

Information structure and language dominance in
subject pronoun resolution in Catalan-Spanish
bilingualism

Núria de Rocafiguera Montanyà

TESI DOCTORAL UPF / 2022

DIRECTOR DE LA TESI

Dra. Aurora Bel

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE



Agraïments

Abans de res, vull expressar el meu agraïment a l'Aurora Bel, directora d'aquesta tesi. Per la generositat, la confiança, el suport i l'acompanyament. Sento que és una sort haver après a fer recerca amb tu. Et vull donar les gràcies per l'exigència i el rigor, per les discussions diverses al voltant de la tesi i més enllà, pel temps que m'has dedicat, comentaris, revisions, per la implicació en tot plegat, per animar-me a participar a tants congressos, per obrir-me les portes d'altres projectes. Gràcies per les moltes estones compartides, com a directora de tesi i com a persona. Gràcies per moltes altres coses. Em sento afortunada d'haver-te tingut de directora.

També em sento afortunada d'haver comptat amb companyes de la universitat que faran que recordi aquesta etapa amb tant d'afecte. M'agradaria agrair a la Rut Benito unes mil coses aproximadament, no arribaria pas mai a dir-les totes. Gràcies per treballar tan bé, per alegrar-te de tot el que m'alegrava i per escoltar-me quan alguna cosa em feia patir; és una sort haver tingut algú com tu amb qui compartir tot això d'una manera tan propera. Gràcies a la Francesca Cerdà, la Mattea Cussel i l'Alba Marín: certament heu fet de la tesi una experiència molt més amable. No puc pensar en aquest temps sense veure-us-hi a vosaltres, gràcies per tot, de tot cor. A la Marússia Pronina, per aquesta curiositat que tens per tot i per tothom, que no s'acaba i que encomanes. A l'Ingrid Vilà, pel suport des del primer dia i tantes converses. A altres veïns com la Florence, en Patrick, l'Ío, la Xiatong, la Júlia, i a la Zi, la Dominika i als companys de la segona planta. A la Tamara Vorobyeva, l'Andrea Biró, l'Estela García-Alcaraz i la Jennifer Ament, estic contenta d'haver coincidit amb vosaltres i d'haver-vos tingut d'exemple. A en Fernando Martín-Villena, per compartir impressions i discussions diverses sobre la tesi en general i sobre els pronoms en particular, també per altres moments menys acadèmics, mil gràcies. I, d'una manera molt especial, gràcies a la Roser Giménez per donar-me confiança, pels comentaris i revisions de la tesi, per fer que tot sigui sempre una mica millor, i pels pastissos.

Vull agrair a la Universitat Pompeu Fabra i al Departament de Traducció i Ciències del Llenguatge els recursos que posen a l'abast dels doctorands per formar-nos, participar a congressos o fer estades. També vull donar les gràcies als directors del Departament amb qui he coincidit, Alex Alsina i Anna Espunya, a la coordinadora del programa de doctorat, Carmen Pérez, i a la meva tutora, Elisenda Bernal, així com a les persones de la secretaria del Departament amb qui he tingut més tracte: Núria Abad, Roger Aranich, Yolanda Bejarano, Susi Bolós, Joan Carles Carrion, Rafa Ordóñez i Núria Sánchez. Us agraeixo molt les ajudes diverses i que sempre hagueu fet els tràmits fàcils.

Després de passar tants anys en aquest departament, vull mostrar també un agraïment molt sincer a tots els professors de qui vaig ser alumna un dia i/o amb qui vaig treballar o impartir docència més endavant. Tinc la sort que heu set molts i que de tots vosaltres n'he après molt, que m'heu despertat l'interès per la lingüística i la investigació des de perspectives ben diferents. Vull agrair d'una manera especial a la Laia Mayol i l'Enric Vallduví els comentaris i suggeriments que em van fer en la recta final d'aquesta tesi. I també vull fer menció a l'Esteve Clua, la Carmen Pérez, la Silvia Perpiñán, la Pilar Prieto i l'Elisa Rosado, perquè us sento propers en aquesta etapa.

Moltíssimes gràcies, de tot cor, a la Nuria Sagarra per acollir-me a Rutgers University amb tanta generositat. Venir a fer una estada és una de les experiències més especials que m'emporto d'aquest doctorat. Gràcies per fer-m'hi un lloc, pels aprenentatges i per les oportunitats. Gracias a Joseph Casillas, Liliana Sánchez y José Camacho. Fue un placer coincidir con vosotros y aprender de vosotros. Gracias a Marcy Schwartz y a Rosy Ruiz por la calidez y por hacerlo todo fácil. Mil gracias a Cristina Lozano y a Laura Fernández, por esperarme con las puertas de vuestra acogedora casita abiertas y por los días en Highland Park, Nueva York y Canadá. Gracias a José Carlos, Abril, Laura, Julio, Katia, Juanjo, Ezequiel, David, Germán, Giselle, y a todos los compañeros del departamento por los buenos momentos compartidos. Gràcies, també, a en Winston i a en Ben. Esta tesis habría sido distinta si no hubiera pasado esos meses con todos vosotros.

Aquesta tesi no hauria set possible sense la col·laboració de tot de gent anònima que vau participar en els experiments. Gràcies a la Irene Baucells per l'ajuda en la recollida de dades. Gràcies a totes les persones que vau fer difusió dels experiments, a Raquel Fernández-Fuertes, María del Carmen Horno, Iraide Ibarretxe-Antuñano, José Luis Mendivil, al Departament de Comunicació de la UPF (a la secretaria i als professors que em van obrir les portes de les seves aules), i al Departament d'Educació Lingüística i Literària, i Didàctica de les Ciències Experimentals i la Matemàtica de la Universitat de Barcelona. Us agraeixo molt la bona predisposició, l'amabilitat i l'ajuda per trobar participants. Aquesta tesi, tal com és, tampoc no hauria set possible sense altres persones com en Joan Borràs, a qui agraeixo molt la paciència, ajuda i consells a l'hora de dissenyar els experiments i analitzar les dades, així com algunes altres converses. També agraeixo l'ajuda de l'Scott en l'anàlisi de dades i en correccions diverses de l'anglès. Per més correccions de l'anglès, també agraeixo el temps dedicat a aquesta tesi a en James, la Ruth i l'Steve.

Finalment, vull agrair la vida fora de la tesi (i el suport amb la tesi quan ha fet falta) a tot d'altres persones que heu set i sou importants per a mi. Gràcies a la Bet i l'Uri per ser-hi, per la companyia, per estar sempre a punt per celebrar les alegries, per fer-ho tot fàcil. Gràcies, amb un afecte molt especial, a l'Anna, a l'Alba, a la Irene, a la Marta, a la Mireia, a l'altra Mireia i a l'altra Mireia, a en Diego, a la Tònia; us valoro molt. Gràcies a la Bet, la Cristina, l'Eva, la Laia, la Laura, la Marta i l'Olga, per ser un lloc on tornar. A les UPF Girls, a les Nenes maques, a les Nenes tmique, a les Teachers 100, a la Xènia, per anar-nos trobant i compartint tan bones estones. A en Pau. Gràcies a la Clara, la Nur i l'Alfred. A la Nuria i família. Gràcies a tota la meva família sencera; espero que tots us sentiu en aquests agraïments i que també hi sentiu els que ja no hi són. Gràcies, sobretot, a les meves tietes (i a les ties), a l'àvia; també als veïns del costat, que, entre d'altres, van fer més agradables els dies de confinament, i a les dues protagonistes de la majoria dels exemples de la tesi, que us he gastat el nom.

En general, dono les gràcies a tothom amb qui he parlat, per una cosa o altra, més estona o menys, d'aquesta tesi, i a tothom que s'alegra que estigui acabada. Gràcies a qui potser no he mencionat directament en aquests agraïments però a qui també estic agraïda.

Gràcies, per tot, a la mare, al pare i a l'Anna, també a en Carles.

Aquesta tesi s'ha realitzat en el marc del Projecte FFI2016-75082-P i amb la beca concedida per l'AGAUR a través de la convocatòria d'ajuts per a la contractació de personal investigador predoctoral en formació (FI), 2018FI_B_00959, 2019FI_B1_00112 i 2020FI_B2_00081.

Abstract

Pronominal anaphora resolution, and particularly the interpretation of third-person null and overt subject pronouns, is a complex linguistic phenomenon of the syntax-pragmatics interface. On the one hand, it has been demonstrated to be sensitive to multiple factors, which have often overlapped in previous research, such as the syntactic function, the information status, or the linear and hierarchic position of the antecedent (Carminati, 2002; de la Fuente, 2015; Torregrossa et al., 2020). On the other hand, it has been defined as especially challenging for bilingual speakers: vulnerable to crosslinguistic influence and to the effects of bilingualism per se, consisting of the use of a bilingual strategy that would give rise to the overextension—or indeterminacy—of overt pronouns (Sorace, 2011). The aim of this study is twofold. Firstly, to examine how and to what extent syntactic, pragmatic, and sequential factors contribute to the interpretation of null and overt subject pronouns in Catalan and Spanish through the manipulation of information structure (comparing canonical sentences, clitic-left dislocations, subject clefts, and object clefts). Secondly, to investigate how language dominance affects pronoun resolution by early and highly functional Catalan-Spanish bilinguals in canonical sentences and when adding complexity with marked information structures.

Using a bidirectional design, Catalan-Spanish bilinguals with different language dominance profiles (Catalan-dominant, balanced, and Spanish-dominant) completed two equivalent forced-choice tasks in Catalan and in Spanish, and a group of Spanish monolinguals completed the task in Spanish. Briefly, the observed impact of information structures on pronoun resolution supports a multifactorial and form-specific approach (Kaiser & Trueswell, 2008), similar in Catalan and in Spanish; no overriding factor was identified (syntactic, pragmatic, or sequential), and the interpretive preferences of each pronominal form were found to be driven by a different combination of factors. Whereas null pronouns tend to prefer subject and topic antecedents (showing sensitivity to syntactic and pragmatic factors), overt pronouns tend to prefer object and second-mentioned, hierarchically lower, antecedents (showing sensitivity to syntactic and sequential factors). In canonical sentences, microvariation was shown between the two typologically similar null subject languages under study: although the two languages comply with the Position of Antecedent Hypothesis (Carminati, 2002), Catalan null and overt subject pronouns show more polarized preferences than both bilingual and monolingual Spanish (Bel & García-Alcaraz, 2018). In these canonical contexts, unidirectional crosslinguistic influence was found from Spanish toward Catalan (Romano,

2019), modulated by language dominance: the higher the dominance in Spanish, the stronger the influence (i.e., the attenuation of the categorical Catalan biases). In marked information structures, no dominance (and no bilingualism) effects were observed, suggesting that Catalan-Spanish bilinguals do not rely on a general bilingual strategy, as would be derived from the Interface Hypothesis (Sorace, 2011). Overall, these findings demonstrate that null and overt subject pronouns are sensitive in different ways to the interaction between syntactic, pragmatic, and sequential factors, that crosslinguistic influence can occur in contexts of microvariation, being modulated by language dominance, and that the use of a bilingual strategy in pronoun resolution cannot be generalized to early and highly functional bilingual populations.

Resum

La resolució de l'anàfora pronominal, i en particular la interpretació dels pronoms de subjecte nuls i explícits, és un fenomen complex de la interfície sintaxi-pragmàtica. Per una banda, s'ha demostrat que és sensible a múltiples factors, que amb freqüència s'han solapat en investigacions prèvies, com ara la funció sintàctica, l'estatus informatiu, o la posició lineal i jeràrquica de l'antecedent (Carminati, 2002; de la Fuente, 2015; Torregrossa et al., 2020). Per una altra banda, s'ha definit com a especialment costosa per als parlants bilingües: vulnerable a la influència translingüística i als efectes del bilingüisme *per se*, consistents en l'ús d'una estratègia bilingüe que donaria lloc a la sobreextensió—o indeterminació—de les propietats dels pronoms explícits (Sorace, 2011). Aquest estudi es proposa, en primer lloc, examinar com i fins a quin punt els factors sintàctics, pragmàtics i seqüencials contribueixen a la interpretació dels pronoms nuls i explícits en català i en espanyol mitjançant la manipulació de l'estructura informativa (comparant oracions canòniques, topicalitzacions de l'objecte via dislocació a l'esquerra, focalitzacions del subjecte via oracions escindides i focalitzacions de l'objecte via oracions escindides). En segon lloc, es proposa investigar si la dominança lingüística afecta la resolució pronominal en bilingües català-castellà primerencs i altament funcionals, tant en oracions canòniques com en oracions amb estructures informatives marcades, que hi afegeixen complexitat.

Fent ús d'un disseny bidireccional, bilingües català-castellà amb diferents perfils de dominança lingüística (catalanodominants, equilibrats i castellanodominants) van completar dues tasques de selecció forçosa equivalents en català i en castellà, i un grup de monolingües de castellà va completar la tasca en castellà. En síntesi, l'impacte observat de l'estructura informativa en l'anàfora pronominal avala una visió multifactorial específica per a cada forma pronominal (Kaiser i Trueswell, 2008), semblant en català i en castellà; no es va identificar un factor predominant (sintàctic, pragmàtic o seqüencial) i les preferències interpretatives de cada forma pronominal es van veure afectades per una combinació de factors diferents. Mentre que els pronoms nuls tendeixen a preferir antecedents subjecte i tòpic (mostrant sensibilitat a factors sintàctics i pragmàtics), els pronoms explícits tendeixen a preferir antecedents objecte i en segona posició, una posició jeràrquicament més baixa (mostrant sensibilitat a factors sintàctics i seqüencials). En les oracions canòniques, es va observar microvariació entre les dues llengües objecte d'estudi, ambdues de subjecte nul i tipològicament semblants: malgrat que ambdues llengües segueixen les prediccions de la Hipòtesi de la Posició de

l'Antecedent (Carminati, 2002), els pronoms de subjecte nuls i explícits en català mostren preferències més polaritzades que en castellà bilingüe i monolingüe (Bel i García-Alcaraz, 2018). En aquestes oracions canòniques, s'hi va observar influència translingüística unidireccional del castellà cap al català (Romano, 2019), modulada per la dominança lingüística: a major dominança del castellà, major influència (i.e., major flexibilització dels biaixos del català). En estructures informativament marcades, no es van apreciar efectes de dominança lingüística (ni de bilingüisme), la qual cosa suggereix que els bilingües català-castellà no recorren a una estratègia bilingüe, com es derivaria de la Hipòtesi de la Interfície (Sorace, 2011). En conjunt, aquests resultats demostren que els pronoms de subjecte nuls i explícits manifesten una sensibilitat diferent a la interacció entre factors sintàctics, pragmàtics i seqüencials, que la influència translingüística pot aparèixer en contextos de microvariació, modulada per la dominança lingüística, i que l'ús d'una estratègia bilingüe en la resolució de l'anàfora pronominal no sembla que es pugui generalitzar en poblacions de bilingües primerencs i altament funcionals.

Resumen

La resolución de la anáfora pronominal, y en particular la interpretación de los pronombres de sujeto nulos y explícitos, es un fenómeno complejo de la interfaz sintaxis-pragmática. Por un lado, se ha demostrado que es sensible a múltiples factores, que se han solapado con frecuencia en investigaciones previas, como la función sintáctica, el estatus informativo, o la posición lineal y jerárquica del antecedente (Carminati, 2002; de la Fuente, 2015; Torregrossa et al., 2020). Por otro lado, se ha definido como especialmente costosa para los hablantes bilingües: vulnerable a la influencia translingüística y a los efectos del bilingüismo *per se*, consistentes en el uso de una estrategia bilingüe que daría lugar a la sobreextensión—o indeterminación—de las propiedades de los pronombres explícitos (Sorace, 2011). Este estudio tiene un objetivo doble. En primer lugar, examinar cómo y hasta qué punto los factores sintácticos, pragmáticos y secuenciales contribuyen a la interpretación de los pronombres nulos y explícitos en catalán y en español mediante la manipulación de la estructura informativa (comparando oraciones canónicas, topicalizaciones del objeto vía dislocación a la izquierda, focalizaciones del sujeto vía oración escindida y focalizaciones del objeto vía oración escindida). En segundo lugar, investigar si la dominancia lingüística afecta a la resolución pronominal en bilingües catalán-español tempranos y altamente funcionales, tanto en oraciones canónicas como en oraciones con estructuras informativas marcadas, que añaden complejidad.

Haciendo uso de un diseño bidireccional, bilingües catalán-español con diferentes perfiles de dominancia lingüística (catalanodominantes, equilibrados y castellanodominantes) completaron dos tareas de selección forzosa equivalentes en catalán y en español, y un grupo de monolingües de español completó la tarea en español. Brevemente, el impacto observado de la estructura informativa en la anáfora pronominal avala una visión multifactorial específica para cada forma pronominal (Kaiser y Trueswell, 2008), similar en catalán y en español; no se identificó un factor (sintáctico, pragmático o secuencial) predominante y las preferencias interpretativas de cada forma pronominal se vieron afectadas por una combinación de factores de distinta índole. Mientras que los pronombres nulos tienden a preferir antecedentes sujeto y tópico (mostrando sensibilidad a factores sintácticos y pragmáticos), los pronombres explícitos tienden a preferir antecedentes objeto y en segunda posición, una posición jerárquicamente más baja en la estructura sintáctica (mostrando sensibilidad a factores sintácticos y secuenciales). En las oraciones canónicas, se observó microvariación entre las dos lenguas objeto de estudio,

ambas de sujeto nulo y tipológicamente similares: a pesar de que ambas lenguas siguen las predicciones de la Hipótesis de la Posición del Antecedente (Carminati, 2002), los pronombres de sujeto nulos y explícitos en catalán muestran preferencias más polarizadas que en español bilingüe y monolingüe (Bel y García-Alcaraz, 2018). En estas oraciones canónicas, se observó influencia translingüística unidireccional del español hacia el catalán (Romano, 2019), modulada por la dominancia lingüística: a mayor dominancia del español, mayor influencia (i.e., mayor flexibilización de los sesgos del catalán). En estructuras informativamente marcadas, no se apreciaron efectos de dominancia lingüística (ni de bilingüismo), lo que sugiere que los bilingües catalán-español no recurren a una estrategia bilingüe, como se derivaría de la Hipótesis de la Interfaz (Sorace, 2011). En conjunto, estos resultados demuestran que los pronombres de sujeto nulos y explícitos manifiestan una sensibilidad diferente a la interacción entre factores sintácticos, pragmáticos y secuenciales, que la influencia translingüística puede ocurrir en contextos de microvariación, siendo modulada por la dominancia lingüística, y que el uso de una estrategia bilingüe en la resolución de la anáfora pronominal no parece poderse generalizar a poblaciones de bilingües tempranos y altamente funcionales.

Contents

1. Introduction	1
1.1 Overview and motivation.....	1
1.2 Aims.....	5
1.3 Research questions	6
1.4 Outline of the remaining chapters	7
2. The linguistic phenomenon: pronominal anaphora resolution in null subject languages	9
2.1 Introduction.....	9
2.2 Anaphora resolution: theoretical framework and factors affecting pronoun interpretation.....	10
2.2.1 Theories of reference and the notion of prominence: the Accessibility Theory, the Givenness Hierarchy, and the Centering Theory	10
2.2.2 Main linguistic and cognitive factors intervening in pronominal anaphora resolution	12
2.2.3 Previous studies confronting semantic, syntactic, pragmatic, and sequential factors	15
2.3 Syntactic factors in anaphora resolution: the syntactic position of the antecedent. 18	
2.3.1 The Position of Antecedent Hypothesis	18
2.3.2 Previous studies on the Position of Antecedent Hypothesis in Catalan and in Spanish.....	22
2.3.3 Microvariation in null subject languages	27
2.4 Pragmatic factors in anaphora resolution: the information status of the antecedent	30
2.4.1 Information structure in the Romance languages.....	30
2.4.1.1 Unmarked structures: canonical sentences.....	33
2.4.1.2 Topicalization constructions: CLLD	34
2.4.1.3 Focalization constructions: it-clefts.....	35

2.4.1.4 On the notions of topic and focus in the present study	36
2.4.2 The impact of information structure on anaphora resolution: evidence from previous experimental studies	38
2.4.3 Previous studies on the influence of information status on pronoun interpretation in Catalan and Spanish	42
2.5 Sequential factors in anaphora resolution: order of mention or surface linear position of the antecedents	47
2.6 Summary and implications for the present study	49
3. The effects of bilingualism and language dominance on anaphora resolution in the Catalan-Spanish context.....	53
3.1 Introduction	53
3.2 Characterizing the bilingual experience.....	54
3.2.1 Defining bilingualism.....	54
3.2.2 Language dominance as a proxy for bilingualism	56
3.2.2.1 Defining language dominance.....	56
3.2.2.2 Factors affecting language dominance.....	57
3.2.2.3 Measuring and operationalizing language dominance.....	59
3.2.2.4 Classifying bilinguals.....	60
3.2.3 The effects of bilingualism per se and the emergence of crosslinguistic influence	61
3.2.3.1 General effects of bilingualism in language processing	62
3.2.3.2 Crosslinguistic influence in bilingual outcomes.....	63
3.2.4 Societal bilingualism and language contact	65
3.3 Bilingualism and language contact in Catalonia	67
3.3.1 Contextualizing bilingualism and language contact in Catalonia	67
3.3.2 Previous studies on language dominance in bilingual Catalan and bilingual Spanish.....	70
3.3.3 Assessing language dominance in adult Catalan-Spanish bilinguals.....	75
3.4 The effects of bilingualism on anaphora resolution in null subject languages	78

3.4.1 Empirical findings on bilingual anaphora resolution in null subject languages	78
3.4.2 The Interface Hypothesis and general effects of bilingualism on anaphora resolution	81
3.4.3 Crosslinguistic influence in anaphora resolution beyond the Interface Hypothesis	84
3.4.4 Previous studies on bilingual anaphora resolution in non-canonical sentences	88
3.5 Summary and implications for the present study	89
4. Methodology	91
4.1 Introduction	91
4.2 Participants	92
4.2.1 General description	92
4.2.2 Language background questionnaire: the Bilingual Language Profile questionnaire	94
4.2.3 Dividing bilingual participants into groups	99
4.2.4 Characterizing each group of Catalan-Spanish bilinguals	101
4.2.4.1 Language history	101
4.2.4.2 Language use	105
4.2.4.3 Language proficiency	107
4.2.4.4 Language attitudes	109
4.2.5 Summary of remarks on the Bilingual Language Profile questionnaire	110
4.3 Materials: the experimental task	112
4.3.1 The nature of the task	112
4.3.2 Experimental items	113
4.3.2.1 Variables and conditions	114
4.3.2.2 Design of the experimental items	115
4.3.2.3 Norming study	118
4.3.3 Practice items, fillers, and distractors	118

4.4 Procedure	119
4.5 Data analyses	121
4.5.1 Data preparation.....	121
4.5.1.1 Data cleaning.....	121
4.5.1.2 Binarization of the responses	123
4.5.2 Statistical analyses	126
4.5.2.1 Variables.....	126
4.5.2.2 Modeling	126
4.6 Summary	130
5. Disentangling the role of syntactic, pragmatic, and sequential factors in pronominal anaphora resolution in Catalan and Spanish: results and discussion.....	133
5.1 Introduction	133
5.1.1 Outline of the research questions	134
5.1.2 Organization of the present chapter.....	136
5.2 The interpretation of subject pronouns in bilingual Catalan and bilingual Spanish: how do syntactic and pragmatic factors shape null and overt subject pronoun resolution?	137
5.2.1 Method	137
5.2.1.1 Participants.....	137
5.2.1.2 Materials	138
5.2.1.3 Reported model	139
5.2.2 General considerations on the results	140
5.2.2.1 Descriptive results.....	140
5.2.2.2 Main effects and interactions.....	141
5.2.2.3 Presentation of the results in the following sections	141
5.2.3 Canonical unmarked structures	142
5.2.3.1 Aims and predictions.....	142

5.2.3.2 Results.....	143
5.2.3.3 Discussion	145
5.2.4 Topicalization structures via clitic-left dislocation	147
5.2.4.1 Aims and predictions.....	147
5.2.4.2 Results.....	149
5.2.4.3 Discussion	152
5.2.5 Focusing subjects and objects via it-cleft structures	155
5.2.5.1 Aims and predictions.....	155
5.2.5.2 Results.....	157
5.2.5.3 Discussion	161
5.2.6 Integrating findings: topicalization and focusing structures compared.....	164
5.2.6.1 Aims and predictions.....	164
5.2.6.2 Results.....	167
5.2.6.3 Discussion	170
5.2.7 Summary of main findings	176
5.2.7.1 Canonical unmarked structures.....	177
5.2.7.2 Object topicalization via clitic-left dislocation structures.....	177
5.2.7.3 Subject focalization via it-cleft structures.....	178
5.2.7.4 Object focalization via it-cleft structures	179
5.2.7.5 The interpretation of null and overt pronouns across structures	180
5.3 The interpretation of subject pronouns in monolingual Spanish: does it differ from Catalan and bilingual Spanish?.....	181
5.3.1 Characterizing pronoun resolution in monolingual Spanish and the effects of marked information structures	181
5.3.1.1 Aims and predictions.....	181
5.3.1.2 Method.....	182
5.3.1.3 Results.....	183
5.3.1.4 Discussion	186

5.3.2 Comparing pronoun resolution in monolingual Spanish and Catalan	190
5.3.2.1 Aims and predictions.....	190
5.3.2.2 Method.....	190
5.3.2.3 Results.....	191
5.3.2.4 Discussion	193
5.3.3 Comparing pronoun resolution in monolingual Spanish and bilingual Spanish in contact with Catalan.....	193
5.3.3.1 Aims and predictions.....	193
5.3.3.2 Method.....	194
5.3.3.3 Results.....	195
5.3.3.4 Discussion	197
5.3.4 Summary of main findings	198
5.4 General discussion	198
5.4.1 The interpretive preferences of null and overt pronouns in canonical sentences in Catalan and Spanish	199
5.4.2 The influence of syntactic, pragmatic and word order factors on null and overt pronoun resolution.....	202
5.4.3 Microvariation between Catalan, bilingual Spanish, and monolingual Spanish	205
6. Bilingualism effects on pronominal anaphora resolution in Catalan and Spanish in contact: results and discussion.....	209
6.1 Introduction	209
6.1.1 Outline of the research questions	212
6.1.2 Organization of the present chapter.....	212
6.2 How bilingualism and language dominance affect pronoun resolution in canonical sentences	213
6.2.1 Aims and predictions	213
6.2.2 Method	216
6.2.2.1 Participants.....	216

6.2.2.2 Materials	217
6.2.2.3 Reported model	218
6.2.3 Results	218
6.2.3.1 Overall results	219
6.2.3.2 Comparing bilinguals with different language dominance profiles (between-group comparisons)	220
6.2.3.3 Comparing the two languages of each group of bilinguals (within-group comparisons)	224
6.2.3.4 Summary of the results	225
6.2.4 Discussion	226
6.2.4.1 On general effects of bilingualism	226
6.2.4.2 On crosslinguistic influence and its directionality	229
6.2.5 Follow-up analysis: Language dominance as a continuous variable	232
6.2.5.1 Motivation of this follow-up analysis	232
6.2.5.2 Method	233
6.2.5.3 Results	234
6.2.5.4 Discussion	236
6.2.6 Summary of main findings	237
6.3 How bilingualism and language dominance affect pronoun resolution in marked information structures	239
6.3.1 Aims and predictions	239
6.3.2 Method	241
6.3.2.1 Participants	241
6.3.2.2 Materials	242
6.3.2.3 Reported model	242
6.3.3 Results	243
6.3.3.1 Overall results	243
6.3.3.2 The effects of marked information structures per group of bilinguals and language	244

6.3.3.3 Summary of the results	247
6.3.4 Discussion	248
6.3.5 Follow-up analysis: Response times.....	251
6.3.5.1 Motivation	251
6.3.5.2 Method.....	252
6.3.5.3 Results and discussion	252
6.3.6 Summary of main findings	255
6.4 General discussion	256
7. Conclusions.....	265
7.1 Summary of the study and conclusions.....	265
7.2 Further research	269
References	273
Appendices	311
Appendix A: Language background questionnaire (based on the BLP questionnaire)	313
Appendix B: Results of the language background questionnaire.....	317
Appendix C: Materials for the experimental task	321
Appendix D: Instructions of the experimental task.....	327
Appendix E: Boxplots for participants' analyses	329
Appendix F: Boxplots for items' analyses.....	333
Appendix G: Model 1, reported in §5.2.....	337
Appendix H: Models 2-4, reported in §5.3.....	337
Appendix I: Model 5, reported in §6.2 and §6.3.....	337
Appendix J: Follow-up analyses, reported in §6.2.5	337
Appendix K: Follow-up analyses, reported in §6.3.5	337

List of figures

2.1 Structure of complex clauses in main-subordinate and subordinate-main clause orders	19
4.1 Participants' global dominance scores in the Bilingual Language Profile (BLP) questionnaire	100
4.2 Scores on the BLP language history module in Catalan and in Spanish, by group of bilinguals.....	102
4.3 Scores on the BLP language use module in Catalan and in Spanish, by group of bilinguals.....	106
4.4 Scores on the BLP language proficiency module in Catalan and in Spanish, by group of bilinguals.....	108
4.5 Scores on the BLP language attitudes module in Catalan and in Spanish, by group of bilinguals.....	110
4.6 Example screen of an item of the task in Spanish.....	112
4.7 Histogram of the responses showing the preference toward subject or object antecedents in Catalan and Spanish in a 0-to-100 visual analogue scale	124
4.8 Degree of certainty of the responses to the experimental task per condition	125
4.9 Screenshot of some of the fixed effects in the output of a mixed-effects logistic regression model with an intercept.....	129
4.10 Screenshot of some of the fixed effects in the output of a mixed-effects logistic regression model with no intercept.....	130
5.1 Predicted subject interpretations of null and overt pronouns in unmarked contexts in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI).....	144
5.2 Predicted subject interpretations of null and overt pronouns in unmarked and CLLD structures, in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI)	150
5.3 Predicted subject interpretations of null and overt pronouns in unmarked, subject cleft and object cleft structures, in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI).....	158

5.4 Predicted subject interpretations of null and overt pronouns in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals, by information structure (± 95 CI).....	168
5.5 Predicted subject interpretations of null and overt pronouns in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals, by pronoun, information structure, and language (± 95 CI)	169
5.6 Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in Spanish.....	185
5.7 Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in (monolingual) Spanish and in Catalan.....	192
5.8 Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in Spanish by monolinguals and by Spanish-dominant bilinguals	196
5.9 Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in canonical sentences in Catalan, bilingual Spanish, and monolingual Spanish.....	199
6.1 Preferred interpretation of null and overt pronouns in canonical unmarked sentences in Catalan, by group of bilinguals ($\pm 95\%$ CI)	221
6.2 Preferred interpretation of null and overt pronouns in canonical unmarked sentences in Spanish, by group of bilinguals ($\pm 95\%$ CI).....	222
6.3 Density plot of raw responses showing the interpretive preferences of null and overt pronouns in canonical sentences in Catalan and Spanish, by group of bilinguals	224
6.4 Preferred interpretation of null and overt pronouns in canonical unmarked sentences in bilingual Catalan and in bilingual Spanish, by language and group ($\pm 95\%$ CI)	225
6.5 Effects of language dominance (BLP global score) on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals	235
6.6 Effects of language history BLP scores on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals	235
6.7 Effects of language use BLP scores on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals	236
6.8 Preferred interpretation of subject pronouns in Catalan and in Spanish, by pronoun, information structure, language, and group of bilinguals ($\pm 95\%$ CI)	245

6.9 Catalan-dominant bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language, and information structure ($\pm 95\%$ CI)	246
6.10 Balanced bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language, and information structure ($\pm 95\%$ CI)	246
6.11 Spanish-dominant bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language and information structure ($\pm 95\%$ CI)	247
6.12 Predicted response times in the interpretation of null and overt pronouns in Catalan and in Spanish, by group of bilinguals.....	247
E.1 Responses of Catalan-dominant bilinguals to each condition in Catalan, by participant.....	329
E.2 Responses of balanced bilinguals to each condition in Catalan, by participant ...	330
E.3 Responses of Spanish-dominant bilinguals to each condition in Catalan, by participant.....	330
E.4 Responses of Catalan-dominant bilinguals to each condition in Spanish, by participant.....	331
E.5 Responses of balanced bilinguals to each condition in Spanish, by participant ..	331
E.6 Responses of Spanish-dominant bilinguals to each condition in Spanish, by participant.....	332
E.7 Responses of Spanish monolinguals to each condition in Spanish, by participant	332
F.1 Responses of all the participants to each condition in Catalan, by item (I)	333
F.2 Responses of all the participants to each condition in Catalan, by item (II)	334
F.3 Responses of all the participants to each condition in Spanish, by item (I)	335
F.4 Responses of all the participants to each condition in Spanish, by item (II).....	336

List of tables

2.1 Main features characterizing subject and object antecedents in each information structure.....	50
3.1 First, identification and habitual language of the population of Catalonia aged 15 years and over (%)	69
3.2 Language uses of the population of Catalonia aged 15 years and over, by area of use (%)	69
3.3 Mean self-assessed knowledge of Catalan and Spanish (from 0 to 10) by the population aged between 15 and 29 years.....	70
4.1 General information about bilingual and monolingual participants.....	93
4.2 Participants' global dominance scores in the Bilingual Language Profile (BLP) questionnaire	99
4.3 Scores on the BLP language history module in Catalan and in Spanish, by group of bilinguals.....	101
4.4 Age of onset of acquisition of participants, by language and group of bilinguals .	103
4.5 Age of onset of acquisition in two slots (from birth and between 2-6 years of age), by language and group of bilinguals	104
4.6 Scores on the BLP language use module in Catalan and in Spanish, by group of bilinguals.....	106
4.7 Scores on the BLP language proficiency module in Catalan and in Spanish, by group of bilinguals.....	108
4.8 Scores on the BLP language attitudes module in Catalan and in Spanish, by group of bilinguals.....	110
4.9 Experimental items per condition in Catalan and in Spanish	114
4.10 Implicit causality of the verbs in Spanish used in the main clause of the experimental sentences	117
5.1 Proportion of subject interpretations in Catalan (by Catalan-dominant bilinguals) and in Spanish (by Spanish-dominant bilinguals), by information structure and type of pronoun (<i>SD</i>)	140

5.2 Summary of the interpretive biases of pronouns in bilingual Catalan and Spanish and the main features characterizing the antecedents in each information structure.....	172
5.3 Proportion of subject interpretations in monolingual Spanish (<i>SD</i>)	184
5.4 Proportion of subject interpretations in bilingual Catalan (by Catalan-dominant bilinguals), by information structure and type of pronoun (<i>SD</i>)	192
5.5 Proportion of subject interpretations in monolingual Spanish and in bilingual Spanish (by Spanish-dominant bilinguals), by information structure and type of pronoun (<i>SD</i>)	195
5.6 Summary of the features that make an antecedent more preferred for a null or an overt subject pronoun	203
6.1 Module scores in the Bilingual Language Profile questionnaire, by group	217
6.2 Proportion of subject interpretations for null and overt subject pronouns in canonical sentences (unmarked information structures) by each group of bilingual speakers (<i>SD</i>)	219
6.3 Proportion of subject interpretations for null and overt subject pronouns in unmarked and syntactically marked information structures, by group of bilinguals (<i>SD</i>)	243
B.1 Results of the history module in the BLP, by language and group of bilinguals...	317
B.2 Results of the use module in the BLP, by language and group of bilinguals	318
B.3 Results of the proficiency module in the BLP, by language and group of bilinguals	319
B.4 Results of the attitudes module in the BLP, by language and group of bilinguals	320
C.1 Experimental items (in the condition of null pronoun, unmarked information structure) used in the two-alternative forced-choice task in Catalan.....	321
C.2 Experimental items (in the condition of null pronoun, unmarked information structure) used in the two-alternative forced-choice task in Spanish	324

Chapter 1

Introduction

1.1 Overview and motivation

The present thesis is concerned with two main themes related to the linguistic phenomenon of interpreting null and overt third-person subject pronouns in Catalan and Spanish: 1) the role of information structure in the interpretive properties of null and overt pronouns and its interaction with syntactic factors, as well as the possible existence of microvariation between two null subject languages, and 2) the role of language dominance in modulating the effects of bilingualism on pronoun resolution by highly functional early Catalan-Spanish bilinguals, in a bidirectional analysis.

