently in the dataset— **personal** (e.g., *a mí* and *nuqata*, ‘me’), **demonstrative** (e.g., *eso* and *chayta*, ‘that’), **universally** quantified, and **existentially** quantified pronouns. The distribution of OV/VO orders across all pronominal categories for each variety is supplied in Table 7.14.

	OV ( <i>n</i> )	OV (%)	VO ( <i>n</i> )	VO (%)	Total
<b>Monolingual Andean Spanish</b>					
personal	9	34.6%	17	65.4%	26
demonstrative	22	25.9%	63	74.1%	85
universal	2	10.5%	17	89.5%	19
existential	1	2.6%	37	97.4%	38
<b>Bilingual Andean Spanish</b>					
personal	18	36.7%	31	63.2%	49
demonstrative	55	43.0%	73	57.0%	128
universal	17	42.5%	23	57.5%	40
existential	2	3.5%	55	96.5%	57
<b>Quechua</b>					
personal	31	70.5%	13	29.5%	44
demonstrative	108	87.1%	16	12.9%	124
universal	16	100.0%	0	0.0%	16
existential	27	87.1%	4	12.9%	31

Table 7.14: Distribution of OV/VO order across pronominal category in Andean Spanish and Quechua

So far it has been the case that each significant linguistic predictor conditions word order variation in the same way in both monolingual and bilingual Andean Spanish. If were the case that each pronominal category also conditioned syntactic variation in the same way, the distribution of OV/VO orders would be expected to be approximately the same across all pronoun types within both groups. However, this is not the case. In both monolingual and bilingual Andean Spanish, personal pronouns are collocated in pre-verbal position in about a third of the tokens, while existential pronouns (e.g., *nada*, ‘nothing’) appear in canonical VO order nearly categorically; however, demonstrative and universal pronouns exhibit a much higher preference for OV order among bilinguals compared to monolinguals. In fact, for bilinguals, demonstrative and universal pronouns exhibit the highest rates of OV order. To determine whether the syntactic behavior of the pronominals is significantly different across groups,

I removed demonstrative and universal pronouns from the monolingual and bilingual datasets, then ran a logistic regression analysis to determine whether there was a significant effect of group (i.e., bilingual versus monolingual) on predicate constituent order variation. When only personal and existential pronouns are included, monolinguals and bilinguals are not significantly different from one another ( $p = 0.581$ ). However, when only demonstrative and universal pronouns are included, there is a significant difference between bilingual and monolingual Andean Spanish groups ( $p = 0.0011$ ).

All pronouns in Quechua exhibit higher rates of OV order because OV is the dominant order overall, and pronouns exhibit an even stronger preference for OV order compared to other objects. The highest rate of OV order for Quechua pronouns corresponds to universal pronouns, followed by demonstratives and existential pronouns. Perhaps the prevalence of OV order among demonstrative and universal pronouns in Quechua has been ‘transferred’ to bilingual Andean Spanish, especially considering that, of all linguistic factor levels evaluated in the present study, demonstrative and universal pronouns exhibit the highest rates of OV order, which is more than three times the proportion of OV order reported for the bilingual group overall (13.2%). Additionally, demonstrative and universal pronouns alone account for 23.2% of all OV tokens, even though they make up only 7.1% of all tokens in the bilingual corpus. To use stronger language, the distribution of OV/VO orders across pronominal category constitute tentative evidence of structural convergence regarding the pre-verbal collocation of demonstrative and universal pronouns in bilingual Andean Spanish and Quechua.

Lastly, through a qualitative observation of demonstrative and universal pronouns configured in OV order, I find several examples of the discursive situations that F. A. Ocampo and Klee (1995) document, particularly summary. For example, it is the case that the medial demonstrative pronouns *eso* and *chay*, ‘that’, are often used to summarize prior discourse in the corresponding corpora. Summarizing is a general function of the demonstrative pronoun in non-contact Spanish as well, but it appears that these pronouns in conjunction with OV order are used as a stylistic device to summarize previous discourse more frequently in bilingual Andean Spanish and Quechua. Consider (123) below:

- (123) a. *buk kuti ñuqa-qa musqu-yku-sqa-ni... buq uh*  
 one time I-TOP dream-INT-PST2-1SG a uh  
*wasi-ta-cha ranti-ri-ku-sqa-ni munaycha-ta*  
 house-ACC-DIM buy-INCP-REFL-PST2-1SG cute-ADV  
*chay animal-cha-kuna-wan jardin-cha-ta planta-cha-kuna-wan*  
 that animal-DIM-PL-INS garden-DIM-ACC plant-DIM-PL-INS  
*me gusta-wa-n ñuqa-man... chay-pi...*  
 OBJ.1SG like-OBJ.1SG-3SG I-DAT that-LOC  
*tiya-sba-ra-ni... feliz... papa-y mama-y...*  
 live-PROG-PST-1SG happy father-POSS.1SG mother-POSS.1SG  
*nishu-ta wikch'u-wa-nku wasi-manta entonces...*  
 very-ADV throw.out-OBJ.1SG-3PL house-ABL then...  
*chaymi **chay-ta-n** musqu-ru-ku-sqa-ni*  
 so that-ACC-EVIDI dream-EXH-REFL-PST2-1SG

‘One time, I dreamt that... I bought myself a little house with those beautiful little animals, a little garden with little plants that I like... there... I was living... happy... my mom and my dad really threw me out from the house then... so I dreamt that.’ [P1]

- b. *Ah pago a la tierra, a ver, yo, yo la verdad no vi tanto... creo que hicieron un pago a la tierra por mi tío el que se fracturó su pierna, ajá, hicieron despacho y todo eso... ah ese despacho se trata de quemar ¿no? quemar ese, una fogata en ahí quemar diferentes plantas, digamos, y un fetus de una oveja, algo por ahí... después lo botan al cerro, lo tienen que armar lo más lejos posible de la este, de de la ciudad... lo botan... así lo botan y según eso, ya está, eso es todo el pago a la tierra... ajá más o menos, así, eso, **eso** hizo mis tíos y mi papá para que se, para que se sane mi tío más o menos*

‘Ah payment to the earth, let’s see, I, I truthfully I didn’t see much... I think that they did a payment to the earth for my uncle that fractured his leg, uh huh, they did an offering and everything... it involves burning the offering, right? burn this, a campfire, in there burn different plants, let’s say, and a sheep’s fetus, something like that... later they throw it out to the mountain, they have to [put] it as far away from the city as possible... they throw it out... they throw it out like this and according to that, that’s it, this is the whole payment to the earth.. uh huh more or less, so my dad and my uncles did that so that, so that my uncle healed more or less’ [Pto]

In both (123a) and (123b), the respective demonstrative pronouns, *chaytan* and *eso*, are used to summarize the preceding discourse. In the Quechua excerpt, *chaytan* summarizes the dream that the participant is describing, and, in the Andean Spanish excerpt, *eso* summarizes the payment to the earth ceremony. Though not exemplified here, the universal pronoun *todo* is often applied with a similar pragmatic function, i.e., to recapitulate previous discourse.

Departing from F. A. Ocampo and Klee (1995), I concede that universal and demonstrative pronouns may also summarize preceding discourse in post-verbal position in all languages varieties investigated in the present study. But, given their frequent collocation in OV order in bilingual Andean Spanish and Quechua, I argue that pre-verbal demonstrative universal pronouns constitute a convergent structure in the Cusco Quechua-Spanish contact situation. That demonstrative and universal pronouns do not exhibit the same syntactic distribution in monolingual Andean Spanish suggests that the direction of this transfer is from Quechua to Andean Spanish. This claim is further supported by the observation that Quechua influence on Andean Spanish is more prevalent in the interior structure, which concerns the semantic and pragmatic organization

of Andean Spanish (Muntendam & Muysken, 2021; Muysken & Muntendam, 2016).

### 7.6.2 A closer look at subject expression

Recall that in both bilingual Andean Spanish and Quechua an overt subject increases the likelihood of VO order, which in Quechua is non-canonical and in Spanish is canonical. That monolingual Andean Spanish is also conditioned in the same way by the same variable may indicate that the direction of transfer is from Andean Spanish to Quechua, unlike the syntactic behavior of pronominals.

A comparison of the raw and relative frequency of all possible word orders in Table 7.15 reveals that there is a strong preference for SVO order over other VO-type orders with an explicit subject, like VSO and VOS. This may substantiate my earlier suggestion that when the subject is explicit, the preferred structure is one in which the verb intervenes between the subject and object to facilitate the disambiguation of subject and object.

	OV Type				VO Type				Total
	OV	SOV	OSV	OVS	VO	SVO	VSO	VOS	
<b>Monolingual Andean Spanish</b>									
<i>n</i>	77	4	5	6	963	256	5	13	<b>1,329</b>
%	5.6%	0.3%	0.4%	0.5%	72.5%	19.3%	0.4%	1.0%	<b>100.0%</b>
<b>Bilingual Andean Spanish</b>									
<i>n</i>	251	12	34	14	1,494	524	9	20	<b>2,358</b>
%	10.7%	0.5%	1.4%	0.6%	63.4%	22.2%	0.4%	0.8%	<b>100.0%</b>
<b>Quechua</b>									
<i>n</i>	659	108	35	25	257	113	5	10	<b>1,212</b>
%	54.4%	8.9%	2.9%	2.1%	21.2%	9.3%	0.4%	0.8%	<b>100.0%</b>

Table 7.15: Clausal constituent order in Andean Spanish and Quechua

The values in Table 7.15 may also suggest that the elevated rate of SVO in the bilingual Cusco Quechua corpus is an outcome of surface-level pattern replication. To elaborate, when only tokens with explicit subjects are considered, SVO order is realized at a rate of 85.5% in bilingual Andean Spanish. Comparing this figure to the rate of SOV order, 1.9%, SVO order is realized 45 times more frequently than the alternative subject-initial order. The discrepancy between subject-initial orders is even higher for Andean Spanish monolingual speakers— SVO is approximately 63 times more frequent than SOV order. If structural borrowing is responsible for the high incidence of SVO structures

in Quechua, it follows that this order is the result of the transfer of the most frequent surface-level pattern, SVO. In other words, if bilingual Quechua is influenced by Andean Spanish, when a subject is explicit, the bilingual speaker has the best chance of aligning with Andean Spanish syntax by employing SVO order.

Furthermore, that SVO is the most frequent order in the Quechua corpus may be signaling a forthcoming typological shift in basic word order from SOV to SVO, or perhaps to a language exhibiting two dominant orders.<sup>259</sup> Previous studies on Quechua suggest that a shift in basic word order may engender other structural alterations. For instance, Sánchez (2003) reports that a high rate of SVO in Lamas and Ulcumayo Quechua among child bilinguals coincides with attrition of the accusative case marker, *-ta*. Kalt and Geary (2021) also report a relationship between stricter word order and morphological loss; however, in their sample of bilingual Chuquisaca Quechua, *ta*-omission correlates with a predilection for prescriptive OV order. Though *ta*-marking was not a significant predictor of syntactic variation in the present study, I found that OV word order is employed more frequently in the absence of other morphological markers, like the negative particle *-chu*. Additionally, in §7.5 I reported a significant correlation between *ta*-omission and age such that *ta*-less objects are associated with a lower mean age than objects marked with *-ta*. Based on the trends uncovered in the bilingual Cusco Quechua corpus, it is quite possible that rates of VO order, including SVO order, will continue to rise in conjunction with *ta*-omission, ultimately yielding a variety of bilingual Quechua that structurally resembles the synthetic language with which it is in contact.

In conclusion, given the existence of both convergent syntactic structures (e.g., pre-verbal demonstrative and universal pronouns and VO order with explicit subjects) and a clear change in progress, I argue that OV/VO variation in bilingual Andean Spanish and Cusco Quechua constitutes contact-induced change. Further support for this conclusion comes from the evidence of intraspeaker convergence revealed by 1) a significant effect of first language and linguistic dominance (BLP scores) among the Quechua and bilingual Andean Spanish groups respectively, and 2) a direct comparison of the proportions of OV and VO orders in the speech samples of both languages among the bilingual speakers who provided an interview in Andean Spanish and Quechua (see Figure 7.6). I do not interpret the finding that OV/VO variation is conditioned by the same linguistic predictors in monolingual and bilingual Andean Spanish as evidence that contradicts the existence of contact-induced change because 1) monolinguals and bilinguals differ with respect to the extralinguistic predictors that motivate syntactic variation and 2) monolingual Andean Spanish is not

<sup>259</sup> Dryer and Haspelmath (2013) classify 67 languages as having two primary alternating orders. Most of these languages alternate between SOV and SVO order (43.3%).

a true non-contact variety of Spanish. Of course, claims of contact-induced change could be further bolstered through a comparison between monolingual and bilingual Cusco Quechua, but logistical challenges, like accessing monolingual Quechua populations, has prevented a comparison of this nature at this time. However, given the claim in various Quechua grammars that SVO is the dominant order in Quechua and the finding that L1 Quechua speakers employ OV more frequently than simultaneous bilinguals, I conclude that OV/VO variation in bilingual Quechua is an instance of contact-induced change.