As null subject languages, both Catalan and Spanish allow the subject to be omitted through null pronouns and for it to be overtly expressed using overt pronouns or lexical subjects. However, the fact that two pronominal forms are syntactically licensed does not mean that they are in free variation, their expression depends on pragmatic conditions. In terms of anaphora resolution, each pronoun has specific interpretive preferences, and there is consensus to claim that these preferences are determined by a combination of factors from different linguistic domains, mainly syntax and pragmatics. However, the extent to which null and overt pronouns are sensitive to these different factors and their interactions has not yet been fully understood.

The interpretive preferences of null and overt pronouns in experimental studies have very often been investigated within the framework of the Position of Antecedent Hypothesis (PAH; Carminati, 2002). According to this hypothesis, null subject pronouns prefer to corefer with subject antecedents (i.e., in Spec,IP), whereas overt subject pronouns preferably corefer with object antecedents or antecedents in a lower structural position than Spec,IP. These preferences are illustrated in example (1.1) below, in Catalan.

- (1.1) a. La Laura_j va espantar la Maria_k quan *pro*_j va entrar a l'habitació. (Null pronoun)
b. La Laura_j va espantar la Maria_k quan ella_k va entrar a l'habitació. (Overt pronoun)
'Laura scared Maria when she went into the room.'

Overall, the PAH predicts that syntactic-structural factors such as the syntactic position of the antecedent guide the resolution of anaphoric dependencies in intrasentential

contexts. However, other studies have highlighted the relevance of pragmatic factors such as the information status of the antecedent (e.g., de la Fuente, 2015; Leonetti, 2021). In fact, pronoun resolution preferences have also been traditionally formulated in terms of topic continuity and topic shift. If we assume that unmarked sentences show a topic-comment discourse configuration, a preverbal subject antecedent tends to convey topical information. This raises a question that has been minimally explored in previous experimental studies: whether the interpretation of null and overt pronouns is primarily guided by syntactic or pragmatic factors, and how these factors interact. Additionally, in unmarked sentences, the topical subject antecedent appears first-mentioned, in initial position, which has been claimed to be a privileged position (e.g., Gernsbacher & Hargreaves, 1988) and which, interestingly, coincides with the highest structural position in the syntactic configuration (see Torregrossa et al., 2020). In fact, another question that derives from the PAH is whether null pronouns prefer subject antecedents (i.e., Spec,IP) or antecedents in the highest specifier position, and whether overt pronouns prefer object antecedents or antecedents in a lower position than Spec,IP. In canonical sentences, which have been the most widely studied contexts in previous research on pronoun resolution within the context of the PAH, subject antecedents convey topical information and appear in initial position (the highest structural position in the syntactic configuration). Therefore, the notions of subjecthood, topicality, and order of mention (or hierarchical height) overlap.

The first aim of the present thesis will be to disentangle these notions in order to assess the impact of syntactic (subject vs. object), pragmatic (topic vs. focus), and sequential (first-mention vs. second-mention) factors on null and overt subject pronoun anaphora resolution. To do so, we will investigate the interpretive preferences of null and overt subject pronouns in Catalan and Spanish for two potential antecedents (in subject and in object position) in various contexts. More specifically, we will use different marked information structures to manipulate the syntactic function, the information status, and the surface linear position of these antecedents: canonical unmarked sentences (1.1), object topicalization via clitic-left dislocation structures (1.2), subject focalization via it-cleft structures (1.3), and object focalization via it-cleft structures (1.4).

- (1.2) a. A la Maria la va espantar la Laura quan *pro/ella* va entrar a l'habitació. (Catalan)
b. A María la asustó Laura cuando *pro/ella* entró en la habitación. (Spanish)
'Maria, Laura scared her when she came into the room.'

- (1.3) a. Va ser la Laura qui va espantar la Maria quan *pro/ella* va entrar a l'habitació.
b. Fue Laura quien asustó a María cuando *pro/ella* entró en la habitación.
'It was Laura who scared Maria when she came into the room.'

- (1.4) a. Va ser a la Maria a qui va espantar la Laura quan *pro/ella* va entrar a l'habitació.
b. Fue a María a quien asustó Laura cuando *pro/ella* entró en la habitación.
'It was Maria whom Laura scared when she came into the room.'

Using a two-alternative forced-choice task (offline data), we will ask participants to interpret the ambiguous null and overt subject pronouns in the subordinate clause of the sentence as coreferring with one of the two antecedents previously introduced in the main clause of the sentence. These syntactically marked information contexts will allow us to test the preference for antecedents with different syntactic functions (subject vs. object) and to manipulate the information status of these antecedents (topical vs. focal), as well as the linear order in which they appear (first-mention vs. second-mention, or SVO vs. OVS). Moreover, the subject antecedent will not always occupy the structurally highest position in the syntactic configuration.

This study examines two languages: Catalan and Spanish. Studying two typologically similar null subject languages will enable us to compare 1) whether they show similar behaviors in pronoun resolution in unmarked contexts and 2) whether the contribution of marked information structures to these interpretive preferences is similar across languages. Carminati (2002) proposed the PAH for Italian and posterior studies confirmed its validity for other null subject languages (e.g., Bel & García-Alcaraz, 2018 in Catalan; Contemori & Di Domenico, 2021 in Spanish; Papadopoulou et al., 2015 in Greek; Rinke & Flores, 2018 in Portuguese; Wolna et al., 2022 in Polish). However, even though the PAH has been proven to explain the interpretive preferences of null and overt pronouns in a wide range of languages, microvariation has been shown between some of these languages. For instance, overt pronouns in Catalan have been argued to show more well-defined PAH-like biases than in Spanish (Bel & García-Alcaraz, 2018). Thus, as part of the first aim of the thesis, we will move beyond a description of pronoun interpretation in Catalan and Spanish and compare these two languages to explore whether microvariation is shown between them.

From the perspective of language acquisition, pronominal anaphora resolution has been characterized as a linguistically and cognitively complex phenomenon, given that it is located at the syntax-pragmatics interface and requires the integration of information from different sources. Moreover, it is a non-categorical and non-univocal phenomenon;

different referring expressions can be more or less valid depending on the specific context in which they appear and there is no one-to-one correspondence between referring expressions and functions. As a complex phenomenon, pronominal anaphora resolution—and specially the interpretation of overt pronouns—has been claimed to be particularly challenging in bilingual acquisition: vulnerable to crosslinguistic influence and the effects of bilingualism per se (Sorace, 2011). With regards to crosslinguistic influence, different approaches have predicted that it should occur from the language with the more categorical biases toward the language with more flexible biases (e.g., Hulk & Müller, 2000), from the language with more flexible biases toward the language with more categorical biases (e.g., Romano, 2019; Tsimpli et al., 2004), or from the dominant toward the non-dominant language (e.g., Yip & Matthews, 2000). Regarding bilingualism per se, in the light of the Interface Hypothesis (Sorace & Filiaci, 2006), it has been suggested that the increased cognitive load that characterizes bilingual processing leads bilinguals to overextend the properties of overt pronouns, resulting in unbiased interpretations of these pronouns. However, the extent to which bilingualism can affect anaphora resolution and the nature of these effects on the interpretive properties of null and overt pronouns is still controversial and not consistent across studies.

Taking into account that Catalan and Spanish have shown microvariation regarding anaphora resolution, and that this phenomenon has been identified as vulnerable in bilingual acquisition, the second main aim of the present thesis will be to investigate the effects of bilingualism and language dominance on anaphora resolution. On the one hand, we will study whether crosslinguistic influence occurs in this setting and in which direction (from Catalan to Spanish, from Spanish to Catalan, or bidirectionally, from the dominant to the non-dominant language), as well as whether it is modulated by language dominance. On the other hand, we will explore whether any effects of bilingualism per se resulting in the overextension of overt pronouns can be observed in the two languages or in the non-dominant language of highly functional early bilinguals that have been raised in a bilingual society. Marked information structures will be especially relevant in this regard, given that they can be considered more complex to process and understand than unmarked structures, and we could therefore expect anaphora resolution in these contexts to be even more challenging for bilingual speakers than unmarked sentences. Most research on bilingualism and anaphora resolution has targeted L2 learners and, to a lesser extent, heritage and attrited speakers. In the present study, we will explore a bilingual population that, to date, has received almost no attention: highly functional early bilinguals that have been raised in a bilingual society (i.e., Catalan-Spanish bilinguals).

Another contribution of this study will be to investigate the effects of bilingualism in Catalan and Spanish, a combination of two typologically similar and closely related languages.

The participants in the present study will be mainly Catalan-Spanish bilinguals that were born and raised in Catalonia, and we will use a two-alternative forced-choice task with two conditions: pronoun (null or overt) and information structure (canonical sentences, clitic-left dislocations, subject clefts, or object clefts). Two equivalent versions of this task, one in Catalan and one in Spanish, will be used to collect data from the two languages of the bilinguals. When investigating the impact of information structure on pronoun resolution, we will be describing the bilingual varieties of Catalan and Spanish in contact. The population that we will take as a reference to investigate pronoun resolution and the effects of marked information structures in Catalan will be Catalan-dominant bilinguals, and Spanish-dominant bilinguals will be the reference group for Spanish. Given that the second main aim of the study is to investigate the effects of bilingualism and language dominance on anaphora resolution, three groups of Catalan-Spanish bilinguals differing in their language dominance profile (measured using the Bilingual Language Profile; Birdsong et al., 2012) will complete the tasks: Catalan-dominant bilinguals, balanced bilinguals, and Spanish-dominant bilinguals. Additionally, we will also consider a group of Spanish monolinguals. We will not be able to include Catalan monolinguals in the study because, as we will see, they do not exist. In any case, in addition to comparing bilingual Catalan and bilingual Spanish between groups and within groups, we will compare the interpretation of null and overt pronouns in unmarked and marked information structures in bilingual Spanish and monolingual Spanish.

1.2 Aims

As previously mentioned, the aim of the present thesis is twofold: 1) it aims to explore the role of information structure in the interpretive preferences of null and overt pronouns in Catalan and Spanish, and 2) it aims to examine the role of language dominance in modulating the effects of bilingualism on the interpretive preferences of null and overt pronouns in Catalan and Spanish by highly functional Catalan-Spanish bilinguals.

Firstly, through the manipulation of information structure, our aim will be to investigate how syntactic, pragmatic, and sequential factors influence the interpretation of ambiguous null and overt subject pronouns. More specifically, we will extricate three

notions that have often overlapped in previous experimental studies: the syntactic function of the antecedent (subject vs. object), the information status of the antecedent (topic vs. focus), and the linear position of the antecedent (first-mention vs. second-mention; SVO vs. OVS). At the same time, we will also compare the two languages under study, Catalan and Spanish, to contrast the resolution preferences of null and overt pronouns and the impact of marked information structures in two typologically similar languages that have shown microvariation.

Secondly, we will aim to explore the effects of bilingualism and language dominance on pronoun resolution by comparing three groups of Catalan-Spanish bilinguals with different dominance profiles (Catalan-dominant, balanced, and Spanish-dominant bilinguals) in their two languages (Catalan and Spanish) in a bidirectional study. More specifically, we will explore whether crosslinguistic influence occurs between Catalan and Spanish, in which direction it occurs, and whether it is modulated by language dominance. Alternatively, we will consider the possibility that, as a consequence of bilingualism per se, the interpretation of overt pronouns by bilingual speakers shows indeterminacy.

1.3 Research questions

According to these two main aims, the present study is guided by the following research questions:

- RQ1: Can the predictions of the Position of Antecedent Hypothesis explain the interpretive biases of null and overt subject pronouns in unmarked structures in bilingual Catalan and bilingual Spanish?
- RQ2: Are the preferences of null and overt pronouns toward subject and object antecedents in bilingual Catalan and bilingual Spanish affected by the manipulation of information structure? How and to what extent do the syntactic function, information status, and linear position of the antecedent shape null and overt pronoun resolution?
- RQ3: Is microvariation shown in the interpretation of null and overt pronouns in canonical structures in Catalan and Spanish? Are the effects of information structure similar in the two languages?

RQ4: How are null and overt pronouns interpreted in canonical and marked information structures in monolingual Spanish? To what extent do these interpretations differ from (bilingual) Catalan and bilingual Spanish?

RQ5: Does language dominance modulate pronoun resolution in canonical sentences by Catalan-Spanish bilinguals?

RQ6: Does language dominance modulate pronoun resolution in marked information contexts by Catalan-Spanish bilinguals? Is there evidence of a general effect related to bilingualism per se?

These questions will be presented again, together with specific predictions, in Chapter 5, which will address RQ1-RQ4, and in Chapter 6, which will answer RQ5 and RQ6. In these chapters, more specific questions will also be presented to guide the presentation of the results and the partial discussions.

1.4 Outline of the remaining chapters

The present thesis is organized as follows. Chapter 2 presents a broad perspective of the linguistic phenomenon of pronominal anaphora resolution and the main factors that have been found to shape subject pronoun interpretation. Special attention is given to syntactic factors and to pragmatic factors related to information structure, as well as to sequential factors. Chapter 3 is dedicated to the main issues regarding bilingualism, language dominance, and the most relevant studies on bilingualism and anaphora resolution that frame the present research. Chapter 4 describes the methodology followed in the study. It characterizes all the participants and discusses how their language dominance profile was defined. Later, it describes the experimental task, its design, the data collection procedure, and the data cleaning and subsequent analysis procedure.

Chapter 5 presents and discusses the findings regarding the role of information structure in anaphora resolution and the interaction between syntactic, pragmatic, and sequential factors. The aims, research questions and predictions of the study are presented, the methodology is briefly recapitulated, the results are reported, and the relevant findings discussed. Chapter 6 follows a similar structure to the previous chapter in order to present and discuss the effects of bilingualism and language dominance on pronominal

anaphora resolution by Catalan-Spanish bilinguals. Finally, Chapter 7 summarizes the main conclusions drawn from this thesis, along with some considerations for further research.

In addition to the appendices that can be found at the end of this document, the dataset, the code used to run the analyses in R, and the full outputs of the reported analyses are available in the Open Science Framework repository (https://osf.io/nqe5j/?view_only=33edd018f33c424891b5f99507384351).

Chapter 2

The linguistic phenomenon: pronominal anaphora resolution in null subject languages

2.1 Introduction

The linguistic phenomenon under study in the present thesis is the interpretation of third person anaphoric subject pronouns in null subject languages. Catalan and Spanish, as null subject languages with rich person and number agreement marking on verbs, allow for both overtly expressing the subject—either using a lexical subject (2.1a) or a pronominal subject (2.1b)—and for omitting it (2.1c) (Chomsky, 1981, 1982; see Camacho, 2013).

- (2.1) a. La Laura sap que el muguet és una flor de muntanya. (Catalan)
b. Ella sap que el muguet és una flor de muntanya.
c. Sap que el muguet és una flor de muntanya.
'Laura/She/*pro* knows that the lily of the valley is a mountain flower.'

Catalan and Spanish have two types of pronouns: phonologically unrealized null pronouns (*pro* in generative grammar; Chomsky, 1982) and overt pronouns (for third person singular subjects: *ell/ella* in Catalan, *él/ella* in Spanish, 'he'/'she'). Although null and overt subject pronouns are both grammatically licensed, their alternation is not totally free but modulated by a combination of factors from different linguistic domains, as well as by cognitive strategies. Speakers of null subject languages need to be able to select the appropriate pronominal form in each context, and comprehenders need to interpret it by appropriately identifying its antecedent. As argued in Camacho (2013, p. 26), the contrast between null and overt pronouns may be used to convey several interpretive differences (see also Leonetti, 2021). In null subject languages, null subject pronouns have been assumed to be the default option, whereas overt pronouns would be more restricted in distribution and show a specialized function (cf. the Avoid Pronoun Constraint, Chomsky, 1981; the Overt Pronoun Constraint, Montalbetti, 1984).

The relationship between pronominal anaphoric expressions and their antecedents, especially that of ambiguous anaphoric subject pronouns, has largely been the focus of linguistic and psycholinguistic research and has revealed a range of language and

cognitive factors that impact antecedent identification (see Arnold, 2010; Kaiser, 2011; von Heusinger & Schumacher, 2019 for overviews). In the following sections, we will review these studies, which are not always in agreement or uncontroversial, in order to offer an overview of the main factors that have been shown to influence pronominal anaphora resolution (§2.2). We will then focus on the factors to be analyzed in the present thesis: syntactic or structural factors, namely the syntactic function of the pronouns' plausible antecedents (§2.3); pragmatic factors, namely the information status of the antecedents (§2.4), and sequential factors, namely word order (i.e., order of mention or linear position of the antecedents) (§2.5). We will conclude this chapter with a summary of implications for the present study (§2.6).

2.2 Anaphora resolution: theoretical framework and factors affecting pronoun interpretation

Several theories of reference have been proposed in terms of explaining how speakers build coreference chains in discourse so that addressees have enough information to easily interpret anaphoric referring expressions. Some of these theories, such as the Accessibility Theory (Ariel, 1991, 2001) or the Givenness Hierarchy (Gundel et al., 1993), as well as the Centering Theory (Grosz et al., 1995), have postulated referential hierarchies that are based on the cognitive accessibility, prominence or salience of potential antecedents for an anaphoric expression. Based on these proposals, numerous studies have investigated anaphora resolution and the role of linguistic and cognitive factors in influencing the prominence of referents as plausible antecedents for different forms of referring expressions. In this section, we will first review the main claims of said theories. Subsequently, we will refer to the main semantic, syntactic, pragmatic, and cognitive factors that have been identified in terms of governing pronoun resolution.

2.2.1 Theories of reference and the notion of prominence: the Accessibility Theory, the Givenness Hierarchy, and the Centering Theory

Firstly, the postulation of the Accessibility Theory (Ariel, 1990, 1991, 2001) has its roots in the idea that the form of a referring expression signals the discourse status of an anaphor's referent: reduced referring expressions (i.e., pronouns) refer to the most salient antecedents, or antecedents that are highly accessible in the addressee's mind, while semantically rich expressions (i.e., noun phrases) refer to less salient or less

accessible antecedents. In other words, the more reduced a form is (i.e., less informative), the more accessible its antecedent needs to be. Ariel's accessibility marking scale is shown in (2.2) (adapted from Ariel, 1990, p. 73). It places in order referring expressions from coreferring with low accessibility to high accessibility referents. In null subject languages, null pronouns should refer to more salient or accessible referents than overt pronouns (likewise demonstrative pronouns).

(2.2) Full name + modifier > full name > long definite description > [...] > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > zero¹

Secondly, the Givenness Hierarchy (Gundel et al., 1993) is concerned with the cognitive status of the intended referent encoded by a referring expression. That is, whether a referent is part of general cultural knowledge or it has been linguistically introduced, as well as whether the addressee already has a mental representation of a referent and whether its attention is focused on this referent. Gundel et al.'s hierarchy (p. 275) is reproduced in (2.3), ordered from the most restrictive cognitive statuses to the least restrictive ones. Unstressed pronouns in non-null subject languages and null pronouns in null subject languages, in contrast to more informative referential expressions, require the intended referent to be "in focus" of attention (and that it is also activated, familiar, and so on, given that more restrictive statuses all entail lower statuses). On the other hand, demonstrative pronouns—at least in non-null subject languages—are used to refer to an entity which is activated but not in focus.

(2.3) in focus > activated > familiar > uniquely identifiable > referential > type identifiable

Finally, the Centering Theory (Grosz et al., 1995) proposes that the speakers' use of referring expressions is conditioned by the "centering" properties of an entity in discourse. Although "centers" are argued to be determined by a combination of factors (such as topic transitions), their grammatical role is signaled as crucial in order to rank these entities: subjects outrank objects and other elements as upcoming 'centers' of discourse (subject > object(s) > other; p. 214). Pronominal expressions are seen as referring to high-ranked entities and as indicating continuity in maintaining discourse coherence.

The hierarchies established in these theories all predict that more accessible, prominent or salient referents, entities in focus of attention, or centers of attention, are more easily

¹ In this hierarchy, zero stands for null subject pronouns in our languages.

pronominalized. However, how the central notion of prominence should be defined is far from clear, as well as what features make an antecedent more prominent or accessible for a subsequent pronoun. In this regard, based on the initial proposal of Himmelmann and Primus (2015) and research on anaphora resolution, von Heusinger and Schumacher (2019) propose three criteria to characterize the notion of prominence in grammar. Firstly, they argue that prominence is a relational notion, given that the status of an entity can only be singled out in relation to that of other entities of the same type and structure. Secondly, they define prominence as dynamic, given that the prominence of an element may change as the discourse unfolds. And thirdly, they claim that prominent referring expressions are structural attractors, so they are more likely to create or continue referential chains—more evident in language production. In the following section, we will review the main factors that have been identified in previous experimental studies as contributing to singling out a referential entity and making it more or less preferred for a pronominal subject ("prominence-lending cues", in terms of von Heusinger & Schumacher, 2019). In the present study, we will refer to antecedents as being more or less *preferred* for null and/or overt subject pronouns, rather than as being more or less *prominent* in discourse.

2.2.2 Main linguistic and cognitive factors intervening in pronominal anaphora resolution

There is general agreement in the literature in two statements on pronominal anaphora resolution in null subject languages: 1) that null pronouns tend to have a preference for referring to more prominent antecedents than overt pronouns, and 2) that multiple cues are involved in reference assignment processes, so that only multifactorial approaches can account for pronominal interpretive preferences. In effect, it is a combination of factors from different linguistic (as well as cognitive) domains that makes an antecedent more or less preferred for a pronoun, and therefore shapes the resolution preferences of anaphoric pronouns in comprehension data. Also, some factors seem to be more determinant than others, as they are derived from studies that have confronted different cues. We will develop these main ideas in more detail in the following paragraphs.

Regarding the role of semantic information in reference assignment, implicit causality has been one of the more widely studied linguistic factors. Transitive verbs' inherent semantic content has been demonstrated to impact the referential biases of subsequent subject pronouns so that some verbs favor subject interpretations (e.g., *to apologize*)

and others, object interpretations (e.g., *to admire*) (Garvey & Caramazza, 1974; Goikoetxea et al., 2008; Koornneef & Sanders, 2013; McKoon et al., 1993; Pyykkönen & Järvikivi, 2010; among others). Similarly, another factor identified as crucial has been thematic role, personal pronouns preferring proto-agents over proto-patients (Schumacher et al., 2017) (see also Arnold, 2001; Goikoetxea et al., 2008; Stevenson et al., 1994). Taken together, the impact of implicit causality biases and thematic role structure demonstrates that the type of verb affects referent accessibility. Additionally, in semantico-pragmatic terms, connectors have been found to be highly influential in activating inherent verb biases. Pronoun preferences are also mediated by clause relationship or coherence relations (e.g., Arnold, 2001; Kehler et al., 2008; Kehler & Rohde, 2013; Portele & Bader, 2020; Stevenson et al., 1994, 2000; Wolf et al., 2004). Semantic factors, however, are not directly addressed in the present thesis. Given that our main interest is to explore syntax-pragmatics factors, we have dealt with globally ambiguous contexts. We will control for semantic cues to be as neutral as possible (by controlling implicit causality, using temporal connectors, and maintaining thematic roles constant) so that the intrinsic preferences of null and overt pronouns arise (see Carminati, 2002; see also §4.3.2.2).

Pronominal subject anaphora resolution has also been shown to be highly sensitive to syntactic factors and, more specifically, to the *principle of subject preference* (Crawley et al., 1990; Frederiksen, 1981). A very strong preference of personal pronouns for subject antecedents has been widely attested in previous studies in non-null subject languages (Arnold, 2010; Bader & Portele, 2019 for German; Kaiser, 2011 for English; Kaiser & Trueswell, 2008 for Finnish; among others). In null subject languages, the Position of Antecedent Hypothesis (PAH; Carminati, 2002 for Italian) deserves special attention. According to this strategy, null pronouns tend to refer to subject antecedents (or antecedents in Spec,IP), which are considered to be more structurally prominent, and overt pronouns to object antecedents (or antecedents in positions lower than Spec,IP). At the same time, these preferences are conditioned by clause order (see Carminati, 2002; Chamorro, 2018; de Rocafiguera & Bel, 2022). The role of syntactic factors on anaphora resolution is further addressed in §2.3, as well as the PAH (§2.3.1), which is central for the purposes of the present study.

In pragmatic or discourse terms, the notions of topicality and givenness have often been intuitively related to prominence (see Cowles et al., 2007; Kaiser, 2011; von Heusinger & Schumacher, 2019). Topical or given information has been considered to be more

prominent and thus more likely to be pronominalized than new information (Chafe, 1976; Givón, 1983; Prince, 1981). In null subject languages, null pronouns have traditionally been assumed to be specialized in conveying topic maintenance (thus preferring to corefer with topical antecedents) and overt pronouns in conveying topic shift (thus preferring to corefer with non-topical antecedents), although null pronouns can also be felicitously used to express topic shift (e.g., see García-Alcaraz & Bel, 2019; Leonetti, 2021; Lozano, 2016; Lubbers-Quesada & Blackwell, 2009).

Involving both syntactic and pragmatic factors, topicalized and focused antecedents via marked information structures have been argued to enhance discourse prominence and accessibility (e.g., Arnold, 1998; Cowles et al., 2007; Foraker & McElree, 2007; Kaiser, 2011a; Runner & Ibarra, 2016). Other studies, however, have found that focusing structures, in contrast to topicalization structures, make an antecedent less accessible for a personal pronoun (Colonna et al., 2012, 2015; de la Fuente, 2015; Patterson et al., 2017; Patterson & Felser, 2020). Pragmatic factors related to information structure and to the notions of topic and focus are also fundamental for the purpose of the present study, and as such they are further addressed in §2.4.

Another factor that has been claimed to influence anaphora resolution is the linear position of the antecedents. In this regard, an especially influential proposal has been that of the first-mention advantage (Gernsbacher & Hargreaves, 1988) (see also Arnold et al., 2000; Contemori & Dussias, 2020). In SVO (Subject-Verb-Object) languages, however, the first-mentioned antecedent often coincides with the grammatical subject. Järvikivi et al. (2005) contrasted the first-mention advantage and the subject-preference accounts in Finnish, using SVO and OVS word orders, and concluded that both the order of mention and grammatical role were relevant for pronoun resolution; they did not attest neither a first-mention advantage nor a systematic preference for subjects. However, Kaiser and Trueswell (2008), in another eye-tracking study on personal pronouns in Finnish, found the subject preference to be stronger than the first-mention preference (see also Bader & Portele, 2019; Fukumura & van Gompel, 2015).

Finally, Arnold (2010) has highlighted that non-linguistic processing constraints also impact the choice and interpretation of referential forms. When speakers experience cognitive load, they might use more explicit referential forms than necessary from the listeners' perspective, according to the discourse status or accessibility of the referent (see also Contemori & Ivanova, 2021; Hendriks, 2016). Taking into account cognitive factors, working memory has also been claimed to play a key role in discourse

representations and referential processing (Almor, 1999; Almor & Nair, 2007). Since the integration of information in discourse when identifying a referent may necessitate increased cognitive resources in bilingual populations, cognitive strategies to overcome increased processing costs have also been shown to influence anaphora resolution (Sorace, 2011, 2016). We further address the role of cognitive factors on pronoun interpretation as general effects of bilingualism in the next Chapter 3, more specifically in §3.4. In this section, we will also address the role of other language-external or individual factors that have been shown to affect the acquisition of reference, such as crosslinguistic influence and language dominance (e.g., Torregrossa et al., 2021).

2.2.3 Previous studies confronting semantic, syntactic, pragmatic, and sequential factors

None of the previously mentioned factors have been found to act as an overarching factor governing anaphora resolution. At the same time, however, not all of them seem to be similarly relevant. That is, the multiple factors involved in pronominal anaphora resolution may be differently ranked or have different weights in defining pronominal interpretive preferences. In the following paragraphs we will review a number of experimental studies that have considered multiple cues in order to identify which features pronominal subjects are more sensitive to.

Kaiser and Trueswell (2008) considered the role of syntactic and information structure factors in Finnish personal and demonstrative pronouns. They found the personal pronoun (*hän*) to be mainly sensitive to the syntactic role of the antecedent (preferring subjects) and the demonstrative pronoun (*tämä*) to be guided by a combination of information structure/word order and syntactic function (preferring discourse-new, postverbal, and object referents). In light of this different sensitivity of each pronominal form to syntax and information structure, Kaiser and Trueswell (2008) proposed the form-specific multiple-constraints approach (see also Bader & Portele, 2019; Schumacher et al., 2017). According to this proposal, different pronoun types (e.g., personal and demonstrative pronouns) may exhibit different degrees of sensitivity to the linguistic factors that influence pronoun resolution. In German, Bader and Portele (2019) similarly found syntactic function to determine the preferences of the personal pronoun *er* (biased toward subject antecedents), and a combination of linear position and topicality to determine the preference of the d-pronoun *der* (biased toward clause-final objects).

Other studies have focused solely on personal pronouns. In a study on English, Kaiser (2011) found subjecthood (and agentivity) to have particularly strong effects on pronoun interpretation, regardless of information structure. These pronouns were sensitive to both topicalization and focusing effects, which boosted the preference for topicalized and focused referents. However, these preferences did not overcome the preference for subjects. Similarly, Blything et al. (2021) found subjecthood/agentivity/ first-mention cues to have a stronger role in the offline interpretation of English subject pronouns in focus position (via prosodic marking and it-clefts). On the other hand, Colonna et al. (2012) found that the preference of personal pronouns in German toward subject antecedents was only attested when the subject coincided with the sentence topic. When the object was topicalized, personal pronouns remained unbiased, indicating a similar weight of syntactic function and information structure.

Other studies by Schumacher et al. (2016, 2017) investigated the interaction between thematic role, grammatical function, and order of mention cues on the interpretation of personal pronouns and d-pronouns in German. Schumacher et al. (2016), using an offline forced-choice task, found personal pronouns (*er*) to prefer subject and agent arguments, and d-pronouns (*der*) to prefer non-subject and non-agent arguments². Order of mention was not found to have a relevant role in this context. In a similar study using both offline and eye-tracking measures, Schumacher et al. (2017) argued that thematic role is a more powerful cue in pronoun resolution in German than grammatical function and linear position (personal pronouns preferring proto-agents and d-pronouns preferring proto-patients). However, these preferences guided by thematic role became weakened when these cues were misaligned with grammatical function and word order cues (i.e., proto-agents not being marked with nominative case, and proto-patients not being postverbal or second-mentioned).

Overall, in non-null subject languages, personal pronouns seem to be especially sensitive to the syntactic function of the antecedent (or thematic role), and possibly to information structure as well. On the other hand, demonstrative pronouns in Finnish and German also seem to show a special sensitivity to syntactic function (or thematic role) and word order or surface linear position. However, mixed findings have also been obtained. Regarding null subject languages, very few studies have confronted multiple cues in a similar way. In light of the Form-specific multiple-constraints approach (Kaiser

² Note that these results show a division of labor, or complementary preferences, of personal and d-pronouns in German, in contrast with the predictions of the *form-specific multiple constraints approach* (Kaiser & Trueswell, 2008).

and Trueswell, 2008), null and overt pronouns, as different pronominal forms, may also, to different extents, be sensitive to syntactic, pragmatic, and sequential factors. This proposal has been highlighted by the authors to be "most relevant for languages that have two (or more) anaphoric forms that cannot be distinguished on the basis of their informativity" (Kaiser and Trueswell, 2008, p. 741). They do not make any reference to null subject languages and, in fact, the possibility that null and overt pronouns are not equally sensitive to different factors has not been the concern of much previous research on pronominal anaphora interpretation following non-variationist approaches.

Many experimental studies contrasting the interpretation of null and overt pronouns have been framed within the PAH and have investigated their sensitivity to a single and isolated factor: the syntactic function of the antecedents. However, Fedele and Kaiser (2014) referred to the Form-specific multiple-constraints approach to account for a different weight of cues related to grammatical factors (i.e., syntactic function of the antecedent), semantic factors (i.e., implicit causality of the main verb) and sentence boundaries (i.e., intrasentential vs. intersentential) in the interpretation of null and overt pronouns in Italian. Using a sentence completion task, they found null pronouns to show a stronger sensitivity to the syntactic function of the antecedent and to sentence boundaries, and overt pronouns to verb semantics.

To our knowledge, only Mayol (2010) has confronted the role of syntactic and information structure factors in the interpretive preferences of null and overt pronouns in Catalan. By comparing canonical contexts and object topicalization via clitic-left dislocation structures, she found the interpretation of null pronouns to be essentially guided by the syntactic function of the antecedent (preferring subjects), and the preferences of overt pronouns to be guided by both syntax and information structure (not showing a clear preference in non-canonical contexts). In Spanish, de la Fuente (2015) also studied, more exhaustively, the role of marked information structures in anaphora resolution. In this case, he used both subject and object topicalization and focusing structures. However, he only concentrated on null pronouns. Therefore, null and overt pronouns have not been contrasted before in a design that allows for fully disentangling the role of syntactic (subject vs. object), pragmatic (topic vs. non-topic/focus), and sequential (SVO vs. OVS) factors in anaphora resolution. Filling this knowledge gap is one of the main purposes of the present thesis. In the following sections, we will discuss in more detail the role of syntactic, pragmatic, and sequential factors by reviewing and synthesizing the main findings attested in the literature.

2.3 Syntactic factors in anaphora resolution: the syntactic position of the antecedent

2.3.1 The Position of Antecedent Hypothesis

Pronominal anaphora interpretation has been found to be particularly sensitive to syntactic or structural factors. In non-null subject languages, a referent in subject position has been described to be more preferred for personal subject pronouns than a referent in a non-subject position (e.g., Arnold, 2010; Bader & Portele, 2019). In null subject languages, capitalizing on the Position of Antecedent Hypothesis (PAH) proposed by Carminati (2002) for intra-sentential anaphora in Italian, null and overt pronouns have been claimed to display a division of labor in their preferential interpretation choices. The PAH predicts that null pronouns have a strong preference for antecedents in Spec,IP (i.e., subject antecedents), whereas overt pronouns show a clear bias for antecedents in lower syntactic positions (i.e., object antecedents). As illustrated in (2.4) by the subindex, when ambiguity is at stake the null pronoun (*pro*) tends to prefer to retrieve the subject antecedent in the preceding clause (*Mario*) and the overt pronoun (*lui*), the object antecedent (*Giovanni*).

(2.4) Quando Mario_j ha telefonato a Giovanni_k, *pro*/*lui*_k aveva appena finito di mangiare.

'When Mario called Giovanni, he had just finished eating.'

(Example from Carminati, 2002, p. 33)

Also in structural terms, resolution preferences of ambiguous pronouns in intra-sentential anaphora have been found to be affected by clause order. When proposing the PAH, Carminati (2002) mainly evaluated complex clauses with the subordinate-main clause order, as in (2.4). However, she also looked at main-subordinate sequences as in (2.5). Even if pronouns showed robust preferences and the PAH was also found to hold for this clause order, the subject bias of the null pronoun seemed weaker than in the opposite subordinate-main clause order.

(2.5) Marta_j scriveva frequentemente a Piera_k quando *pro*/*lei*_k era negli Stati Uniti.

'Marta wrote frequently to Piera when she was in the United States.'

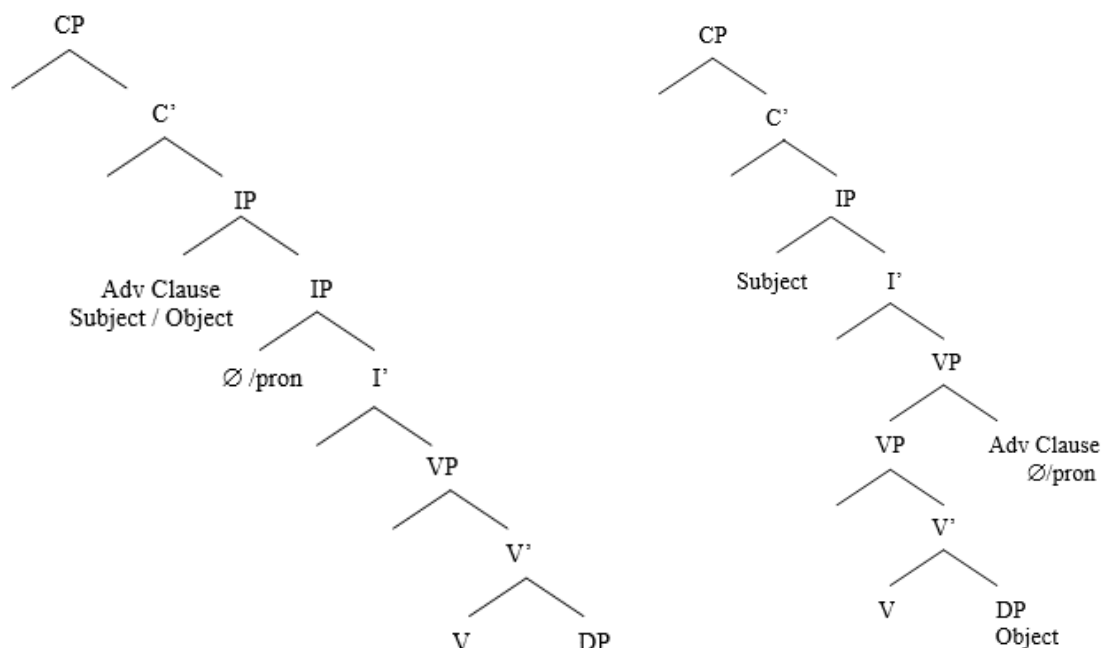
(Example from Carminati, 2002, p. 45)

Recently, de Rocafiguera and Bel (2022) overtly contrasted main-subordinate and subordinate-main pronominal resolution preferences in Spanish. Their study demonstrated that relative clause order crucially affects pronoun interpretation, null pronouns only showing a subject antecedent bias in subordinate-main sequences (see

also Bel & García-Alcaraz, 2015, 2018; Chamorro, 2018; Filiaci, 2011). Based on Carminati (2002), the fact that temporal subordinate clauses in main-subordinate sequences are attached to the VP (Figure 2.1), as well as the object antecedent, could explain that the object of the main clause more easily recovered for the subsequent subject pronoun. This would account for weaker subject biases of null pronouns and reinforced object biases of overt pronouns in these contexts. In subordinate-main clause order, on the other hand, the subordinate (temporal) clause appears in an IP-initial position (Figure 2.1). In this position, subject and object antecedents would be similarly accessible, which would make it the perfect scenario for intrinsic interpretative preferences of null and overt subject pronouns to arise. Beyond structural terms, discourse expectations and memory resources could also play a role in this regard. Whereas an initial subordinate clause indicates that a main clause will follow, an initial main clause does not indicate that more discourse text is coming. In main-subordinate contexts, the potential antecedents would thus not be equally active, and the syntactic parser could rely more on other strategies, such as referring to the most recent antecedent.

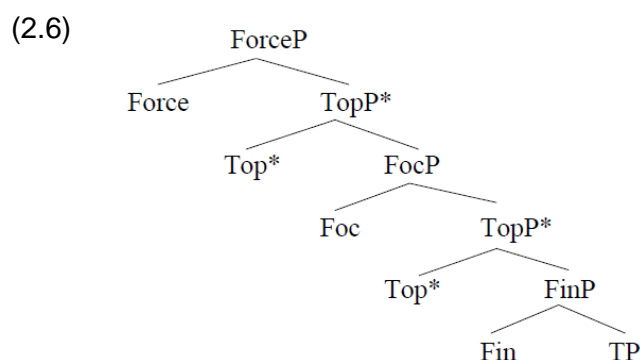
Figure 2.1

Structure of complex clauses in main-subordinate (left) and subordinate-main (right) clause orders



Note. Adapted from Carminati (2002, pp. 39, 46)

Although Carminati (2002) considers the interpretation of pronominal subjects to be essentially guided by syntactic/structural factors, she recognizes that pronouns may also be sensitive to discourse factors such as the information status of the antecedents. Assuming Rizzi's (1997) articulated structure of the CP-domain shown in (2.6) (from Fábregas, 2016, p. 4), Carminati (2002, p. 184) further hypothesizes that referents introduced in positions higher than Spec,IP, such as topic positions (i.e., Spec,TopP), may be similarly preferred for null pronouns as antecedents in Spec,IP. She speculates that topicalized referents would compete with subject antecedents without overriding the preference of null pronouns for antecedents in Spec,IP. Focalized referents (i.e., in Spec,FocP), on the other hand, would not compete with subject referents. She attributes this asymmetry between topicalized and focalized referents to discourse factors. Null pronouns typically retrieve old referents in discourse, associated with topicalized antecedents, whereas focalized referents often introduce new information into the discourse. In this regard, she notes that Spec,IP is the position of the default topic and Spec,TopP the position of the marked topic, both positions often conveying topical information.



Taking into account the different position of subjects and topics, Torregrossa et al. (2020) propose reformulating the PAH in a way that defines the relative prominence of a constituent in terms of hierarchical height. They base their proposal on Rizzi (2018)'s claim that null subjects in Italian are mainly sensitive to "aboutness", a property shared by both subject and topic positions, rather than to "subjecthood". More specifically, Torregrossa et al. (2020, p. 9) propose the principles in (2.7) to account the interpretation of null pronouns.

- (2.7) a. A null subject is expected to have the referent of a prominent DP.
 b. A DP is more prominent than another DP if the former is hierarchically higher than the latter.

- c. Prominence of a DP depends on other factors beyond syntax (e.g., verb-type, coherence relations, discourse topicality, prosody, etc.).

In this proposal, the hierarchic position of an antecedent is granted special importance in terms of making an antecedent prominent, or more preferred, for subsequent null pronouns. As summarized in (2.7b), the preferred antecedent of a null pronoun will be the one appearing in a hierarchically higher position, typically occupied by subjects or topics. In this regard, they add that "the greater the difference between constituents in terms of hierarchical height, the more evident this bias is [the bias of null pronouns toward the hierarchically higher antecedent]" (p. 9). As such, the bias of null pronouns is argued to be more pronounced when the object occurs in situ than when it is, for instance, left-dislocated. When the object occurs in situ there is more distance in terms of hierarchical height between Spec,IP (i.e., the subject) and the object constituent. However, these observations were not empirically tested in their study.

Finally, in (2.7c), Torregrossa et al. (2020) recognize the need of a multifactorial approach to account for the possibility that null subjects corefer with object antecedents, given that pronominal preferences are not categorical. However, they do not have sufficient evidence to define how, or to what extent, these multiple factors interact (p. 22). Note that, in contrast to Carminati (2002), Torregrossa et al. (2020) predict that null pronouns should not similarly prefer subject antecedents in Spec,IP and antecedents in Spec,TopP. Instead, the preference for the hierarchically higher position should prevail (i.e., Spec,TopP).

In the present study, we will empirically test Carminati's intuitions on whether the preference of null pronouns for referents in Spec,IP competes with their preference for referents in higher positions than Spec,IP, such as topic and focus positions (i.e., Spec,TopP and Spec,FocP). In pragmatic terms, we will also consider whether topics are the preference for null pronouns rather than foci. Beyond null pronouns, it is important to highlight that we will investigate the preferences of overt pronouns in these same contexts to examine whether the two pronominal forms are similarly or differently sensitive to different cues, as argued in §2.2.3. In order to do so, we will compare the interpretation of null and overt pronouns in four different contexts: canonical sentences (where the subject antecedent appears in the highest structural position, Spec,IP); in object CLLD structures (where the object antecedent appears in the highest structural position, Spec,TopP); in subject cleft structures (where the subject antecedent appears in the highest position, Spec,FocP); and in object cleft structures (where the object

antecedent appears in the highest position, Spec,FocP). At the same time, this comparative design in which the position of the antecedent is manipulated will also allow us to empirically test Torregrossa et al. (2020)'s proposal. Overall, we will be studying how the information status of the antecedent interacts with its syntactic position.

Given that subordinate clauses lack the topic and focus positions in CP, the only suitable order to test the interpretive preferences of null and overt pronouns in syntactically marked information structures is the main-subordinate clause order. In §2.3.2, we will review previous studies on the PAH in Catalan and in Spanish, and in §2.3.3 we will refer to microvariation among null subject languages. Later, in §2.4, we will further address the role of pragmatic factors, such as information status, in anaphora resolution.

2.3.2 Previous studies on the Position of Antecedent Hypothesis in Catalan and in Spanish

Studies assessing the validity of the PAH in other null subject languages have not always replicated the well-defined biases of Italian pronouns. As further developed in §2.3.3, these findings have recently been interpreted as null subject languages showing microvariation (Contemori & Di Domenico, 2021; Giannakou & Sitaridou, 2020; Torregrossa et al., 2020). Focusing now on the languages under study in the present thesis, Catalan and Spanish seem to differ in the extent to which they comply with the PAH in totally ambiguous contexts³ (Bel & García-Alcaraz, 2018).

Broadly speaking, in Catalan, null and overt pronouns have been found to display clear-cut PAH-like biases toward subject and object antecedents, respectively (Bel & García-Alcaraz, 2018; Mayol & Clark, 2010). In monolingual Spanish, on the other hand, while null pronouns have generally been found to prefer subject antecedents, overt pronouns have been claimed not to display such a clear bias toward object antecedents (e.g., Bel & García-Alcaraz, 2018; de Rocafiguera & Bel, 2022; Filiaci et al., 2014). However, contradicting evidence has been found in this regard, given that other studies in Spanish have attested clear PAH-like biases in both null and overt pronouns (e.g., Contemori & Di Domenico, 2021; de la Fuente, 2015; García-Alcaraz, 2015), or even other patterns. Said previous findings on the PAH in Catalan and in Spanish are further explained below.

³ We consider globally ambiguous contexts to be the ones in which the inherent interpretive preferences of subject pronouns will more easily emerge.

Concerning Catalan, only two studies have previously experimentally examined null and overt pronoun interpretive biases in globally ambiguous contexts within the framework of the PAH. Bel and García-Alcaraz (2018) analyzed intra-sentential contexts using an acceptability judgement task and found Catalan to show well-defined PAH-like biases of both null and overt pronouns in subordinate-main clause order, as illustrated in (2.8). In main-subordinate sequences, null pronouns showed no biases, whereas overt pronouns maintained their clear preference for object antecedents, as shown in (2.9). Importantly, they controlled for the implicit causality of the verbs in the main clause in order to avoid semantic preferences. The other study, by Mayol and Clark (2010), assessed inter-sentential contexts using a two-alternative forced-choice task and a self-paced reading task. Although these are not the specific contexts for which the PAH was formulated, null pronouns were also found to prefer subject antecedents and overt pronouns to prefer object antecedents, as in (2.10) and in line with the PAH, expanding its predictions formulated for intrasentential contexts. Addressing production data, Bel et al. (2010) looked at anaphoric third person subject pronouns in oral and written narratives. In their results they observed that null pronouns mainly expressed topic maintenance and preferred to refer to antecedents in subject position, whereas the scarce number of overt pronouns did not show a clear coreference pattern.

(2.8) *Intra-sentential anaphora in subordinate-main clause order*

Mentre en Cèsar_j desmentia en Joaquim_k, *pro*_j/*ell*_k es va posar vermell.

'While Cèsar refuted Joaquim, he turned red.'

(Example taken from Bel & García-Alcaraz, 2018, p. 46)

(2.9) *Intra-sentential anaphora in main-subordinate clause order*

La Irene_j va saludar la Catalina_k quan *pro*_j/*ella*_k va entrar a la botiga.

'Irene greeted Catalina when she entered the store'

(Example from Bel & García-Alcaraz, 2018, p. 46)

(2.10) *Inter-sentential anaphora*

La Marta_j escrivia sovint a la Raquel_k. *pro*_j/*Ella*_k vivia als Estats Units.

'Marta wrote frequently to Raquel. She lived in the United States.'

(Example from Mayol & Clark, 2010, p. 784)

In consideration of Spanish, several studies assessing intra-sentential anaphora resolution in subordinate-main contexts—where the PAH predictions should more clearly arise—have not found clear PAH-like biases. Null pronouns have been said to clearly prefer to retrieve subject antecedents, but overt pronouns do not show well-defined

preferences (in acceptability judgements tasks: Bel & García-Alcaraz, 2018; de Rocafiguera & Bel, 2022; in self-paced reading tasks: Filiaci, 2011; Filiaci et al., 2014). These patterns are illustrated in (2.11a). However, these findings are not uncontroversial. Other studies have found the predictions of the PAH to hold for Spanish subordinate-main contexts (Bel & García-Alcaraz, 2015 in an acceptability judgements task; García-Alcaraz, 2015 in a forced-choice task; Keating et al., 2016 in a self-paced reading task), as shown in (2.11b).

(2.11) *Intra-sentential anaphora in subordinate-main clause order*

a. Cuando Tomás_j vio a Alberto_k, *pro*_j/*él*_{j/k} estaba nervioso.

'When Tomás saw Alberto, he was nervous.'

(Example from de Rocafiguera and Bel, 2022, p. 12)

b. Cuando Juana_j recogió a Anaïs_k, *pro*_j/*ella*_k se empezó a encontrar mal.

'When Juana picked up Anaïs, she started to feel sick.'

(Example from García-Alcaraz, 2015, p. 147)

Mixed results have also been attested in main-subordinate clause order. Some studies have found null pronouns to show flexible interpretations and overt pronouns to show a clear object bias (in acceptability judgement tasks: Bel & García-Alcaraz, 2015, 2018; Chamorro, 2018; Chamorro et al., 2016; de Rocafiguera & Bel, 2022; in a self-paced reading task: Filiaci, 2011; in an offline sentence interpretation task and an online eye-tracking task: Schimke et al., 2018)⁴, as shown in (2.12a). In other studies, the opposite pattern has been attested: null pronouns have been found to display a clear bias toward subject antecedents and overt pronouns to show more flexible interpretations (Clements & Domínguez, 2017 in a picture verification task; Jegerski et al., 2011; Keating et al., 2011 in two-alternative forced-choice tasks), as in (1.12b). Another pattern that has been attested is the one predicted by the PAH: null pronouns clearly preferring subject antecedents, and overt pronouns clearly preferring object antecedents (Bel et al., 2016; Contemori & Di Domenico, 2021; de la Fuente, 2015; García-Alcaraz, 2015 in sentence interpretation/forced-choice tasks), as in (2.12c). Finally, Giannakou and Sitaridou (2020), in an offline self-paced listening task with open comprehension questions, found both null and overt pronouns to have flexible coreference patterns, as shown in (2.12d). This lack of pronominal biases—or flexibility in interpretations—in Spanish was also attested by Leonetti-Escandell and Torregrossa (under review) in an interpretation task

⁴ In Schimke et al.'s (2018) online eye-tracking task, null pronouns were even found to significantly prefer coreference with object antecedents, in main-subordinate clause order stimuli.

using a 5-point Likert-scale (in sentences such as *El doctor pagó al arquitecto mientras pro/él cerraba la cartera*).

(2.12) *Intra-sentential anaphora in main-subordinate clause order*

a. Ana_j esperó a Olga_k cuando *pro*_{j/k}/ella_k llegó de viaje.

'Ana waited for Olga when she came back from a trip.'

(Example from de Rocafiguera & Bel, 2022, p. 12)

b. La mujer_j empuja a la niña_k mientras *pro*_{j/k}/ella_{j/k} se come un helado.

'The woman pushes the girl on the swing while she eats an ice-cream.'

(Example from Clements & Domínguez, 2017, p. 43)

c. Jorge_j vio a Luis_k cuando *pro*_{j/k}/él_k iba a la cafetería.

'Jorge saw Luis when he was going to the coffee shop.'

(Example from Contemori & Di Domenico, 2021, p. 1005)

d. La abuela_j besaba a la enfermera_k cuando (ya) *pro*_{j/k}/ella_{j/k} se ponía el abrigo⁵

'The old lady was kissing the nurse when ([adverb]) she was putting on her coat.'

(Example from Giannakou and Sitaridou, p. 26)

In inter-sentential contexts, Alonso-Ovalle et al. (2002), in an offline sentence interpretation task, found null pronouns to display a clear bias toward subject antecedents but overt pronouns to be flexible in their interpretations, as in (2.13). Conversely, Gelormini-Lezama and Almor (2011), in an online self-paced reading task, found the PAH to apply in these contexts.

(2.13) *Inter-sentential anaphora*

Juan_j pegó a Pedro_k. *pro*_{j/k}/Él_{j/k} está enfadado.

'Juan hit Pedro. He is angry.'

(Example from Alonso-Ovalle et al., 2010, p. 3)

No comprehensive explanation has been proposed yet to account for these mixed findings attested across contexts and across studies. De Rocafiguera and Bel (2022) confirmed that clause order has an impact on anaphora resolution and suggested that the type of task or experimental methodology could also be influencing anaphora resolution, although mixed evidence has also been found in this regard. As noted by Contemori and Di Domenico (2021), the type of variety of Spanish spoken by participants

⁵ The adverb *ya* only appeared in the null pronoun condition.

cannot explain mixed evidence either (see also Comínguez et al., 2017). As these authors suggest, it could be possible that pronouns in Spanish have less robust biases than in other null subject languages and that this lack of strength contributes to the inconsistency of the attested results. Other factors that could be influencing subject pronouns' biases and that have not been controlled for in all the studies are those affecting the degree of ambiguity of the stimuli, such as implicit causality of the verbs (e.g., Järvikivi et al., 2017), agentivity (Schumacher et al., 2017), tense and aspect, or the connector used to introduce the subordinate temporal clause (Martín-Villena et al., 2021). To date, relatively few studies have addressed the role of these factors in anaphora resolution in null subject languages.

Overall, more consistent PAH-like preferences have been found in Catalan than in Spanish. In the sole study that has overtly contrasted pronoun resolution preferences in Catalan and in Spanish, Bel and García-Alcaraz (2018) found overt pronouns in Catalan to significantly display more polarized resolution patterns than in monolingual Spanish. Interestingly, Bel and García-Alcaraz (2018) also found bilingual Spanish to differ from monolingual Spanish, suggesting that the bilingual variety of Spanish in contact with Catalan displays well-defined PAH-like biases like Catalan. We will return to these findings in the next chapter, when addressing the role of bilingualism on anaphora resolution (see §3.4.3).

In the present study, we will provide further evidence on anaphora resolution biases in main-subordinate temporal sequences in Catalan and Spanish and we will compare the two languages to see whether the findings in Bel and García-Alcaraz (2018) can be replicated and whether the two languages show microvariation. In addition to assessing unmarked canonical sentences, which have already been studied to some degree (although with unclear results), we will manipulate the information structure of the sentences. In doing so, we will disentangle the notion of subject from the notions of topic and first-mention to look at the role that other non-syntactic factors play in anaphora resolution. Moreover, this will enable us to test the role of the syntactic position in terms of hierarchical height. To date, we are not aware of any study that has experimentally tested this. In §2.4, we will center on pragmatic factors that may be influencing anaphora resolution in Catalan and Spanish. Before doing so, however, in §2.3.3 we refer in more detail to studies that have attested microvariation between null subject languages. Of special interest is how these studies account for crosslinguistic differences; explanations which have been mainly related to syntactic factors.

2.3.3 Microvariation in null subject languages

Recent comparative studies have demonstrated that variation exists in the resolution preferences of subject pronouns among (Romance) null subject languages. As previously mentioned, Bel and García-Alcaraz (2018) compared Catalan and Spanish using an acceptability judgements task and found Catalan to show more well-defined and polarized PAH-like biases than monolingual Spanish, both in intra-sentential main-subordinate and subordinate-main clause orders. Whereas in Catalan null pronouns are clearly biased toward subject antecedents and overt pronouns toward object antecedents, overt pronouns in monolingual Spanish show significantly more mitigated biases.

This difference between Spanish and Catalan may be similar to the one attested by Filiaci et al. (2014) and Contemori and Di Domenico (2021) between Spanish and Italian, by Giannakou and Sitaridou (2020) between Spanish and Greek⁶, and by Leonetti-Escandell and Torregrossa (under review) between Spanish, Italian, and Greek. In all these studies, Spanish was found to follow the predictions of the PAH to a lesser degree than Italian and Greek⁷. In Filiaci et al. (2014) and in Giannakou and Sitaridou (2020), overt pronouns differed between Spanish and Italian/Greek. In contrast to the strong bias toward object antecedents in Italian and Greek, overt pronouns in Spanish did not show any interpretive preferences. These results echo those found in Bel and García-Alcaraz (2018). In Contemori and Di Domenico (2021), on the other hand, both null and overt pronouns showed significant PAH-like biases in the two languages. In this case, the interpretive biases of both pronouns in Spanish were significantly more tenuous than in Italian. In Leonetti-Escandell and Torregrossa (under review), null and overt pronouns in Spanish showed unbiased interpretations, not different from chance, and significantly different from those of Greek and of Italian. Overall, there seems to be enough evidence to claim that subject pronouns in Spanish show more flexible interpretations than in other null subject languages such as Catalan, Italian, and Greek, especially regarding overt pronouns. Whether null pronouns are also sensitive to crosslinguistic differences remains less clear, but recent evidence points in this direction. Another null subject language in which a well-defined PAH-like behavior has been attested in Basque (Iraola Azpiroz et al., 2017). However, no studies have addressed the comparison between Basque and Spanish.

⁶ Recall that Greek is not a Romance language.

⁷ Note that null and overt pronouns did not show significant interpretive biases in Leonetti-Escandell and Torregrossa (under review).

Torregrossa et al. (2020) is another study on microvariation that should be referred to, although it does not involve comparisons with Spanish. This paper compared Greek and Italian coreference patterns. Italian was found to comply with the PAH to a greater extent than Greek, which showed more flexible coreference patterns for both null and overt pronouns. Madeira et al. (2021) also observed microvariation between European Portuguese and Italian. Differences regarding the strength of the pronominal biases based on the syntactic position of the antecedent were only attested for null pronouns, European Portuguese showing more pronounced biases toward subject antecedents than Italian in main-subordinate sequences. These two studies posit that null subject languages may exhibit a differential sensitivity to different factors influencing anaphora resolution. Torregrossa et al. (2020, pp. 3, 22) suggest that Greek may be more sensitive to morphological cues (due to the morphological complexity of its nominal paradigm), and Italian to the syntactic position of the antecedent, as explained in the following paragraph. Madeira et al. (2021) argue that the syntactic position of the antecedent could be a more relevant factor in European Portuguese than in Italian.

Several explanations have been proposed to account for these microparametric crosslinguistic differences leading to subtle differences among null subject languages. To explain differences in the interpretation of overt subject pronouns in Spanish and in Italian, Filiaci (2011) and Filiaci et al. (2014, p. 219) referred to Cardinaletti and Starke (1999)'s cross-linguistic typology of deficient forms. They argued that, in contrast to the Italian overt pronouns *lui/lei*, the Spanish overt pronouns *él/ella* are weak elements or structurally deficient, similar to *egli/ella* in Italian (Cardinaletti & Starke, 1999). This explanation could similarly account for the wider scope of overt pronouns in Spanish compared to Greek, given that the overt pronoun in Greek (*aftos*) is a strong pronominal form, as the Italian pronoun (*lui/le*) (Giannakou & Sitaridou, 2020). However, it would not account for differences in the interpretation of null pronouns, neither for differences present between Greek and Italian.

More recently, Torregrossa et al. (2020) and Leonetti-Escandell and Torregrossa (under review) have explained microvariation in pronominal subject interpretation in null subject languages as referring to the availability of VSO word order and of differential object marking (DOM). Torregrossa et al. (2020) attributed microvariation across Greek and Italian to crosslinguistic differences in word order. Contrary to Greek, VSO order is not available in Italian in broad focus sentences (Roussou & Tsimpli, 2006). According to these authors, in Greek VSO sentences, the subject and the object constituents can

appear in the same clausal domain because of the morphological complexity of the nominal paradigm in Greek, so that the distance in terms of hierarchical height between the subject and the object is reduced. The fact that Greek allows for VSO sentences could make the syntactic position of antecedents a less reliable cue in this language compared to Italian, where subject and object positions always show pronounced differences in terms of hierarchical height (see Torregrossa et al., 2020, pp. 10–11, 21–22). If this is the case in Catalan, which does not allow for VSO order either, Italian-like patterns of coreference would be expected in Catalan.

Leonetti-Escandell and Torregrossa (under review) expanded the findings in Torregrossa et al. (2020) by comparing null and overt pronoun resolution in Greek and Italian to Spanish. As previously stated, they found Spanish pronouns to show non-significant interpretive biases, meaning that both null and overt pronouns were found to indistinctly prefer subject and object antecedents. Their interpretive preferences were weaker than those of Greek, and even weaker than those of Italian. In light of these results, the syntactic position of the antecedent does not seem to be as reliable a cue for pronoun resolution in Spanish as in other languages. Similarly to Greek and in contrast to Italian, VSO order is available in broad focus sentences in Spanish. However, Spanish does not only differ from Italian, but it also shows weaker structural biases than Greek. Leonetti-Escandell and Torregrossa (under review) relate the differences between Spanish and Greek to differential object-marking (DOM), which is used in Spanish but not in Greek. Given that Spanish uses DOM to mark object constituents, they argue that hierarchical height may not be as a reliable cue for Spanish speakers when interpreting null and overt subject pronouns.

It should be noted that the lack of biases of Spanish pronouns in Leonetti-Escandell and Torregrossa (under review) has not been generally attested in the previous literature on pronoun resolution in Spanish (see the studies reviewed in previous §2.3.2), although it is true that PAH-like biases do not emerge as clearly in main-subordinate contexts such as the ones used in this study (de Rocafiguera & Bel, 2022). If the syntactic position of the antecedent was not used as a reliable cue to solve anaphoric relations, however, interpretive biases in Spanish should be similarly weak across the studies in question. In any case, it is possible that Spanish is less sensitive than other null subject languages to syntactic-structural factors such as the syntactic position of the antecedent.

In the present study, of particular interest is comparing pronoun resolution in Catalan and in Spanish to see whether Bel and García-Alcaraz's (2018) results showing

microvariation between these two languages can be replicated. Similarly to Italian and in contrast to Spanish, Catalan does not allow for VSO word order in broad focus sentences (see Colomina, 2019; Ordóñez, 1998; Picallo, 1998; Solà Pujols, 1992). This can be seen in the examples in (2.14) below from Picallo (1998, p. 228).

- (2.14) a. Hojeaba Juan el periódico [Spanish]
b. (??)*Fullejava en Joan el diari [Catalan]
'Joan was browsing the newspaper'

Furthermore, the distribution of DOM also differs between Catalan and Spanish. In standard Catalan, the uses of DOM are much more restricted than they are in Spanish, only being obligatory in terms of marking personal pronouns (IEC, 2016). As a language that does not license VSO word order and lacks DOM, Catalan could rely more on syntactic-structural cues in anaphora resolution and show more pronounced PAH-like biases than Spanish. However, recent experimental studies on the acceptability and use of DOM in Catalan have revealed it to be used and accepted in a wider range of contexts (e.g., Benito & Bel, 2022). In any case, the distribution of DOM is still wider in Spanish.

In summary, in light of the predictions that can be derived from Leonetti-Escandell and Torregrossa's (under review) proposal, together with Bel and García-Alcaraz's (2018) findings of crosslinguistic differences between Catalan and Spanish, microvariation between Catalan and Spanish should also be attested in the present study. It will therefore be our aim to test whether further evidence can be provided on microvariation between Romance null subject languages.

2.4 Pragmatic factors in anaphora resolution: the information status of the antecedent

2.4.1 Information structure in the Romance languages

Beyond syntactic factors, pronominal subject anaphora resolution has been shown to be affected by pragmatic factors, such as information structure and the information status (or discourse status) of referents. The term *information structure* (see Cruschina, 2016; Domínguez, 2018 for recent overviews) was first used by Halliday (1967) and has generated a large body of literature from various theoretical backgrounds (see Chafe, 1976; Erteschik-Shir, 2007; Fábregas, 2016; Frascarelli, 2007; Krifka, 2008; Lambrecht, 1994; López, 2009; Prince, 1981; Reinhart, 1982; Rizzi, 1997; Vallduví, 1992; among

others). It can be defined as "how new and old information is packed in a sentence" (Domínguez, 2018, p. 372), or as "the way in which linguistic expressions interact with discourse functions, so that the information conveyed within the sentence is packaged in accordance with given discourse contexts and with the mental states of the interlocutors" (Cruschina, 2016, p. 596). In the present study, we are interested in investigating how the interpretation of third-person subject pronouns interacts with the information status of their plausible antecedents.

The two principal notions that identify the information status of a constituent are those of *topic* and *focus*. These concepts have been approached in very different ways in the literature and there is no broad consensus in the terminology used or in their definitions (see Cruschina, 2016, 2021a, 2022; Domínguez, 2018; Fábregas, 2016; Gutiérrez-Bravo, 2008; Olarrea, 2012; Rizzi & Bocci, 2017 for overviews). Broadly speaking, *topic* refers to given or presupposed information in a sentence, or the information that is already shared by the speaker and the addressee, whereas *focus* refers to new or non-presupposed information. With respect to a particular discourse context, the notion of topic has also been understood in topic-comment articulation (e.g., Reinhart, 1982). The topic establishes what the sentence is about, and the comment makes a statement about the topic, as in (2.15). There are topics or comments that may contain a focus, as in (2.16). Concerning focus, it has also been claimed to be understood in relation with a focus-presupposition (or focus-background) sentence partition, as in (2.17).

(2.15) [Jacqueline Kennedy]_{Topic} [married Aristotle Onassis]_{Comment}

(2.16) A: When did [Aristotle Onassis]_{Topic} marry Jackeline Kennedy?

B: [He]_{Topic} [married her [in 1968]_{Focus}]_{Comment}

(Examples from Krifka, 2008, p. 265-266)

(2.17) Q: Who stole the picture?

a. [_{Focus} The thief] [_{Presupposition} stole it] (English)

b. [_{Presupposition} Lo robó] [_{Focus} el ladrón] (Spanish)

'The thief stole it.'

(Example from Domínguez, 2018, p. 372)

Regarding the types of topics and foci, two types of topics have generally been identified according to the aforementioned studies: 1) aboutness topics, which typically identify what a sentence is about, and 2) given topics, which convey information that has already

appeared in the previous discourse⁸. As for foci, two main types of focus can be distinguished: 1) broad focus, when a whole sentence receives a focal interpretation, and 2) narrow focus, when only one constituent in a sentence expresses focus. A third type of focus, 3) predicate focus, refers to the predicate in topic-comment sentence partitions. Different subtypes or interpretations of narrow focus can also be distinguished: 1) contrastive or corrective focus (also identificational focus), when explicit contrast is established between the focused constituent and a closed set of alternatives contextually determined; 2) informative focus (also presentational focus), when no specific alternatives are mentioned in the context; and 3) mirative focus, when it expresses unexpectedness or surprise with respect to more likely alternatives.