## 7.7 Summary

This chapter has supplied responses to the research questions guiding the present study through a discussion of the statistical results. In §7.2, I concluded that Andean Spanish bilinguals exhibit higher rates of non-canonical OV order (13.2%) than monolinguals, who employ OV order at a rate comparable to those reported for non-contact varieties, like Río Platense Spanish (7.1%). In comparing the results of the present study to the previous literature on OV/VO variation in Andean Spanish, I find that OV order is noticeably less frequent among the bilingual participants of the present study. I partially attribute the numerical discrepancy to methodological differences, but hypothesize that the decrease in non-canonical OV order over time is an ancillary outcome of contemporary social processes like globalization and urbanization. This theory is corroborated in §7.3 through the finding that age and OV order frequency are positively correlated. In fact, upon removing the youngest generation of speakers from the bilingual Andean Spanish corpus, the overall rate of OV order is equivalent to those reported in studies conducted one to three decades prior to the present study.

In §7.3, I address the extralinguistic and linguistic variables selected as significant predictors of predicate constituent order variation in monolingual and bilingual Andean Spanish. Incidentally, OV/VO variation is conditioned by discrete extralinguistic predictors in each group. For bilinguals, both age and linguistic dominance are inversely correlated with VO order likelihood respectively, suggesting 1) a change in progress characterized by a burgeoning partiality for the prescriptive norm and 2) an idiolectal effect of language contact. On the other hand, the syntactic behavior of predicate constituents is conditioned by the place of residence and level of education among Andean Spanish monolinguals. Participants residing outside the city limits of Cusco (i.e., more rural locales) employ OV order more frequently than inhabitants of metropolitan Cusco, ostensibly because these individuals dwell within primarily Quechua-

speaking communities, and thus have stronger social ties with bilingual individuals whose speech is impacted by Quechua contact. Though education was also selected as a significant predictor, I ultimately concluded that this result is likely attributable to unbalanced factor levels, especially considering that the monolingual group is highly educated overall.

Moreover, it was determined that the same four linguistic predictors condition predicate constituent order variation in both the bilingual and monolingual groups— information status of the object, definiteness, polarity, and subject expression. In essence, given, pronominal, affirmative, and null subject factor levels favor non-canonical OV order. The findings regarding information status of the object and subject expression are mostly corroborated by previous studies, but the reported conditioning effect of definiteness and polarity are novel, to my knowledge. Definiteness is the strongest predictor of non-canonical order, which is evidenced by the fact that pronominals are associated with the highest rates of OV order in both bilingual and monolingual groups compared to all other factor levels. The statistical analysis of linguistic predictors ultimately reveals that though bilinguals employ OV order at a significantly higher rate than monolinguals, the linguistic contexts that trigger non-canonical order are congruous.

In §7.4, I relay the proportions of OV/VO orders reported for Cusco Quechua bilinguals. According to the frequency criteria of determining dominant order, bilingual Cusco Quechua may be classified as an OV-dominant language overall; however, nearly half (46.4%) of the participants employ VO order at a rate that suggests a lack of dominant order on the idiolectal level. Another critical observation discussed in this section is that, when sentences with all explicit clausal constituents (i.e., subject, object, and verb) are considered, SOV is not the most frequent order, much less the dominant order. In fact, SVO is the most frequent order, followed closely by prescriptive SOV. Based on the reported frequencies, I conclude that bilingual Cusco Quechua is lacking a dominant order of these three clausal elements. The structural consequences of this apparent typological shift in progress will be addressed in the final chapter.

Regarding the effect of independent predictors on syntactic variation in the bilingual Cusco Quechua corpus, I uncovered the same inverse correlation between age and VO order likelihood that was uncovered in bilingual Andean Spanish in §7.5. On the basis of this observation, I again presume that there is a change in progress such that non-canonical VO order is displacing canonical OV order over time in Quechua. Additionally, based on the finding that L1 Quechua bilinguals are associated with a higher likelihood of canonical order than simultaneous bilinguals, I conclude that language contact, is partially

responsible for the increase in non-canonical structures in bilingual Quechua. The evaluated linguistic predictors that were found to condition predicate constituent order variation in bilingual Quechua are information status, definiteness, subject expression, polarity, and absolute weight of the object. In general, VO order likelihood increases when the object referent is heavy, new, and non-pronominal in an affirmative clause with an overt subject. In comparison to Andean Spanish, some of these predictors exercise the opposite effect on syntactic variation in Quechua, possibly as a result of its head-directionality.<sup>260</sup> For instance, the given before new preference established in head-initial languages, does not apply in Quechua, as new referents are more likely to appear in earlier positions. However, other factors, like weight and definiteness, influence predicate constituent order in the manner that has been documented for head-initial languages— heavy objects tend to be collocated in later (i.e., post-verbal) positions, and more definite referents (e.g. pronominals) appear in earlier positions.

The last section (§7.6) features a comparison of the three varieties. First, I reiterate that monolingual and bilingual Andean Spanish are conditioned in the same way by the same linguistic predictors, suggesting that the most prominent difference between these two varieties is the frequency with which non-canonical structures are employed, not the contexts that invite non-canonical structures. Second, I ascertain that two linguistic predictors— definiteness and subject expression— exert the same effect on predicate constituent order in bilingual Andean Spanish and Cusco Quechua. Upon taking a closer look at pronominal objects, I find that OV order is highly preferred when the pronominal object is demonstrative or universally quantified. Because this same preference does not pertain to monolingual Andean Spanish, I theorize that the collocation of demonstrative and universal pronouns in pre-verbal position is a structure originating in Quechua. Similarly, VO order is the preferred arrangement of predicate constituents when there is an expressed subject in all three varieties. Because explicit subjects inhibit syntactic variation in monolingual Andean Spanish as well, I claim that the direction of transfer is from Spanish to Quechua, which is bolstered by the fact that SOV is reportedly the dominant order of Quechua. Ultimately, I argue that language contact in this region has resulted in contact-induced syntactic change in bilingual Cusco Quechua and Andean Spanish. This is evidenced by the emergence of two convergent structures (e.g., pre-verbal demonstrative/universal pronominals and SVO order) and intra-speaker convergence, which is conveyed through a positive correlation between VO order in Andean Spanish and Quechua and OV order in both languages among bilingual participants. The latter finding suggests that the

<sup>260</sup> Though it has reported that there is no dominant order of clausal constituents in this corpus of bilingual Quechua, because OV order remains dominant, I assume that Quechua may still be classified as a head-final language.

two syntactic systems available to the bilingual speaker influence one another at the level of the individual.

# CHAPTER 8

## CONCLUSION

### 8.1 Introduction

In this final chapter, I offer my concluding marks. First, I highlight the novel contributions of the present study in §8.2, which include 1) the identification of predicate constituent order variation in Andean Spanish and Quechua as a change in progress, 2) the evidence of convergence in bilingual speech, and 3) the consideration of linguistic predictors that govern word order variation in other languages but had not yet been evaluated in the previous literature on Andean Spanish and Quechua (animacy, definiteness, and weight). In §8.3, I discuss the implications of a hypothetical typological shift in the basic word order of Quechua, as the rising prevalence of VO-type orders overtime justifies such a discussion. Next, I propose directions for future scholarship, which are motivated by some of the challenges encountered in the data codification process applied in the present study. The chapter concludes with a cohesive summary of the dissertation.

### 8.2 Contributions

Though the discrete findings revealed by the statistical analyses are numerous, in this section I highlight only the novel contributions of the present study.

#### 8.2.1 Change in progress

The first of these novel contributions is the discovery of a significant negative correlation between age and likelihood of VO order in both contact languages. In bilingual Andean Spanish, this relationship indicates that relatively younger speakers realize canonical order (VO) comparatively more frequently than older

speakers. Conversely, in bilingual Cusco Quechua, relatively younger speakers are associated with higher rates of non-canonical order (VO) than older speakers, who employ canonical order (OV) more frequently. I interpret these significant trends as evidence of a change in progress characterized by two concurrent processes: the patterns of use regarding predicate constituent order in Andean Spanish are approximating those of the prescriptive norm exhibited in non-contact varieties and the patterns of use in Quechua are approximating those of Andean Spanish, the language with which it is in contact.

I will now contextualize this finding within the Quechua-Spanish contact situation, as described in §1.3.2. Recall from Chapter 1 that situations of language shift typically invoke the mechanism of substratum transfer (i.e., group second-language acquisition) whereby learner varieties of the shifting speakers are imitated by subsequent generations (i.e., offspring of the shifting group) and native speakers of the target language (Muysken, 1984; Thomason & Kaufman, 1988; Winford, 2003). On the other hand, in situations of language maintenance, structures are borrowed from the source language into the recipient language via the agency of recipient language speakers. The linguistic outcomes of language shift (substratum influence) and language maintenance (borrowing) do not follow the same path. Situations of language shift are theorized to engender structural transfer first, while in situations of language maintenance, the lexicon is the first linguistic element to be borrowed, followed by structural components. The intensity of contact, particularly the duration of contact, yields disparate results that depend directly on the type of contact situation. In language shift, a quick shift (e.g., one generation) is likely to yield a higher degree of structural influence on the recipient language than a longer shift (e.g., several generations) because the latter timeline permits an intermediary stage of bilingualism, and consequently, perfect learning on behalf of the shifting group. In situations of language maintenance, on the other hand, a longer duration of contact may yield extensive bilingualism, which promotes borrowing. This is because bilingualism is a necessary condition for more intense (i.e., structural) borrowing, since recipient language speakers must be proficient enough in the source language to isolate and transfer these structures.

Based on the directionality of the change in progress regarding predicate constituent order variation in both contact languages, I advance the theory that the Quechua-Spanish contact situation may be classified as both language shift and language maintenance.<sup>261</sup> I contend that the mechanism of contact-induced change guiding transfer from Quechua to Andean Spanish is substratum influence, while structural borrowing is responsible for transfer in the reverse direction. Recall that intense contact between Quechua and Spanish did not

<sup>261</sup> Refer to the example of Yiddish and English contact described in §1.2.1.

begin until the mid-20th century (Escobar, 2011; Klee & Lynch, 2009), at which point Quechua monolingualism in the region was quickly displaced by Spanish monolingualism and Quechua-Spanish bilingualism. According to census data, within the last few decades, the displacement of Quechua has slowed, and the current contact situation is characterized by stable bilingualism. In short, an initial shift from Spanish to Quechua has been interrupted by an intermediary stage of bilingualism.

This characterization of the Quechua-Spanish contact situation may be consulted in predicting the types of contact-induced changes that prevail in each language. It may be predicted that Andean Spanish— a contact variety originally formed through the process of group L2 acquisition— was initially characterized by a preponderance of Quechua-like structures (e.g., OV order) at the beginning of the shifting period. However, as the displacement of Quechua eased, bilingualism emerged, and Andean Spanish began gaining native speakers of its own, who were also exposed to non-contact varieties of Spanish through contemporary processes and conventions (e.g., greater access to mass media and the modern invention of the internet), the incidence of Quechua features in the contact language likely began to recede. This theory is supported by the gradual decline in the rate of non-canonical OV order evidenced by both the bilingual Andean Spanish age analysis and the comparison of OV order rates between the present study and studies conducted one or more decades ago (Klee, 1996; Muntendam, 2008, 2009, 2010, 2013; Muysken, 1984; F. A. Ocampo & Klee, 1995). This theory of Quechua-Spanish contact is further supported by studies on dialect contact in Lima, which have concluded that most linguistic features of Andean Spanish disappear within a generation upon migration to an urban center (Caravedo & Klee, 2012; Klee & Caravedo, 2005, 2006; Klee et al., 2011; Rojas-Sosa, 2008).

Regarding the effect of contact on Quechua, there was likely little change to its linguistic system at first given the direction of the shift (i.e., toward Spanish, away from Quechua). However, as bilingualism emerged and strengthened in the region, the proficiency of bilingual speakers would have facilitated the transfer of Andean Spanish elements into Quechua, beginning with the integration of lexical items and followed by structural borrowings. This trajectory is borne out by the results of the Quechua age analysis, which demonstrate that older speakers of Quechua employ canonical OV order more frequently than their younger counterparts. That Spanish-like structures have become more pervasive in Quechua overtime supports the notion that borrowing is the relevant mechanism of contact-induced change regarding cross-linguistic transfer from Andean Spanish to Quechua.

<sup>262</sup> This figure should not be interpreted as the general percentage of loanwords in bilingual Cusco Quechua, as it represents only the presence of loanwords within the predicate.