There is no direct or one-to-one mapping of the topic and focus semantic-pragmatic categories onto grammatical encoding. Information structure is mainly expressed through syntax, prosody, or through a combination of these means (see Vallduví, 1992, 2002 for Catalan; Zubizarreta, 1998, 1999 for Spanish). As in other Romance languages, in Catalan and in Spanish unmarked SVO word orders⁹ the canonical subject is often understood as conveying topical or presupposed information, or broad focus. The expression of narrow focus is often associated with the rightmost stressed constituent of a clause. The alteration of canonical word orders, so that the focused constituent is aligned with the main stress of the clause, is one of the main mechanisms, at the syntax-prosody interface, to mark the information structure of a sentence. Each language shows its own preferences in the expression of information structure, and variation in the preferred mechanisms is found among Romance languages and among language varieties (e.g., Feldhausen & Vanrell, 2014; Leonetti, 2017; Vanrell & Fernández-Soriano, 2018, 2013).

In the following subsections, we will review the main syntactic mechanisms available in Catalan and Spanish to topicalize and focalize constituents. We will also outline the main syntactic and pragmatic features of the four constructions that we will scrutinize in the present thesis: unmarked canonical sentences, object topicalization via clitic-left dislocation structures, and subject focalization via it-cleft structures, and object focalization via it-cleft structures. Using syntactically marked information structures will

⁸ In this thesis, we will work with *sentence topics* (like de la Fuente, 2015; Colonna et al., 2012), and not with *discourse topics* (see van Dijk, 1977).

⁹ There is an ongoing debate surrounding the unmarked word order in Romance languages. SVO word order is considered the neutral unmarked order in Spanish (see Zubizarreta, 1998; Domínguez, 2013, Chapter 2) and in Catalan (see Vallduví, 2002), although postverbal subjects constitute the neutral unmarked word order under certain circumstances (see Olarrea, 2012, p. 608).

allow us to ensure that subject and object antecedents occupy a topic or a focus position and are read and interpreted as conveying topical or focal information as intended.

2.4.1.1 Unmarked structures: canonical sentences

The relationship between topics and subjects in pragmatically unmarked sentences has long been a controversial issue. On the one hand, it has been generally considered that decontextualized sentences have no topic or a broad focus interpretation (also referred to as all-focus orthetic sentences, e.g., Olarrea, 2012). On the other hand, preverbal subjects have also been assumed to have a default topical status in null subject languages. Sentence topics usually correlate with preverbal subjects or sentence-initial positions, and declarative sentences tend to follow the topic-focus order (e.g., Casielles-Suárez, 2004). Moreover, the 'aboutness' feature of topics (often defined as identifying what a sentence is about), makes the subject position akin to the topic position (Chafe, 1976; Reinhart, 1982; see Leonetti, 2021; Rizzi, 2018 for recent discussions). Thus, there is a tendency to associate preverbal subjects with topics, which may have influenced pronominal coreference patterns in out-of-the-blue experimental sentences in studies within the framework of the Position of Antecedent Hypothesis. Even if decontextualized sentences can be read as having no topic, they also admit a reading that identifies a preverbal subject as conveying topical information.

Regarding foci in canonical sentences, in Romance languages it can be assumed that in unmarked SVO word orders with neutral intonation, the focus position is assigned to the rightmost constituent through default stress assignment (e.g., Cruschina, 2022). Thus, unmarked canonical sentences can be associated with different focus structures, such as broad focus in (2.18a), predicate focus in (2.18b), or narrow focus in (2.18c). In the absence of a discourse or pragmatic context, all these interpretations of focus are possible.

- (2.18) a. [En Quico va trencar el got]_{Focus}
b. En Quico [va trencar el got]_{Focus}
c. En Quico va trencar [el got]_{Focus}
'Quico broke the vase'

(Example adapted from Cruschina, 2022)

2.4.1.2 Topicalization constructions: CLLD

Regarding the mechanisms that mark topic expression, we can distinguish between three topicalization constructions, as shown in (2.19): hanging topic left dislocation (HTLD), clitic left dislocation (CLLD), and clitic right dislocation (CLRD) (in terms of Vallduví, 1992, left and right detachment).

(2.19) a. Les puputs, m'encanten aquests ocells. (HTLD)

'Hoopoes, I love those birds.'

b. La puput la vaig veure ahir. (CLLD)

'The hoopoe, I saw it yesterday.'

c. La vaig veure ahir, la puput. (CLRD)

'I saw it yesterday, the hoopoe.'

In these constructions, the topical constituent appears in a peripheral sentence position (sentence-initial in left dislocations, or sentence-final in right dislocations) and, in Catalan and Spanish, it is obligatorily retaken by a resumptive clitic in CLLD and CLRD. These structures allow for marking the dislocated element as a topic, but they can also be used to mark the non-dislocated material as a focus, given that it appears in a position where it can receive stress. This is shown in the following examples, adapted from Domínguez (2018, p. 383-384; example (2.20) was adapted from Vallduví, 1992).

(2.20) a. [F Ficarem el ganivet al calaix]

'We will put the knife in the drawer.'

b. Hi ficarem [F el ganivet], al calaix

'We will put the knife, in the drawer.'

(2.21) Q: Who paid the bill?

a. #_[F] Susana pagó la cuenta (non-felicitous)

'Susana paid the bill.'

b. La cuenta, la pagó _[F] Susana (felicitous)

'Susana, she paid the bill.'

The most evident example of variation between Catalan and Spanish is that CLRD is much more common in the former than in the latter, and Catalan also seems to make greater use of dislocations than Spanish (Feldhausen & Vanrell, 2014; Vallduví, 1992; Villalba, 2011).

As mentioned, in the present study we will use CLLD structures to mark an object antecedent with a topical information status and to make it appear first-mentioned. At the same time, the subject antecedent in these constructions appears second-mentioned and in a postverbal position and has a focal (or non-topical) information status. An example is the following (2.22).

- (2.22) a. A la Sabina la va interrompre l'Olívia quan va començar a parlar. (Catalan)
b. A Sabina la interrumpió Olivia cuando empezó a hablar. (Spanish)
'Sabina, Olivia interrupted her when she started speaking.'

2.4.1.3 Focalization constructions: it-clefts

Regarding the expression of focus, it largely depends on the type of focus, and it may also vary depending on the focused constituent. The most common strategies to mark narrow focus in Romance languages include focus fronting (2.23), postverbal focalization (2.24), and cleft constructions (2.25) (Cruschina, 2021b, 2022; see also Leal et al., 2018). Both focus fronting and postverbal focalization may coincide with prosodically marked focus in situ, as in (2.23a) and (2.24b), when focused constituents remain in their unmarked surface position. Focus fronting implies syntactic displacement of the focused constituent to the left periphery, as clearly observed in (2.23b), and postverbal focalization typically involves movement to a postverbal position within the core of the sentence (i.e., ignoring right-dislocated material, if any), as in (2.24a).

- (2.23) a. LA MARIA va veure l'òliba. (Focus Fronting or focus in situ)
'Maria saw the barn owl.'
b. L'ÒLIBA va veure la Maria. (Focus Fronting)
'The barn owl, Maria saw.'

- (2.24) a. La va veure LA MARIA. (Postverbal focalization)
'Maria saw it.'
b. La Maria va veure L'ÒLIBA. (Postverbal focalization or focus in situ)
'Maria saw the barn owl.'

- (2.25) a. Va ser LA MARIA qui va veure l'òliba. (Cleft structure)
'It was Maria who saw the barn owl.'
b. Era L'ÒLIBA el que va veure la Maria. (Cleft structure)
'It was the barn owl that Maria saw.'

Cleft constructions are composed of two clauses: a matrix copular clause, which includes a copula and the clefted constituent, and a subordinate relative or *relative-like* clause (De Cesare, 2017; Kiss, 1998; Lambrecht, 2001). From a pragmatic perspective, clefts typically follow a focus-presupposition information structure articulation; the clefted constituent expresses focus and the information in the subordinate clause is presupposed. Although clefted constituents have generally been said to mainly express contrastive or corrective focus, Feldhausen and Vanrell (2014) also identified clefting as a frequent strategy in production data to express informative focus in Catalan (only for subject antecedents) and Spanish (for both subject and object antecedents).

In the present study we will use it-cleft constructions to mark subject and object antecedents as conveying focal information in a sentence-initial position. The second-mentioned antecedent, in the relative clause, conveys presupposed (topic-like) information. The following sentences show an example of subject focalization (2.26) and object focalization (2.27) through it-clefts¹⁰.

- (2.26) a. Va ser l'Olívia qui va interrompre la Sabina quan *pro/ella* va començar a parlar.
b. Fue Olivia quien interrumpió a Sabina cuando *pro/ella* empezó a hablar.
'It was Olivia who Sabina interrupted when she started speaking.'

- (2.27) a. Va ser a la Sabina a qui va interrompre l'Olívia quan *pro/ella* va començar a parlar.
b. Fue a Sabina a quien interrumpió Olivia cuando *pro/ella* empezó a hablar.
'It was Sabina whom Olivia interrupted when she started speaking.'

2.4.1.4 On the notions of topic and focus in the present study

Given that we want to look at the intrinsic interpretive biases of null and overt pronouns, we consider targeting ambiguous pronouns to be the best way to examine how syntactic, pragmatic, and word order factors define the properties of pronominal subjects from an empirical perspective (see §4.3.2.2 in Chapter 4). The pronouns will appear in decontextualized experimental sentences to keep them as semantically ambiguous as possible. This is because we want to avoid adding any contextual material that could introduce other cues for pronoun interpretation that we do not control for that could intervene in our results. However, we are aware that information structure and

¹⁰ Given that it is the most common term in the literature, I will use the anglocentric label *it-cleft structure*. However, to shorten the labels "subject focalization via it-cleft structure" and "object focalization via it-cleft structure" I will also use the terms *subject cleft* and *object cleft*.

information status are highly dependent on the discourse context. Unavoidably, prescind of the discourse context may cause us to be unsure of whether participants interpret a decontextualized canonical clause as conveying broad focus or as having a topic-comment configuration. We are aware that equating the notions of subject and topic, and those of object and focus in out-of-the-blue canonical sentences may not be entirely accurate. Nor we will be sure of whether a cleft construction is expressing informative or contrastive focus, but the nuances that different types of focus may introduce in pronoun resolution is not within the aims of the present study.

In the interest of presentational clarity, and consistent with previous studies on the role of pragmatic factors in anaphora resolution, when discussing the information status of the antecedents in the present study we will use the topic-focus dichotomy (see also de la Fuente, 2015 for a similar choice). On the one hand, we will consider that subject antecedents convey topical information and object antecedents convey focal information in unmarked sentences and in object clefts. On the other hand, we will consider that subject antecedents convey focal information and object antecedents convey topical information in object CLLDs and subject clefts. As observed by Olarrea (2012, p. 607), the clear parallelism between topic-comment and focus-presupposition dichotomies makes one think whether these partitions are equivalent to the topic-focus dichotomy, "without the need of further primitives" (Erteschik-Shir, 2007, p. 27; see also Casielles-Suárez, 2004).

We are aware that it is not straightforward to consider that a constituent in the background/presupposition position of a cleft sentence is a topic (although it conveys non-focal information). Also, not all authors necessarily agree that CLLD structures can be considered to mark the non-topicalized and postverbal constituent as a focus (although it conveys non-topical information). In any case, the four constructions under analysis—pragmatically unmarked canonical sentences, topicalization of the object via CLLD, focalization of the subject via it-cleft, and focalization of the object via it-cleft—allow us for effectively and unambiguously tease apart the notions of topic (and focus), subject (and object), and first-mention (and second-mention).

2.4.2 The impact of information structure on anaphora resolution: evidence from previous experimental studies

The information status of an antecedent has been used to define its "prominence" in previous literature on anaphora resolution. Both topicalized and focused entities, which often appear fronted in an initial position, have been intuitively highlighted as being more prominent and, consequently, more preferred for subject pronouns. In terms of attention allocation and working memory, psycholinguistic research has identified linguistic focus to increase attention, boosting the availability of focused entities in memory (Foraker & McElree, 2007; Káldi & Babarczy, 2021). However, regarding anaphora resolution, the fact that this more available focused antecedent is ultimately chosen as the preferred antecedent to interpret a subject pronoun is not as straightforward (see Blything et al., 2021; Patterson & Felser, 2020).

Regarding topicality, most studies are in agreement that antecedents in topic position tend to be good candidates for subject pronouns in reference assignment processes (e.g., Colonna et al., 2012; de la Fuente, 2015; Kaiser, 2011). However, the preference for topic antecedents does not seem to override the preference for subjects, only compete with it. Regarding focalization constructions, and more specially it-cleft structures, mixed findings have been obtained. Whereas some studies have found these constructions to enhance coreference with clefted antecedents (e.g., Blything et al., 2021 for object clefts; Colonna et al., 2015 for intersentential contexts; Cowles et al., 2007 for subject clefts), other studies have not found clear effects of it-clefts (e.g., Blything et al., 2021 for subject clefts; Colonna et al., 2012, 2015 for intrasentential contexts; Järvikivi et al., 2014; Kaiser, 2011), or to even decrease the accessibility of clefted antecedents (e.g., de la Fuente, 2015; Patterson et al., 2017; Patterson & Felser, 2020). Similarly to topicalization structures, clefting an antecedent neither seems to override the baseline preference of subject pronouns for subject antecedents. Overall, as described in §2.2.2, information structure is one of the factors that, along with syntax, semantics, or linear position, shapes anaphora resolution. Few studies have contrasted the effects of information structure on different pronominal forms (see §2.2.3), but studies such as Kaiser and Trueswell (2008) on Finnish or Bader and Portele (2019) on German found that only demonstrative pronouns were sensitive to topicality (or sequential position of the antecedents).

The first studies on information structure and pronoun resolution were concerned with English subject pronouns. Cowles et al. (2007) found both topical and focused referents

(discourse topics, sentence topics, and clefted foci) to be more preferred for subject pronouns. However, both topical and clefted antecedents in Cowles et al.'s study were in subject position. Topichood and subjecthood also overlapped in Arnold (1998), who found topical subjects to be more preferred than focal objects for subject pronouns. Building on these studies, Kaiser (2011) confronted the positions of subject and topic of the plausible antecedents for a subject pronoun. Both in online and offline data, she found that the preference for subject antecedents was strengthened when the subject antecedent was also the topic and weakened when the topic was the object antecedent. Importantly, in the topical object condition (2.28), where subjecthood and topicality were misaligned, these two cues seemed to be similarly weighted during online comprehension, although subjecthood had a stronger effect on the offline final interpretive choices.

- (2.28) a. *Mike* did very well in last month's tennis tournament.
 b. *John* congratulated *him* enthusiastically yesterday. (Critical sentence)
 c. The prizes for the best-ranked tennis players were about to be announced, and
 d. *he* was holding a new yellow tennis racket. (Test sentence)
 e. Everyone was in a good mood that day.

(Examples from Kaiser, 2011, p. 1635)

In a second experiment, Kaiser (2011) analyzed the impact of contrastive focus in contexts such as the ones in (2.29) and found an overarching preference for subject antecedents in all four experimental conditions (2.29b). She also found that the proportion of object choices was slightly higher in canonical SVO sentences with focal objects than when the focused object was clefted (OVS).

- (2.29) Speaker A: I heard that Greg congratulated Mike enthusiastically yesterday.
 Speaker B:
 a. No, that's not quite right.
 b. (i) He congratulated John. (Critical sentence: [SVO.Object_focus])
 (ii) John congratulated him. [SVO.Subject_focus]
 (iii) It was John that he congratulated. [Cleft.Object_focus]
 (iv) It was John who congratulated him. [Cleft.Subject_focus]
 c. The prizes for the best-ranked tennis players were about to be announced, and
 d. he was holding a new yellow tennis racket. (Test sentence)
 e. Everyone was in a good mood that day.

(Examples from Kaiser, 2011, p. 1648)

Overall, the findings in Kaiser (2011) suggest that subjecthood (or the syntactic function of the antecedent) is a more heavily weighted cue in anaphora resolution than information structure (and word order) (see also Bader & Portele, 2019; Kaiser & Trueswell, 2008). However, topicality was also found to play a role in increasing the preference for topics (although not in contrastive focus contexts).

Similar findings regarding the "prominence" increase of topicalized antecedents were attested by Colonna et al. (2012) in an offline forced-choice task for German pronouns. Their baseline subject preference was maintained in the subject topicalization condition (via HTLD), not enhancing the already preferred option. The preference for subjects, however, significantly decreased when the object was topicalized (2.30). In this context, in contrast to Kaiser (2011), neither subjecthood nor topicality defined the final interpretation of the pronoun. These two factors seemed to be similarly weighted. The experimental sentences in Colonna et al. (2012), however, appeared with no context and used a topicalization construction that involved an alteration of the baseline SVO word order, the topicalized object being left-dislocated and first-mentioned.

(2.30) a. Peter hat Hans geohrfeigt, als er jung war. (Baseline)

'Peter slapped John when he was young.'

b. Was Peter betrifft, er hat Hans geohrfeigt, als er. (Topicalized subject)

'As for Peter, he slapped John when he was young.'

c. Was Peter betrifft, Hans hat ihn geohrfeigt, als er. (Topicalized object)

'As for Peter, John slapped him when he was young.'

(Examples from Colonna et al., 2012, p. 1001)

Colonna et al. (2012) also analyzed it-cleft structures, which were further investigated in Colonna et al. (2015). In the latter study, the authors compared the effects of focalization on anaphora resolution in intrasentential (2.31) and intersentential (2.32) contexts (in two different forced-choice offline tasks). In intersentential conditions, clefting both the subject and the object antecedents made these clefted referents more "accessible" for the subsequent pronoun, although they do not clarify whether these pronouns kept their subject preference despite the increase of object choices. In intrasentential contexts, the preference for a clefted antecedent was not found to be enhanced compared to the baseline condition, and personal subject pronouns maintained their subject preference across contexts (similar to Kaiser, 2011). These offline results were replicated in an eye-tracking task.

- (2.31) a. *Peter hat Hans geohrfeigt, als er jung war.* (Baseline)
 'Peter slapped John when he was young.'
- b. *Es ist Peter, der Hans geohrfeigt hat, als er.* (Subject cleft)
 'It is Peter who slapped John when he was young.'
- c. *Es ist Peter, den Hans geohrfeigt hat, als er.* (Object cleft)
 'It is Peter whom John slapped when he was young.'
- (2.32) a. *Peter hat Hans geohrfeigt. Er war damals Lehrling.* (Baseline)
 'Peter slapped John when he was young.'
- b. *Es ist Peter, der Hans geohrfeigt hat. Er war damals Lehrling.* (Subject cleft)
 'It is Peter who slapped John when he was young.'
- c. *Es ist Peter, den Hans geohrfeigt hat. Er war damals Lehrling.* (Object cleft)
 'It is Peter whom John slapped when he was young.'

(Examples from Colonna et al., 2015, p. 1309)

Colonna et al. (2015) attested a significant difference between the different effects of clefting in intra- and intersentential contexts, interpreting these differences in terms of "discourse coherence". More specifically, they proposed that cleft constructions signal a potential topic-shift, which would reduce coherence if it occurs within the same sentence (or discourse unit), but is possible in a new sentence (or discourse unit) "without negatively affecting discourse coherence" (Colonna et al., 2015, p. 1318). Although they do not attest a significant decrease in the preference for clefted subject or object antecedents with respect to the baseline condition, attributed to the preference of subject pronouns to keep the sentence topic constant within sentences, Colonna et al. (2012) label this slight decrease as an "anti-focus effect". This anti-focus effect was attested only for offline data in within-sentence pronoun resolution in German (Patterson et al., 2017; Patterson & Felser, 2020) and in Spanish (de la Fuente, 2015; see §2.4.3).

Authors such as Blything et al. (2021), however, have questioned that focus is regularly interpreted as indicating topic shift in the upcoming sentence, considering that there is also a general tendency to express sentence continuity (Givón, 1983; Lambrecht, 1994). Indeed, Järvikivi (2014) did not attest any evidence for an anti-focus effect—or any focus effects—in German personal pronouns in intrasentential contexts. Blything et al. (2021) did not find it-clefts to enhance the preference of English subject pronouns for clefted subject antecedents in intersentential contexts, although clefted objects did become more preferred for subject pronouns compared to canonical objects.

In this section, we have referred only minimally to studies on the influence of information structure on anaphora resolution in null subject languages such as Catalan and Spanish. As previously mentioned, very few studies have explored these factors. We will provide an overview of these studies in the following section.

2.4.3 Previous studies on the influence of information status on pronoun interpretation in Catalan and Spanish

The study by de la Fuente (2015) is crucial in terms of this thesis, as it is the only study we are aware of that has empirically and exhaustively addressed the role of information structure in pronoun resolution in Spanish (and in a null subject language, together with Mayol, 2010 in Catalan). Importantly, de la Fuente only studied null subject pronouns, so he did not include overt pronouns in his design. To investigate the role of information status in the interpretation of null subject pronouns, he examined the effects of two marked information contexts that manipulated the topical or focal status of the subject and the object antecedent: topicalization via Hanging Topic Left Dislocations (HTLD) and focalization via *it*-cleft structures. The following are examples of the five experimental conditions tested in de la Fuente (2015, p. 112, Experiment 1):

(2.33) a. *Baseline*

Eduardo llamó a Samuel cuando estaba en la oficina.

'Eduardo called Samuel when he was in the office.'

b. *Dislocated subject (via HTLD)*

Hablando de Eduardo, él llamó a Samuel cuando estaba en la oficina.

'Speaking of Eduardo, he called Samuel when he was in the office.'

c. *Dislocated object (via HTLD)*

Hablando de Samuel, Eduardo lo llamó cuando estaba en la oficina.

'Speaking of Samuel, Eduardo called him when he was in the office.'

d. *Clefted subject*

Fue Eduardo quien llamó a Samuel cuando estaba en la oficina.

'It was Eduardo who called Samuel when he was in the office.'

e. *Clefted object*

Fue a Samuel a quien Eduardo llamó cuando estaba en la oficina.

'It was Samuel whom Eduardo called when he was in the office.'

The interpretation of the ambiguous null subject pronoun in the subordinate clause was assessed in a sentence interpretation task, in which participants were asked to complete a prompt such as "..... estaba en la oficina" ('..... was in the office'). This task resembles a forced-choice task, and de la Fuente does not mention any cases in which participants referred to antecedents other than the subject or the object in the main clause (*Eduardo* and *Samuel*, respectively, in the previous examples).

The results in this task revealed that, compared to the baseline condition, topicalizing an antecedent (subject or object) made it more preferred for the null subject pronoun, whereas focalizing an antecedent (subject or object) made it less preferred for the null subject pronoun. Although the preference of null pronouns toward subject antecedents did not significantly increase in HTLD constructions, the preference of null pronouns for object antecedents significantly increased when the object was left-dislocated. In it-cleft constructions, the interpretations of null pronouns as corefering with both subject and object antecedents decreased when these antecedents were focused. Instead, null pronouns showed a higher preference for the non-clefted antecedent, conveying presupposed, topic-like, information. Similar to Colonna et al. (2012) or Hemforth et al. (2010), de la Fuente (2015) concluded that, in intrasentential contexts, topicalizing an antecedent increases its prominence for a null pronoun, whereas focusing an antecedent leads to an anti-focus effect.

Regarding focalization constructions, de la Fuente (2015, Experiment 4) further explored the effects of it-clefts in contextualized items conveying either informative focus¹¹, as in (2.34a), or contrastive/corrective focus, as in (2.34b). Again, and regardless of focus type, he found a general dispreference of null pronouns for clefted antecedents. Interestingly, he also attested similar anti-focus effects in focalization constructions different from it-clefts, using the focus-sensitive particles *solo* ('only'), *incluso* ('even'), and *también* ('also'), as in (2.35) (see de la Fuente, 2015, Experiment 5). Focused antecedents within the scope of these particles were also less preferred with respect to the baseline. Overall, he concluded that antecedents conveying new, unknown, and unexpected information do not seem to be good antecedents for a null subject pronoun in intrasentential contexts.

¹¹ De la Fuente (2015) refers to this type of focus as "narrow focus". To be consistent with the terminology used in the present thesis, we will label this type of focus as "informative focus".

(2.34) a. *Informative focus (Subject)*¹²

A: Al parecer, Samuel, Eduardo y Cristina vuelven a llevarse bien.

'It seems that Samuel, Eduardo, and Cristina are getting along well again.'

B: ¿Quién llamó a Samuel?

'Who called Samuel?'

A: Fue Eduardo quien llamó a Samuel cuando estaba en la oficina.

'It was Eduardo who called Samuel when he was in the office.'

b. *Contrastive focus (Subject)*

A: Al parecer, Samuel, Eduardo y Cristina vuelven a llevarse bien.

'It seems that Samuel, Eduardo, and Cristina are getting along well again.'

B: Lo sé. Cristina llamó a Samuel.

'I know. Cristina called Samuel.'

A: No, fue Eduardo quien llamó a Samuel cuando estaba en la oficina.

'No, it was Eduardo who called Samuel when he was in the office.'

(Examples from de la Fuente, 2015, p. 137)

(2.35) a. *Baseline*

Juan llamó a Pedro cuando estaba en la oficina.

'Juan called Pedro when he was in the office.'

b. *Focused subject*

También/Incluso/Solo Juan llamó a Pedro cuando estaba en la oficina.

'Also/Even/Only Juan called Pedro when he was in the office.'

c. *Focused object*

Juan llamó también/incluso/solo a Pedro cuando estaba en la oficina.

'Juan called also/even/only Pedro when he was in the office.'

(Examples from de la Fuente, 2015, p. 142)

Regarding topicalization structures, reference to Mayol (2010) is needed. In this study, in contrast to de la Fuente (2015), both null and overt pronouns were investigated, in this case in Catalan, but only object antecedents were topicalized, via CLLD. In a two-alternative forced-choice task, participants were asked to interpret the ambiguous null or overt subject pronoun in unmarked sentences, as in (2.36a) or in CLLD contexts such as the ones in (2.36b). These two structures allowed her to partly disentangle the role of syntactic factors, namely the syntactic function of the antecedent, and pragmatic factors,

¹² See de la Fuente (2015, p. 137) for the examples on focused object antecedents and baseline structures, we only report the dialogues referring to focused subject antecedents.

encoded as word order, in pronoun resolution. Importantly, Mayol tested intersentential contexts.

(2.36) a. *Unmarked sentences (SVO)*

A: Què li va passar a la Marta?

'What happened to Marta?'

B: La Marta escrivia sovint a la Raquel. *Pro/Ella* vivia als Estats Units.

'Marta wrote frequently to Raquel. (She) lived in the United States.'

b. *CLLD sentences (OVS)*

A: Què li va passar a la Marta?

'What happened to Marta?'

B: A la Raquel, l'escrivia sovint la Marta. *Pro/Ella* vivia als Estats Units.

'To Raquel, Marta wrote (to her) frequently. (She) lived in the United States.'

Examples from Mayol (2010, p. 129)

Based on her results, Mayol (2010) argues that null pronouns are sensitive to syntactic factors but not to word order manipulation. Topicalizing an object antecedent via CLLD, and thus appearing first-mentioned, did not affect their general preference for subject antecedents (similarly to Kaiser, 2011). On the other hand, overt pronouns showed sensitivity to both syntactic and pragmatic factors, preferring object antecedents only in the unmarked SVO condition (non-subject, non-topic constituents¹³), and not showing a clear preference in the CLLD OVS condition. In line with other authors such as Kaiser and Trueswell (2008), and similarly to de la Fuente (2015), Mayol (2010) advocates for the need of multifactorial approaches to account for pronoun resolution.

It is worth noting that, in contrast to de la Fuente (2015)'s results in Spanish, Mayol (2010) did not find topicalized object antecedents to compete with subject antecedents in the resolution preferences of null pronouns. These results also differ from the predictions of Carminati (2002), who hypothesized that null pronouns should be sensitive to both subjecthood and topicality cues. Several confounding factors could be intervening in these contradictory results. Firstly, de la Fuente (2015) tested intrasentential contexts, those analyzed by Carminati (2002) and for which the PAH was proposed, whereas Mayol (2010) studied intersentential relationships. In the light of Colonna et al. (2012, 2015), sentence relationship should be considered as another factor intervening in subject pronoun resolution, affecting the impact that the information status of the

¹³ Using Vallduví (1992)'s terms, Mayol (2010) does not refer to *topics* but to *links*.

antecedent has on the interpretive preferences of null and overt pronouns. Secondly, de la Fuente (2015) and Mayol (2010) used different topicalization constructions, HTLD and CLLD, respectively. A priori, however, we would expect both structures to lead to similar results, considering that it is the information status of the antecedent that influences pronoun interpretation and that, in both cases, the left-dislocated antecedent has a similar topical status. Thirdly, the two studies targeted different languages (Spanish and Catalan), although we have no evidence that leads us to expect different results between them, nor across Romance null subject languages.

Finally, we should also consider the possibility that testing null and overt pronouns in the same experimental task leads to slightly different results than if the participants only see one type of pronoun. This was an interesting observation made by de la Fuente (2015). He attested PAH-like biases in an experimental task including both a null pronoun and an overt pronoun condition (null pronouns coreferring with subject antecedents and overt pronouns with object antecedents in unmarked sentences, see de la Fuente's Experiment 3), while in another task using the same sentences but in which only null pronouns were tested, these pronouns remained unbiased (see de la Fuente's Experiment 1). Based on this observation across tasks, he hypothesized that pronominal anaphora resolution could be to some extent affected by metalinguistic strategies favored by the same task design¹⁴. In this regard, whereas de la Fuente (2015) only tested null pronouns and included both topicalization and focalization constructions, Mayol (2010) tested both null and overt pronouns but did not include any focalization constructions.

In the present study we will follow a similar design to the one by de la Fuente (2015), but it will include both null and overt pronouns, and we will have two equivalent tasks, one in Catalan and another one in Spanish. Given that Carminati's (2002) PAH was proposed for intrasentential contexts, these are the contexts that we will be analyzing, in contrast to Mayol (2010). We intend our design to aid in clarifying the mixed findings by the only two published studies that, to our knowledge, have addressed the role of information structure in pronoun resolution in null subject languages. Our study will expand the findings of these two studies by exhaustively investigating the interpretation of both null

¹⁴ Note, however, that de la Fuente (2015) tested only null pronouns but included five different information structures in Experiment 1, and tested both null and overt pronouns in Experiment 3, but only one information structure (the baseline condition, unmarked canonical sentences). The two experiments did not only differ in the levels of the Pronoun condition, but also in the Information structure condition.

and overt pronouns in both topicalization and focusing structures, clearly extricating not only syntactic and pragmatic factors, but also sequential factors, which we will address in the following section.

2.5 Sequential factors in anaphora resolution: order of mention or surface linear position of the antecedents

One consequence of manipulating the syntactic position and the information status of antecedents through syntactically marked information structures is that the surface linear position of the antecedents (or the order of mention of the antecedents), is also altered. As we have previously mentioned, while unmarked sentences and subject clefts in the present study follow an SVO word order (subject antecedents are first-mentioned), object CLLD and object clefts follow an OVS word order (object antecedents are first-mentioned).

In addition to syntactic and pragmatic factors, sequential factors have also been recurrently identified as possibly influencing the interpretation of ambiguous subject pronouns. In previous studies on canonical SVO sentences, the subject does not only overlap with the sentence topic, but it also coincides with the initial position, which has been argued to be a privileged position for an antecedent to be picked up by a subsequent pronoun. The preference for antecedents in initial position regardless of other factors has been referred to as the first-mention advantage (Gernsbacher & Hargreaves, 1988) and attributed to general cognitive processes (see also Carreiras et al., 1995).

Studies such as Kaiser and Trueswell (2008) or Bader and Portele (2019) on personal pronouns in Finnish and German have demonstrated that the preference for subject antecedents is stronger than the preference for first-mentioned antecedents (see also Cowles et al., 2007 for English). Other studies such as Järvikivi et al. (2005) have shown that there is no advantage for the first-mentioned antecedent but that both syntactic function and order of mention play a role in the interpretation of personal pronouns in Finnish. Therefore, and consistent with the studies reviewed in previous sections demonstrating the relevance of language-related factors on anaphora resolution, the first-mention advantage does not seem to be a determining factor in anaphora resolution. The results of de la Fuente (2015) and Mayol (2010) in Catalan and Spanish likewise do not support the presence of a systematic first-mention advantage for null pronouns, across contexts, in null subject languages.

However, as noted by Järvikivi et al. (2005) the order of mention—or surface linear position—of the antecedents may still play a role in anaphora resolution. For instance, Bader and Portele (2019) found a combination of syntactic function, topicality, and also linear position to shape the resolution preferences of d-pronouns in German, which could behave similarly to overt pronouns in null subject languages. However, few studies have successfully disentangled the notions of subject and first-mention, as well as those of topic and first-mention. The design of the present study will allow us to clarify whether the order of mention of the antecedents contributes to determining the preferences of null and overt pronouns.

We need to take into account, however, that the surface linear position of the antecedents in our experimental sentences will overlap with the hierarchical position of the antecedent in the syntactic configuration. As previously explained in §2.3.1, in SVO canonical sentences and SVO subject clefts the first-mentioned subject will also appear in a higher position in the syntactic structure than the object antecedent (Spec,IP or Spec,FocP, respectively). In OVS CLLD structures and OVS object clefts, the first-mentioned object antecedent will also appear in a higher position in the syntactic structure than the subject antecedent (Spec,TopP or Spec,FocP, respectively). Therefore, the design of the present study will not allow us to extricate the initial surface position and the hierarchically higher syntactic position of the antecedents.

In the interest of clarity, we will consider syntactic factors to refer to the syntactic function of the antecedent (as most studies in the previous literature have understood syntactic factors in anaphora resolution), and sequential factors to refer to the order of mention of the antecedents or word order. However, as we have seen, the linear position of the antecedents in canonical sentences, CLLDs and clefts in Catalan and Spanish is also explained in terms of structural configurations. In other words, we are arguably able to state that sequentiality reproduces, at the linear level, the structural position of the antecedent in the syntactic configuration. Similarly, first-mentioned (vs. second-mentioned) or preverbal (vs. postverbal) antecedents will coincide with hierarchically, structurally, or configurationally higher (vs. lower) antecedents in the syntactic structure.

When discussing the results, based on the lack of evidence in favor of a consistent first-mention advantage and considering the fact that we will be dealing with offline data (and not processing data), we will be more inclined to interpret a preference for first-mentioned (or second-mentioned) antecedents as being related to a preference for antecedents in

a hierarchically higher (or lower) position in the syntactic structure. We will remind the reader that sequential and structural factors cannot be completely teased apart in our design.