The hypothesis that contact-induced change is mediated by the borrowing mechanism in Quechua and by substratum influence in Andean Spanish is further supported by the disparate proportion of loanwords in each language. In the bilingual Cusco Quechua dataset, 42.7% of all predicates contain a Spanish loanword in the object constituent, verb constituent, or both.<sup>262</sup> Examples of Spanish loanwords in the Quechua corpus include novel terms that do not have a Quechua equivalent, like *turismo*, ‘tourism’, and more basic terms that do have a Quechua equivalent (e.g., *tomay*, ‘to drink’ on the model of Spanish *tomar* in lieu of Quechua *ukyay*). In the bilingual Andean Spanish dataset, on the other hand, the presence of Quechua loanwords was ultimately not included as an independent variable due to the paucity of lexical borrowings. Less than 2% ( $n = 45$ ) of all tokens in the bilingual dataset contained a loanword in one or both of the predicate constituents. Where Quechua lexical borrowings did appear in the Andean Spanish corpus, they were typically terms related to cultural concepts, e.g., *ayni*, ‘reciprocal labor’; *Inti Raymi*, ‘the Sun Festival’; and *Apu*, ‘mountain deity’. Again, the ubiquity of Spanish lexical borrowings in Quechua lends credence to the view that transfer from Spanish to Quechua is best described as contact-induced change via borrowing, as structural changes are preceded and mediated by lexical transfer in situations of language maintenance (Thomason & Kaufman, 1988; Winford, 2003). The modest presence of Quechua loanwords in bilingual Andean Spanish does not indicate a lack of contact-induced change, rather provides evidence of contact-induced change via substratum influence, as shifting speakers tend to first acquire the target language vocabulary and map the proper lexicon onto source language grammatical patterns (Thomason & Kaufman, 1988; Winford, 2003).

To summarize, I have applied the insights offered by an age analysis to suggest that predicate constituent order variation in bilingual Andean Spanish and Cusco Quechua is guided by a change in progress such that the structure of Andean Spanish is moving toward the non-contact, prescriptive norm, while bilingual Cusco Quechua is moving away from the prescriptive norm and toward a higher prevalence of the Spanish-like structure. I have cited this change in progress in the characterization of the Quechua-Spanish contact situation as one in which there was an initial quick shift from Quechua to Spanish that has been interrupted by a period of bilingualism, which may be partially attributed to modern efforts toward Quechua language maintenance by institutions and individuals (Escobar, 2011). The initial shift to Spanish resulted in the formation of the Andean Spanish contact variety, which was likely originally characterized by a more notable presence of Quechua-like structures (e.g., OV order); however, as the native speakers of Andean Spanish participate in dialect mixing with

speakers of non-Andean varieties of Spanish in the present day, consequently, the incidence of non-canonical, Quechua-like features declines. The opposite trajectory defines Quechua— canonical order has become less frequent over time, suggesting that the Quechua language is subject to structural borrowing mediated by both lexical borrowing and robust bilingualism. On the basis of these trends, I hypothesize that predicate constituent order variation in Andean Spanish will continue to approximate the non-contact standard over time, but canonical order in Quechua will continue to be displaced by Spanish-like structures, possibly to the point of a change in basic word order, as is the case in Lamas Quechua (Sánchez, 2003). In §8.3, I briefly discuss the typological consequences of such a shift.

## 8.2.2 Convergence in bilingual speech

### Convergent structures

Through a comparative analysis of the significant linguistic predictors governing predicate constituent order variation, I identified two convergent structures. By convergent structures I refer to instances in which a particular linguistic feature (e.g., definiteness or subject expression) exerts a similar effect on the syntax of predicate constituents in both languages. This is the case for demonstrative and universally quantified pronouns, which exhibit a preference for OV order in both bilingual Andean Spanish and Quechua. Though, admittedly, this syntactic preference is not categorical in Andean Spanish, demonstrative and universally-quantified pronouns (e.g., *eso*, ‘that’ and *todo*, ‘everything’, respectively) are configured in OV order nearly four times as frequently as all other tokens (42.9% versus 10.9%). This disparity is substantial, especially considering that VO is the canonical order of Andean Spanish and that adherence to canonical order is stricter in Andean Spanish than in Quechua. In Quechua, demonstrative and universally-quantified pronouns are realized in OV order at a rate of 88.6%, which is higher than the rate of OV order among the remaining tokens, 65.6%. Of course, this contrast is not as stark as that of Andean Spanish given that OV is the canonical order, but the difference between these two categories is significant ( $X^2 = 30.20, p < 0.05$ ). In monolingual Andean Spanish, personal pronouns are associated with the highest rate of pre-verbal collocation, followed by demonstrative, then universal pronouns. Because the rate of OV order from highest to lowest does not exhibit the same arrangement with respect to pronomial category in the monolingual and bilingual Andean Spanish groups (i.e., OV order rate is highest for personal pronouns among monolinguals but demonstrative pronouns among bilinguals), I speculate that the direction of

this contact-induced change is from Quechua to Andean Spanish. I recognize, however, that demonstrative and universal pronouns are still associated with elevated rates of OV order in monolingual Andean Spanish in comparison to the overall rate of OV order (23.1% versus 7.1%). A cross-variety comparison of monolingual Andean Spanish and non-contact Spanish is required to clarify whether these elevated rates are characteristic of Andean Spanish specifically, or Spanish more generally.

The second convergent structure identified in the present study involves subject expression. In all three varieties examined— monolingual Andean Spanish, bilingual Andean Spanish, and bilingual Cusco Quechua— the presence of an explicit subject exerts a significant effect on predicate constituent ordering such that the odds of VO order are higher for overt subjects than for null subjects. Because monolingual and bilingual Andean Spanish exhibit the same pattern regarding subject expression, I theorize that this structure originates in Spanish and has subsequently been transferred into Quechua. This theory is corroborated by Muysken's (1984) observation that OV order is restricted by the presence of a subject in all varieties of Spanish. He argues that the relatively high rate of OV order in Andean Spanish without an explicit subject results from the imitation of a surface-level word order pattern in Quechua through a syntactic mechanism already available in Spanish— preposing of the object constituent. I speculate that a similar mechanism may account for the high incidence of SVO order in Quechua— the transfer of a surface-level syntactic pattern through the imitation of frequent structures.

### **Intra-speaker convergence**

Convergence in bilingual speech is also found at the idiolectal level, that is, within the bilingual individual. In §7.6 of the previous chapter, I found that there is a significant correlation between the rate of OV order in bilingual Andean Spanish and OV order in Cusco Quechua,<sup>263</sup> which suggests that the two linguistic systems within the bilingual speaker impact one another in that they converge on similar structural patterns. Indispensable to a discussion on intra-speaker convergence is the application of this term by Bullock and Toribio (2004): "it is appropriate to apply the notion of convergence to bilingual grammars as the structural congruencies surfacing from the contact between a bilingual's two grammars can and do provide synchronic evidence of the processes that ultimately lead to linguistic convergence, even if those processes are incomplete or variable in their application" (p. 91). Thus, the authors contend that the trend of structural convergence exhibited within the bilingual speaker may tentatively foreshadow a growing structural convergence between general

<sup>263</sup> The same relationship applies to VO order rate in Andean Spanish and Quechua.

linguistic systems. It does not appear that this is the case for Andean Spanish and bilingual Cusco Quechua, however, because, as predicate constituent order in Quechua approximates Andean Spanish, Andean Spanish moves in the opposite direction toward the prescriptive Spanish norm.

### 8.2.3 Analysis of new linguistic predictors

The last novel contribution of the present study that I highlight in this section is the consideration of certain linguistic predictors that have not been evaluated quantitatively in Andean Spanish and Quechua. These variables— animacy, definiteness,<sup>264</sup> and weight— have been identified as conditioning factors of syntactic variation in other languages. Animacy was not selected as a significant predictor of predicate constituent order variation in either Andean Spanish or bilingual Cusco Quechua despite the findings of past studies, which state that animate entities appear in earlier positions due to their heightened predictability and conceptual accessibility (J. K. Bock & Warren, 1985; Branigan et al., 2008; McDonald et al., 1993). The lack of an animacy effect may be interpreted in a couple ways. First, it is possible that the animacy of a referent governs the relative order of two-constituent configurations outside the verbal domain (e.g., scrambling in the pre-verbal space in Dutch and the genitive phrase in English), but not within the verbal domain (i.e., the relative position of verb and object) (Rosenbach, 2005; van Bergen & de Swart, 2009, 2010). Second, animacy may simply be a factor that does not affect syntactic ordering preferences in the specific languages under investigation. Additional research on the effect of animacy on variable order in other domains in Andean Spanish and Quechua are need to confirm this possibility.

Though weight was not found to exert a significant effect on predicate constituent order variation in Andean Spanish,<sup>265</sup> I observe a positive correlation between the absolute weight of an object constituent and the likelihood of VO order in bilingual Cusco Quechua. Particularly, heavier objects are more likely to exhibit non-canonical, post-verbal collocation than lighter objects. The contributions of this finding to the existing cross-linguistic literature on the effects of syntactic weight are two-fold. First, this finding challenges comprehension-oriented accounts of weight-based variation (e.g., the Minimize Domain principle proposed by Hawkins, 1994, 2004), which predict a long-before-short preference for head-final languages, but offers evidence in support of production-oriented accounts, which suggest that lighter constituents are produced earlier irrespective of head directionality. According to these theories, considering that sentence planning is an online task in which speech production occurs incrementally, shorter constituents appear earlier because they are more accessible

<sup>264</sup> Since I cover the effect of definiteness in the previous section, I discuss only animacy and weight in this section.

<sup>265</sup> However, weight is a significant predictor of syntactic variation in the post-verbal domain regarding the relative order of two sets of two-constituent configurations in Spanish: a subject-oriented depictive and object NP or locative NP (Heidinger, 2013).

and, thus, more easily retrieved (J. K. Bock & Warren, 1985; McDonald et al., 1993; Yao, 2018). Second, the present study provides an investigation of weight effects in an agglutinative language, which is scarce in the extant literature on weight-based phenomena. Due to the high degree of synthesis that typologically characterizes Quechua (Hintz, 2009, 2016), I ultimately chose to operationalize weight as the number of syllables. Upon comparing the effect of weight using two different metrics— number of words and number of syllables— I found that both exert an independent significant effect on predicate constituent order variation in Quechua. However, only weight conceptualized as number of syllables was chosen as a significant predictor in the logistic regression analysis. This finding contradicts the assertion that all weight-based measurements are essentially equivalent (Heidinger, 2013; Wasow, 1997, 2002) and ultimately points to the need for further research on the operationalization of weight across languages of various typologies.

### **8.3 Implications of potential typological shift**

The findings of the present study clearly do not indicate a shift in basic constituent order for Andean Spanish. In fact, according to the results of the age analysis recapitulated above, the distribution of predicate constituent orders is trending toward that of non-contact varieties of Spanish (i.e., the rate of VO order is increasing over time). However, this is not the case for bilingual Cusco Quechua. Though the current distribution of OV/VO orders indicates that this variety of Quechua is still most appropriately classified as OV-dominant, an inspection of the distribution of clausal constituent orders suggests that bilingual Cusco Quechua lacks a dominant order of subject, object, and verb, at least on the basis of the frequency criteria for determining dominant word order (Dryer, 2013a). This observation, taken together with the results of the Quechua age analysis, suggests that word order in Quechua appears to be shifting from an SOV-dominant and OV-dominant language to one lacking dominant order. As mentioned before, it is unclear at this point if word order in bilingual Cusco Quechua has reached its final destination or if the data presented in this study represent merely a brief stop on the way to a complete shift to SVO- and VO-dominant orders, as documented in bilingual Lamas Quechua (Sánchez, 2003). Either way, the presented trajectory of predicate constituent order variation in Quechua warrants a general discussion of the possible typological implications of such a shift.

To exemplify the potential typological implications of word order shift in Quechua, I briefly review past literature on auxiliaries and verbal suffix forma-

tion (Hintz, 2009, 2016). In Quechua, tense, aspect and other grammatical categories are encoded via verbal suffixes, and auxiliation is a crucial step in the generation of these verbal suffixes. The process of verbal suffix formation is exemplified by the below schema (124) adapted from Hintz (2016, p. 322):

(124) stem-nominalizer auxiliary → stem-nominalizer=auxiliary → stem-suffix

This schema illustrates that the process of verbal suffix formation begins with a periphrastic verb phrase that consists of a lexical verb stem affixed with a nominalizing suffixing followed by an inflected auxiliary verb. Next, the auxiliary “may attach to the preceding nominalized complement as an enclitic” (Hintz, 2016, p. 322). Eventually, the nominalized suffix and stem of the auxiliary may be reanalyzed as a new verbal suffix. This is the supposed process by which the inchoative suffix, *-gri*, arose in Northern Quechua. In Cusco Quechua, the structure remains in its periphrastic form. Taken together, the examples of Cusco Quechua (125) and Northern Quechua (126) below exemplify the first and last stage of the formation of new verbal suffixes, respectively.

(125) *Manuel-cha-qa qhawa-q ri-rqa-n*  
 Manuel-DIM-TOP see-PURP go-PST-3SG  
 ‘Manuelito went to see.’

(Hintz, 2016, p. 324)

(126) *miku-gri-ni*  
 eat-INTENTION-1SG  
 ‘I am going to eat (I will eat).’

(Hintz, 2016, p. 324)

In (125), the nominalizing suffix, *-q*, is affixed to a lexical verb root and followed by an inflected auxiliary verb, *riy*, ‘to go’, to express purpose and movement. That is, ‘seeing’ is the purpose for ‘going’ in the above example. This structure has undergone a grammaticalization process of verbal suffix formation whereby the nominalizing suffix and auxiliary verb root, *-q + -ri* have been reanalyzed as a single verbal suffix, *-gri*, that encodes incipience or futurity.

Hintz (2009, 2016) argues that contact with an analytic language like Spanish may simultaneously beget a preference for periphrastic constructions containing auxiliaries (125) and interrupt the process of verbal suffix formation in Quechua. Regarding the former point, he claims that “while only one or more

auxiliaries have been reconstructed for Proto Quechua, at least seventeen auxiliaries are now attested across the language family” (p. 329). Regarding the latter point, in Spanish, the auxiliary verb in periphrastic verbal constructions precedes the nominalized lexical verb. For example, future actions may be expressed in Spanish via a periphrastic structure in which the auxiliary verb, *ir*, is followed by a lexical verb in its infinitive form, as illustrated in (127).

(127) AUX        NOM  
*voy a compra-r-lo*  
 go.ISG to buy-INF-3SG

‘I will buy it.’

(Hintz, 2016, p. 330)

As is the case with predicate constituents, the order of the auxiliary and nominalized verb in Spanish is the mirror-image of the canonical order of these elements in Quechua. However, Hintz observes that, as a result of language contact, the order auxiliary-nominalized verb is also observed in contemporary Quechua alongside the prescriptive order. This contact-induced order of verbal elements is observed in the purposive constructions removed from the bilingual Cusco Quechua dataset, as well. For instance, in (128a), the inflected auxiliary, *riy*, precedes the verb affixed with the purposive nominalizing suffix, *-q*. This same order is exhibited in (128b) with an additional interrupting element—the object licensed by the lexical verb intervenes between the two elements of the periphrastic verb phrase. In the examples below, the nominalized verb is bolded and the inflected auxiliary is underlined.

(128) a. *niña ka-spa, huk kuti ri-ra-ni        **pastea-q***  
 girl be-SUB one time go-PST-1SG graze-PURP  
*waka-kuna-ta*  
 cow-PL-ACC

‘when I was a girl, one time I went **to graze** the cows’ [P1]

b. *binaspa manta-y        ri-ra-n        llama **maskha-q,***  
 thusly mother-POSS.1SG to-PST-3SG llama look.for-PURP  
*paqucha-yku        maskha-q*  
 alpaca-POSS.1PL.EXCL look.for-PURP

‘thusly, my mother went **to look** for [the] llama [and] look for our alpaca’ [P19]

Hintz (2016) summarizes the typological consequence of contact-induced variation in this syntactic domain in Quechua in the following manner: “[the] intervening material effectively blocks the fusion of the two elements into a single suffix, neutralizing this formerly productive environment for the renewal of verbal suffixes. As a result, the expression of aspect via periphrasis can no longer produce new verbal suffixes” (p. 333). These changes engender not only a typological shift in the syntactic system of Quechua, but also the morphological system. The non-canonical order of elements in periphrastic constructions, as well as their increased frequency, interrupts the process of new suffix formation, which is a hallmark of a polysynthetic language like Quechua. Thus, suffixes become less integral in marking grammatical relationships, and the ultimate result is a “greater reliance on the analytic morphological expression” (p. 334).

The most important takeaway from Hintz’s work for the present study is that changes in the syntactic patterning of a language may have larger typological implications. A shift from OV-dominant order to VO-dominant order (or no dominant order) is a typological reconfiguration in and of itself, which may also bring about other typological changes. Recall that, for example, a shift to SVO order in Lamas Quechua coincides with a high rate of omission of the accusative marker, *-ta* (Sánchez, 2003). Though Sánchez does not offer evidence to substantiate a causal relationship between these two processes, it is reasonable to hypothesize that, as VO order becomes more predominant in Quechua, its syntax takes on the burden of encoding grammatical relationships, which motivates the loss of accusative case-marking in order to minimize redundancy. In bilingual Cusco Quechua, *ta*-marking remains fairly robust (91.3%); however, as reported in §7.5, there is a significant difference between *ta*-marking and *ta*-omission with respect to the average age of the speaker such that the average age associated with *ta*-omission is lower than the average age associated with *ta*-marking. This suggests that, parallel to predicate constituent order variation, *ta*-marking may be declining over time. Consequently, I theorized that the syntactic mechanism for marking grammatical relationships (i.e., VO order) must be established prior to the loss of morphological markers that prescriptively bear this grammatical responsibility.

A shift to VO-dominant order may also cause or co-occur with syntactic reorganizations in other domains, like the reordering of the elements of periphrastic verbal constructions (Hintz, 2009, 2016). As mentioned in Chapter 2, Greenberg (1963) finds that basic word order is often associated with other typological properties within and across languages. For example SOV languages are often postpositional, while SVO languages are often prepositional. Furthermore, languages with prepositions tend to exhibit noun-genitive order in the

genitive phrase, and languages with postpositions tend to exhibit genitive-noun order. Thus, as the rate of VO and SVO order rises, the changing syntax at the clausal level may apply pressure on the syntax of other systems to achieve cross-category harmonization (Hawkins, 1980, 1982). This line of investigation is beyond the scope of the present study, but certainly offers an intriguing area for future scholarship.

## 8.4 Other considerations and future directions

Prior to providing my final remarks, I suggest a few areas for future scholarship motivated by some methodological challenges encountered in the data analysis process. Though the distinct typological profiles between the two languages under investigation is what motivates the present study, it is also the cause of certain limitations in the applied coding procedures. This was evidenced early on in both the exclusion of particular Quechua structures from the analysis that did not have a counterpart in Andean Spanish (e.g. purposive constructions and infinitive verb objects) and the codification of certain independent variables in the Quechua analysis that did not pertain to Andean Spanish (e.g., argument type and accusative marking). Though these adjustments are meant to improve the comparability of the Andean Spanish and Quechua datasets, they also illustrate that structurally, languages do not correspond to one another in a one-to-one fashion. However, in a quantitative variationist study, it is necessary to dilute some of the nuances of a language in the interest of conducting a coherent statistical analysis.

Codification challenges were most readily evident in determining the definiteness of a referent, particularly distinguishing definite from indefinite referents. In the Andean Spanish data, the presence of definite, demonstrative, and possessive determiners identified a referent as definite. Though demonstrative determiners and possessive suffixes may also convey definiteness in Quechua, the language is not equipped with a definite determiner. Sánchez (2003) claims that the accusative marker *-ta* has likely acquired “some determiner-like properties” (p. 100); however, it is not the case that *-ta* on its own marks definiteness or indefiniteness consistently, at least not in bilingual Cusco Quechua. For the sake of efficiency, all objects without a possessive suffix, demonstrative determiner, or topic marker (in some cases),<sup>266</sup> were considered indefinite in the codification process of the bilingual Cusco Quechua corpus. The simplification of this process may have resulted in the misidentification of certain referents, particularly the categorization of a definite referent as indefinite. Consider example (129):

<sup>266</sup> See §4.4.3.

(129) *qbari chay-ta sigui-pa-sqa, sigui-pa-sqa,*  
 man that-ACC follow-REG-PST2 follow-REG-PST2  
*sigui-pa-sqa uray, ukbu-man mayu pata-n-ta,*  
 follow-REG-PST2 below inside-ALL river above-POSS.3SG-ACC  
*apa-sqa chay puma... arrastra-sqa warmi-ta... arrastra-spa*  
 carry-PST2 that puma drag-PST2 woman-ACC drag-SUB  
*apa-yu-sqa*  
 carry-INT-PST2

‘the man followed, followed, followed that [path] above, inside, to the top of the river, [and] that puma carried... [he] dragged **the woman**, dragging [the woman], he carried her’ [P22]

In this example, the object, *warmi*, ‘woman’, is not accompanied by an overt marker of definiteness, so it was coded in the Quechua dataset as *indefinite*. However, in the larger discursive context, the participant is relaying a local lore about a pregnant woman who left her home at night and was subsequently snatch and killed by a puma. By zooming out, it is clear that the referent in this example, *warmi*, refers to a particular, identifiable woman. Thus, though the default interpretation in the absence of certain suffixes may be generic or indefinite (Mannheim et al., 2010), this is not the only possible interpretation. A more accurate disambiguation of definite from indefinite referents may require a nuanced qualitative analysis. This is not a practical procedure to apply to the evaluation of a large number of tokens for many predictors, which is characteristic of a quantitative, variationist analysis.

It is important to note that the same complication did not apply to the codification of pronouns, which are more straightforwardly classified. Furthermore, the rate of OV order reported for definite and indefinite referents in the Quechua dataset are quite similar, 61.3% and 66.7%, and the statistical model did not report a significant difference between these two factor levels. Thus, it may not be the case that these codification challenges had much bearing on the final results. Regardless, this study highlights the need for a better understanding of definiteness in Quechua and, potentially, other determiner-less languages, which may be addressed in future studies.<sup>267</sup>

As mentioned earlier, the operationalization of weight in Quechua constitutes another methodological challenge of the present study. Weight was determined using syllable counts in the Quechua dataset due to the assertion that weight metrics tend to correlated with one another (Wasow, 1997, 2002). In retrospect, it seems that another appropriate metric for measuring weight

<sup>267</sup> It is important to recognize that Kalt and Geary (2021) and Sánchez (2003) have both established that the numeral *buk* may be emerging as an indefinite determiner in certain regional varieties of Quechua.

in Quechua is the number of morphemes, given its classification as an agglutinative language (Cerrón-Palomino, 1987a). However, some suffixes or suffix combinations may prove challenging for this definition of weight. For example, would *-ni*— a phonological connector suffix that is used to join a suffix or root that ends in a consonant with another suffix that begins with a consonant— be considered a morpheme in its own right or combined with the suffix that precedes it? Furthermore, the habitual past in Quechua is formed by affixing the purposive suffix, *-q*, to a lexical verb root followed by the present tense conjugation of the copulative verb, *kay*, as illustrated in (130):

- (130) *puri-q*                      *ka-ni*  
 walk-HAB.PST be-1SG  
 ‘I would walk’

In counting the morphemes of this verbal structure, would *-q* and *ka-* count as two individual morphemes or a single morpheme, since together they express a single aspectual category? These inquiries illustrate some of the challenges that may accompany defining weight as the number of morphemes, even in an agglutinative language. Future research on determining a functional definition of weight for agglutinative languages, or perhaps even a definition of weight that may be universally applied, would certainly facilitate weight-related research.

Lastly, in the present study, I consider OV/VO variation in only main clauses in keeping with the methodological architecture of previous studies on Andean Spanish word order (Klee, 1996; Klee et al., 2011; Muntendam, 2009; F. A. Ocampo & Klee, 1995). However, future scholarship on word order variation in subordinate clauses, particularly in Quechua, may prove to be an insightful area of inquiry concerning the degree to which contact with Spanish has affected the syntactic structure of Quechua. According to Cerrón-Palomino (1987a), Quechua is an SVO-dominant language with relatively free word order in main clauses, but with rigid OV order in subordinate clauses. Given that Thomason and Kaufman (1988) posit that word order changes beyond the simple rearrangement of clausal constituents indicate more intense contact, inspecting the order of predicate constituents within the subordinate clause may offer a clearer picture of the nature of structural contact-induced change in Quechua.

## 8.5 Summary and final remarks

This investigation has provided an in-depth examination of predicate constituent order variation in the varieties of bilingual Quechua and Andean Spanish spoken in Cusco, Peru. I used conversational speech data elicited via sociolinguistic interviews to construct three separate corpora corresponding to the three varieties under investigation— monolingual Andean Spanish, bilingual Andean Spanish, and bilingual Cusco Quechua. From these corpora, I compiled three datasets comprised of all instances of predicates containing an explicit object and verb that conformed to the exclusion and inclusion criteria offered in Chapter 4. I conducted separate variationist analyses on each dataset to determine the extralinguistic and linguistic predictors that condition predicate constituent order variation, and later compared the results of these analyses to determine the role of contact-induced language change.