2.6 Summary and implications for the present study

In this chapter we have reviewed previous studies on anaphora resolution that demonstrate that the interpretation of third-person subject pronouns is sensitive to a wide range of linguistic factors and multiple constraints must be taken into account (e.g., Kaiser & Trueswell, 2008; see §2.2). Although a number of studies have confronted different factors in order to determine their relative weight on anaphora resolution in non-null subject languages, very few studies have investigated the interaction between syntactic, pragmatic, and sequential cues in null subject languages (de la Fuente, 2015; Mayol, 2010). One of the main aims of the present study will be to investigate how, and to what extent, syntactic (i.e., syntactic function of the antecedents), pragmatic (i.e., information status of the antecedents), and sequential factors (i.e., linear position of the antecedents or position of the antecedents in the syntactic structure) influence the interpretive properties of null and overt subject pronouns in Catalan and in Spanish.

In order to do so, we will use marked information structures to disentangle the notions of subject, topic, and initial position in an experimental task (see §4.3 in Chapter 4). More specifically, we will manipulate the information status of the plausible antecedents of the ambiguous pronouns through syntactic means, as described in §2.4. To topicalize object antecedents, we will use clitic-left dislocation (CLLD) structures, as in (2.37). To focus subject and object antecedents, we will use it-cleft structures, as in (2.38) and (2.39). This means that both topical and focal subjects and objects will be in sentence-initial position at different points. These marked structures will be compared to a baseline condition: pragmatically unmarked canonical SVO sentences.

Firstly, in CLLD structures (OVS word order), object antecedents are topicalized and appear first-mentioned, in the hierarchically highest position of the syntactic structure (Spec,TopP). Subject antecedents convey focal (or non-topical) information and appear second-mentioned, in a hierarchically lower position than objects.

(2.37) *Clitic-left dislocations*

- a. A la Sabina la va interrompre l'Olívia quan va començar a parlar. (Catalan)
- b. A Sabina la interrumpió Olivia cuando empezó a hablar. (Spanish)
'Sabina, Olivia interrupted her when she started speaking.'

Secondly, in subject it-cleft structures (SVO word order), subject antecedents are focused and appear first-mentioned, in the hierarchically highest position of the syntactic structure (Spec,FocP). Object antecedents convey presupposed (topical) information and appear second-mentioned, in a hierarchically lower position than subjects.

(2.26) *Subject clefts*

- a. Va ser l'Olívia qui va interrompre la Sabina quan va començar a parlar.
 - b. Fue Olivia quien interrumpió a Sabina cuando empezó a hablar.
- 'It was Olivia who Sabina interrupted when she started speaking.'

Finally, in object it-cleft structures (OVS word order), object antecedents are focused and appear first-mentioned, in the hierarchically highest position in the syntactic structure (Spec,FocP). Subject antecedents convey presupposed (topical) information and appear second-mentioned, in a hierarchically lower position than objects.

(2.27) *Object clefts*

- a. Va ser a la Sabina a qui va interrompre l'Olívia quan va començar a parlar.
 - b. Fue a Sabina a quien interrumpió Olivia cuando empezó a hablar.
- 'It was Sabina whom Olivia interrupted when she started speaking.'

The following table summarizes the main features of subject and object antecedents in each of the information structures under analysis.

Table 2.1

Main features characterizing subject and object antecedents in each information structure

Information structure	Canonical unmarked structure		Topicalized object via clitic-left dislocation		Focused subject via it-cleft		Focused object via it-cleft	
	Subject	Object	Object	Subject	Subject	Object	Object	Subject
Syntactic function	Subject	Object	Object	Subject	Subject	Object	Object	Subject
Information status	Topical	Focal	Topical	Focal	Focal	Topical	Focal	Topical
Order of mention¹⁵	First	Second	First	Second	First	Second	First	Second

¹⁵ Remember that first-mentioned antecedents are also the ones that appear in the hierarchically highest position of the syntactic configuration.

Canonical sentences, in which subjecthood, topicality and first-mention overlap, have been the most widely studied contexts in the literature on the Position of Antecedent Hypothesis (PAH; Carminati, 2002) (see §2.3.1). In the present study, canonical or unmarked sentences will also be used as a baseline condition for comparisons with marked topicalization and focus structures.

Firstly, contrasting these unmarked contexts with marked information structures in which the subject antecedent does not necessarily appear in the hierarchically higher position (or first-mentioned, see §2.5) will provide novel evidence on the PAH. Although this hypothesis was formulated in structural terms, in most studies the syntactic function of subject antecedents coincides with the hierarchically highest position in the syntactic structure. According to Torregrossa et al. (2020), this is the preferred position for the antecedent of a null pronoun. Within Carminati's (2002) framework, we aim to empirically test whether null pronouns prefer to be interpreted as coreferring only with subject antecedents (in Spec,IP) or antecedents in the hierarchically highest structural position (Spec,IP in unmarked contexts, Spec,TopP in CLLD, and Spec,FocP in it-clefts), as predicted by Torregrossa et al. (2020). Moreover, we will test whether overt pronouns prefer to be interpreted as coreferring with object antecedents or antecedents in the hierarchically lowest structural position (e.g., lower than Spec,IP).

Secondly, comparing unmarked and marked information structures will enable us to disentangle the notions of subject and topic (see §2.4.1). Regarding the interpretation of null pronouns, Carminati (2002) also hypothesized that antecedents in topic position may compete with antecedents in subject position if topicality and subjecthood are not aligned, weakening—but not overriding—the preference of null pronouns for subject antecedents. The role of the information status of the antecedent in pronoun resolution is not yet clear (see §2.4.2), and how it influences the interpretation of null and more especially overt pronouns has not yet been comprehensively studied (see §2.4.3). To our knowledge, the effects of topicalization and focalization in Romance null subject languages have only been investigated by de la Fuente (2015), in Spanish. He found that topicalization structures increase the preference of null pronouns for topical antecedents, and focalization it-cleft structures decrease the preference of null pronouns for clefted antecedents. The comparison between null and overt pronouns has only been addressed by Mayol (2010), who tested object topicalization constructions in Catalan. She found that overt pronouns decrease their preference for topicalized objects and, in contrast to de la Fuente (2015), she did not find any effects on null pronouns, which

maintained their subject bias. Another relevant contribution of the present study will be to investigate how syntactic and pragmatic factors interact in anaphora resolution and to what extent the information status of the antecedent intervenes in the interpretation of null and overt pronouns. In addition, the importance of informational status in Romance languages has been investigated in Spanish and in Catalan, but with the two languages being considered separately. In our study, we will analyze and compare both languages to further investigate a potential microvariation between them.

Broadly speaking, both Catalan and Spanish have been demonstrated as complying with the predictions of the PAH (Bel & García-Alcaraz, 2018; see §2.3.2): null pronouns tend to prefer to corefer with subject antecedents (in Spec,IP), and overt pronouns tend to prefer to corefer with object antecedents (in lower syntactic positions). However, differences in the strength of pronoun resolution biases seem to arise between Catalan and Spanish, as well as across null subject languages (see §2.3.3). In the present study, we will investigate the role of syntactic, pragmatic, and sequential/structural factors in anaphora resolution in Catalan and in Spanish. We also aim to compare the two languages to confirm whether they show microvariation in the interpretive properties of subject pronouns. Importantly, given that in the present thesis we are not only concerned with language-internal factors (described in the present chapter) but also with language-external factors (i.e., bilingualism and language dominance, as described in Chapter 3), in addition to comparing Catalan and Spanish, we will also compare pronoun interpretation in two varieties of Spanish: in bilingual Spanish in contact with Catalan and in monolingual Peninsular Spanish. The specific research questions related to the mentioned aims, together with our predictions, will be presented and addressed along Chapter 5.

Chapter 3

The effects of bilingualism and language dominance on anaphora resolution in the Catalan-Spanish context

3.1 Introduction

The present research addresses pronominal anaphora resolution by early Catalan-Spanish bilinguals in both Catalan and Spanish. Inevitably and intentionally, bilingualism is one of the core themes of analysis in this thesis. Bilingual individuals provide an excellent testing ground to explore how interactions between syntax and other domains take place since these speakers show varying mastery in different modules (Benmamoun et al., 2013; Meisel, 2011; Slabakova, 2013), something which is in principle impossible to test in monolingual speakers. Furthermore, early bilinguals who have been raised in bilingual societies also constitute a population with a valuable profile—different from more widely studied L2 speakers, L1 attriters or heritage speakers.

The study of highly functional Catalan-Spanish bilinguals provides an unexplored and privileged context within which to examine the effects of bilingualism on anaphora resolution. It will allow for investigation into the role of language dominance in bilinguals who—as will be argued—are highly proficient in both of their languages. This study will provide valuable insights into the effects of bilingualism on a widely studied linguistic phenomenon, as well as into how these effects may be modulated by or depend on language dominance. In addition to studying the interpretation of third person anaphoric subject pronouns in Catalan and Spanish and the relative weight of syntactic, pragmatic, and word order factors, the present thesis aims to investigate the impact of bilingualism and language dominance on shaping anaphora resolution in early bilingual speakers, using a bidirectional design.

Pronoun interpretation is governed by intuitions, rather than by explicit knowledge or categorical and univocal factors. It is located at the interface between syntax and pragmatics and requires the integration of information from different linguistic domains (see Chapter 2). As will be described over the course of the present chapter, this makes it a cognitively challenging and difficult-to-acquire phenomenon, vulnerable to the effects of bilingualism (e.g., Sorace and Filiaci, 2006). Although anaphora resolution has been extensively studied in bilingual populations, almost no research has targeted early

bilinguals that have been born, grown up, and lived in a society in which two languages are used daily and across different environments.

In the present chapter, we will firstly refer to the bilingual experience from a global perspective (§3.2). We will define how we understand bilingualism (§3.2.1) and language dominance (§3.2.2.1). Given that we will use language dominance as a proxy for the bilingual experience, we will also attempt to identify the main factors affecting this construct (§3.2.2.2), how it can be measured and operationalized (§3.2.2.3), and how it can be used to more or less homogeneously classify bilinguals according to their profile (§3.2.2.4). We will then refer to two main effects of bilingualism: 1) general effects of bilingualism per se (§3.2.3.1) and 2) crosslinguistic influence (§3.2.3.2). In addition to bilingualism within the individual, we will further refer to societal bilingualism and language contact situations (§3.2.4). Secondly, given that the present research is framed within the Catalan context, we will refer to bilingualism and language contact in Catalonia (§3.3). In this section, we will briefly contextualize the sociolinguistic context of Catalonia (§3.3.1), we will provide an overview of previous experimental studies on Catalan-Spanish bilingualism that have taken language dominance into account (§3.3.2), and we will analyze how these studies have assessed the language dominance of their participants, in order to explain briefly how we are going to do so (§3.3.3). Thirdly, we will devote a section to how bilingualism affects anaphora resolution (§3.4). We will provide an overview of previous empirical findings (§3.4.1) and discuss how these results have been interpreted as showing effects of bilingualism itself (§3.4.2) or as showing crosslinguistic influence (§3.4.3). Lastly, we will summarize the most relevant findings reviewed in the present chapter as a recapitulation of the main implications for the present research (§3.5).

3.2 Characterizing the bilingual experience

3.2.1 Defining bilingualism

Several definitions of bilingualism—and multilingualism—have been proposed according to different approaches and have broadened over time. A bilingual was long defined as someone who is fully competent in two languages, in keeping with Bloomfield (1933, p. 56). This focus on language proficiency later shifted to language use, as illustrated in Weinreich's (1953) and Mackey's (1962, p. 52) view of bilingualism as "the alternate use of two or more languages by the same individual". Similarly, Grosjean (1989, 1998, 2010)

has defined bilinguals as those who use two or more languages (or dialects) in their everyday lives. Beyond this definition, he further proposed the Complementary Principle, according to which bilinguals "acquire and use their languages for different purposes, in different domains of life, with different people" (Grosjean, 1998, p. 132). As such, he implied that the characterization of bilinguals is necessarily more complex and could depend on the functions and domains of use of each of the bilinguals' languages. Moreover, he insisted on the fact that bilinguals should not be seen as the sum of two monolinguals and rarely have the same competency in their two languages across domains and abilities. This is the reason why the image of the "ideal" or "balanced" bilingual—whatever meaning this notion may have—does not represent the vast majority of the bilingual population (e.g., Grosjean, 1989).

Grosjean (1998) already identified the need to consider multiple factors that could impact language processing and representation differently in bilinguals, such as biographical data (age, sex, and socioeconomic and educational status), age of onset of acquisition, context of acquisition, contexts of use, proficiency in different skills, or language mode. Taking into account the multiple dimensions necessary to consider when assessing the bilingual experience, bilingualism has been reconceptualized recently as a gradient (or continuous) and dynamic construct (e.g., Dunn & Fox-Tree, 2009; Luk & Bialystok, 2013). At present, it is becoming more frequent to better describe bilingualism as, for instance, "a multidimensional spectrum of experiences composed of several continuous constructs associated with language usage and exposure across multiple languages" (Gullifer et al., 2021, p. 271). However, how to measure and operationalize bilingualism as a multidimensional, gradient, and dynamic construct so as to efficiently capture the bilingual experience of different populations in experimental studies is still widely debated and remains an open question in ongoing research. In the present thesis, we will weigh these aforementioned multiple factors influencing the profile of bilingual speakers within a comprehensive measure of bilingualism, through the assessment of language dominance (see §3.3.3 below).

3.2.2 Language dominance as a proxy for bilingualism

3.2.2.1 Defining language dominance

There is consensus that bilingual individuals tend to have a *dominant language*, or stronger language, and a *non-dominant language*, or weaker language. The notion of language dominance, however, has been interpreted in very different manners. By and large, language proficiency and language use—in many cases understood as also including language exposure—have been identified as the two key components of this construct (see Treffers-Daller, 2019 for a review). More specifically, language dominance has been defined as referring to "observed asymmetries of skill in, or use of, one language over the other" (Birdsong, 2014, p. 1) or as "the relative weight and relationship of the two languages of a bilingual in terms of language use and degree of proficiency" (Montrul, 2016, p. 16). Other studies have understood language dominance in "its narrow sense of relative language proficiency" (Unsworth, 2016a; Unsworth et al., 2018, p. 14).

In other studies, mainly on bilingual children, the amount of exposure to each language has been identified as the determining factor when establishing the dominant language, defined as "the language in which the bilingual child obtains more input on a regular basis" (Argyri & Sorace, 2007, p. 83). In Argyri and Sorace's study, however, the language of the community was considered the dominant language. In fact, language dominance has also been often understood to refer to societal language dominance. In this case, the dominant language would be the predominant language in a given setting where different languages are used (Meisel, 2001, 2007; see also Treffers-Daller, 2019). Similarly, other studies have identified the dominant language of a bilingual as the majority or socially dominant language in a society or in a specific environment (e.g., Davidson, 2022; Polinsky, 2008). However, studies such as Schmeißer et al. (2016), in which bilingual children raised in bilingual families in Germany speaking the language of the environment (i.e., German) and a foreign Romance language, have demonstrated that the language spoken in the wider community is not a reliable predictor for (individual) language dominance (see also Hervé et al., 2016; Unsworth et al., 2018).

Overall, relative language dominance—as bilingualism—can probably be more accurately defined as multidimensional, gradient, and dynamic (Birdsong, 2016). However, there is no general agreement on identifying the variables that need to be considered to best define a bilingual's language dominance profile. The main factors or variables that have been argued to form our understanding of language dominance are overviewed in the following section (§3.2.2.2).

3.2.2.2 Factors affecting language dominance

Beyond her definition, Montrul (2016) identifies the following components of language dominance: biographical variables (age of acquisition, place of birth, place of residence, languages of the environment), estimations of language input, degree of language use, and proficiency in each language. These factors can also be identified from the definitions of language dominance previously presented. However, if language dominance is conceived as multidimensional, all these components—and not only one of them—should be taken into account when defining, measuring, and operationalizing language dominance. In the following paragraphs, we describe age of onset of acquisition, language exposure, language use, and language proficiency as key factors affecting language dominance in bilingual speakers.

Frequently, bilinguals have been characterized based on the age at which their languages are acquired (De Houwer, 2009; Meisel, 2009, 2021). The acquisition of two languages from birth has been defined as *simultaneous bilingual acquisition* (2L1), or *bilingual first language acquisition*, whereas the acquisition of one language from birth and the other at a later stage has been defined as *sequential* (or *successive*) *bilingual acquisition*, or *bilingual second language acquisition*. With respect to sequential bilingualism, a further distinction has been made between *early* or *child second language acquisition* (cL2, when the child acquires an L2 before an age around 6, after having partially acquired an L1) and *late* or *adult second language acquisition* (aL2). As reviewed in Meisel (2021), a wealth of research on child acquisition has shown that, whereas successive bilingual children differ in substantial ways from simultaneous bilinguals and monolinguals in the acquisition of their L2, simultaneous bilinguals can resemble monolingual children in both language development and ultimate attainment of grammatical knowledge. However, an early age of onset of acquisition is highlighted as a necessary but insufficient condition for the development of a native competence. In this regard, Meisel emphasizes the crucial importance of other factors related to input or 'adequate exposure'.

Language exposure, or quantity and quality of input, is also a fundamental factor in bilingual language development (Hoff et al., 2012; La Morgia, 2015; Meisel, 2021; Paradis, 2011; Unsworth, 2016b). In part, it is determined by language-external conditions related to place of birth and residence, or the languages used at home, of the environment, and of education. The amount of input, as well as the contexts of exposure, the sources of this input, or its variety and richness have been identified as relating to

variability in bilingual acquisition and to the rate and the ultimate state of acquisition. This is why language exposure or input has been used in a number of studies to determine the dominant language of bilingual children as the language they are more exposed to (e.g., Hervé et al., 2016) or as the interaction language at home (e.g., Foroodi-Nejad & Paradis, 2009). In bilingual communities with a clearly dominant language, input levels at home and at school have been revealed to show a strong link with the development of the minority language, its timing of acquisition and its ultimate attainment (Gathercole & Thomas, 2009).

Another fundamental aspect of bilingual language development refers to language use, or output (e.g., Unsworth, 2016b). This factor is closely related to language exposure. Indeed, language dominance has been often seen as a combination of language exposure and language use (e.g., Bedore et al., 2012; see also Grosjean, 2010; Treffers-Daller, 2019). Early bilinguals have been defined as *balanced* when they have received a comparable amount of input in each language and have used both languages to a similar extent. When they have received substantially more input in one of the languages and have used it to a substantially greater extent compared to the other, they have been considered to be *dominant*, or *unbalanced*, bilinguals.

Finally, language dominance has been very frequently defined based on language proficiency (e.g., Kupisch, 2012; Wei, 2007; among many others). Whereas *balanced bilinguals* have been typically described as having a similar proficiency—and thus a similar dominance—in their two languages, *unbalanced bilinguals* have been described as having a higher proficiency in their dominant language and a lower proficiency in their non-dominant language. Nevertheless, highly proficient bilinguals—and even early simultaneous bilinguals—can also show different abilities in their two languages and be dominant in one of them, which is a reason why dominance should be defined beyond proficiency (Gertken et al., 2014; Kupisch & van de Weijer, 2016; Montrul, 2016; Puig-Mayenco et al., 2018). Unsworth (2016a) found a strong relationship between proficiency and amount of input, amount of use, and age of acquisition. In light of these results, she proposed language exposure and language use as valid means of operationalizing language dominance—understood as relative proficiency—in bilingual children (see also Unsworth et al., 2018). As previously mentioned, in the present study we will use a comprehensive and multidimensional measure of language dominance (see §3.3.3). To adequately measure and operationalize language dominance taking into account multiple dimensions, however, is challenging and still controversial.

3.2.2.3 Measuring and operationalizing language dominance

It seems clear that language experience is necessarily multidimensional, and that language proficiency, language use, and language exposure can all be considered as different components of language dominance (e.g., Montrul, 2016). Also, it should be conceived as gradient or continuous. However, how the complex components that form the construct of language dominance should be measured and operationalized is far from obvious (see Treffers-Daller, 2019). It is not clear which specific variables should be taken into account, nor what is their relative weight when quantifying language dominance. Measuring and operationalizing language dominance is therefore challenging, and far from agreed upon.

Given that language dominance has very often been equated to language proficiency, standardized proficiency tests typically used for second language acquisition have been widely used to measure language dominance. Identifying dominance with proficiency, however, has important limitations, even more so in cases of bilingual speakers who have grown up in bilingual communities. Despite this tendency towards measures of proficiency, there are more fine-grained questionnaires that have aimed to capture the bilingual experience of adult bilinguals across different domains and dimensions. Some of these questionnaires include the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), the Bilingual Dominance Scale (Dunn & Fox-Tree, 2009), the Language History Questionnaire (Li et al., 2006, 2014), the Bilingual Language Profile questionnaire (Birdsong et al., 2012), and the Language and Social Background Questionnaire (Anderson et al., 2018). However, the manner these authors define and measure dominance (or, in some cases, bilingual experience) varies.

Finally, it is worth mentioning the review by Kaščelan et al. (2022). This study illustrates the variation in the operationalization and measurement of several overarching components of the bilingual experience, such as language exposure and use, activities in each language, and current language skills. However, they only focused on questionnaires quantifying children's bilingual experience, and do not directly address the construct of language dominance.

In §3.3.3, we will review how language dominance has been measured and operationalized in previous studies on adult Catalan-Spanish bilinguals. In this section, we will also address the decision to use the Bilingual Language Profile (Birdsong et al., 2012; Gertken et al., 2014) to characterize the participants in the present study.

3.2.2.4 Classifying bilinguals

As mentioned, there is nowadays little controversy in the understanding of bilingualism as a multidimensional, dynamic, and gradient construct, affected by a number of different factors. However, the use of a combination of categorical and continuous models of bilingualism has also been proposed (e.g., Kremin & Byers-Heinlein, 2021). In fact, many studies still tend to classify bilinguals into subgroups, and this view is not incompatible with understanding the bilingual construct as continuous. This facilitates systematic comparisons within and between more or less homogeneous groups of bilinguals with different profiles (see Perpiñán & Soto-Corominas, 2021; Treffers-Daller, 2016). Even though there is a risk of simplification, grouping participants may also help visualizing and identifying common and distinguishing elements within each group. At any rate, it is also true that the criteria with which to classify bilinguals into groups may be arbitrary and controversial (see Birdsong, 2016).

Bilinguals have been—and are still being—classified and characterized by following a number of different criteria (see Wei, 2007, p. 4 for a variety of terms used to describe bilingual speakers). These categorizations have mainly been based on age of onset of acquisition (e.g., Meisel, 2021), or on language dominance (e.g., Treffers-Daller, 2019). Although there is solid evidence to claim that the age of onset of acquisition is a determining factor in bilingual acquisition, as discussed above (e.g., Bylund et al., 2021; Meisel, 2009, 2011), the scope and the locus of its impact are still a matter of controversy, even more so regarding the development of minority languages in bilingual societies (see Perpiñán & Soto-Corominas, 2021, pp. 1464-1465 and references within). In fact, age of onset of acquisition, although crucial, is not sensitive enough to capture the full range of bilingual populations (e.g., Birdsong, 2016; Dunn & Fox-Tree, 2009). In the same manner, measures based on proficiency also offer a limited view of a bilingual profile. Individual experiences of bilingualism vary over the years, are influenced by language exposure and language use from a diachronic and a synchronic perspective, and—as previously argued—are more intricate than what a single measure can reflect (e.g., de Bruin, 2019; Gullifer et al., 2021; Luk & Bialystok, 2013).

Language dominance, as with bilingualism, can be understood as a multifaceted and gradient construct (e.g., Gertken et al., 2014), and has been identified as a relevant factor determining bilingual outcomes (e.g., Perpiñán & Soto-Corominas, 2021; Yip & Matthews, 2006; see also §3.3.2 for studies on Catalan-Spanish bilingualism). Under these considerations, and similar to other studies on Catalan-Spanish bilinguals (see

§3.3.3), we decided to use language dominance to distinguish between three subgroups that we consider to reliably represent the Catalan-Spanish bilingual continuum: Catalan-dominant bilinguals, balanced bilinguals, and Spanish-dominant bilinguals. Before describing the sociolinguistic context of Catalonia and previous studies that have dealt with language dominance in this community, in the following sections we will further characterize the bilingual experience, and we will refer to the main effects of bilingualism and language contact from an individual and a societal perspective.

3.2.3 The effects of bilingualism per se and the emergence of crosslinguistic influence

Having more than one linguistic system in a single mind has implications on language representation, processing, and use, from both a cognitive and a linguistic perspective (e.g., see Di Pisa et al., 2021 for a recent review). Importantly, both languages have been shown to be simultaneously active in the bilingual mind, even in contexts where only one language is required (Costa & Sebastián-Gallés, 2014; Dijkstra, 2005; Green & Abutalebi, 2013; Hopp, 2017; Kroll et al., 2012, 2015; Kroll & Bialystok, 2013; Kroll & Navarro-Torres, 2018; Marian & Spivey, 2003; among others).

For the purpose of the present study, we consider two main consequences of bilingualism that may affect linguistic outcomes such as the interpretation of null and overt pronouns, based on Sorace (2011): 1) general effects of bilingualism under conditions of increased processing demands, and 2) crosslinguistic influence (CLI). Firstly, the parallel activation of the languages of a bilingual results in the competition for finite cognitive resources. The non-target language needs to be inhibited and bilinguals' processing demands increase. This competition for resources makes bilingual processing cognitively demanding, and potentially costly. Secondly, the interaction of more than one linguistic system in the same individual results in CLI across languages.

Phenomena at the syntax-discourse interface, as anaphora resolution, have indeed been shown to be especially sensitive to both processing limitations and CLI. Overall, however, it is still not totally understood how language-specific and general cognitive factors interact and influence the linguistic outcomes of bilinguals in language contact situations or in bilingual societies (Sorace, 2016).

3.2.3.1 General effects of bilingualism in language processing

Research on adult bilingual sentence processing has explained differences between native and non-native language processing in terms of qualitative differences in the underlying parsing mechanisms, or in terms of quantitative differences related to cognitive resources (see Cunnings, 2017; Hopp, 2022 for recent reviews). Theoretical proposals on language processing in L2 acquisition such as the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018) have suggested essential differences between L1 and L2 sentence processing. Clahsen and Felser claim that non-native processing, even by highly proficient L2 speakers, is less efficient at the level of syntactic representations. Non-native speakers rely more on semantic, pragmatic, probabilistic or surface-level information than on syntactic cues. This strategy is effective but may lead to instability or variability, which in this case would not be related to inefficiencies on computational resources. However, these observed less efficient processing strategies of L2 learners should not be automatically generalized to early bilinguals (Clahsen & Felser, 2018; Felser, 2020). How early bilinguals process their dominant and non-dominant languages has not been widely studied and we lack evidence on whether societal bilinguals are affected by these processing limitations in any of the languages in which they are highly proficient.

On the other hand, other authors have proposed that L1 and L2 sentence processing are fundamentally similar and defend that differences are quantitative, rather than qualitative, and related to cognitive resources, rather than to an under-reliance on grammatical information (Hopp, 2010, 2017, 2022). According to this approach, native-like L2 processing and native-like use of syntactic cues are possible with sufficient proficiency, given that underlying processing mechanisms of bilingual populations are essentially the same as those of monolingual speakers. Syntactic constructions that are cognitively demanding for bilinguals are also challenging for monolinguals. Hopp (2022, p. 245), after a comprehensive review on studies into second language processing, further suggests that the "effects of bilingualism can lead to non-native sentence processing in an L2 because the parser operates on representations in interlanguage systems that are by definition distinct from monolingual systems". In any case, the effects of bilingualism and parallel activation on language representation, processing, and use have been found to be varied, which is not surprising considering that the bilingual experience is necessarily more complex than that which dichotomous characterizations can capture (see Kroll & Navarro-Torres, 2018).

Also, it is worth mentioning that neuroimaging research has demonstrated that the bilingual experience affects not only language processing, but also nonverbal cognitive performance and brain structure (Abutalebi & Green, 2016; Bialystok, 2017; Del Maschio et al., 2020; DeLuca et al., 2020; Mishra & Abutalebi, 2020; Pliatsikas, 2019; among others). Thus, bilingualism affects general cognitive processes and the neural networks underlying them. However, little is known yet about how the variety of bilingual experiences can change the mind and brain, as well as how this impact is reflected in linguistic outcomes (see Zirnstein et al., 2019). In any case, the effects of bilingualism on cognition go beyond the scope of the current thesis.

In the present study we will be interested in studying whether evidence of general bilingual processing strategies is found in the final interpretive preferences of null and overt pronouns (see §3.4). Thus, we will not be directly investigating the relationship between bilingualism and language processing. Preferences in offline comprehension, the kind of data we will gather, can provide insights into issues that matter in the processing of anaphoric dependencies, as well as indirect evidence of how bilingual speakers process anaphors. In light of a possible less efficient bilingual processing, the linguistic performance of bilinguals in interface phenomena such as anaphora resolution has been attributed to bilingualism itself. General effects of bilingualism have been identified as a possible explanation for the variability shown mainly by L1 attriters and near-native L2 speakers and accounted for via the Interface Hypothesis (see Sorace, 2011 and §3.4.2; see also Teixeira, 2020). Interestingly, this has not been investigated with early bilinguals from social bilingual settings. We will review in more detail these effects of bilingualism per se on anaphora resolution and the Interface Hypothesis in §3.4.2.

3.2.3.2 Crosslinguistic influence in bilingual outcomes

Although bilinguals are known to have separate and differentiated linguistic systems for each of their languages (De Houwer, 1990, 2005; Genesee, 1989; Meisel, 1989, 2001) (see also Kroll & Tokowicz, 2005), parallel activation and systematic contact lead to interaction between the bilingual's languages, and possibly to CLI (Hopp, 2017; Hulk & Müller, 2000; Nicoladis, 2006; Paradis & Navarro, 2003; Serratrice, 2013, 2016; among others). As defined in Serratrice (2013, p. 4), we understand CLI as referring to "instances in which there is evidence for the effect of one language on the other". However, the contexts in which CLI is most likely to emerge still remain unclear. Van Dijk

et al. (2021), in a recent meta-analysis of studies on CLI in early bilingual children, identify four factors that have frequently been examined as predictors of the emergence and the strength of CLI: surface overlap between the bilingual's languages, language domains involved in the phenomenon under study, language dominance, and age.

We firstly refer to language-internal factors: language overlap and language domain. Hulk and Müller (2000) and Müller and Hulk (2001) proposed two necessary conditions for CLI to occur in a specific linguistic structure: 1) that it involve an interface between two modules of grammar (mainly between syntax and other modules such as pragmatics), and 2) that the two systems overlap at the surface level. Several studies have confirmed this proposal and have attested CLI in the presence of surface overlap in structures both at the syntax-pragmatics and syntax-semantics interface (see Serratrice, 2013 for an overview). However, other studies have found evidence of CLI in the absence of surface overlap or for non-interface phenomena (e.g., Argyri & Sorace, 2007; Bosch & Unsworth, 2021). Also, there are studies that have found evidence of CLI in a direction not predicted by Hulk and Müller proposals. Hulk and Müller (2000) and Müller and Hulk (2001) assume that CLI is predictably unidirectional. More specifically, they predict that a language A is likely to influence a language B if language B allows for more than one grammatical analysis and language A reinforces one of these possible analyses (see also Döpke, 1998). However, CLI has also been found to occur bidirectionally (e.g., Foroodi-Nejad & Paradis, 2009; López Otero, 2022; Meir et al., 2017; Pavlenko & Jarvis, 2002; Soto-Corominas, 2021).

Regarding language-external factors, language dominance has been demonstrated to be more determinant in shaping cross-language interactions than age at testing (e.g., Bosch & Unsworth, 2021; van Dijk et al., 2021)¹⁶. More specifically, the stronger or dominant language has been found to influence bilinguals' weaker or non-dominant language (e.g., Argyri & Sorace, 2007; Foroodi-Nejad & Paradis, 2009; Kupisch, 2007; Serratrice et al., 2009; Yip & Matthews, 2000, 2007)¹⁷. Kupisch (2012, p. 740) even specifies that language dominance can overrule language-internal factors so that CLI only occurs from the stronger towards the weaker language and not in the opposite direction. However, it has been noted that not only can the L1 or the dominant language

¹⁶ In the present study we will be more concerned with language dominance effects than age effects, given that our participants are early bilingual adults and not children, where age could be a more relevant factor (see also §3.2.2.2).

¹⁷ Note that language dominance has been operationalized in various manners depending on the study (e.g., Silva-Corvalán & Treffers-Daller, 2016; see also §3.2.2)

influence the bilinguals' L2 or the weaker language, but the L2 can also influence the L1 provided that bilinguals are proficient enough in their L2 (see also Kroll et al., 2015). Overall, language dominance has not been found to predict the occurrence of CLI affecting a particular structure, which is modulated by language-internal factors, but instead to affect the directionality and also the strength of CLI when occurring (e.g., Kupisch, 2007, 2012, 2014). In terms of Genesee and Nicoladis (2007), language dominance may act as a mitigating factor.

As previously mentioned, in the present study we will try to distinguish between whether general effects of bilingualism arise in bilingual pronoun interpretation (not language-specific) or whether CLI is observed (bound to language-specific interpretations). Moreover, we will look at the performance of highly proficient or functional bilinguals who only differ in relative language dominance. This will allow us to observe whether language dominance modulates the appearance of general bilingualism effects or the directionality and/or strength of CLI.

3.2.4 Societal bilingualism and language contact

In the previous sections, we have understood bilingualism as a property of the individual. However, it is important to take into account that Catalan-Spanish bilinguals in Catalonia also live in a bilingual society (see §3.3.1). Therefore, the interaction between bilinguals' two linguistic systems does not only take place at the individual level, but also at the societal level. In addition to language-internal factors such as surface overlap or language domain, and language-external factors such as language dominance (see §3.2.3), societal factors may also play a determinant role in modulating the effects of crosslinguistic interaction. Systematic and prolonged language contact situations may result in diachronic processes of language change (e.g., Muysken, 2013; Silva-Corvalán, 1994, 2008; Thomason, 2001, 2020; Thomason & Kaufman, 1988).

Language contact, as a driver of language change, has been identified as giving rise to two main outcomes: language convergence or transfer. Firstly, language convergence has been understood as a process by which the structural distance in a certain paradigm of the grammar between two languages in contact is reduced, a process through which the two languages in contact influence each other mutually (Heine & Kuteva, 2005) (see also Blas Arroyo, 2015; Grant, 2020; Sánchez, 2003, 2004). More specifically, Bullock and Toribio (2004, p. 91) refer to convergence as "the enhancement of inherent structural

similarities found between two language systems". Secondly, another possible result of language contact is language transfer, which implies unidirectionality and "the imposition of a structural property from a foreign source language" (Bullock & Toribio, 2004, p. 91; see also Matras, 2009; Thomason, 2001, 2020; van Coetsem, 1988). Silva-Corvalán (2008, p. 215) identifies the bilingual individual as the locus of transfer. In this regard, she argues that innovative patterns first permeate the grammar of the individual, then spread across individuals, and may eventually lead to syntactic change. The linguistic outcomes of language change, however, can be explained in many cases by externally motivated causes in combination with the internal evolution of a language (see Heine & Kuteva, 2005; Thomason, 2020; among others). Ultimately, language contact can result in the emergence of contact varieties, sometimes characteristic of a specific community (e.g., Boix-Fuster & Sanz, 2008; Bullock & Gerfen, 2004; Otheguy et al., 2007; Silva-Corvalán, 1994, 2008).