Through a general analysis of Andean Spanish, I found that the frequency of non-canonical predicate constituent order (i.e., OV) in monolingual Andean Spanish is comparable to that of non-contact varieties of Spanish. The results of the bilingual Andean Spanish analysis demonstrate that, though OV order is realized more frequently in bilingual versus monolingual speech (7.1% versus 13.2%), the rate of OV order associated with bilinguals of the present study is considerably lower than those cited in other studies conducted one to three decades ago, which range from 17.3% to 28.3% (Klee, 1996; Klee et al., 2011; Muntendam, 2009; Muysken, 1984; F. A. Ocampo & Klee, 1995). Through an examination of the relationship between age and predicate constituent order variation, I reasoned that this discrepancy may be attributed to the fact that younger speakers employ canonical VO order at a higher rate than relatively older speakers. Upon removing the youngest generation of speakers from the dataset, the resulting OV order rate (19.6%) closely resembled those of previous studies. As mentioned in §8.2.1, the results of this apparent-time analysis corroborates the existence of a syntactic change in progress in Andean Spanish whereby non-canonical structures are falling out of favor over time. Overall, the extralinguistic predictors that significantly condition syntactic variation were distinct for each group— increased VO order likelihood is associated with urban monolinguals<sup>268</sup> but older, Quechua-dominant bilinguals. That Quechua-dominant bilinguals are associated with lower VO order likelihood suggests an effect of language contact on ordering preferences.

A comparison of the linguistic predictors across monolinguals and bilinguals revealed that both varieties are conditioned in the same way by the same predictors— definiteness, information status, polarity, and subject expression.

<sup>268</sup> That is, monolinguals residing in the city of Cusco.

The statistical analysis suggests that predicate constituent order variation in monolingual Andean Spanish is more strongly conditioned by information status (i.e., increased odds of VO with given referents) but less strongly conditioned by definiteness and subject expression compared to bilingual Andean Spanish. The discordant magnitudes associated with the predictors indicate that bilingual Andean Spanish is indeed less restricted by pragmatic constraints (information structure), though these constraints still apply (Camacho, 1999; Muntendam, 2009). However, this is not the end of the story, as predicate constituent order variation is also conditioned more strictly by definiteness and subject expression (i.e., pronominal objects increase the odds of non-canonical order, while explicit subjects have the opposite syntactic effect) for bilinguals. Incidentally, these are the two predictors that condition variation in a parallel manner in bilingual Cusco Quechua. Notwithstanding these minor differences in the linguistic analyses, the primary difference between Andean Spanish monolinguals and bilinguals is that the latter group realizes OV order more frequently, but not in different linguistic contexts.

Though Quechua is an OV-dominant language, the statistical analysis of bilingual Cusco Quechua revealed that OV order is realized slightly more than twice as frequently as VO order (68.2% and 31.8% respectively). The rate of OV order provided in the present study falls within the range of rates cited for previous studies on child bilinguals, which fluctuate between 14.9% to 89.2% depending on the regional variety (Kalt & Geary, 2021; Sánchez, 2003). Upon inspecting the distribution of only clausal word orders with an explicit subject, I found that SVO is the most frequent order in the sample (38.2%), followed closely by SOV (36.5%). Thus, a dominant order of clausal constituents is lacking for bilingual Cusco Quechua. I discussed the potential typological implications of shift in basic word order in §8.3, which include a possible decline in morphological case-marking and word order shift in other syntactic domains to achieve cross-category harmony.

As is the case in Andean Spanish, there is a negative correlation between age and OV order rate in Quechua, which again points to a change in progress. In §8.2, I relied on this finding to offer a description of the Quechua-Spanish contact situation as one in which there was an initial shift to from Quechua to Andean Spanish that has since been interrupted by a period of stable bilingualism. The first language of the speaker was also found to govern OV/VO variation in Quechua such that L1 Quechua bilinguals are more inclined to employ canonical order than simultaneous bilinguals. This finding constitutes further evidence that predicate constituent order variation in Cusco Quechua is contact-induced. The output of the linguistic analysis identified definiteness,

information status, polarity, subject, weight, and argument type as significant predictors of syntactic variation. In general, a higher probability of canonical OV order is associated with pronouns, new referents, affirmative clauses, null subjects, and non-goal arguments. Argument type was the only ‘Quechua only’ factor selected as a significant predictor. Upon closer inspection, I found that there is no difference in the distribution of predicate constituents across argument type for L1 Quechua speakers, but simultaneous bilinguals employ non-canonical VO order twice as frequently as OV order with goal arguments and OV order twice as frequently as VO order with non-goal arguments. Because goal arguments are not considered accusative objects in Spanish, I speculated that simultaneous bilinguals rely on the syntax to differentiate goal from non-goal arguments in Quechua. This is a tentative conclusion that requires additional research.

In comparing the Quechua and Andean Spanish results, I found that some predictors, like information status and polarity, exercise language-specific effects. Specifically, both new referents and negative polarity precipitate a stronger adherence to canonical norms in each language. Similarly, weight was selected as a predictor of predicate constituent order variation only in Quechua. In §8.2.3, I discussed how this finding both challenges comprehension-oriented accounts of weight-based phenomena and motivates further exploration of weight-based phenomena in agglutinative languages. Other predictors, like definiteness and subject expression, exert the same conditioning effect in both languages— pronominals are associated with a higher rate of OV order and explicit subjects with a higher rate of VO order. Based on these findings, I argued that word order in bilingual Andean Spanish and Cusco Quechua is characterized by two convergent structures, which are detailed in Chapter 7.

I close the present study with the acknowledgement that neither of the languages under investigation is monolithic. Though Andean Spanish was initially formed through the learner varieties of shifting L1 Quechua speakers, in the present day, Andean Spanish boasts native speakers of its own, many of whom are monolingual speakers of the contact variety. Thus, Andean Spanish may be conceptualized as a continuum reflecting the sociolinguistic profiles of its speakers. On one end of the spectrum are L1 Quechua bilinguals who acquired Andean Spanish in adolescence through exposure to formal education and on the other end of the spectrum are monolingual speakers raised by monolingual Spanish-speaking parents who have intimate social ties with speakers of non-contact varieties, sometimes as a result of their relocation to urban centers, like Lima. Of course, a similar continuum emerges for speakers of bilingual Cusco Quechua. The diversity of these speakers is reflected in the high degree of inter-

and intra-speaker variation documented for bilinguals in the present study. Ultimately, variation in the order of predicate constituents is simply one of many features that contributes to the multifaceted linguistic identities of speakers of Andean Spanish and bilingual Quechua in Cusco, Peru.

# APPENDIX A

This appendix provides a list of the glossing abbreviations used in the present study.

<b>Abbreviation</b>	<b>Term</b>
I	first person
2	second person
3	third person
ABL	ablative
ACC	accusative
ADV	adverbial
ALL	allative
ASP	aspect
ATT	attributive
CAUS	causative
CL	clitic
CLA	classifier
COND	conditional
CONN	connector (phonological)
CONTR	contrastive
COORD	coordinator
COP	copulative
DAT	dative
DEF	definite
DIM	diminutive
DIR	directional
DM	discourse marker
DOM	differential object marker

EMO	emotive
EMP	emphatic
EVID1	first-hand evidential
EVID2	second-hand evidential
EXCL	exclusive
EXH	exhortative
F	female
FOC	focus
FUT	future
GEN	genitive
GER	gerund
HAB.PST	habitual past
IMP	imperative
INCP	incipient
INCL	inclusive
INF	infinitive
INS	instrumental
INT	intensifier
INTER	interactive
IPFV	imperfective
LIM	limitative
LOC	locative
NEG	negation
NMLZ	nominalizer
NOM	nominative
OBJ	object
OBLG	obligative
PFV	perfective
PH	placeholder
PL	plural
POSS	possessive
PRS	present
PROG	progressive

PST	past tense
PST2	past tense, unexperienced (- <i>sqa-</i> )
PTCP	participle
PURP	purposive
Q	question particle
REFL	reflexive
REG	regressive
REP	repetitive
SG	singular
SR	switch referent
SUB	subordinator
TOP	topic

## APPENDIX B

This appendix contains the Sociolinguistic Interview Schedule used to collect spoken data for the present study. Each question is given in English (black), Spanish (blue), and Quechua (red).

1. How are you doing today? Can you tell me a little bit about yourself?  
¿Cómo está usted hoy? ¿Me puede contar un poco de Ud., por favor?  
Imaynallan kashanki? Ama hina kaychu, qanmanta pisillata willarikuway.
2. Can you tell me about your childhood? What were you like as a child? Can you share some particular experiences you remember from childhood. For example, what life was like in your neighborhood or community?  
¿Me puede contar un poco sobre su juventud? ¿Cómo era Ud. como niño/a? ¿Me puede compartir algunas experiencias particulares que recuerda? Por ejemplo, ¿cómo era su vida en su comunidad?  
Irqi kashasqaykimanta willarikuway. Irqi kashaspa, imaynan kasharanki? Wakin yuyariynykita ancha chaninniyuq willarikuway. Por ejemplo, ¿imaynan kawsayniyki llaqtaykipi kasharan?
3. Please tell me about your parents. Talk about a special memory that you have with your mother and/or father when you were a child.  
Cuénteme por favor de sus padres. Por favor, explíqueme un recuerdo especial que Ud. tiene con su madre y/o padre cuando era niño/a.  
Taytamamaykimanta willarikuway. Ama hina kaychu, irqi kashaspa, chanin yuyariynykita taytamamaykimanta willarikuway.
4. Please recount a memory (good or bad) during your time in primary school, e.g., your first day of class, a day you got in trouble with your teachers, a school event, etc.  
Por favor cuénteme un recuerdo de su tiempo en la escuela primaria. (Por ejemplo, su primer día de clase, un día en que se comportó mal en frente de los profesores, un evento escolar único, etc.)

Ama hina kaychu, yachay wasipi kashaspa, huk allin utaq mana allinchu yuyariynykita willarikuway. (Por ejemplo, ñawpaqta p'unchay yachay wasipi kasharanki, huk p'unchay mana allinchu rikuranki amautawan, un evento escolar único, etc.)

5. Do you have a best friend? Please describe a memorable event you experienced with your best friend.

Tiene Ud. un(a) mejor amigo/a? Por favor descríbame una experiencia que Ud. recuerda con esa persona.

Astawan allin khumpayuyqchu kanki? Huk yuyariynykimanta chay khumpaykiwan willarikuway.

6. Do you have any children? When did you give birth for the first time?/When was your first child born? Please recount that experience.

¿Tiene Ud. hijos? ¿Cómo era la experiencia de estar embarazada por primera vez? ¿Qué tal el día del nacimiento? ¿Me puede contar todo sobre aquel día, por favor?

Wawayuyqchu kanki? Ñawpaq kuti wiksayuq kashaspa, imaynallan kasharanki? Imaynallan wachakuy kasharan qanpaq? Chay p'unchaypi wachakusqaykita willarikuway?

7. Where do you work? Please tell me about a frustrating day at work.

¿Dónde trabaja Ud.? Por favor cuénteme sobre un día frustrante del pasado en el trabajo.

Maypin llank'ashanki? Ama hina kaychu, llank'ayniykiipi huk mana allinchu p'unchaymanta willarikuway.

8. Do you have a significant other? What are they like? How did you and your partner meet? Can you describe that day for me? If you are married, describe the day of your wedding.

¿Tiene Ud. esposo/a o novio/a? ¿Cómo es? ¿Cómo se conocieron Uds.? Explíqueme aquel día, por favor. Descríbame por favor el día de su boda.

Qhariyuyqchu/Warmiyuyqchu utaq yana urpiyuq kanki? ¿Imaynan kan? Imaynan riqsiriranakunkichis? Haqay p'unchaypi risqirinakusqaykichismanta willarikuway. Haqay p'unchaypi sawakunakusqaykichismanta willarikuway.

9. Do you remember a time that you were ill? In as much detail as possible, describe that experience to me.

¿Se acuerda Ud. un momento cuando estaba muy enfermo/a? Por favor descríbame esa experiencia.

Huk kuti anchata unqusqaykita yuyarinkichu? Chay kutiqamanta willarikuway.

10. Please tell me about a moment when you found yourself in a dangerous situation (e.g., assault, robbery, car accident).  
Por favor cuénteme sobre un momento cuando Ud. se encontró en una situación muy peligrosa (e.g. un asalto, un robo, un choque/accidente del carro).  
Ama hina kaychu, huk kutipi manchakusqaykimanta willarikuway.(e.g. hap'ipariymanta, suwaymanta, carro accidente nisqamanta).
11. Please tell me about a vivid dream that you remember.  
Por favor cuénteme un sueño muy intenso que Ud. recuerda.  
Huk musquyniykimanta willarikuway.
12. Do you celebrate the Inti Raymi festival? How do you celebrate Inti Raymi? When was the most recent time you celebrated Inti Raymi? Describe that day.  
¿Celebra Ud. las fiestas del Inti Raymi? ¿Cómo lo celebra? ¿Cuándo fue la última vez que Ud. celebró Inti Raymi? Descríbame aquel día, por favor.  
Inti Raymita celebrankichu? Imaynan celebranki? Hayk'aqmi Inti Raymita celebraranki? Inti Raymita celebrasqaykimanta willarikuway.
13. Do you know of a traditional legend or story that is important in your community? Please tell me that story. ¿Conoce Ud. algún cuento tradicional de su pueblo? Cuéntemelo, por favor.  
Ancha chanin willakuyta llaqtaykimanta riqsinkichu? Chay willakuyta willarikuway.

## APPENDIX C

This appendix provides the Language Background Questionnaire (*Cuestionario del fondo lingüístico*) in Spanish, which was employed to collect demographic and linguistic information about the participants of the present study.

Sitio: \_\_\_\_\_ Fecha: \_\_\_\_\_ N° de Participante # \_\_\_\_\_

Cuestionario del fondo lingüístico

(TODA LA INFORMACIÓN SE MANTENDRÁ CONFIDENCIAL Y ANÓNIMA.)

\*\*\*\*\*

**I. Información biográfica**

1. ¿Cuál es su sexo? Varón / Mujer
2. ¿Cuántos años tiene Ud.? \_\_\_\_\_ años
3. ¿Fue Ud. a la escuela? Sí / No  
>> ¿Cuál es el nivel de educación más alto que Ud. ha alcanzado?  
\_\_\_\_\_
4. ¿Cuál es su profesión? \_\_\_\_\_
5. ¿Dónde vive Ud.? \_\_\_\_\_  
>> ¿Está ubicado su vecindario en un área urbana o rural? urbana / rural  
>> ¿Por cuánto tiempo vive Ud. allí? \_\_\_\_\_

\*\*\*\*\*

**II. Historia lingüística**

1. ¿De dónde son sus padres?  
Madre: \_\_\_\_\_ Padre: \_\_\_\_\_
2. ¿Qué idioma(s) hablan sus padres?  
Madre: \_\_\_\_\_ Padre: \_\_\_\_\_
3. ¿En qué trabajan sus padres?  
Madre: \_\_\_\_\_ Padre: \_\_\_\_\_
4. ¿Cuál es el nivel de educación más alto que sus padres han alcanzado?  
Madre: \_\_\_\_\_ Padre: \_\_\_\_\_
5. ¿A cuántos años empezó Ud. a hablar CASTELLANO? \_\_\_\_\_ años
6. ¿A cuántos años empezó Ud. a hablar QUECHUA? \_\_\_\_\_ años
7. ¿Qué idioma(s) usaban mayormente sus padres al hablar con Ud.?  
>> Madre: quechua castellano ambos otro \_\_\_\_\_  
>> Padre: quechua castellano ambos otro \_\_\_\_\_
8. ¿Qué idioma(s) usaba mayormente Ud. al hablar con sus padres?  
>> Con la madre: quechua castellano ambos otro \_\_\_\_\_

>> Con el padre:      quechua      castellano      ambos      otro \_\_\_\_\_

9. ¿Le animaban a Ud. sus padres a hablar quechua en casa?      Sí / No

10. ¿Qué idioma(s) usaba Ud. al hablar con sus hermanos/as?

quechua      castellano      ambos      otro \_\_\_\_\_

11. ¿Qué idioma(s) usaba(n) su(s) abuelo(s) al hablar con Ud.?

quechua      castellano      ambos      otro \_\_\_\_\_

12. ¿Qué idioma(s) usaba Ud. al hablar con su(s) abuelo(s)?

quechua      castellano      ambos      otro \_\_\_\_\_

13. ¿Qué idioma(s) usaba Ud. con otros niños?

quechua      castellano      ambos      otro \_\_\_\_\_

-----**Historia lingüística: La escuela primaria**-----

14. ¿Con qué frecuencia usaba Ud. el quechua entre los 6 y los 11 años?

nunca      poco      mucho      siempre

15. Entre los 6 y los 11 años, ¿con quién(es) hablaba Ud. quechua? (*Elija cada opción que sea relevante*)

madre      padre      hermanos      amigos      otros \_\_\_\_\_

16. ¿Asistió Ud. a la escuela primaria en Cusco?      Sí / No

17. ¿Cuál era el idioma de instrucción?

quechua      castellano      ambos      otro \_\_\_\_\_

18. ¿Tenía que tomar Ud. una clase para aprender el quechua en la escuela primaria?      Sí / No

19. ¿Qué idioma(s) hablaba Ud. con sus amigos quechua-hablantes en la escuela primaria?

quechua      castellano      ambos      otro \_\_\_\_\_

-----**Historia lingüística: La escuela secundaria**-----

20. ¿Con qué frecuencia usaba Ud. quechua entre los 12 y 17 años?

nunca      poco      mucho      siempre

21. Entre los 12 y 17 años, ¿con quién(es) hablaba Ud. quechua?

madre      padre      hermanos      amigos      otros \_\_\_\_\_

22. Asistió Ud. a la escuela secundaria en Cusco?      Sí / No

23. ¿Cuál era el idioma de instrucción?

quechua      castellano      ambos      otro \_\_\_\_\_

24. ¿Tenía que tomar Ud. una clase para aprender el quechua en la escuela secundaria? Sí / No

25. ¿Qué idioma(s) hablaba Ud. con sus amigos quechua-hablantes en la escuela secundaria?

quechua          castellano          ambos          otro \_\_\_\_\_

\*\*\*\*\*

### III. Uso lingüístico corriente

1. ¿Qué idioma(s) usa Ud. con los amigos?

- a. solamente quechua
- b. mayormente quechua
- c. ambos equitativamente
- d. mayormente castellano
- e. solamente castellano

2. ¿Qué idioma(s) usa Ud. con la familia?

- a. solamente quechua
- b. mayormente quechua
- c. ambos equitativamente
- d. mayormente castellano
- e. solamente castellano

3. ¿Qué idioma(s) usa Ud. en la escuela/el trabajo?

- a. solamente quechua
- b. mayormente quechua
- c. ambos equitativamente
- d. mayormente castellano
- e. solamente castellano

4. ¿En qué idioma piensa Ud.?

- a. solamente quechua
- b. mayormente quechua
- c. ambos equitativamente
- d. mayormente castellano
- e. solamente castellano

\*\*\*\*\*

### IV. Competencia lingüística

1. Indique su capacidad corriente en general en CASTELLANO (*elija una opción*)

- 0 = No entiendo ni puedo hablar
- 1 = Entiendo pero no puedo hablar para nada
- 2 = Entiendo pero me dificulta mucho hablar
- 3 = Entiendo pero me dificulta un poco hablar
- 4 = Entiendo y hablo normal con poca dificultad
- 5 = Entiendo y hablo muy bien sin problema como nativo/a hablante

2. Indique su capacidad corriente en general en QUECHUA (*elija una opción*)

- 0 = No entiendo ni puedo hablar
- 1 = Entiendo pero no puedo hablar para nada
- 2 = Entiendo pero me dificulta mucho hablar
- 3 = Entiendo pero me dificulta un poco hablar
- 4 = Entiendo y hablo normal con poca dificultad
- 5 = Entiendo y hablo muy bien sin problema como nativo/a hablante

3. En general, ¿qué idioma prefiere usar Ud.?

- castellano      quechua      ambos      no tengo preferencia

\*\*\*\*\*

### V. Actitudes lingüísticas

1. Indique las siguientes oraciones en una escala de 0 (*no estoy de acuerdo*) a 6 (*estoy completamente de acuerdo*).

- a. Me siento como yo mismo/a cuando hablo CASTELLANO..... 0 1 2 3 4 5 6
- b. Me siento como yo mismo/a cuando hablo QUECHUA..... 0 1 2 3 4 5 6
- c. Soy hablante de CASTELLANO..... 0 1 2 3 4 5 6
- d. Soy hablante de QUECHUA..... 0 1 2 3 4 5 6
- e. A mí me importa usar CASTELLANO como hablante nativo/a... 0 1 2 3 4 5 6
- f. A mí me importa usar QUECHUA como hablante nativo/a..... 0 1 2 3 4 5 6
- g. Quiero que otros sepan que soy hablante nativo/a de CASTELLANO 0 1 2 3 4 5 6
- h. Quiero que otros sepan que soy hablante nativo/a de QUECHUA 0 1 2 3 4 5 6

2. Desea Ud. mejorar sus destrezas lingüísticas en CASTELLANO?      Sí / No

3. Desea Ud. mejorar sus destrezas lingüísticas en QUECHUA?      Sí / No

4. ¿A Ud. le importa el CASTELLANO?      Sí / No

5. ¿A Ud. le importa el QUECHUA?      Sí / No

**¡MUCHAS GRACIAS por completar este cuestionario!**

**Por favor, devuelva este formulario a la(s) investigadora(s).**

# APPENDIX D

This appendix supplies the modifications made to the Bilingual Language Profile in constructing the Language Background Questionnaire. Each modification is accompanied by a justification.

## I. Biographical Information

- Added a question about the occupation of the individual
  - **Justification:** In previous literature on Andean Spanish, occupation has been used as a metric for reporting the socioeconomic status of the participant (Klee, 1996; Klee et al., 2011; F. A. Ocampo & Klee, 1995)

## II. Language History

- Added questions pertaining to the linguistic and socioeconomic background of the individual's parents including place of residence, languages spoken, occupation, and highest level of education
  - **Justification:** Identify the linguistic input of the individual as a child and their socioeconomic status in childhood and adolescence
- Omitted question about the age that the individual 'became comfortable' using Quechua or Spanish
  - **Justification:** Removed for sake of efficiency and clarity
- Added questions about the quality of linguistic interaction in primary and secondary school (i.e., with whom the individual spoke Spanish and Quechua) in lieu of the number of years of instruction in Spanish/Quechua
  - **Justification:** Provide more detailed information about the quality of linguistic interaction; number of years of Spanish or Quechua instruction were inferred from the individual's

answer to the question ‘What was the language of instruction in your primary/secondary education?’ (e.g., Quechua instruction reported in primary only is equivalent to 6 years of Quechua instruction)

- Omitted question about the number of years the participant resided in a region in which Quechua and Spanish was spoken and the number of years that they had spent in a family where Quechua and Spanish was spoken
  - **Justification:** All participants have spent their entire lives residing in a region where both languages are spoken. Familial information was extracted from additional questions in the Biographical Information section about the language that was spoken by their parents and additional questions about the quality of linguistic interaction in the Language History section (e.g., if a participant over 20 years of age indicated that both parents spoke Quechua, the number of years the participant spent in a family where Quechua was spoken would be considered 20+)

### III. Language Use

- In lieu of reporting percentage of language use in 10% increments, individuals reported their current language use in several domains by selecting among the options ‘only Quechua’, ‘mostly Quechua’, ‘both equally’, ‘mostly Spanish’, ‘only Spanish’
  - **Justification:** Options were adapted for the sake of efficiency. The options ‘mostly Quechua’ or ‘mostly Spanish’ were recorded in the Bilingual Language Profile as 70% use of the respective language when calculating dominance scores.
- Omitted question that asked ‘how often do you count in Quechua or Spanish’
  - **Justification:** Many Quechua-dominant speakers use Spanish numbers, so this question may have skewed dominance results.

### IV. Language Proficiency

- Question about speaking and understanding Quechua and Spanish was collapsed into a single question asking about their ‘working

linguistic capacity'. Instead of a likert scale from 0 to 6, participants could select a number 0 through 5, each accompanied with a qualitative description of what this number indicates (e.g., next to 2, the description read 'I understand but it is difficult for me to speak')

- **Justification:** Questions were combined for sake of efficiency, and descriptions were provided to enhance the clarity of options.
- Omitted questions about the reading and writing competencies of the individual.
  - **Justification:** Quechua-dominant individuals are rarely proficient in reading and writing. This question may have skewed dominance results.

## V. Language Attitudes

- Two questions were added. The first asked, 'Would you like to improve your linguistic skills in Quechua/Spanish?', and the second asked 'Is Quechua/Spanish important to you?'
  - **Justification:** Questions were added to extract additional qualitative information about the linguistic attitudes of the individual.