From a different perspective, and as a third possible outcome of language contact, Enrique-Arias (2010) has argued that language contact can act as an inhibitor of language change. More specifically, he defends that when there is a change in progress in a monolingual variety, the existence of a parallel structure in the contact language of bilingual speakers may favor the retention of the traditional feature in the contact variety. As such, language contact would slow down a change that is taking place in non-contact varieties (see also Enrique-Arias, 2019; Enrique-Arias & Méndez Guerrero, 2020).

Investigating social factors or studying language change as a result of language contact at the societal level falls outside of the scope of the present thesis. Rather, it is our aim to investigate the effects of language dominance at the individual level on a specific linguistic phenomenon, the interpretation of null and overt pronouns in Catalan and in Spanish. However, individual bilingualism is nondetachable from societal bilingualism and language contact. Thus, we must keep in mind that bilingual participants in the present study have grown up and live in a bilingual setting, Catalonia.

3.3 Bilingualism and language contact in Catalonia

3.3.1 Contextualizing bilingualism and language contact in Catalonia

The location of the present study is Catalonia, a bilingual autonomous community in the northeast of Spain. Catalonia is a multilingual society with two co-official languages, Catalan and Spanish¹⁸, that have coexisted for centuries and are nowadays widely used in everyday life. Despite the historical language contact situation, however, the main demolingistic changes in the knowledge and use of Catalan and Spanish took place after the 19th century (see Vila, 2020a), bilingualism not being widespread until the second half of the 20th century (Argenter, 2020, p. 603). Indeed, the status of Catalan and Spanish in Catalonia—as well as in the Catalan-speaking territories—has been strongly influenced by governmental policies and social changes throughout recent history (for recent reviews, see Argenter, 2020; Flors-Mas et al., 2021; Soler-Carbonell et al., 2016; Soto-Corominas, 2018; Vila, 2020a and references within).

Under the Francoist dictatorship (1939-1975), the public use of Catalan was prohibited, which led to the use of Catalan being restricted to the private sphere. It was not until the approval of the Spanish Constitution (1978) and the Statute of Autonomy of Catalonia (1979) that Catalan was established as a co-official language in Catalonia. After the democratic transition, and with a considerable political and societal consensus, the regional government (Generalitat de Catalunya) developed several language policies to support and normalize the use of Catalan. These measures allowed for a successful revitalization of Catalan in Catalonia (e.g., Fishman, 1991; Strubell & Boix-Fuster, 2011; Woolard, 2016). Together with the emergence of both public and private mass media broadcasting in Catalan, a very successful policy that came with broad societal support was establishing the use of Catalan as the vehicular language in schools (e.g., Strubell, 1996; Vila, 2020b).

Nowadays, and in the current global context, these linguistic policies that had contributed to securing the status and the continuity of the everyday use of the Catalan language are facing several challenges (e.g., Soler-Carbonell et al., 2016; Woolard & Frekko, 2013). Crucially, the increase in the linguistic and cultural diversity after waves of transnational migration at the beginning of the 21st century reconfigured the Catalan sociolinguistic

¹⁸ Aranese is also a co-official language in Catalonia, although its use is mainly restricted to the Aran region (see Suïls & Huguet, 2001). Catalan Sign Language (LSC) has been recognized by the Catalan Parliament (Llei 17/2010).

scenario (e.g., Flors-Mas et al., 2021; Vila, 2020a), giving rise to new needs regarding language and educational policies (e.g., Corona et al., 2013; Flors-Mas, 2021; Newman et al., 2013; Pujolar, 2020). Also, the use of Catalan, especially in schools, has been threatened by the recent political struggles and acts of judicialization against, for instance, the educational policies of immersion (e.g., Carbonell, 2019; Solsona-Puig et al., 2021).

Referring to the present situation of both Catalan and Spanish in Catalonia, we summarize some of the relevant results of the last Survey on Language Uses of the Population (Idescat & Language Policy Secretariat, 2018)¹⁹. This survey provides an overview of the current sociolinguistic situation of Catalan and Spanish in Catalonia and is therefore relevant for the contextualization of the participants in the present study.

As can be observed in Table 3.1, around half of the population of Catalonia over 15 years of age states that Spanish is their first language (understood as the first language spoken at home), and only a third of the population has Catalan as their first language. As argued in Flors-Mas et al. (2021), the absolute number of speakers having Catalan as their first language has remained stable in recent years (between 2003 and 2018), but it has declined in relative terms. Regarding the population's language of identification and habitual language, the percentages are also lower for Catalan than for Spanish, although both languages are shown to be widely used in Catalonia. Interestingly, few people report having both Catalan and Spanish as first, identification, or habitual languages. It is worth noting, however, that outside the home environment almost half of the population declares using both languages, even though the extent to which speakers use each language varies (Table 3.2). Furthermore, the extent of use of Catalan and Spanish also diverges depending on the region. In the Barcelona metropolitan area, Spanish stands as predominant in terms of first language and habitual language. In the other regions of Catalonia, the number of speakers who habitually use Catalan and who count it as their first language is higher and exceeds the figures for Spanish (except for the region of Tarragona) (Boix-Fuster & Sanz, 2008; Davidson, 2022; Idescat & Language Policy Secretariat, 2018).

¹⁹ A survey on the linguistic uses, knowledge, and identification of the Catalan population aged over 15 years old which has been published every five years since 2003 by Idescat (the Statistical Institute of Catalunya) and the Catalan Ministry of Culture (Departament de Cultura) under the Language Policy Secretariat (Secretaria de Política Lingüística).

Table 3.1*First, identification and habitual language of the population of Catalonia aged 15 years and over (%)*

	Catalan	Spanish	Both	Other
First language	31.5	52.7	2.8	13
Language of identification	36.3	46.6	6.9	10.2
Habitual language	36.1	48.6	7.4	7.9

Note. From the Survey on Language Uses of the Population, by Idescat and Secretaria de Política Lingüística (2018).

Table 3.2*Language uses of the population of Catalonia aged 15 years and over, by area of use (%)*

	Only Catalan	More Catalan than Spanish	Both Catalan and Spanish	More Spanish than Catalan	Only Spanish	Other combinations
At home	27.2	5.6	7.8	10.1	37.7	11.4
With friends	13.4	16.5	18.3	13.1	28.9	9.8
With schoolmates	18.2	18.1	19.1	12.9	19.9	11.7
With workmates	15.3	15.1	20.2	13.2	26.0	10.0

Note. From the Survey on Language Uses of the Population, by Idescat and Secretaria de Política Lingüística (2018).

Regarding language competence, although 94.4% of the population of Catalonia reports being able to understand Catalan, only 64.7% consider themselves to have knowledge of Catalan in all skills (oral and written comprehension and production). For Spanish, the percentages are higher, 97.5% of the population reporting being competent in all language abilities. These figures, however, refer to the whole population of Catalonia aged 15 or more. When looking specifically at the age group between 15 and 29 years old—the one that encompasses the vast majority of our participants—their declared knowledge of Catalan in all competencies rises to 88.4%, and peaks at 99% in Spanish.

Self-assessed knowledge of Catalan and Spanish is also high for both languages (see Table 3.3). However, knowledge of Spanish still receives higher scores than that of Catalan. In this regard, the results of several tests at the end of primary and secondary education, as well as the results of the exams to access university, show that the Catalan

educational system guarantees an equivalent and effective competence of Catalan and Spanish (Consell Superior d'Avaluació del Sistema Educatiu, 2017, 2021b, see also 2021a for a sociodemographic and sociolinguistic longitudinal study on the students in Catalan secondary schools). Overall, we will consider that the participants in this thesis, early Catalan-Spanish bilinguals who have been educated in Catalan schools, are functional bilinguals and highly proficient in their two languages.

Table 3.3

Mean self-assessed knowledge of Catalan and Spanish (from 0 to 10) by the population aged between 15 and 29 years

	Catalan	Spanish
Understanding	9.1	9.8
Speaking	8.4	9.6
Reading	8.9	9.7
Writing	8.1	9.3

Note. From the Survey on Language Uses of the Population, by Idescat and Secretaria de Política Lingüística (2018).

3.3.2 Previous studies on language dominance in bilingual Catalan and bilingual Spanish

In the bilingual context of Catalonia, societal language contact and individual bilingualism have led to the emergence of crosslinguistic influence (CLI) or effects of bilingualism in bilingual Catalan and bilingual Spanish (e.g., Bel & García-Alcaraz, 2018; Cuza & Guijarro-Fuentes, 2018; Jiménez-Gaspar et al., 2020a, 2020b). Furthermore, several experimental studies on adult Catalan-Spanish early bilinguals have attested robust effects of language dominance on both bilingual Catalan and bilingual Spanish, and across different linguistic domains (e.g., Arnaus Gil, 2021; Perpiñán, 2017, 2018; Perpiñán & Soto-Corominas, 2021; Puig-Mayenco et al., 2018; Soto-Corominas, 2018, 2021).

At the phonological level, for instance, a considerable number of studies have found language dominance to determine the behavior of Catalan-Spanish bilinguals. Spanish-dominant bilinguals have consistently been found to be less accurate in the production and discrimination of Catalan phonemic contrasts that do not exist in Spanish, such as mid-vowel contrasts (/e/-/ɛ/, /o/-/ɔ/) or the /ʎ/-/ɜ/ contrast (e.g., Mora & Nadeu, 2012;

Navarra et al., 2005; Pallier et al., 2001; Ramírez & Simonet, 2018; Simonet, 2011, 2014). Although research into bilingual Spanish has been more limited than that into bilingual Catalan, language dominance also affects bilingual Spanish. For instance, it has been proved to influence the production of lateral velarization (/l/) (Davidson, 2022).

In contrast with this very fruitful research on the phonological domain, fewer studies have been concerned with the effects of language dominance on morpho-syntactic, lexico-syntactic, or syntactico-pragmatic linguistic phenomena. In these domains involving syntax, as explored in the following paragraphs, language dominance has generally been attested to affect the performance of Catalan-Spanish bilinguals (e.g., Perpiñán, 2017, 2018; Perpiñán & Soto-Corominas, 2021; Puig-Mayenco et al., 2018). However, in other studies on different linguistic phenomena no evidence of language dominance effects has been attested, although this lack of effects has been explained by different reasons (e.g., Arnaus Gil, 2021; Perpiñán, 2018; Puig-Mayenco et al., 2018). Hence, the role of language dominance, as well as the consequences of bilingualism or language contact, may be different depending on the linguistic phenomenon under study.

Referring to bilingual Catalan in contact with Spanish, language dominance has been found to affect the processing and production of DOM (differential object marking) (Benito & Bel, 2022; Jiménez-Gaspar, 2020; Perpiñán, 2018; Puig-Mayenco et al., 2018), the use of *estar* in locative and existential constructions (Perpiñán & Soto-Corominas, 2021), the acceptability and production of the non-personal pronominal clitics *en* and *hi* (Perpiñán, 2017), and the use of the indefinite accusative clitic *ho* (Perpiñán, 2018). In general, Catalan-dominant and Spanish-dominant bilinguals tend to show different behaviors in Catalan, whereas balanced bilinguals fall between the two groups. Although they sometimes align with Catalan-dominant speakers, they more often match Spanish-dominant bilinguals. These differences between groups are generally compatible with the influence of Spanish on Catalan, mainly observed in the Catalan of Spanish-dominant bilinguals and, to a lesser extent, in that of balanced bilinguals. Thus, language dominance has been found to play a role in modulating the occurrence of CLI between Catalan and Spanish.

CLI, however, has not been the only explanation for the effects of language dominance in the aforementioned studies. Clitic omission (and the acceptance of clitic omission) of obligatory non-personal clitics by balanced and Spanish-dominant bilinguals in Catalan has also been attributed to variability in the input (Perpiñán, 2017; see also Soto-Corominas, 2018; Tarrés Larrègola, 2021). The expansion of the use of DOM by all

groups of bilinguals has been related to the internal evolution of Catalan, which may be accelerated by language contact (Perpiñán, 2018). Regarding existential predicate selection, Perpiñán and Soto-Corominas (2021, p. 1488) identified *indirect* CLI. In this case, CLI would surface in unexpected contexts in which bilinguals *perceive* a certain overlap of uses in two structures. Overall, Perpiñán and Soto-Corominas (2021, p. 1490), building on Perpiñán (2017), speculate on the existence of a bilingualism continuum with the possibility of a divergent production, understood as the reflection of a divergent mental representation.

On the other hand, there are studies that have found no evidence of effects of language dominance in bilingual Catalan. In some cases, this lack of effect has been explained as bilinguals having distinct representations for Catalan and Spanish not influenced by language contact, as in Puig-Mayenco et al. (2018) for the acceptance and processing of negative concord items²⁰. In other cases, the absence of a determining role of language dominance has been attributed to bilinguals having different representations from the expected ones for the standard variety and, hence, indicative of language change (Arnaus Gil, 2021 for locative copula in Catalan; Perpiñán, 2018 for the acceptance of DOM in Catalan). More specifically, Perpiñán (2018) found all groups of bilinguals to similarly accept the ungrammatical presence, as well as the grammatical absence, of DOM in Catalan. This generalized optionality in the bilingual behavior was also attested by Arnaus Gil (2021) in her study on copula selection in locative predicates with non-eventive subjects in Catalan, in which both *ser* and *estar* were similarly preferred across groups²¹. Both Perpiñán and Arnaus Gil interpreted their results as indicative of language change. In contrast to other constructions, in this case, language contact may have resulted in a new bilingual variety that has already been spread across bilinguals, explaining why no effects of dominance are observed. Similarly, in a study on DOM in bilingual Catalan in contact with Spanish, in which language dominance was not assessed, Benito (2020) suggests that a Catalan variety showing optionality in DOM may exist, either as a result of prolonged language contact with Spanish, of an internal linguistic process, or of a convergence of both processes.

²⁰ In their study, Puig-Mayenco et al. (2018) investigated the co-occurrence of sentential negation with a negative concord item in pre-verbal position (allowed in Catalan, but not possible in Spanish).

²¹ Although *ser* should be the preferred and more natural option, both *ser* and *estar* are possible in Catalan, whereas only *estar* is allowed in Spanish

Referring to Spanish in contact with Catalan, effects of language dominance interpreted as resulting in CLI from Catalan to Spanish have also been attested, although the studies are scarce. Influence from Catalan was attested in the acceptability of DOM in Spanish across bilinguals and in the processing of DOM in the case of Catalan-dominant bilinguals (Puig-Mayenco et al., 2018), and participants that had been more exposed to Catalan also showed influence from Catalan in the production of DOM in Spanish (Jiménez-Gaspar, 2020). Similarly, effects of language dominance have been found in pronominal and partitive clitics, both in acceptability and in production data, with Catalan-dominant speakers the bilinguals who present more influence from Catalan (Soto-Corominas, 2018, 2021 for child acquisition of clitics in Catalanian Spanish; see also Jiménez-Gaspar et al., 2020a for clitics in Majorcan Spanish). In studies on language contact, the influence between languages has not always been found to lead to contact-driven innovations, but also to the preservation of structures that would have otherwise disappeared. For instance, some syntactic features that were recessive in monolingual Spanish, such as the deontic periphrasis *haber de* + infinitive, have been reinforced in the bilingual variety of Spanish in contact with Catalan by the existence of a parallel Catalan structure (e.g., Blas Arroyo, 2007, 2015; Enrique-Arias, 2010, 2014; Garachana, 2018, 2021). Again, the influence from Catalan into Spanish has been proposed to be higher as more Catalan-dominant the group is, and where Catalan is the environmental majority language (Blas Arroyo, 2007).

We will now refer to the few studies that have contrasted the effects of language dominance on both bilingual Catalan and bilingual Spanish bidirectionally. Especially relevant is the work by Puig-Mayenco et al. (2018), who studied sentential negation with a negative concord item in preverbal position and differential object marking (DOM). As previously reported, no effects of language dominance were attested regarding the first phenomenon, but they did affect DOM. The over-acceptance of ungrammatical sentences in Catalan by Spanish-dominant bilinguals and the optionality in the Spanish grammar of Catalan-dominant bilinguals were interpreted as Spanish influencing the Catalan of Spanish-dominant bilinguals and as Catalan influencing the Spanish of Catalan-dominant bilinguals. Similar findings were obtained by Soto-Corominas (2021) when studying child acquisition of the quantitative clitic *en*. Catalan-dominant bilinguals—and balanced bilinguals, to a lesser extent—were found to use this clitic in Spanish (ungrammatical). This ungrammatical use was interpreted as an effect of CLI from Catalan, their dominant language, on their non-dominant language. On the other hand, the ungrammatical omissions of the clitic *en* in Catalan by balanced and Spanish-

dominant bilinguals were partially attributed to CLI from Spanish. However, in this case the role of CLI is not as straightforward since there are other factors, such as input variability, that could also account for optionality. In summary, evidence of language dominance determining the directionality of CLI has been found in bidirectional studies (e.g., Puig-Mayenco et al., 2018; Soto-Corominas, 2021).

Finally, it is worth mentioning the work by Boix-Fuster and Sanz (2008), also a bidirectional study. Although they did not directly assess language dominance, they compared oral narratives in Catalan and in Spanish by Catalan-Spanish bilinguals with three different profiles (from Catalan-speaking families, from Spanish-speaking families, and from bilingual families). In their data from Catalan, they identified transfer across linguistic domains (i.e., lexical, morphosyntactic, and phonological), using qualitative measures. This transfer was attested to a varying degree depending on the bilingual's profile. These findings led them to suggest a distinction between a Catalan variety used by bilinguals raised in Catalan-speaking families, and a Catalan variety that would characterize the other two groups and that seems to present more transfer from Spanish. Regarding their data in Spanish, they observed the use of a bilingual Spanish variety that would be shared by all bilinguals irrespective of their background profile and that would not be characterized by transfer from Catalan. In light of this evidence, they proposed the existence of a common variety of Spanish that is spoken in Catalonia. Overall, Boix-Fuster and Sanz interpreted their findings as showing a more prevalent influence from the 'majority language', Spanish, on the 'minority language', Catalan²².

In a similar vein, from a language contact perspective, Davidson (2020) studied the directionality of CLI in the production of intervocalic fricatives in Catalan and Spanish in Catalan-Spanish bilinguals from two communities: Barcelona and Valencia. He found that CLI was bidirectional and asymmetric. He attested CLI from the "L1" towards the "L2" of the bilinguals, and he found CLI from Catalan to Spanish to be stronger in Barcelona and CLI from Spanish to Catalan to be stronger in Valencia. Davidson attributed these asymmetries in language contact outcomes to the role of social factors (i.e., the asymmetric sociopolitical and sociolinguistic relationships between the two languages in each community), and not specifically to language dominance.

²² In a study on Persian-English bilinguals, Foroodi-Nejad and Paradis (2009, p. 426) also speculated on the role of the social status of a language, suggesting that "the impact of dominance might be different depending on the majority–minority status of the languages".

All in all, a variety of non-linguistic factors have been proposed to explain language outcomes in bilingual situations, although the debate is still open as to whether language dominance influences the target language system. Thus, in the present study, the role of language dominance on Catalan-Spanish bilinguals will be assessed in the study of a syntactico-pragmatic phenomenon, pronominal anaphora resolution. We will use a bidirectional design so that the same phenomenon is scrutinized in the bilinguals' two languages. If language dominance effects emerge, they will be attributable to different bilingual profiles and not to specific linguistic properties of the studied phenomena. We therefore expect to be able to provide further evidence that contributes to establishing the role that language dominance may play in modulating the effects of bilingualism in different linguistic domains in bilingual Catalan and in bilingual Spanish.

Finally, it is worth bearing in mind that the abovementioned studies have conceptualized and measured language dominance differently. In the following section (§3.3.3), we review the different manners in which this construct has been measured in the study of Catalan-Spanish bilinguals.

3.3.3 Assessing language dominance in adult Catalan-Spanish bilinguals

As previously mentioned, there is no single definition of language dominance and no widely accepted method for operationalizing and measuring it in bilingual populations (see Silva-Corvalán & Treffers-Daller, 2016). Also, no comprehensive questionnaire has been designed specifically targeting bilinguals born and raised in a bilingual community such as Catalonia. In this community, Catalan and Spanish are both official languages, are used and taught along the whole schooling, and are present—to a greater or lesser extent—in the life of any person that has been raised and educated in it (see §3.3.1)²³. In previous research accounting for language dominance in Catalan-Spanish bilingual participants, dominance has been understood in different manners and has been measured using a variety of tools and methods.

Many of the studies on Catalan-Spanish bilinguals assessing language dominance have used language background questionnaires designed ad hoc (e.g. Arnaus Gil, 2021; Ferré

²³ As discussed in studies such as Perpiñán (2017) and Puig-Mayenco et al. (2018), correlating language proficiency and language dominance in the context of Catalan-Spanish bilingualism might not prove useful. Young adults that have attended the Catalan education system, as is the case with the participants in the present study (mainly university students), have been educated in both Catalan and Spanish and they have needed to prove that they have a high competence in both languages to access university (see also §3.3.1).

& Brysbaert, 2017; Perpiñán, 2017, 2018; Perpiñán & Soto-Corominas, 2021), and sometimes administrated as oral interviews (e.g., de Prada Pérez, 2009). Other studies have resorted to widely-used available questionnaires such as the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007), as in Puig-Mayenco et al. (2018), or the Bilingual Language Profile questionnaire (BLP; Birdsong et al., 2012), which has been more and more used in recent studies on Catalan-Spanish bilinguals (e.g., Amengual, 2016; Puig-Mayenco et al., 2020; Ramírez & Simonet, 2018; Renwick & Nadeu, 2019; Simonet, 2014; Timmer et al., 2021). In all these questionnaires, language dominance was understood as multidimensional and different variables were taken into account (referring to concepts such as the first and/or native language, exposure to each language within the family, the educational and the social environment, the current use of each language in different contexts, language identification or preference, among others). On the other hand, there are studies that give few details on how participants were classified into Catalan-dominant and Spanish-dominant groups or describe them as essentially differing in a single component of language dominance. For instance, the dominant language has been defined as the language of the family or environment in which bilinguals were raised until the ages of 3 or 4 years old (e.g., Navarra et al., 2005; Pallier et al., 2001), or as a combination of the bilinguals' first language and habitual language (Blas Arroyo, 2007).

In the present study, we will use the BLP (Birdsong et al., 2012) to measure language dominance. This questionnaire was created to assess and reliably operationalize language dominance in bilinguals who have been raised in a variety of bilingual situations. As will be described in §4.2.2, it assesses four modules (language history, use, proficiency, and attitudes) and provides a continuous measure of language dominance. The main argument for using the BLP is that it does not focus on a single index but on different dimensions of bilingual experience, while providing a continuous measure instead of treating bilingualism as a categorical variable. Hence, it conceives language dominance as a multifaceted and gradient construct, which has been claimed to be preferable when representing the bilingual experience (Luk & Bialystok, 2013). At the same time, obtaining a continuous measure based on multiple dimensions of bilingualism facilitates the classifying of the bilinguals into groups if needed, as will be the case in the present study. Furthermore, given that the BLP is becoming widely used in bilingualism research, using this tool will allow our results to be more readily comparable to other experimental studies on different bilingual populations and language

combinations. Finally, the fact that the BLP is a self-reported questionnaire made it an easy-to-use and efficient tool for the data collection process.

The BLP questionnaire, however, has some drawbacks as well. Its main limitation in relation to the present study, for instance, is that it has not been specifically designed to characterize bilinguals in a bilingual society such as Catalonia, and maybe not in bilingual societies in general. Even though there is a Catalan version of this questionnaire, it is a mere translation of the English version that has not been adapted in any way to the specific bilingual Catalan-Spanish context. Instead, the authors aimed precisely to create a single questionnaire that targeted bilinguals from contexts as varied as L2 acquisition, heritage learning, attrition, migration contexts, and sequential or simultaneous bilingualism (Gertken et al., 2014, p. 221). This generalization facilitates comparability across studies but has important limitations given that it does not consider the specificities of the context surrounding the bilinguals, which may not be comparable. To overcome this limitation, some complementary questions were added in the language background questionnaire and some of the questions in the BLP were modified slightly, with minimal adaptations that affected neither the essential target content of the questions nor the number of questions or scoring of the results. Although no widely-used questionnaire has been designed to assess language dominance in bilinguals born and raised in bilingual communities, Perpiñán and Soto-Corominas (2021) recently attempted to fill this gap by creating the Language Dominance Questionnaire (LaDoQ). Similar to the BLP, the answers to this questionnaire are transformed into numerical values to create a scale of bilingualism that facilitates the identification of the language dominance profile of bilinguals by using a gradient and compound measure. The language background questionnaire that our bilingual participants completed—the BLP, with some adapted questions and some additional questions—will be described in detail in §4.2.2.

Finally, some authors complement self-reports such as the ones obtained through the BLP with other, more objective, performance measures in order to group and characterize bilinguals more reliably (e.g., de Bruin et al., 2017 for Basque-Spanish bilinguals). This seems to be an interesting approach. However, very few tests are available and, importantly, they are restricted to very few languages (see Treffers-Daller, 2016). Therefore, we assessed the possibility of using a picture-naming test (e.g., the MINT test, Gollan et al., 2012) or a vocabulary recognition test (e.g., the LexTale test, Lemhöfer & Broersma, 2012). However, these tasks refer to lexical measures. Ferré and

Brysbaert (2017) demonstrated that the performance in the LexTale-ESP (Izura et al., 2014) successfully discerns between Catalan-dominant and Spanish-dominant bilinguals, but it remains unclear how more balanced bilinguals—a profile in which we are also interested—would perform in it. Furthermore, there is no Catalan version of the LexTale, so in a bidirectional study like the present one, we would have only gained access to half of the picture. For these reasons, we finally decided not to use multiple tasks and to base our measure of language dominance on the self-reports in the BLP, which we consider to be enough for the purpose of the present study. In fact, the BLP does not target a single domain (e.g., proficiency) but instead refers to multiple dimensions of language dominance.

As a final remark, we could mention the fact that Bonvin et al. (2021) revealed that the metrics of BLP dominance and the LexTale lexical indices show a strong linear association. Another interesting finding in this study is that balanced bilingualism, according to the BLP, was found to be different from balanced proficiency according to lexical decision tasks. Given that our main interest is to analyze how language dominance—and not lexical measures—modulates anaphora resolution, we consider the BLP to be the most appropriate tool with which to assess language dominance according to the needs of the present study, even though it was not designed to test bilinguals raised in bilingual societies.

3.4 The effects of bilingualism on anaphora resolution in null subject languages

3.4.1 Empirical findings on bilingual anaphora resolution in null subject languages

The linguistic phenomenon under study in the present thesis, pronominal subject anaphora resolution, has been identified as a vulnerable domain in bilingual language acquisition (see Sorace, 2011 for an overview). Probably, the most recurrent finding across bilingual populations has been the attesting of non-target-like interpretations of overt subject pronouns. Overt pronouns have shown optionality or indeterminacy in their associations with subject and object antecedents, irrespective of L1-L2 similarities and even in highly proficient bilinguals. These patterns have been identified for near-native L2 speakers with a non-null subject language as an L1 (e.g., for L2-Italian L1-English: Belletti et al., 2007; Sorace & Filiaci, 2006; Sorace & Serratrice, 2009), L2 speakers of

combinations of null subject languages (e.g., Bel & García-Alcaraz, 2015 for L2-Spanish L1-Moroccan Arabic; Bini, 1993 for L2-Italian L1-Spanish; Georgopoulos, 2017 for L2-Spanish L1-Greek and L1-English; Margaza & Bel, 2006 for L2-Spanish L1-Greek; Sorace & Serratrice, 2009 for L2-Italian L1-Spanish), L1 attrited speakers (e.g., Kaltsa et al., 2015 for L1-Greek L2-Swedish; Tsimpli et al., 2004 for L1-Italian L2-English), heritage speakers (e.g., Kaltsa et al., 2015 for Greek heritage speakers living in Sweden; Keating et al., 2016 for Spanish heritage speakers living in the US), and early bilinguals (Bel, García-Alcaraz, et al., 2016 for Moroccan Arabic-Spanish bilinguals in Spanish; de Rocafiguera, 2017 for Basque-Spanish bilinguals in Spanish). This overextension of the properties of overt pronouns has not only been attested in interpretive data, but an overuse of overt pronouns has also been found in L2 production data (e.g., for L2-Catalan Sign Language L1-Catalan: Bel et al., 2015; for L2-Italian L1-Greek: Di Domenico et al., 2020; for L2-Spanish L1-Greek: Lozano, 2009, 2016; Margaza & Bel, 2006; for L2-Spanish L1-English: Martín-Villena & Lozano, 2020; for L2-Basque L1-Spanish: Rodríguez-Ordóñez & Sainzmaza-Lecanda, 2018). In most of the aforementioned studies, the interpretation of null pronouns as coreferential with subject antecedents has been widely replicated, whereas the acquisition of overt pronouns has been found to be more vulnerable and flexible. This overextension and higher vulnerability of the interpretation and use of overt pronouns has been explained, or partially explained, by referring to the Interface Hypothesis (Sorace, 2011, 2012; Sorace & Filiaci, 2006; see §3.4.2).

Other studies, however, have not found these asymmetries in the acquisition of the properties of null and overt pronouns. For instance, some authors have questioned the fact that null pronouns are seen as unproblematic in acquisition or not as pragmatically complex as overt pronouns (Clements & Domínguez, 2017; Teixeira et al., 2022). On the one hand, non-native-like uses of null subjects have also been attested in production and corpora data (Domínguez, 2013 for an overview; García-Alcaraz & Bel, 2019 for Moroccan Arabic-Spanish early bilinguals; Lozano, 2016; Margaza & Bel, 2006 for L1-Greek L2-Spanish; Montrul & Rodríguez-Louro, 2006 for L1-English L2-Spanish). On the other hand, it has been demonstrated that the syntax-pragmatic interpretive constraints of both null and overt pronouns can be acquired by L2 speakers with advanced levels of proficiency (Bel, Sagarra, et al., 2016 for L2-Spanish L1-Moroccan Arabic or L1-English; Clements & Domínguez, 2017 for L2-Spanish L1-English; Kraš, 2008, 2016 for L1-Croatian L2-Italian; Rothman, 2009 for L2-Spanish L1-English) and by child heritage speakers (Rinke & Flores, 2018 for heritage speakers of European Portuguese with

German or Spanish/Catalan as environmental languages). Thus, target-like behaviors have also been identified in the interpretive patterns of both null and overt pronouns by bilingual populations.

To our knowledge, few studies have addressed bilingual anaphora resolution by early bilinguals raised in bilingual societies. An overextension in the interpretive biases of overt pronouns in Spanish was found in Basque-Spanish bilinguals in the Basque Country but not in Catalan-Spanish bilinguals in Catalonia, who showed clear PAH-like biases (de Rocafiguera, 2017; de Rocafiguera & Bel, 2019). Bel and García-Alcaraz (2018) found bilingual Spanish by Catalan-Spanish bilinguals to differ from monolingual Spanish, suggesting the presence of bilingualism effects that are different from optionality, such as CLI. Rodríguez-Ordóñez and Sainzmaza-Lecanda (2018) studied contact effects on pronoun expression in Basque by early bilinguals and L2-Basque speakers of L1-Spanish. They found that L2 speakers (also referred to as Spanish-dominant speakers) 1) overproduce overt pronominal subjects compared to natives, and 2) transfer their Spanish system to Basque, showing the same patterns in Basque as monolingual speakers of Spanish in Spanish. Early bilinguals often showed strong similarities with Basque native speakers. Finally, Iraola (2014) found a delay in children in the acquisition of the properties of overt subject pronouns by cL1 and cL2 Basque-Spanish bilinguals when compared with adult grammar.

Overall, the results on bilingual interpretation and the use of pronominal subjects across different populations show a mixed picture and have been attributed to several different reasons. It is generally agreed that the anaphoric referential properties of null and overt pronouns are challenging to acquire and that persistent deficits may be attested regarding certain pragmatic functions, especially those of overt pronouns. Within the framework of the Interface Hypothesis (see §3.4.2 below), Sorace and Serratrice (2009, p. 198), mention several non-mutually-exclusive factors that can make these structures particularly challenging for bilingual speakers: underspecification in grammatical representations, increased competition for processing resources, crosslinguistic influence (CLI), quality and quantity of input, and executive control limitations related to language coactivation. As further developed in the sections to follow, these explanations can mainly be summarized as referring to general effects of bilingualism (§3.4.2)—non-dependent on the language combination—or to CLI (§3.4.3)—involving differences between the bilingual's two languages.

All the reviewed studies on null subject languages in the present section have only analyzed canonical contexts. In §3.4.4, we will refer to the few existing studies that have tested anaphora resolution in non-canonical structures by bilingual populations (in L2-French and L2-German). In the present study, we will further explore anaphora resolution in highly proficient early bilinguals. At the same time, we will also analyze how these bilinguals integrate and use different types of cues in marked information contexts.

3.4.2 The Interface Hypothesis and general effects of bilingualism on anaphora resolution

The most influential hypothesis to explain optionality in the interpretation of overt pronouns has been the Interface Hypothesis (IH) (Sorace, 2011, 2012; Sorace & Filiaci, 2006). This hypothesis predicts that language structures involving an interface between syntax and other cognitive domains, such as pragmatics, are a locus of permanent optionality and may be challenging for bilinguals due to the difficulty of integrating information from different linguistic domains. The IH was first proposed by Sorace and Filiaci (2006), building on previous studies that had found patterns of residual first language effects, indeterminacy, or optionality involving interface phenomena on non-native grammars (Sorace, 2000, 2005). Based on these observations, the IH has been defined as "the hypothesis that narrow syntactic properties are completely acquirable in a second language, even though they may exhibit significant developmental delays, whereas interface properties involving syntax and another cognitive domain may not be fully acquirable" (Sorace & Filiaci, 2006, p. 340). The IH was first proposed to account for non-native-like anaphora resolution in L2 language acquisition at advanced proficiency levels. In light of the studies reported in §3.4.1, it was afterward extended to L1 language attrition, heritage language acquisition, and bilingual first language acquisition (although few studies have investigated adult early bilinguals in bilingual settings).

Sorace and Filiaci (2006) suggest two main explanations for the causes of the optionality or instability of overt pronouns' interpretation attested in bilingual populations: the first one refers to underspecification at the level of knowledge representations, and the second one to less efficient processing by bilinguals, which is related to the increased cognitive load required to integrate information from different domains in structures at the syntax-pragmatics interface. Sorace (2011), distilling the major findings on IH up until

that point, labels these two explanations as the *representational account* and the *processing resources account*.

The *representational account* relates optionality to the underspecification of interpretable features that are linked to different parametric choices within the bilinguals' L1 and L2. This account is based on Tsimpli et al. (2004), who demonstrated attrition effects in L1 Greek and L1 Italian under the influence of L2 English in the distribution and interpretation of overt pronominal subjects, regulated by pragmatic constraints such as the interpretable features of [topic-shift] and [focus]. This proposal follows on from the Interpretability Hypothesis, according to which uninterpretable features not present in the L1 are insurmountable in L2, whereas interpretable ones can be acquired (and resetted) (see also Prentza & Tsimpli, 2013; Tsimpli & Mastropavlou, 2007).

In Tsimpli et al. (2004), attrited speakers allowed more interpretations of overt pronouns conveying topic continuity than their monolingual counterparts. The representational account explains that these findings are due to one of the grammatical systems influencing the representations of the other system. More specifically, it predicts that the language with more "economical" grammar—with the less complex pronominal system or, in terms of Sorace (2011), with the less restrictive option—will be the one to influence the other language. Thus, the non-null subject language is predicted to affect the null subject language. This influence is assumed to be unidirectional, regardless of language dominance (Sorace, 2004; Tsimpli et al., 2004).

A limitation of the representational account is that it does not suffice to explain why optionality has also been attested in speakers of two null subject languages, which do not differ in the null subject parameter (e.g., Bel & García-Alcaraz, 2015; e.g., Bini, 1993; Lozano, 2006; Margaza & Bel, 2006; Serratrice et al., 2009; Sorace & Serratrice, 2009). Although Sorace and Serratrice (2009) and Sorace (2011) recognize the possibility that null subject languages are not identical regarding the scope of their overt pronouns' interpretation, null subject languages exhibit the same parametric values and share the same grammatical representations. For this reason, it has been claimed that this account does not explain anaphora resolution across language combinations. In view of this limitation, the main explanation to the optionality of overt pronouns in the literature was recast as referring to general effects of bilingualism related to processing and cognitive resources, without excluding the possibility of crosslinguistic influence (CLI). Sorace and Serratrice (2009, p. 198) propose that CLI can occur "in representations and/or parsing

strategies", suggesting that it may not take place at the level of grammar but at the level of disambiguation strategies. However, this possibility is not further developed.

The *processing resources account* relies on processing differences related to bilingualism itself, regardless of the combination of the bilingual's languages. Under this account, the overextension of overt pronouns is explained as a default strategy of bilinguals. As argued by Sorace (2011, 2016), bilinguals could rely on this default strategy to relieve language-processing demands during the real-time computation of syntactic and pragmatic mappings, a process that is consuming in terms of cognitive load and increases competition for resources in bilingual populations (see also §3.2.3.1). Interestingly, Sorace (2016) hypothesizes that an increased inhibitory control of bilingual populations could form part of a trade-off relationship with a less efficient ability to integrate information from different sources. Although these effects refer to real-time computation and stronger effects may be observed using online experimental tasks, the processing resources account can also be tested in comprehension or acceptability tasks.

In terms of the processing account, the amount of exposure (or the age of onset of acquisition) and/or language dominance may affect bilinguals' processing abilities differently. As argued by Sorace (2016), inhibiting their more dominant language may require more resources than inhibiting their less dominant language. These asymmetric inhibition effects could explain differences on the extent to which bilinguals rely on the default strategy to overextend overt pronouns. Similarly, early bilinguals may differ from late bilinguals, or balanced bilinguals may differ from unbalanced bilinguals (see Costa & Santesteban, 2004).

The underspecification of grammatical representations and the increased competition for cognitive resources in bilingual processing, factors underpinning the so-called representational and processing accounts, have not been the only factors proposed in the literature to account for bilingual anaphora resolution. Language input or exposure have been shown to have a significant impact on pronominal patterns, at least in bilingual L1 acquisition (Argyri & Sorace, 2007; Serratrice et al., 2009; Sorace et al., 2009). The low frequency of overt pronouns in the input could also be crucial in this regard (Iraola Azpiroz, 2014; see also Iraola Azpiroz et al., 2017). However, little research has focused on the role of variables related to language experience in pronoun interpretation. By analyzing bilinguals with different language dominance profiles reflecting different bilingual experiences, especially regarding language exposure and use, we will analyze

the extent to which the predictions of the IH are attested in early bilingual populations—either explained through the representational or through the processing accounts.

3.4.3 Crosslinguistic influence in anaphora resolution beyond the Interface Hypothesis

Other recent approaches to pronominal anaphora in language contact situations, such as the one proposed by Romano (2019), have subjected the variability in the distribution of pronominal subjects by bilinguals to crosslinguistic influence (CLI). Within this framework, CLI is predicted to occur between languages with more flexible and more categorical distributions, not necessarily affecting only language pairs differing in the null subject parameter, and not necessarily affecting only overt pronouns. More specifically, Romano (2019, p. 14) defines CLI as taking place "from the language with more relaxed biases for antecedents towards the language with more categorical ones". However, CLI occurring from the language with more categorical biases towards the language with more flexible biases has also been predicted in the previous literature (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001 see §3.2.3.2).

In his study, Romano (2019) aimed at reinterpreting and comprehensibly accounting for the results in experimental studies of pronoun interpretation in contexts of language contact such as L2 acquisition (Belletti et al., 2007; Jegerski et al., 2011; Keating et al., 2011; Sorace & Filiaci, 2006) and L1 attrition (Kaltsa et al., 2015; Tsimpli et al., 2004). All these studies involve combinations of languages differing in the null subject parameter (pronoun interpretation in Italian, Spanish, and Greek in contact with English or Swedish). Within the representational account (see §3.4.2 above), Sorace (2004, 2011) and Tsimpli et al. (2004) had already proposed that CLI could be taking place from the language with a more economical system (i.e., the non-null subject language) on the language with a more complex system (i.e., the null subject language). Recently, in a bidirectional study, Quesada (2021) also attested evidence of unidirectional influence of English on Spanish by L1 English-L2 Spanish bilinguals, and no influence of Spanish on English by L1 Spanish-L2 English bilinguals. Romano's (2019) proposal extends the possible occurrence of CLI to contexts of microvariation, that is, between languages that share the null subject parameter. This possible outcome was also suggested in Bel and García-Alcaraz (2018) in light of the differences observed between null subject languages (see also García-Alcaraz, 2015, p. 166).

Assuming that interpretive biases in Spanish are more flexible than in Greek and in Italian, Romano suggests that CLI from these more relaxed biases in Spanish could explain the overextension of pronouns attested in previous studies of Greek and Italian in contact with Spanish (Filiaci et al., 2014; Margaza & Bel, 2006; Sorace et al., 2009; Sorace & Serratrice, 2009). All in all, Romano (2019) proposes that CLI can occur between pairs of null subject languages that differ in the flexibility of the interpretive biases of null and overt pronouns. Regarding the directionality of CLI, he states that influence should take place from the language with more flexible biases to the language with more categorical biases (predictably from Spanish to Catalan in the present study, see §2.3.2 and §2.3.3). This would give rise to a more flexible resolution system in the more categorical language in bilingual speakers and, possibly, to referential optionality regarding overt pronouns.

From a different but complementary view, it is worth mentioning the Vulnerability Hypothesis (VH), proposed by de Prada Pérez (2019). The VH establishes a hierarchy of permeability to CLI that predicts that "structures that show variable distributions are permeable while those that exhibit categorical distributions are not" (p. 670). In other words, it predicts that "more variable distributions, where more than one form can be used, are more susceptible to cross-linguistic influence than categorical distributions, where a specific form is used (nearly) exclusively" (p. 689). In contrast to Romano's (2019) proposal, however, the VH does not explicitly make any predictions regarding the directionality of CLI. From the VH, we can interpret that categorical distributions will be less strongly influenced by crosslinguistic differences, whereas variable distributions will be more vulnerable to the effects of language contact²⁴.

In her study, de Prada Pérez (2019) compared the expression of subject pronouns in Spanish by Catalan-Spanish bilinguals from Menorca and by Spanish monolinguals. A Catalan control group of Catalan-dominant speakers was also included. Although two groups of bilinguals were tested—Catalan-dominant and Spanish-dominant bilinguals—no effects of language dominance were highlighted in her results. She contrasted three different variables: speech connectivity (i.e., in same vs. different referent contexts), constrained by pragmatic factors; the presence or absence of verb form ambiguity,

²⁴ We would like to draw attention to the fact that Romano's and de Prada Pérez's proposals are based on data of a different nature (interpretation vs. production) and are framed in different—although compatible—linguistic approaches. Whereas Romano (2019) reinterpreted the role of CLI in previous studies as accounting for bilingual pronoun interpretation within the framework of the PAH, de Prada Pérez (2019) analyzed bilingual subject pronoun expression from a variationist and language-contact approach.

constrained by morphology; and verb type, constrained by lexico-semantics. Bilingual's anaphoric pronoun expression did not differ from monolingual's pronoun use regarding speech connectivity, the variable that showed more categorical distributions in both Catalan and Spanish. Conversely, differences between monolingual and bilingual speakers were attested regarding verb form ambiguity, where pronoun uses showed more variable distributions in Catalan and more categorical distributions in monolingual Spanish.

De Prada Pérez (2019) argues that these results would have not been predicted by the IH, which would have expected monolingual and bilingual speakers to differ mainly in contexts regulated by the syntax-pragmatics interface (i.e., speech connectivity). She proposes that the effects of language contact do not depend on the notion of interfaces—which is not measurable and often controversial—but on how variable or categorical a distribution is in a specific language, based on relative frequency. Instead of interfaces, de Prada Pérez considers that the complexity of a linguistic phenomenon, defined as variability, is a better factor to explain CLI selectivity. Compatible with this view, Rinke & Flores (2018) suggest that variability and complexity explain differences in interpretation between monolingual and bilingual children (heritage speakers), an account that they label as the *conflicting input/complexity hypothesis*²⁵.

Bel and García-Alcaraz (2018), when comparing monolingual Spanish to bilingual Spanish in contact with Catalan, attested a stronger bias of overt pronouns towards object antecedents in bilingual Spanish. This pattern was attested in an acceptability judgements task and was also confirmed by processing data, through a self-paced reading task. To explain this difference between monolingual and bilingual Spanish, the authors hypothesized that CLI could modulate the interpretive patterns of anaphora resolution in bilingual Spanish: the stronger biases displayed by Catalan influencing the weaker biases of overt pronouns in Spanish in contact with Catalan. This is not the directionality of CLI that would be predicted by Romano (2019) but by the VH (de Prada Pérez, 2019). Catalan, the language with a less variable distribution, did not seem to show traces of CLI from Spanish. However, in addition to the explanation based on CLI, Bel and García-Alcaraz (2018) also suggest that language convergence could be taking

²⁵ Rinke and Flores (2018), who investigate heritage speakers of European Portuguese with German or Spanish/Catalan as environmental languages, find the two groups of heritage bilinguals to behave alike. This leads them to dismiss the emergence of CLI, instead arguing that variability in the grammar of bilingual children can be explained by the complexity of the phenomenon and variability in the target system.

place in a language-contact situation, the object bias of overt pronouns in bilingual Spanish being reinforced by contact with Catalan (see §3.2.4).

In sum, Romano (2019) and de Prada Pérez (2019) put forward two testable and mutually exclusive hypotheses that predict different effects of CLI on bilingual anaphora resolution involving two null subject languages. While Romano predicts the influence of the language with less categorical biases upon the language with more categorical biases (i.e., predictably from Spanish to Catalan in the present study), de Prada Pérez predicts that the language with more categorical biases (i.e., Spanish) will be less susceptible to the effects of CLI. Furthermore, a third possibility is that CLI occurs from the dominant language towards the non-dominant language (e.g., Yip & Matthews, 2007). As explained in §3.2.3.2, language-external factors such as language dominance may influence the directionality of CLI, as well as its occurrence or its strength.

Almost no experimental studies on bilingual anaphora resolution in pairs of null subject languages have been designed to specifically address whether CLI emerges in contexts of microvariation (see, however, Giannakou, 2018). Moreover, studies on null subject language pairs have instead assumed no differences between null subject languages and have dismissed explanations related to CLI (Rinke & Flores, 2018; Sorace & Serratrice, 2009). In this regard, Romano (2019) evokes the need for bidirectional studies to reliably conclude whether CLI takes place between two languages. This is, indeed, a problematic aspect for most studies on L2 acquisition or L1 attrition that identify CLI as a source of divergence from native speakers, given that so few studies have tested bilinguals in their two languages.

In the present study, the same participants will be tested in Catalan and in Spanish. Using a bidirectional design will allow us to investigate more accurately the role of CLI in pronoun resolution by early functional Catalan-Spanish bilinguals. Moreover, we will test three groups of Catalan-Spanish bilinguals with different dominance profiles to further address whether language dominance conditions the occurrence of CLI. Importantly, the bilinguals in Bel and García-Alcaraz (2018) were identified as balanced bilinguals. In the present study, we will be able to further explore the comparison between bilingual Catalan and Spanish by studying bilinguals with different profiles to better understand the possible effects of CLI on anaphora resolution.

One of the aims of the present study is to test whether evidence of CLI between Catalan and Spanish is found in anaphora resolution and whether it can be explained by Romano's (2019) predictions. At the same time, we explore whether the VH (de Prada

Pérez, 2019) can be refined and extended to interpretation data—where, to our knowledge, it has not been tested before. Lastly, we are also interested in identifying whether language-external factors (i.e., language dominance) condition the vulnerability of null and overt pronouns due to CLI. These results will be interpreted in conjunction with predictions based on the IH and the processing resources account presented above, in §3.4.2. Therefore, we will contrast whether the effects of bilingualism on anaphora resolution can be better explained by CLI or by bilingualism itself.

3.4.4 Previous studies on bilingual anaphora resolution in non-canonical sentences

The studies that we have reported and reviewed so far have investigated canonical sentences (i.e., informatively unmarked structures). Although research on anaphora resolution in non-canonical contexts is scarce, non-native-like behaviors by L2 learners have also been attested in anaphora resolution in marked information structures. However, we are not aware of any study that has investigated bilingual populations and null subject languages. Schimke and Colonna (2016) analyzed canonical, subject dislocation, and object dislocation structures in native and non-native French. They found L2 French-L1 Turkish bilinguals to rely more on discourse cues (i.e., topicality) when resolving subject pronouns than L1 French natives. Whereas non-native speakers preferred antecedents in the initial topicalized position, French natives relied more on subjecthood regardless of the structure. In another study, Patterson et al. (2017) analyzed focusing structures and also found non-native-like interpretations by L2 German-L1 Russian speakers. In this case, German natives preferred to relate personal subject pronouns with non-focused antecedents ('anti-focus' effect). L2 speakers, on the other hand, interpreted subject pronouns as coreferring with focused antecedents in initial position. Patterson et al. interpreted these findings as indicating that non-native speakers rely more on surface-level cues than on (native-like) information structure cues. This explanation seems to be compatible with Schimke and Colonna's (2016) results.

In the present study, we will be able to assess how early bilingual Catalan-Spanish speakers interpret null and overt pronouns in canonical and informatively marked contexts in their two languages. To date, we are not aware of any study that has contrasted the interpretative preferences surrounding null and overt pronouns in non-canonical contexts by bilingual populations (neither by monolingual populations, except for the few studies reported in §2.4 that have made contributions in this regard). Despite

this, in the previous studies by Schimke and Colonna (2016) and Patterson et al. (2017) optionality or indeterminacy was not attested in preferences surrounding non-native pronoun resolution, nor CLI from their L1s.

3.5 Summary and implications for the present study

The linguistic phenomenon under study in the present thesis—pronominal subject anaphora resolution—has been identified as a vulnerable domain in bilingual language acquisition (Sorace, 2011; see §3.4.1). Although some studies have attested target-like pronoun resolution patterns in near-native bilingual speakers, the interpretation of null and especially overt pronouns has been proven to be complex to acquire and cognitively challenging for bilingual populations. Interestingly, no studies have addressed pronoun resolution in non-canonical marked information structures, which may be even more cognitively challenging for bilinguals (see §3.4.2). Moreover, this phenomenon has not been investigated with early and highly functional bilinguals from social bilingual settings such as Catalan-Spanish bilinguals in Catalonia (see §3.3.1). One of the main aims of the present study will be to investigate the role of language dominance in modulating the effects of bilingualism on the interpretive preferences of null and overt pronouns in Catalan and in Spanish by Catalan-Spanish bilinguals.

Referring to bilingualism, we have understood it as a multifaceted, gradient, and dynamic construct (§3.2.1). Given that our bilingual population, Catalan-Spanish early bilinguals, are highly functional bilinguals, we use language dominance as a proxy for bilingualism, following previous studies on this population (see §3.2.2 and §3.3.2). To compare anaphora resolution patterns of bilinguals with different profiles, we will include three groups of bilinguals of varying language dominance: Catalan-dominant, balanced, and Spanish-dominant bilinguals. To operationalize language dominance we will use the Bilingual Language Profile questionnaire (BLP; Birdsong et al., 2012). This tool allows for the weighting of multiple factors influencing bilingual experience (e.g., language history, use, proficiency, and attitudes) in a comprehensive measure of bilingualism (see §3.3.3).

In reference to the linguistic phenomenon, in Catalan and in Spanish null pronouns have been shown to preferably corefer with subject antecedents in canonical sentences, and overt pronouns with object antecedents, in line with the PAH (Carminati, 2002; Bel & García-Alcaraz, 2018). These biases seem to be stronger in Catalan than in Spanish. As

with previous studies, we consider two possible outcomes of the effects of bilingualism—which may be modulated by language dominance—on anaphora resolution (see §3.2.3, §3.4.2 and §3.4.3). Firstly, the effects of bilingualism could be attributable to general bilingual strategies—bilinguals resorting to the overextension of properties of overt pronouns to alleviate increased processing loads in both of their languages or in their weaker language (Sorace & Filiaci, 2006; Sorace, 2011). Secondly, the effects of bilingualism could result in CLI—language dominance modulating CLI's strength or directionality. More specifically, three scenarios where CLI can emerge are contemplated: 1) Spanish by Catalan-dominants could show a stronger CLI from Catalan if influence occurs from the more categorical to the more flexible language (de Prada Pérez, 2019; Hulk and Müller, 2000), 2) Catalan by Spanish-dominant bilinguals could show a stronger CLI from Spanish if influence occurs from the more flexible to the more categorical language (Romano, 2019), or 3) language dominance could modulate CLI's directionality if it can occur bidirectionally from the dominant toward the non-dominant language (Yip & Matthews, 2006). The lack of effects of bilingualism and language dominance would be either indicative of 1) each group having similar anaphora resolution preferences not affected by bilingualism, or of 2) the presence of a general bilingual strategy irrespective of the bilinguals' profile (see §3.3.2).

Overall, we intend to provide evidence from an understudied population that will help to elucidate how bilingualism and language dominance influence a language system and, more specifically, the interpretation of pronominal anaphoric dependencies in the two languages of a bilingual. The specific research questions related to the effects of bilingualism on anaphora resolution by early functional Catalan-Spanish bilinguals, together with our predictions, will be presented and addressed along Chapter 6.

Chapter 4

Methodology

4.1 Introduction

In this chapter, we will thoroughly describe the methodology followed in the present study. This thesis is all based on two equivalent forced choice tasks (one in Catalan and one in Spanish). These tasks target at analyzing the interpretation of two types of pronouns (null and overt third person subject pronouns) as coreferring with subject or object antecedents in sentences with four different information structures (unmarked canonical sentences, topicalization of an object antecedent via clitic-left dislocations, focalization of a subject antecedent via it-clefts, and focalization of an object antecedent via it-clefts). The two tasks were completed by Catalan-Spanish bilinguals with different language dominance profiles (Catalan-dominant, balanced and Spanish-dominant bilinguals), as well as by a group of Spanish monolinguals, who only completed the task in Spanish.

This complex design will allow for analyzing whether the Position of Antecedent Hypothesis (PAH; Carminati, 2002) can explain null and overt pronouns' interpretive preferences beyond the 'classical' design, which has mainly tested syntactically unmarked sentences with canonical preverbal subjects and postverbal objects as plausible antecedents. It is worth highlighting that the same participants give an answer to all the experimental conditions in each language under the same circumstances, because the different conditions referring to the linguistic phenomenon (pronoun and information structure) are tested in the same task. Importantly, this design will also allow for within-subject comparisons.

At the same time, data was collected from two different languages, Catalan and Spanish, to be able to make cross-linguistic comparisons, and from bilinguals differing in language dominance. Thus, the role of language dominance on anaphora resolution is assessed in a bidirectional design, which makes a novel contribution to the field. Importantly, we will not only compare three group of bilinguals to see whether pronoun interpretation in each language is modulated by language dominance, between subjects, but we will also investigate whether resolution patterns differ between the two languages of the bilinguals, within subjects.

In the following sections, the participants in the study will be first presented along §4.2. To begin with, a general description of the participants will be provided (§4.2.1), as well as a description of the questionnaire used for determining their language dominance profile (§4.2.2). Having described the language background questionnaire, each group of Catalan-Spanish bilinguals will be more in-depth characterized: the criteria followed to divide them into three dominance groups will be first described (§4.2.3), and a more detailed portrait of each group of bilinguals is provided afterward (§4.2.4). After this description of participants, we will thoroughly describe the methodological design of the experimental tasks (§4.3), as well as the procedure followed to collect data in the experimental sessions (§4.4). Finally, we will account for data analyses: firstly, the preparation and cleaning of data for subsequent statistical analyses will be specified (§4.5.1); secondly, the variables of the study will be briefly recapitulated, and the conducted analyses will be outlined (§4.5.2). Given that several mixed effects logistic regression models were run including the relevant data to answer different research questions, this chapter will only describe the general procedure for these analyses. The specificities and the formula of each computed model will be reported together with its results in the relevant sections in Chapter 5 and Chapter 6.

4.2 Participants

4.2.1 General description

The participants in the present study are 94 Catalan-Spanish bilinguals (mean age: 22.25, range: 18-35; 67 females) that have been born and raised in Catalonia, and 40 monolingual Spanish speakers²⁶ (mean age: 20.12, range: 18-30; 32 females) that have been born and raised in monolingual regions of Spain, mainly in Aragon. Crucially, no monolingual speakers with Catalan as their first and only language participated in the study as there are no speakers with such a language background. Catalan-dominant speakers are also exposed to Spanish in the society and end up acquiring both languages (see §3.3.1).

²⁶ Even though we use the term *monolingual* to refer to Spanish speakers that have no knowledge of Catalan and that have not grown up in a bilingual society, we assume that they all have knowledge of one or more L2s, so we are aware that they are not true monolinguals. Similarly, we are also aware that *bilinguals* are in fact multilinguals. We will therefore use these terms in a broad sense.

All participants were young adults and were mainly studying a university-level degree or had already graduated from university. Some other participants were studying vocational education ($n = 13$). We assume that all our participants have knowledge of at least one L2, English, as it is taught in school and is now present in the day-to-day life of most young adults. The inclusion criteria for bilingual speakers to participate in the study was that they had not grown up in families speaking languages other than Catalan or Spanish, that they had always been living in Catalonia, and that they had been educated in the Catalan education system. As for monolingual speakers of Spanish, we only selected participants who had been raised monolingually from birth, who had not been living abroad, and whose parents were native speakers of Peninsular Spanish.

Sixteen additional participants were excluded from the study after completing the experiment for two different reasons: twelve bilingual participants were excluded when choosing a cut-off point to group participants according to their language dominance profile, in order to have more well-defined and clearly distinguished groups of bilinguals (for more details, see §4.2.3), and four participants—two bilinguals and two monolinguals—were excluded during the process of inspecting and cleaning the data for statistical analyses (for more details, see §4.5.1.1). In this way, 108 bilinguals and 42 monolinguals had initially participated in the task, but only 94 bilinguals and 40 monolinguals were considered in the analyses of the results and in this description and characterization of participants.

As previously mentioned, Catalan-Spanish bilingual participants were divided into three groups according to language dominance: Catalan-dominant bilinguals, balanced bilinguals, and Spanish-dominant bilinguals. In Table 4.1 below, general information on the age and gender of each group of participants has been included.

Table 4.1

General information about bilingual and monolingual participants

	N	Females	Age			
			<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Catalan-dominant bilinguals	34	26	22.32	3.57	18	35
Balanced bilinguals	31	21	21.16	2.37	19	26
Spanish-dominant bilinguals	29	20	23.28	3.74	19	31
Monolinguals	40	32	20.12	3.67	18	30

As further argued in previous §3.3.3, to identify the language dominance profile of bilingual participants, they completed a background questionnaire based on the Bilingual Language Profile (BLP; Birdsong et al., 2012). This questionnaire is described in detail in the next section (§4.2.2) and the criteria used to divide participants into groups is presented afterward (§4.2.3). Given the need to provide a thorough description of the bilingual profile of each group of participants, the results of the BLP questionnaire are more exhaustively reported in §4.2.4.

Regarding geographical distribution, Catalan-dominant bilinguals were from Barcelona ($n = 12$) or from central regions in Catalonia ($n = 15$), from the province of Barcelona. The remaining participants were from the provinces of Tarragona ($n = 4$) and Lleida ($n = 3$). Balanced bilinguals were mainly from Barcelona ($n = 16$) or its metropolitan area ($n = 10$), and the remaining participants were from other cities in the province of Barcelona ($n = 3$), Tarragona ($n = 1$), and Lleida ($n = 1$). Finally, Spanish-dominant bilinguals were also from Barcelona ($n = 19$) or its metropolitan area ($n = 8$), and the remaining participants were from Tarragona ($n = 1$) and a city in the province of Barcelona ($n = 1$). The great majority of participants were therefore speakers of Central Catalan, although there were also some speakers of Occidental Catalan ($n = 7$). As already stated, Spanish monolingual participants were mainly from Aragon ($n = 30$) and some participants were from Navarra ($n = 5$), Madrid ($n = 3$), La Rioja ($n = 1$), and Castilla y León ($n = 1$). All participants, bilinguals and monolinguals, had been exposed to Peninsular Spanish at home.

All participants reported not having reading disorders or language impairments. They all gave informed written consent regarding the data use and received a stipend for their participation in the study. Ethical approval was received from the Institutional Committee for Ethical Review of Projects in the Universitat Pompeu Fabra (CIREP-UPF; Reference code 0032).

4.2.2 Language background questionnaire: the Bilingual Language Profile questionnaire

The language background questionnaire that bilingual participants answered, in Spanish, was the Bilingual Language Profile (BLP; Birdsong et al., 2012). The main reasons for using this questionnaire were the following: it has been specifically designed to measure and operationalize language dominance, it conceives this construct as

multidimensional and gradient, it is quick and easy to use, the obtained data is easy to codify afterward, and it has been increasingly used in recent studies on Catalan-Spanish bilinguals (see §3.3.3). Given that the BLP was not specifically designed for bilingual populations living in a bilingual society such as Catalonia, we adapted some questions (with minimal changes) to make them more appropriate to our bilingual context. In the present section, this questionnaire will be thoroughly described, as well as the additional background questions and the small adaptations we introduced to some of its questions. The slightly modified BLP questionnaire that participants completed can be found in Appendix A. All participants completed it in Spanish.

The Bilingual Language Profile (BLP; Birdsong et al., 2012) is a self-reported questionnaire specifically designed to assess different dominance degrees of bilinguals' languages to place them within a continuum, as thoroughly described in Gertken et al. (2014). It was designed to produce a continuous dominance score from the responses to 19 questions that consider multiple dimensions and contexts of bilingual experience. At the beginning of the questionnaire, participants are asked about some "biographical information" that is not used in the computation of the language dominance score. Afterward, the 19 questions that are used to obtain the numerical language dominance score are organized in four modules: language history (6 questions), language use (5 questions), language proficiency (4 questions) and language attitudes (4 questions).

Regarding the biographical information section, the BLP asks participants about their age, sex, current place of residence and highest level of formal education. We added some complementary questions to this background section, inspired by questionnaires such as the ones used in Bel and García-Alcaraz (2018) and Ferré and Brysbaert (2017). The only aim of these additional questions was to be able to further characterize the bilingual participants and did not affect their dominance scores. Firstly, in addition to the current place of residence, we asked participants to state their place of birth. If they lived in a different place from where they were born, they were also asked for how long they had been living in their current place of residence. Secondly, we added a multiple-choice question in which participants had to indicate the language in which they feel more comfortable—Catalan, Spanish, or another language ("¿En qué lengua/s te sientes más cómodo/a?"). Finally, we added a set of multiple-choice questions referring to the language that bilinguals usually use to talk to their parents, siblings, and partner—Catalan, Spanish, another language ("Indica la/s lengua/s que utilizas habitualmente para hablar con: a) tu madre, b) tu padre, c) tu(s) hermano/a(s), d) tu pareja"). If a

speaker chose the option 'another language' in any of these questions, they were not selected to participate in the study²⁷.

We will now refer to the questions in the four modules used to compute the language dominance score. First, the items in the language history module (questions Q1-Q6 in Appendix A) include information about the age of onset of acquisition, the age at which participants started to feel comfortable using each language, the years of schooling in each language, and the years spent in a country or region, in a family, and in a work environment in which each language was used. Second, the language use module (questions Q7-Q11) measures the percentage of time using Catalan, Spanish, or another language, in five different settings: with friends, with family, at work, when talking to themselves and when counting (the total use for all languages in each context needed to equal 100%). Third, regarding the language proficiency module (questions Q12-Q15), participants rate their competences in listening, reading, speaking, and writing in Catalan and Spanish on a 6-point scale (0 = 'not well at all', 6 = 'very well'). Finally, in the language attitudes module (questions Q16-Q19), participants rate four statements on a 6-point scale (0 = 'disagree', 6 = 'agree') about whether they feel like themselves when using each language, whether they identify with each culture, the importance they give to using each language as a native speaker, and the importance they give to being perceived as a native speaker. In each question, participants had to give an answer for Catalan, and an answer for Spanish, so that a language-particular score is separately obtained for each language.

As already anticipated, minimal adaptations were made to some of three questions in the BLP to make them more appropriate to the bilingual context in Catalonia. For instance, the vehicular language of the education system in Catalonia is, in principle, Catalan, and both Catalan and Spanish are taught and studied in language grammar classes. Taking this context into account, we decided to modify the question asking for the years of classes taken in Catalan and Spanish. Instead of asking "How many years of classes (grammar, history, math, etc.) have you had in the following languages (primary school through university)?", we asked "How many years of classes (history, math, etc., excluding grammar classes) have you had in the following languages?" (Q3 in Appendix A). In this way, we asked for the languages of schooling of the bilinguals, given that they all took grammar classes in both Catalan and Spanish at least until they

²⁷ Participants were asked to complete the BLP questionnaire before taking part in the experimental sessions.

started studying a degree at university or vocational education²⁸. We also split this question to separate pre-school and primary school (ages 3-12) (Q3a) from secondary school and higher levels of education (after the age of 12) (Q3b). This question is the only one that we modified in the language history module.

However, it would have been interesting to have modified another question: instead of asking for the years spent living in a country/region where each language is spoken (Q4), it would have been a good idea to ask about the years spent living in a town/city/area/neighborhood where each language is spoken. Given that all the selected participants had always lived in Catalonia, more nuances regarding the languages spoken in their living area could have been captured. We also wonder whether asking for age of onset of acquisition (Q1) as the age at which they started "learning" each language was the most accurate way to obtain this information.

In the language proficiency module, instead of asking for language proficiency with the questions "How well do you speak/understand/read/write Catalan/Spanish?", we reformulated the questions as "Which is your level of oral comprehension/written comprehension/oral production/written production in the following languages" (Q12-Q15 in Appendix A). We assumed that all our highly educated participants would be familiar with this terminology, which we considered to be more precise.

Finally, the other two questions that we modified are part of the attitudes' module. We consider our participants to be native speakers of both Catalan and Spanish, given that they have been born and raised in a bilingual society. Therefore, we changed the statement "It is important to me to use (or eventually use) Catalan/Spanish like a native speaker" for "It is important to me to use Catalan/Spanish adequately and correctly" (Q18 in Appendix A). The other statement that we modified regarding attitudes was "I want others to think I am a native speaker of Catalan/Spanish". We formulated this statement as "I want others to think that I do not have an accent in Catalan/Spanish" (Q19). In the first case we were asking about how they feel about being perceived as native speakers of Catalan and Spanish. In the second case we were asking about how they feel about others thinking that they have an accent in their languages. We are not sure of how these two questions contribute to defining language dominance in a bilingual, because they may be answered as resorting to non-language-specific attitudes regarding bilingualism. However, we did not want to alter the number of questions in the BLP and, at the same

²⁸ All the bilingual participants in this study were educated in the Catalan education system.

time, we wanted to ask relevant questions for the participants. In any case, anticipating the results of the questionnaire, we did not find that they altered the dominance score; they did just not make a relevant contribution.

Based on the participants' responses to all the described questions in the BLP, a language-particular module score is first obtained, on a different scale for each module. Afterward, the module scores for each language are transformed so that each module receives equal weight, and the four language-particular module scores are added to obtain a language-particular global score for Catalan, and another for Spanish. These punctuations finally yield to a global dominance score. Importantly, the background or biographical questions at the beginning of the questionnaire are not taken into account when computing the dominance scores in the BLP. In the following paragraph we give more specific information on how these global dominance scores are obtained.

First, all the questions in the BLP modules are assigned a numerical value following the authors' guidelines: the 6 language history questions are worth between 0 and 20 points each (maximum score of 120 points per language), the 5 language use questions are worth between 0 and 10 points each (maximum score of 50 points per language), the 4 language proficiency questions are worth between 0 and 6 points each (maximum score of 24 points per language), and the 4 language attitudes questions are also worth between 0 and 6 points each (maximum score of 24 points per language). To calculate the language-particular module scores (for Catalan and for Spanish), the total points of the responses to the questions in each module are added for each language separately. Subsequently, the score for each module is multiplied by a specific factor so that all modules receive equal weighting. These module scores (maximum module score for each language: 54.5) are added to generate a global language score (maximum global score for each language: 218). Once obtained the global language score for Catalan and for Spanish, one language score is subtracted from the other score, producing the global dominance score, whose limits are, thus, -218 and +218. In the present study, the Catalan score was subtracted from the Spanish score. Therefore, negative values are indicative of Catalan dominance, and positive values indicative of Spanish dominance. The scores close to zero suggest balanced bilingualism.