## BIBLIOGRAPHY

- Adelaar, W. F. H. (2004). *The languages of the Andes*. Cambridge University Press.
- Appel, R., & Muysken, P. (2005). *Language contact and bilingualism*. Edward Arnold.
- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Audacity Team. (2022). Audacity(R): Free audio editor and recorder. <https://audacityteam.org/>
- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Baker, M. C. (1996). *The polysynthesis parameter*. Oxford University Press.
- Bancu, A. (2017). Word order variation and change in Transylvanian Saxon. 23(2), 11–20. <https://repository.upenn.edu/pwpl/vol23/iss2/3>
- Bateman McDonald, B. D. (2022). *Contact-induced subjectivization in the compound past: An examination of PP/PRET variation in Cusco Spanish* (Dissertation). University of Georgia. Athens, GA.
- Bennett, R., Elfner, E., & McCloskey, J. (2016). Lightest to the right: An apparently anomalous displacement in Irish. *Linguistic Inquiry*, 47(2), 169–234. [https://doi.org/10.1162/ling\\_a\\_00209](https://doi.org/10.1162/ling_a_00209)
- Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. <https://sites.la.utexas.edu/bilingual/>

- Birner, B. J. (2006). Inferential relations and noncanonical word order. In B. J. Birner & G. L. Ward (Eds.), *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn* (pp. 31–51). John Benjamins.
- Birner, B. J. (2013). *Introduction to pragmatics*. Wiley-Blackwell.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formation. *Cognition*, 21, 47–67.
- Bosque, I., & Gutiérrez-Rexach, J. (2008). *Fundamentos de sintaxis formal*. Akal.
- Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118, 172–189. <https://doi.org/10.1016/j.lingua.2007.02.003>
- Brody, J. (1995). Lending the 'unborrowable': Spanish discourse markers in indigenous American languages. In C. Silva-Corvalán (Ed.), *Spanish in four continents: Studies in language contact and bilingualism* (pp. 132–147). Georgetown University Press.
- Bullock, B. E., & Toribio, A. J. (2004). Introduction: Convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition*, 7(2), 91–93. <https://doi.org/10.1017/S1366728904001506>
- Camacho, J. (1999). From SOV to SVO: The grammar of interlanguage word order. *Second Language Research*, 15, 115–132.
- Caravedo, R., & Klee, C. A. (2012). El contacto de dialectos en Lima: Un análisis del pretérito perfecto en las cláusulas narrativas. *Lengua y Migración*, 4, 5–24.
- Casielles-Suárez, E. (2004). *The syntax-information structure interface: Evidence from Spanish and English* (1st ed.). Routledge.

- Cerrón-Palomino, R. (1987a). *Lingüística quechua*. Centro de Estudios Rurales Andinos "Centro Bartolomé de las Casas".
- Cerrón-Palomino, R. (1987b). Multilingüismo y política idiomática en el Perú. *Allpanchis*, 19, 17–44.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic: A new typology of language* (pp. 25–55). Academic Press.
- Comrie, B. (1989). *Language universals and linguistic typology* (2nd ed.). The University of Chicago Press.
- Courtney, E. H. (1998). *Child acquisition of Quechua morphosyntax* (Doctoral dissertation). The University of Arizona. Tucson, AZ.
- Croft, W. (2012). Grammatical hierarchies and the semantic map model. *Typology and universals* (Second). Cambridge University Press.
- Cusihuamán, A. (1976). *Gramática Quechua: Cuzco-Collao*. Centro Bartolomé de las Casas.
- Cusihuamán, A. (2001). *Gramática Quechua: Cuzco-Collao* (2nd ed.). Centro Bartolomé de las Casas.
- Delforge, A. M. (2008). Unstressed vowel reduction in Andean Spanish. In L. Colantoni & J. Steele (Eds.), *Selected Proceedings of the Third Conference on Laboratory Approaches to Spanish Phonology* (pp. 107–124). Cascadilla Proceedings Project.
- Delforge, A. M. (2012). 'Nobody wants to sound like a provinciano': The recession of unstressed vowel devoicing in the Spanish of Cusco, Perú. *Journal of Sociolinguistics*, 16(3), 311–335.
- Dryer, M. S. (2010). Word order. In T. Shopen (Ed.), *Language typology and syntactic description* (2nd ed., pp. 61–131). Cambridge University Press.
- Dryer, M. S. (2013a). Determining dominant word order. *The world atlas of language structures online*. Max Planck Institute for Evolutionary

- Anthropology. Retrieved February 24, 2021, from <http://wals.info/chapter/s6>
- Dryer, M. S. (2013b). Order of object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. Retrieved February 24, 2021, from <http://wals.info/chapter/83>
- Dryer, M. S. (2013c). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. Retrieved February 24, 2021, from <http://wals.info/chapter/81>
- Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. Retrieved February 24, 2021, from <http://wals.info>
- ELAN. (2023). <https://archive.mpi.nl/tla/elan>
- Escobar, A. M. (2011). Spanish in contact with Quechua. In M. Díaz Campos (Ed.), *The Handbook of Spanish Sociolinguistics* (pp. 323–352). Blackwell.
- Escobar, A. M., & Roy, J. (2022). Grammaticalization in the evidential pathway of the Andean Spanish present perfect: Language contact as a trigger for language change (J. MacDonald, Z. Fagyal, A. Beristain, & R. Turner, Eds.). *RLLT18: Special Issue of Isogloss. Open Journal of Romance Languages*, 8(4/14), 1–38. <https://doi.org/https://doi.org/10.5565/rev/isogloss.184>
- Faghiri, P., & Samvelian, P. (2020). Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A Journal of General Linguistics*, 5(1), 1–33. <https://doi.org/https://doi.org/10.5334/gjgl.1078>
- Faller, M. (2007). The Cusco Quechua reportative evidential and rhetorical relations. In P. K. Austin & A. Simpson (Eds.), *Endangered languages* (pp. 223–251). Helmut Buske Verlag.

- Fortescue, M. (1993). Eskimo word order variation and its contact-induced perturbation. *Journal of Linguistics*, 29(2), 267–289. Retrieved October 3, 2023, from <https://www.jstor.org/stable/4176235>
- García Tesoro, A. I., & Jang, J. S. (2018). El pretérito perfecto compuesto en el español andino peruano: Usos innovadores y extensión a contextos de aoristo. *Forma y Función*, 31(1), 93–123.
- Givón, T. (2001). *Syntax: An introduction* (Vol. 1). John Benjamins Publishing Company.
- Gómez Rendón, J., & Adelaar, W. (2009). Loanwords in Imbabura Quechua. In M. Haspelmath & U. Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook*. De Gruyter, Inc.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language* (pp. 73–113). MIT Press.
- Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. A. Moravcsik, & J. Wirth (Eds.), *Studies in syntactic typology* (pp. 209–242). John Benjamins.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274–307. <https://www.jstor.org/stable/416535>
- Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*. Springer.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 199–244. <https://www.jstor.org/stable/4174965>
- Haugen, E. (1953). *The Norwegian language in America: A study in bilingual behavior*. University of Pennsylvania Press.
- Hawkins, J. A. (1980). On implicational and distributional universals of word order. *Journal of Linguistics*, 16(2), 171–338.

- Hawkins, J. A. (1982). Cross-category harmony, X-bar and the predictions of markedness. *Journal of Linguistics*, 18(1), 1–35.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge University Press.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.
- Hayashi, M., & Yoon, K.-E. (2010). A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In N. Amiridze, B. H. Davis, & M. MacLagan (Eds.), *Fillers, pauses and placeholders* (pp. 33–66). John Benjamins Publishing Company.
- Heath, S. B., & Laprade, R. (1982). Castilian colonization and indigenous languages: The cases of Quechua and Aymara. In R. L. Cooper (Ed.), *Language spread: Studies in diffusion and social change* (pp. 117–147). Indiana University Press.
- Heidinger, S. (2013). Information focus, syntactic weight and postverbal constituent order in Spanish. *Borealis: An International Journal of Hispanic Linguistics*, 2(2), 159–190. <https://doi.org/http://dx.doi.org/10.7557/1.2.2.2742>
- Heine, B. (2008). Contact-induced word order change without word order change. In P. Siemund & N. Kintana (Eds.), *Language contact and contact languages* (pp. 36–60). John Benjamins.
- Hickey, R. (2010). Language contact: Reconsideration and reassessment. *The handbook of language contact* (pp. 1–28). Wiley-Blackwell.
- Hintz, D. J. (2009). Reordenamiento sintáctico en construcciones verbales analíticas del quechua por el contacto con el castellano. *Linguística*, 22, 187–201.
- Hintz, D. J. (2016). Auxiliation and typological shift: The interaction of language contact and internally-motivated change in Quechua. In A. L. Berez-Kroeker, D. M. Hintz, & C. Jany (Eds.), *Language contact and*

- change in the Americas: Studies in honor of Marianne Mithun* (pp. 315–338). John Benjamins.
- Hoffman, M. (2014). Sociolinguistic interviews. In J. Holmes & K. Hazen (Eds.), *Research methods in sociolinguistics: A practical guide*. John Wiley & Sons.
- Hornberger, N. H., & Coronel-Molina, S. M. (2004). Quechua language shift, maintenance, and revitalization in the Andes: The case for language planning. *International Journal of the Sociology of Language*, 167, 9–67.
- Howe, C. (2018). El pasado compuesto en el español peruano. Hacia una explicación unificada de su significado. In R. Risco (Ed.), *Estudios de variación y contacto lingüístico en el español peruano* (pp. 169–197). Universidad Nacional de La Plata: Facultad de Humanidades y Ciencias de la Educación.
- Hubbel, S. (2023). Variation in predicate constituent order in Southern Peruvian Quechua. In M. Waltermire & K. Bove (Eds.), *Mutual influence in situations of Spanish language contact in the Americas* (pp. 117–133). Routledge.
- Instituto Nacional de Estadística e Informática. (2017). Cusco: Resultados definitivos. *Censos nacionales 2017: XII de población, VII de vivienda y III de comunidades indígenas*. [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1544/](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1544/)
- Kagan, O. (2020). Differential object marking. *The semantics of case* (pp. 147–188). Cambridge University Press. 10.1017/9781108236867
- Kalt, S. E. (2002). *Second language acquisition of Spanish morpho-syntax by Quechua-speaking children* (Dissertation). University of Southern California. Los Angeles, California.
- Kalt, S. E. (2021). Acquisition, loss and innovation in Chuquisaca Quechua—What happened to evidential marking? *Languages*, 6(2), 76. <https://doi.org/https://doi.org/10.3390/languages6020076>

- Kalt, S. E., & Geary, J. A. (2021). Typological shift in bilinguals' L1: Word order and case marking in two varieties of child Quechua. *Languages*, 6(42). <https://doi.org/https://doi.org/10.3390/languages/6010042>
- Kiss, K. É. (1998). Identificational focus versus informational focus. *Language*, 74(2), 245–273. <https://doi.org/https://doi.org/10.1353/lan.1998.0211>
- Klee, C. A. (1990). Spanish-Quechua language contact: The clitic pronoun system in Andean Spanish. *Word*, 41(1), 35–46.
- Klee, C. A. (1996). The Spanish of the Peruvian Andes: The influence of Quechua on Spanish language structure. In A. Roca & J. B. Jensen (Eds.), *Spanish in contact: Issues in bilingualism* (pp. 73–91). Cascadilla Press.
- Klee, C. A., & Caravedo, R. (2005). Contact-induced language change in Lima, Peru: The case of clitic pronouns. In D. Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 12–21). Cascadilla Proceedings Project.
- Klee, C. A., & Caravedo, R. (2006). Andean Spanish and the Spanish of Lima: Linguistic variation and change in a contact situation. In C. Mar-Molinero & M. Stewart (Eds.), *Globalization and language in the Spanish-speaking world*. Palgrave.
- Klee, C. A., & Lynch, A. (2009). Contacto del español con lenguas indígenas en Hispanoamérica. *El español en contacto con otras lenguas*. Georgetown University Press.
- Klee, C. A., & Ocampo, A. M. (1995). The expression of past reference in Spanish narratives of Spanish-Quechua bilingual speakers. In C. Silva-Corvalán (Ed.), *Spanish in four continents: Studies in language contact and bilingualism* (pp. 52–70). Georgetown University Press.
- Klee, C. A., Rogers, B. M. A., Mónica, d. I. F. I., & Ramsburg, J. (2023). Intervocalic /s/ voicing in the Andean Spanish of southern Peru. In M. Waltermire & K. Bove (Eds.), *Mutual influence in situations of Spanish language contact in the Americas* (pp. 93–116). Routledge.

- Klee, C. A., Tight, D. G., & Caravedo, R. (2011). Variation and change in Peruvian Spanish word order: Language contact and dialect contact in Lima. *Southwest Journal of Linguistics*, 30(2), 5–31.
- Komen, E. (2013). *Finding focus: A study of the historical development of focus in English* (Dissertation). Radboud University.
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse references*. Cambridge University Press.
- Lanning, E. P. (1967). *Peru before the Incas*. Prentice-Hall, Inc.
- Leonetti, M., & Escandell-Vidal, M. V. (2009). Fronting and verum focus in Spanish. In A. Dufter & D. Jacob (Eds.), *Focus and background in Romance Languages* (pp. 155–204). John Benjamins Publishing Company.
- Levshina, N., Namboodiripad, S., Allasonnière-Tang, M., Kramer, M. A., Talamo, L., Verkerk, A., Wilmouth, S., Garrido Rodriguez, G., Gupton, T., Kidd, E., Liu, Z., Naccarato, C., Nordlinger, R., Panova, A., & Stoyanova, N. (2023). Why we need a gradient approach to word order. *Linguistics: An Interdisciplinary Journal of the Language Sciences*, 61(4), 825–883.
- Ley para la educación bilingüe intercultural. (2002). <https://cdn.www.gob.pe/uploads/document/file/4049767/Ley%2027818%20-%20Educaci%C3%B3n%20Biling%C3%BCe%20Intercultural.pdf?v=1673978037>
- Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, 791–824.
- Longacre, R. (1983). *The grammar of discourse*. Plenum.
- Luján, M., Minaya, L., & Sankoff, D. (1984). The universal consistency hypothesis and the prediction of word order acquisition stages in the speech of bilingual children. *Language*, 60(2), 343–371. <https://www.jstor.org/stable/413644>

- Manley, M. (2007). Cross-linguistic influence of the Cuzco Quechua epistemic system on Andean Spanish. In K. Potowski & R. Cameron (Eds.), *Spanish in contact: Policy, social, and linguistic inquiries* (pp. 191–209). John Benjamins Publishing Company.
- Mannheim, B. (1991). *The language of the Inka since the European invasion*. University of Texas Press.
- Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2010). A developmental analysis of generic nouns in Southern Peruvian Quechua. *Language Learning and Development*, 7, 1–23.
- Mayer, E., & Sánchez, L. (2016). Object agreement marking and information structure along the Quechua-Spanish contact continuum. *Revista Española de Lingüística Aplicada*, 29(2), 544–581.
- McDonald, J., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, 25, 188–230.
- Meillet, A. (1921). *Linguistique historique et linguistique générale*. Champion.
- Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of English*. Blackwell.
- Ministerio de Educación. (2023). Registro nacional de instituciones educativas EIB. <https://app.powerbi.com/view?r=eyJrIjoiMTEyZGQ4MTEtZj%20goMSooMzNiLTliOGMtOWZlZTYyMmVkOTAoIiwidCI6IjE3%20OWJkZGE4LWQ5NjQtNDNmZihZDNiLTY3NDE4%20NmEy%20ZmEyOCIsImMiOiR9&pageName=ReportSection>
- Monroy, C., & Mackie, C. (2022). Education in Peru. *World Education News and Reviews*. <https://wenr.wes.org/2022/03/education-in-peru-2>
- Müller, F. M. (1875). *Lectures on the science of languages* (Vol. 1). Scribner, Armstrong.
- Muntendam, A. (2008). Crosslinguistic influence in Andean Spanish: Word order and focus. In M. Bowels, R. Foote, S. Perpiñán, & R. Bhatt (Eds.),

- Selected Proceedings of the 2007 Second Language Research Forum*. Cascadilla Proceeding Project.
- Muntendam, A. (2009). *Linguistic transfer in Andean Spanish: Syntax or pragmatics?* (Doctoral dissertation). University of Illinois at Urbana-Champaign. Urbana, Illinois.
- Muntendam, A. (2010). The syntax and pragmatics of language contact: A case study of Andean Spanish. *LSA Annual Meeting Extended Abstracts*, 1.
- Muntendam, A. (2013). On the nature of cross-linguistic transfer: A case study of Andean Spanish. *Bilingualism: Language and Cognition*, 16(1), 111–131. <https://doi.org/10.1017/S1366728912000247>
- Muntendam, A. (2015). Discourse deixis in Southern Quechua: A case study on topic and focus. In M. Manley & A. Muntendam (Eds.), *Quechua expressions of stance and deixis* (pp. 208–258). Brill.
- Muntendam, A., & Muysken, P. (2021). Situaciones y fenómenos de contact lingüístico en los Andes: La relación entre el quechua y el español. In E. M. Eckkrammer (Ed.), *Manual del español en América* (pp. 281–299). De Gruyter.
- Muntendam, A., & Torreira, F. (2016). Focus and prosody in Spanish and Quechua. In M. Armstrong, N. Henriksen, & M. del Mar Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 69–89). John Benjamins Publishing Company.
- Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In A. Highfield & A. Valdman (Eds.), *Historicity and variation in creole studies* (pp. 52–78). Karoma Publishers.
- Muysken, P. (1984). The Spanish that Quechua speakers learn: L2 learning as norm-governed behavior. In R. W. Andersen (Ed.), *Second languages: A cross-linguistic perspective* (pp. 101–119). Newbury House.
- Muysken, P. (1995). Focus in Quechua. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 375–393). Oxford University Press.

- Muysken, P. (2012). Spanish affixes in the Quechua languages: A multidimensional perspective. *Lingua*, 122, 481–493. <https://doi.org/10.1016/j.lingua.2011.10.003>
- Muysken, P., & Muntendam, A. (2016). Interfacing interfaces: Quechua and Spanish in the Andes. In S. Fischer & C. Gabriel (Eds.), *Manual of grammatical interfaces in Romance* (pp. 607–633). De Gruyter, Inc.
- Naccarato, C., Panova, A., & Stoyanova, N. (2021). Word-order variation in a contact setting: A corpus-based investigation of Russian spoken in Daghestan. *Language Variation and Change*, 33, 387–411. <https://doi.org/10.1017/S095439452100017X>
- Ocampo, F. A. (1992). The word order of constructions with a verb, a subject, and a direct object in spoken Spanish. In J. Amastae, G. Goodall, M. Montalbetti, & M. Phinney (Eds.), *Contemporary Research in Romance Linguistics* (pp. 291–305). John Benjamins Publishing Company.
- Ocampo, F. A. (1995). The word order of two-constituent constructions in spoken Spanish. In P. Downing & M. Noonan (Eds.), *Word order in discourse* (pp. 425–447). John Benjamins Publishing Company.
- Ocampo, F. A., & Klee, C. A. (1995). Spanish OV/VO word-order variation in Spanish-Quechua bilingual speakers. In C. Silva-Corvalán (Ed.), *Spanish in four continents: Studies in language contact and bilingualism* (pp. 71–82). Georgetown University Press.
- O'Rourke, E. (2009). Phonetics and phonology of Cuzco Quechua declarative intonation: An instrumental analysis. *Journal of the International Phonetic Association*, 39(3), 291–312.
- O'Rourke, E. (2023). Depalatalization and delateralization in the Ecuadorian Amazon: Spanish /l/ in Contact with Kichwa. In M. Jara, R. Zariquiey, P. Valenzuela, & A. M. Escobar (Eds.), *Spanish diversity in the Amazon* (pp. 185–217). Brill.
- Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83(4), 770–802.

- Paredes, L., & Valdez, M. L. (2008). Language contact and change: Direct object léísmo in Andean Spanish. In M. Westmoreland & J. A. Thomas (Eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics* (pp. 140–148). Cascadilla Proceedings Project.
- Parker, G. J. (1963). La clasificación genética de los dialectos Quechuas. *Revista del Museo Nacional*, 32, 241–252.
- Pasquale, M. (2002). *Quechua and Spanish language contact: Influence on the Quechua phonological system* (Dissertation). Michigan State University. Lansing, MI.
- Pasquale, M. (2009). Phonological variation in Peruvian Quechua speech community. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages*. John Benjamins Publishing Company.
- Penny, R. (2000). *Variation and change in Spanish*. Cambridge University Press.
- Poplack, S., & Levey, S. (2010). Contact-induced grammatical change. In P. Auer & J. E. Schmidt (Eds.), *Language and space— An international handbook of linguistic variation* (pp. 391–419). Mouton de Gruyter.
- Poplack, S., Zentz, L., & Dion, N. (2011). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition*, 15(2), 203–225.
- Porter, J. M., & Morrison, A. R. (2024). From pilots to policies: Challenges for implementing intercultural bilingual education in Latin America. *International Review of Education*, 1–18. <https://doi.org/https://doi.org/10.1007/s11159-023-10039-5>
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223–255). American Academic Press.
- Puerma Bonilla, J. (2019). Orden básico de constituyentes en construcciones transitivas sintácticas del español actual. *Études Romanes De BRNO*, 40, 51–67. <https://doi.org/10.5817/ERB2019-1-4>

- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English*. Longman.
- R Core Team. (2023). R: A language and environment for statistical computing. <http://www.rstudio.com/>
- Rochemont, M. S. (1986). *Focus in generative grammar*. John Benjamins.
- Rojas-Sosa, D. (2008). The use of the past tenses in the Spanish of Lima: Variation in a situation of contact. In J. Bruhn de Garavito & E. Valenzuela (Eds.), *Selected Proceedings of the 10th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project.
- Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *81*(3), 613–644.
- Salas Cruz, A. (1993). *Manual de gramática Quechua: Pautas fundamentales*. Instituto de Pastoral Andina.
- Sánchez, L. (1999). Null objects and Do features in contact Spanish. In J.-M. Authier, B. E. Bullock, & L. A. Reed (Eds.), *Formal perspectives on Romance linguistics: Selected papers from the 28th Linguistic Symposium on Romance Languages (LSRL XXVIII), University Park, 16-19 April 1998* (pp. 227–242). John Benjamins Publishing Company.
- Sánchez, L. (2003). *Quechua-Spanish bilingualism: Interference and convergence in functional categories*. John Benjamins Publishing Company.
- Sánchez, L. (2004). Functional convergence in the tense, evidentiality and aspectual systems of Quechua Spanish bilinguals. *Bilingualism: Language and Cognition*, 7(2), 147–162.
- Sánchez, L. (2010). *Morphology and syntax of topic and focus: Minimalist inquiries in the Quechua periphery*. John Benjamins Publishing Company.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(3), 209–231.
- Shepard, D., & Padman, M. (2021). Esther Perel returns. *Armchair Expert*.

- Silva-Corvalán, C. (1993). On the permeability of grammars: Evidence from Spanish and English contact. In W. J. Ashby, M. Mithun, G. Perissinotto, & E. Raposo (Eds.), *Linguistic perspectives on the Romance Languages: Selected papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, California, 21-24 February 1991* (pp. 19–44). John Benjamins Publishing Company.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford University Press.
- Silva-Corvalán, C. (1998). On borrowing as a mechanism of syntactic change. In A. Schwegler, B. Tranel, & M. Uribe-Etxebarria (Eds.), *Romance Linguistics: Theoretical Perspectives* (pp. 225–246). John Benjamins.
- Silva-Corvalán, C. (2008). The limits of convergence in language contact. *Journal of Language Contact*, 2(1), 213–224.
- Sorlin, S., & Gardelle, L. (2018). Anthropocentrism, egocentrism and the notion of animacy hierarchy. *International Journal of Language and Culture*, 5(2).
- Stallings, L. M., MacDonald, M. C., & O’Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, 39, 392–417.
- Struik, T., & Schoenmakers, G.-J. (2023). When information structure exploits syntax: The relation between the loss of VO and scrambling in Dutch. *Journal of Linguistics*, 59, 655–690.
- Struik, T., & Van Kemenade, A. (2018). On the givenness of OV word order: A (re)examination of OV/VO variation in Old English. *English Language and Linguistics*, 24(1), 1–22.
- Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. In G. Purnelle, C. Fairon, & A. Dister (Eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10-12, 2004* (pp. 1032–1039). Louvain-la-Neuve, Presses universitaires de Louvain.

- Tagliamonte, S. A. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.
- Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*, 318–330. <https://doi.org/10.1016/j.jml.2011.04.009>
- Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333. <https://doi.org/10.1016/j.cognition.2006.09.011>
- Thomason, S. G. (2001). Contact-induced language change: Results. *Language contact: An introduction* (pp. 59–98). Edinburgh University Press.
- Thomason, S. G. (2010). Contact explanations in linguistics. In R. Hickey (Ed.), *The handbook of language contact* (pp. 31–47). John Wiley & Sons, Incorporated.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. University of California Press.
- Torero, A. (1964). Los dialectos quechuas. *Anales Científicos de la Universidad Agraria*, *2*, 446–478.
- Torero, A. (1974). *El quechua y la historia social andina*. Universidad Ricardo Palma.
- Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, *34*(3), 459–520. <https://doi.org/https://doi.org/10.1515/ling.1996.34.3.459>
- van Bergen, G., & de Swart, P. (2009). Definiteness and scrambling in Dutch: Where theory meets practice. In A. Schardl, M. Walkow, & M. Abdurrahman (Eds.), *Proceedings of NELS 38* (pp. 113–124). GLSA.
- van Bergen, G., & de Swart, P. (2010). Scrambling in spoken Dutch: Definiteness versus weight as determinants of word order variation. *Corpus Linguistics and Linguistic Theory*, *6*(2), 267–295.

- van Coetsem, F. (1988). *Loan phonology and the two transfer types in language contact*. Foris.
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9, 81–105.
- Wasow, T. (2002). *Postverbal behavior*. CSLI.
- Weinreich, U. (1953). *Languages in contact: Findings and problems*. New York.
- Winford, D. (2003). *An introduction to contact linguistics*. Blackwell Publishing.
- Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head final language. *Cognition*, 81, B45–B55.
- Yao, Y. (2018). NP weight effects in word order variation in Mandarin Chinese. *Lingua Sinica*, 4(5), 1–31.
- Zagona, K. (2002). *The syntax of Spanish*. Cambridge University Press.
- Zavala, V. (2001). Borrowing evidential functions from Quechua: The role of puses as a discourse marker in Andean Spanish. *Journal of Pragmatics*, 33, 999–1023.
- Zdrojewski, P., & Sánchez, L. (2014). Variation in accusative clitic doubling across three Spanish dialects. *Lingua*, 151, 162–176.
- Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. MIT Press.