4.2.3 Dividing bilingual participants into groups

As already stated, the Catalan-Spanish bilingual participants were divided into three groups based on their global dominance scores resulting from the BLP: Catalan-dominant bilinguals, balanced bilinguals, and Spanish-dominant bilinguals. Only the BLP global dominance score was used to define the groups, and not the information provided in any specific question or module. Global dominance scores for participants in the present study ranged from -112.30 to 87.19²⁹. As mentioned, the minimum and maximum numerical scores in the BLP are -218 and +218, negative values indicating Catalan dominance in the present study, and positive values indicating Spanish dominance. As shown in Table 4.2, the dominance scores of Catalan-dominant bilinguals ranged from -112.30 to -59.22, those of balanced bilinguals from -34.96 to 23.71, and those of Spanish-dominant bilinguals from 41.69 to 87.19. Since choosing a cut-off point that separates balanced bilinguals from dominant ones is not an obvious decision, twelve additional participants with BLP scores between -59 and -35 or between 24 and 41 were excluded from the analyses. As previously mentioned in §4.2.1, the analyses in the present study consider data from 94 bilingual participants but other bilinguals participated in the study completing the language background questionnaire and the experimental tasks. As shown in Figure 4.1 below, this decision allowed to create three clearly distinguishable groups and to avoid having participants at the edges. The three groups were relatively homogeneous in number.

Table 4.2

Participants' global dominance scores in the Bilingual Language Profile (BLP) questionnaire

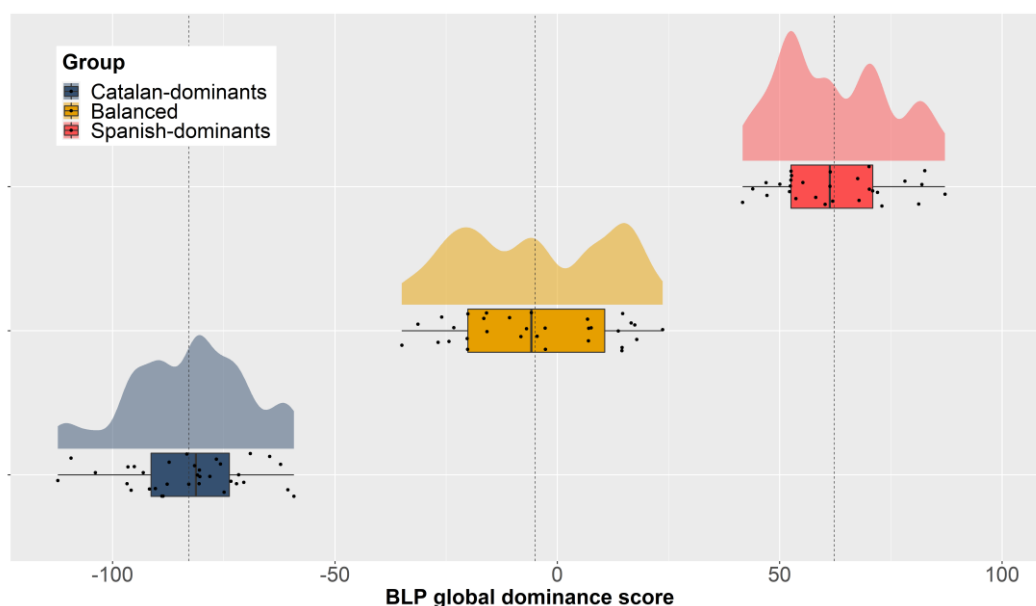
	N	M	SD	min	max
Catalan-dominant bilinguals	34	-82.83	13.19	-112.30	-59.22
Balanced bilinguals	31	-5.00	17.07	-34.96	23.71
Spanish-dominant bilinguals	29	62.31	12.68	41.69	87.19

Note. The minimum and maximum scores in the BLP are -218 and +218 (negative values are indicative of Catalan dominance and positive values indicative of Spanish dominance).

²⁹ Participants do not occupy the full range of the BLP (from -218 to +218) because they are all bilinguals. Strictly speaking, the values at the edges would be indicative of monolingualism.

Figure 4.1

Participants' global dominance scores in the Bilingual Language Profile (BLP) questionnaire



Note. The boxplots show the distribution of participants' BLP scores in the three bilingual groups. Each point corresponds to one participant and the split violins indicate density. The vertical discontinuous lines mark the mean dominance score for each group.

More specifically, Figure 4.1 illustrates the distribution of Catalan-dominant, balanced and Spanish-dominant bilinguals in the bilingual continuum generated by the BLP. The specific descriptive statistics of each group's dominance scores can be found in Table 4.2 above. The three bilingual groups clearly differ from each other, representing different profiles in the bilingual continuum. Based on the boxplots and density plots, as well as standard deviation, we can observe that Catalan-dominant and Spanish-dominant bilinguals seem to be more homogeneous than balanced bilinguals. These groups will be thoroughly described in the following section (§4.2.4) by scrutinizing their responses to the questions in the BLP and the background questionnaire.

Finally, it is worth highlighting that although we used the dominance scores in the BLP to group participants, we also think that it would be interesting to use these numerical scores as a continuum to further explore bilingualism effects. Therefore, an exploratory and complementary analysis using the BLP dominance scores as a continuous variable will also be included in Chapter 6 (§6.2.5).

4.2.4 Characterizing each group of Catalan-Spanish bilinguals

Describing participants based on an overall dominance score that subtracts the global scores of each language may fall short of capturing the full picture of participants' bilingual experience. Thus, the results of the language background questionnaire will be scrutinized in this section to describe the three groups of bilingual participants in detail and to show how they differ. These results will be presented following the four modules in which the BLP is divided: language history, use, proficiency, and attitudes. In order to have the complete picture, each groups' results for Catalan and for Spanish will be separately presented. The complementary questions added as background questions will be used to complete the information obtained through the adapted BLP when relevant. The descriptive results for each of the questions in the BLP can be found in Appendix B.

4.2.4.1 Language history

The language history module included questions on the age of onset of acquisition, the age they started to feel comfortable using each language, and the languages of schooling, the country, the family, and the work environment (see §4.2.2). Regarding the language-particular scores of this module (illustrated in Figure 4.2 and Table 4.3), a Kruskal-Wallis test³⁰ revealed significant differences between the three groups of bilinguals in their history scores (in Catalan: $H(2) = 39.821$, $p < .001$; in Spanish: $H(2) = 42.899$, $p < .001$). Nevertheless, as shown by the post-hoc analyses that were performed using Dunn test for multiple comparisons, balanced and Spanish-dominant bilinguals did not differ in their scores for Spanish.

Table 4.3

Scores on the BLP language history module in Catalan and in Spanish, by group of bilinguals

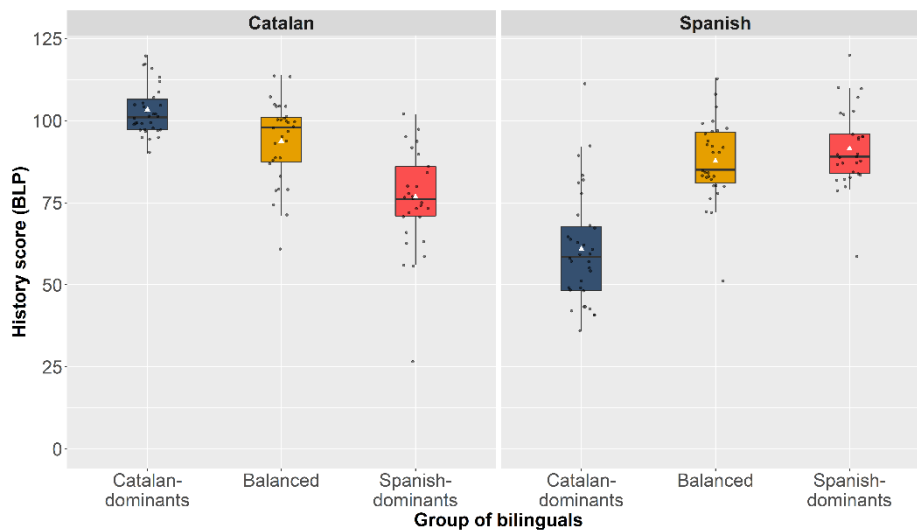
	Catalan				Spanish			
	M	SD	Min	Max	M	SD	Min	Max
Catalan-dominant	103.29	7.90	90	120	60.88	17.14	36	111
Balanced	93.74	12.39	61	114	87.71	12.20	51	113
Spanish-dominant	76.79	17.31	27	120	91.45	11.99	59	120

Note. The maximum score for each language in this module was 120.

³⁰ Both the normality and the homogeneity of variance assumptions for a repeated measures ANOVA were not met.

Figure 4.2

Scores on the BLP language history module in Catalan and in Spanish, by group of bilinguals



Note. The white triangle represents the mean score of each group. The maximum score for each language in this module was 120.

As expected, Catalan-dominant bilinguals have been further exposed to Catalan than balanced bilinguals ($p < .001$), who have been overall more exposed to Catalan than Spanish-dominant bilinguals ($p < .001$). In Spanish, the picture is slightly different. Whereas Catalan-dominant bilinguals have been less exposed to Spanish than the other two groups (both $p < .001$), balanced and Spanish-dominant bilinguals have obtained similar scores ($p = .415$). More variability is observed in the scores of balanced bilinguals, indicating that they may not be homogeneous in terms of language history. Variability is also attested in the non-dominant language of Catalan- and Spanish-dominant bilinguals, showing individual variation in the extent to which they have been exposed to their weaker language.

Interestingly, the results are slightly skewed toward Catalan; the exposure to Catalan of Spanish-dominant bilinguals is higher than the exposure to Spanish of Catalan-dominant bilinguals. This higher presence of Catalan across groups may be attributable to the language of schooling, which received higher scores in Catalan than in Spanish (see Table B.1 in Appendix B). It is surprising, however, to see in the results of the mentioned table that Catalan was not the only vehicular language at school for many of the bilinguals. In any case, Catalan- and Spanish-dominant bilinguals have been overall more exposed to their dominant language (both $p < .001$), whereas balanced bilinguals

can be said not to clearly differ regarding language history in their two languages ($p = .062$)³¹.

Some of the questions included in this module are of particular interest to describe the bilingual profile of the participants, such as the age of onset of acquisition, or the languages spoken within the family environment. We have therefore further explored their responses to these questions.

First, regarding the age of onset of acquisition of each language, Catalan-dominant bilinguals declare that they began acquiring Catalan earlier than Spanish, and both balanced and Spanish-dominant bilinguals started acquiring Spanish earlier than Catalan, as shown in Table 4.4. Given that this table includes mean results, together with the fact that participants may have interpreted the question on age of acquisition in several ways, we will also refer to this information by grouping the data in percentages for two different time slots, from birth or later on (between 2 and 6 years of age), in Table 4.5. More specifically, all Catalan-dominant bilinguals state they began acquiring Catalan from birth and Spanish either from birth (52.94%) or at a later stage (47.06%), between the age of 2 and 6. Concerning balanced bilinguals, most of them declare having begun acquiring both Catalan and Spanish from birth (74.19%), only a few of them (22.58%) claim having begun acquiring Catalan between 2 and 6, and one participant started acquiring Spanish after Catalan (3.23%). Parallel to Catalan-dominant bilinguals, all Spanish-dominant bilinguals began acquiring Spanish from birth and Catalan either from birth (51.72%) or at a later stage (48.28%).

Table 4.4

Age of onset of acquisition of participants, by language and group of bilinguals

	Catalan				Spanish			
	<i>M</i>	<i>SD</i>	Min	Max	<i>M</i>	<i>SD</i>	Min	Max
Catalan-dominant	0	0	0	0	1.82	2.21	0	6
Balanced	0.77	1.36	0	4	0.1	0.54	0	3
Spanish-dominant	1.41	1.66	0	6	0	0	0	0

³¹ The comparison between the two languages of each group of bilinguals was obtained through a paired samples Wilcoxon Test.

Table 4.5

Age of onset of acquisition in two slots (from birth and between 2-6 years of age), by language and group of bilinguals

	Acquisition of Catalan		Acquisition of Spanish	
	From birth	Between 2-6 years of age	From birth	Between 2-6 years of age
Catalan-dominant	100%	-	52.94%	47.06%
Balanced	77.42%	22.58%	96.77%	3.23%
Spanish-dominant	51.72%	48.28%	100%	-

Second, complementary to the observations on age of onset of acquisition, it is also interesting to refer to the years that participants had spent in a family environment speaking Catalan, on the one hand, and Spanish, on the other hand (see Table B.1 in Appendix B). As an additional question to those of the BLP, we asked participants what languages they used to talk to their parents, siblings, and partners, to have more information on the languages spoken within the family environment. Consistent with previous data on the age of acquisition, most Catalan-dominant bilinguals (76.47%) state they were raised in Catalan-speaking families and the rest of participants declare having been raised in families where both languages were used. However, in the complementary question, all the Catalan-dominant bilinguals say that they only use Catalan to talk to their parents and/or siblings. The majority of balanced bilinguals (70.97%) grew up in bilingual families, a quarter of them (25.81%) grew up in Spanish-speaking families, and only one participant in the balanced group was raised in a Catalan-speaking family (3.23%). When asked about the languages they use with their mother, father and/or siblings, around half of them declare to use both languages (45.16%), to preferably use Spanish (38.71%), and few of them to preferably use Catalan (16.12%). Thus, there is an asymmetry in the number of balanced bilinguals that have grown up in Catalan-speaking families compared to Spanish-speaking families, who make a more noticeable contribution to this group of bilinguals. Finally, Spanish-dominant bilinguals were mostly raised in Spanish-speaking families (72.41%) and only some of them (27.59%) say they grew up in bilingual families. In this case, only three participants in the Spanish-dominant group declare using both Catalan and Spanish to talk to their parents and/or siblings (10.34%). Overall, using the two languages at home—or growing up in bilingual families—seems to promote balanced bilingualism, at least in the way measured with the BLP.

Given that language dominance is a dynamic construct that may change across the lifespan, not all unbalanced bilinguals can be said to be early sequential bilinguals and balanced bilinguals, simultaneous bilinguals. In the present study, most balanced bilinguals are simultaneous bilinguals, since they state that they have been exposed to both languages from birth within their family environment. However, it is not the case for all of them. Regarding Catalan-dominant and Spanish-dominant bilinguals, half of the participants in each group state having acquired their weaker language from birth, at the same time as their dominant language, even though not all these participants grew up in bilingual families. The other half of participants can be more clearly defined as early sequential bilinguals that consider they acquired their weaker language later than their dominant language, probably when they started being further exposed to it at school.

Although the information provided by the BLP is enough for the purposes of the present study, it can be highlighted that it does not include any questions on the context in which participants acquired each language or to what extent they were exposed to each language in different environments and across time (during childhood, adolescence, etc.). Moreover, the amount of exposure to the languages under study is measured in years, instead of a percentage or an ordinal scale, which may be easier to answer by participants. Also, it does not directly address whether the early bilingual participants have acquired Catalan and Spanish as simultaneous or sequential bilinguals, even though we could deduce this information from a combination of questions (including an additional complementary question to the BLP).

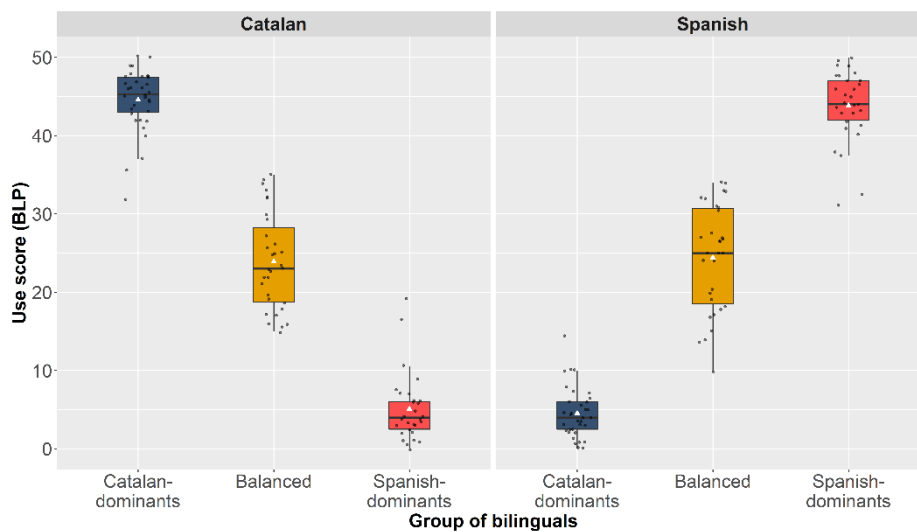
4.2.4.2 Language use

Language use is the module score that seems to better distinguish the three groups of bilinguals. The questions included in this module referred to five different settings: friends, family, work, inner thinking, and counting. Significant differences were attested between groups in the results of a Kruskal-Wallis test (in Catalan: $H(2) = 81.156$, $p < .001$; in Spanish: $H(2) = 81.306$, $p < .001$). These differences were further explored performing Dunn's tests. As very clearly illustrated in Figure 4.3 and Table 4.6, a progression can be observed in the overall uses of Catalan and Spanish by each group, in a very symmetric picture. Catalan-dominant and Spanish-dominant bilinguals use their dominant language more often than the other two groups of bilinguals (all $p < .001$). Also, these dominant groups seem quite homogeneous, showing a more or less limited use of their non-dominant languages. Balanced bilinguals, on the other hand, show a lot of

variability, even more in the use of Spanish. However, they clearly use Catalan more often than Spanish-dominants and less often than Catalan-dominants, and Spanish more often than Catalan-dominants and less often than Spanish-dominants (all $p < .001$). Importantly, despite the variability, they are not skewed toward one of the two languages, and, at the end, they show similar uses of Catalan and Spanish ($p = .793$)³².

Figure 4.3

Scores on the BLP language use module in Catalan and in Spanish, by group of bilinguals



Note. The white triangle represents the mean score of each group. The maximum score for each language in this module was 50.

Table 4.6

Scores on the BLP language use module in Catalan and in Spanish, by group of bilinguals

	Catalan				Spanish			
	M	SD	Min	Max	M	SD	Min	Max
Catalan-dominant	44.59	4.01	32	50	4.53	3.27	0	14.5
Balanced	23.94	6.16	15	35	24.42	6.81	10	34
Spanish-dominant	5.09	4.35	0	19	43.83	4.66	31	50

Note. The maximum score for each language in this module was 50.

³² The comparison between the two languages of balanced bilinguals was obtained through a paired samples Wilcoxon Test.

The question-by-question responses providing information on the different contexts of use (see Table B.2 in Appendix B) show that there are two questions that unquestionably identify the dominant language of Catalan-dominant and Spanish-dominant bilinguals: the language they use to count and to talk to themselves. However, in these questions, balanced bilinguals show a lot of variability: some of them are inclined toward one of the two languages, and some of them state they use both languages to a similar extent. The use of each language in the family context also polarizes participants in a very explicit way, similar to the former two questions. Regarding uses with friends and at university and/or work environments, clear differences between groups can be appreciated as well, but the three groups appear less polarized.

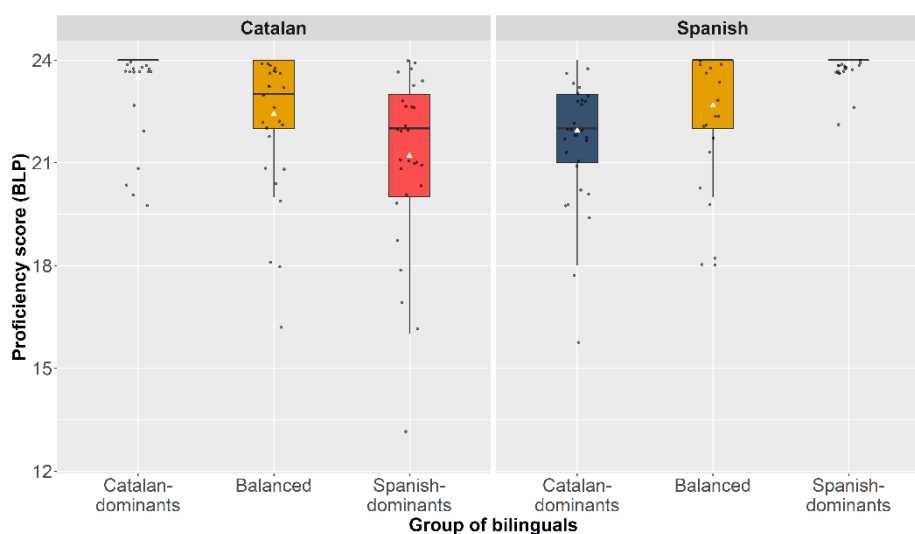
4.2.4.3 Language proficiency

Given the sociolinguistic context, we have assumed that all the bilingual participants of this study are highly proficient in their two languages. They have grown up in Catalonia and they completed post-obligatory secondary education or higher education in the Catalan education system. In fact, language proficiency in Catalan and Spanish after secondary education has been shown to be relatively high and very similar in both languages (Consell Superior d'Avaluació del Sistema Educatiu, 2017, 2021) (see also §3.3.1). However, slight differences between groups are appreciated when the self-assessment scores of the three groups are compared, as shown in Figure 4.4 and Table 4.7. In fact, results of a Kruskal Wallis test revealed significant differences between groups (in Catalan: $H(2) =$, $p < .001$; in Spanish: $H(2) =$, $p < .001$), and all post-hoc comparisons using Dunn tests were also found to be significant, all groups differing between them in both languages (all $p < .001$).

Although most bilinguals rate their proficiency as being 'very good' in both languages, not all bilinguals self-assess their proficiency in Catalan and Spanish at ceiling. Catalan- and Spanish-dominant bilinguals do rate their proficiency in their dominant language at ceiling, which constitutes an additional indicator of dominance in the corresponding language, and as being higher than in their non-dominant language. Whereas they are very homogeneous when scoring their dominant language, more variability is attested in their non-dominant language, especially in the case of Spanish-dominant bilinguals.

Figure 4.4

Scores on the BLP language proficiency module in Catalan and in Spanish, by group of bilinguals



Note. The white triangle represents the mean score of each group.

Table 4.7

Scores on the BLP language proficiency module in Catalan and in Spanish, by group of bilinguals

	Catalan				Spanish			
	M	SD	Min	Max	M	SD	Min	Max
Catalan-dominant	23.47	1.26	20	24	21.94	1.87	16	24
Balanced	22.42	2.11	16	24	22.68	1.97	18	24
Spanish-dominant	21.21	2.62	13	24	23.90	0.41	22	24

Note. The maximum score for each language in this module was 120.

As shown in Table B.3 in Appendix 3, the lowest ratings of Spanish proficiency by Catalan-dominants are those of oral expression skills. On the other hand, the self-perceived proficiency of Spanish-dominants in Catalan mirrors that of Catalan-dominants oral expression in Spanish, but they also give lower scores to their written expression skills. Balanced bilinguals overall show more variability, but to a similar extent in their two languages. Additionally, they evaluate their competence in both languages and in the four proficiency skills in a similar way, maybe as slightly higher in Spanish, and showing similar variability in each language as well (see Table B.3 in Appendix B). The lowest scores in balanced bilinguals' proficiency in both Catalan and Spanish correspond to written expression.

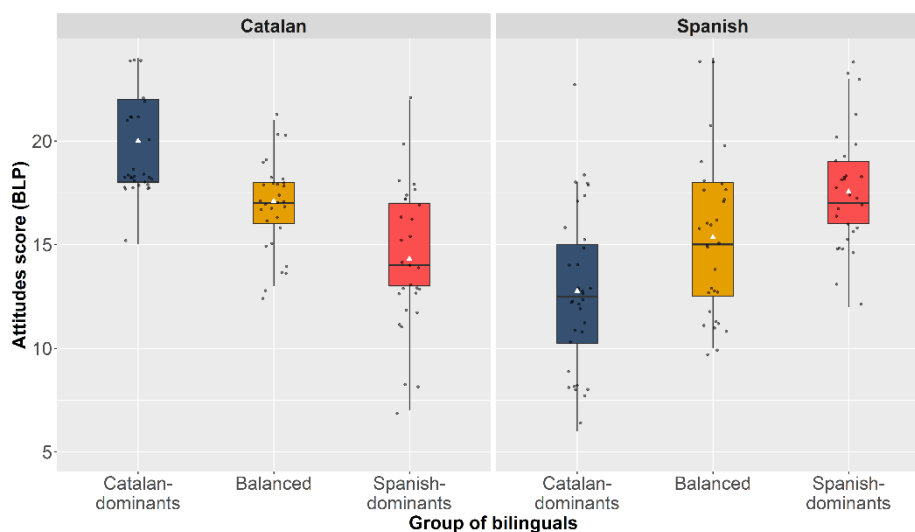
In any case, we can define the Catalan-Spanish bilinguals in the present study as being highly proficient in both of their languages. The reason why slight differences can be observed when the scores of the three groups and their two languages are compared—even though all participants self-assess their proficiency in Catalan and Spanish mainly as good or very good—may be related to the use of self-assessments. As Grosjean (1989, p. 5) states, bilinguals do not always assess their language competencies as adequate: "They often assume and amplify the monolingual view and hence criticize their language competence: how many times have bilinguals reported that they neither speak nor write their different languages adequately!".

4.2.4.4 Language attitudes

Finally, language attitudes also allow for identifying a gradation in the three groups in both of their languages, although variability is quite high in all groups. In this case, a repeated measures ANOVA found a significant interaction between Group \times Language ($F(2) = 41.160, p < .001$). As illustrated in Figure 4.5 and Table 4.8, the scores in Spanish show slightly more dispersion than in Catalan, in which slightly more polarization is observed. Catalan-dominant bilinguals are the ones that show more polarized attitudes, having higher scores in Catalan than in Spanish. They also show a higher identification with the Catalan language and culture than the other two groups (compared to Spanish-dominant bilinguals, $p < .001$; compared to balanced bilinguals, $p = .005$). Balanced bilinguals also show a higher identification with Catalan than Spanish-dominant bilinguals ($p = .013$). In contrast, Spanish-dominant bilinguals show a higher identification toward the Spanish language and culture than Catalan-dominant bilinguals ($p < .001$), but similar to that of balanced bilinguals ($p = .100$). Balanced bilinguals are less homogeneous regarding their attitudes toward Spanish, and a lot of variability is attested in the Spanish-dominant bilinguals' attitudes toward Catalan. The statement in this module of the BLP that mainly allowed for distinguishing the three groups of bilinguals is the one referring to the feeling like being oneself when speaking each language (see Table B.4 in Appendix B). The groups also differ regarding their identification with Catalan-speaking and Spanish-speaking cultures, but to a lesser extent.

Figure 4.5

Scores on the BLP language attitudes module in Catalan and in Spanish, by group of bilinguals



Note. The white triangle represents the mean score of each group.

Table 4.8

Scores on the BLP language attitudes module in Catalan and in Spanish, by group of bilinguals

	Catalan				Spanish			
	M	SD	Min	Max	M	SD	Min	Max
Catalan-dominant	20.00	2.66	15	24	12.76	3.85	6	23
Balanced	17.10	2.43	12	24	15.35	3.83	10	24
Spanish-dominant	14.31	3.51	7	22	17.55	2.84	12	24

Note. The maximum score for each language in this module was 120.

4.2.5 Summary of remarks on the Bilingual Language Profile questionnaire

After using the Bilingual Language Profile (BLP; Birdsong et al., 2012) to thoroughly describe the bilingual profile of our three groups of participants, we would like to summarize some remarks on the use of this tool to measure language dominance on Catalan-Spanish bilinguals. As we already anticipated (see §3.3.3), this questionnaire presents some limitations when measuring language dominance in bilingual populations that have grown up in a bilingual society such as Catalonia. It very clearly discriminates the two groups at the two poles of language dominance (i.e., Catalan-dominant and

Spanish-dominant bilinguals), which are quite homogeneous, but the group of balanced bilinguals shows a lot of variation. In this regard, we could also point to the concept of "balanced bilingualism". Perfectly balanced bilinguals are rare, but it is also true that there is a group of bilinguals that cannot be clearly identified nor as Catalan-dominant neither as Spanish-dominant.

According to our results, the module in the BLP that more clearly discriminates between our bilinguals is the module of language use, which analyzes the synchronic uses of each language. Two questions that made a particularly relevant contribution in this regard were the ones asking for the language used when talking to oneself and when counting.

The language history module, conversely, seems to show many overlapping scores between groups, particularly between balanced and Spanish-dominant bilinguals. We could relate this overlapping to the small contribution of questions in which Catalan-Spanish bilinguals do not differ much. Concretely, all bilinguals started acquiring both languages at early ages (range: 0-6), they also show small differences regarding the language of schooling (whole range 0-10, but they mainly occupy a very thin range between 8 and 10), due to the fact that Catalan is legally the vehicular language of schooling, and they have all spent most of their life in a region where both languages are used. The question on the work environment also provided few information, given that our participants are mainly university students and most of them have spent little time in this context (whole range 0-20, but they mainly occupy a range between 0 and 5). Nevertheless, there are two questions that do made a more relevant contribution within this module: the question on the age at which they became comfortable using each language and the years spent in a family where each language is spoken.

Regarding the language proficiency module, it does not aid much at differentiating the three groups of bilinguals. All bilinguals obtained scores near the ceiling level, so they generally consider themselves as being highly proficient in both of their languages. This is not surprising given that our participants belong to a specific population, young adults with good literacy skills in both languages.

Finally, the attitudes module also showed some differences between groups that could be attributable to the question on the degree to which they feel like themselves when using each language. In this case, we are reluctant to think that the questions on attitudes should have the same weighting as the other modules (e.g., history and use, which

include more fine-grained questions) in the final computation of the global dominance scores.

Taking these observations into account, we think that the BLP could be improved to better portray the nuanced dominance profile of Catalan-Spanish bilinguals. However, it has been proved useful to broadly classify bilinguals into three distinguished dominance groups, which we believe to be accurate enough for the purposes of the present study.

4.3 Materials: the experimental task

4.3.1 The nature of the task

An offline written two-alternative forced choice plus slider task was designed for the purpose of the present study. In this task, participants had to assign an interpretation to an ambiguous subject pronoun (null or overt) by choosing between two plausible antecedents that had been previously introduced in the same sentence (in subject or in object position). Different from a two-alternative forced choice task, which has been widely used for many of the offline experiments in the reported literature, participants had to use a slider bar to make a choice and to express at the same time their degree of certainty, as shown in Figure 4.6. Thus, participants were asked to choose an antecedent in a continuous visual scale, in a combined forced choice plus slider task that could allow for obtaining, in a single task, a referential choice and a rating scale reporting the strength of this choice (Marty et al., 2020).

Figure 4.6

Example screen of an item of the task in Spanish

Mariona evitó a Isabel cuando estuvo en el pueblo.

¿Quién estuvo en el pueblo?

Isabel

Mariona



Note. In a combined forced choice plus slider task, participants had to choose between two referents using a slider bar to express their certainty.

The main reason to use this task was that obtaining fine-grained answers by means of a continuous scale may uncover subtle but reliable nuances in interpretive biases that could remain imperceptible in a categorical dichotomous decision (Langsford et al., 2018; Schütze & Sprouse, 2014; Sorace & Keller, 2005; Sprouse & Almeida, 2017). However, as will be more detailly argued in §4.5.1.2, participants gave answers that were mainly at the edges of the continuous scale: they expressed an interpretive choice, but the task did not seem to capture a gradation in the strength or certainty of these choices. Therefore, the answers were finally reduced to a binary choice (i.e., as subject or object choices), as would have been in a traditional two-alternative forced choice task. The nature of the experimental sentences and subsequent questions may have also favored dichotomous answers. In any case, we decided to use this more open task to favor more nuanced responses, even if these were not finally obtained.

4.3.2 Experimental items

The experimental items constructed for this task were globally ambiguous sentences that consisted of two clauses, as shown in (4.1), an example for the baseline condition. The first clause was a main clause in which two characters of the same gender—feminine for half of the items, masculine for the other half—were introduced using proper names in subject and object position. The second clause was a temporal subordinate clause containing an ambiguous pronoun (null or overt) in subject position, followed by a verb and a prepositional complement. Depending on the condition, the information structure of the main clause and the type of subject pronoun in the subordinate clause were manipulated, as will be explained below. Each sentence was followed by a question asking for an interpretation of the ambiguous pronoun. Using a slider bar, as shown in previous Figure 4.6, participants had to show their preference toward the subject (e.g., *Mariona*) or the object antecedent (e.g., *Isabel*) in the main clause.

- (4.1) a. La Mariona va evitar la Isabel quan *pro/ella* va estar al poble. ¿Qui va estar al poble? (Catalan)
- b. Mariona evitó a Isabel cuando *pro/ella* estuvo en el pueblo. ¿Quién estuvo en el pueblo? (Spanish)
- 'Mariona avoided Isabel when she was in the town. Who was in the town?'

4.3.2.1 Variables and conditions

Two within-subject independent variables were manipulated to create the experimental items in a 4 × 2 factorial design: 1) information structure of the main clause (unmarked structure, object topicalization via clitic left dislocation, subject focalization via it-cleft, or object focalization via it-cleft), and 2) type of pronoun (null or overt). Another within-subject independent variable included in the design of the tasks was language (Catalan or Spanish)³³. An example for each condition is shown in Table 4.9 below, in Catalan and in Spanish. Mainly translation equivalents were used in each language so that the two tasks would be as comparable as possible. In this way, two tasks with equivalent items, one in Catalan and one in Spanish, were completed by participants in two separate sessions, as will be explained in §4.4. The materials used in each task, along with the instructions given to the participants, can be found in Appendix C. In the appendix, experimental items have been included only in one condition (unmarked structure, null subject pronoun), in each language.

Table 4.9

Experimental items per condition in Catalan and in Spanish

Information Structure	Pronoun	Example
Unmarked structure	Null	L'Anna va esperar l'Alba quan va arribar a l'estació. Ana esperó a Alba cuando llegó a la estación. 'Anna waited for Alba when (she) arrived at the station.'
	Overt	L'Anna va esperar l'Alba quan ella va arribar a l'estació. Ana esperó a Alba cuando ella llegó a la estación. 'Anna waited for Alba when she arrived at the station.'
Dislocated object (CLLD)	Null	A l'Alba la va esperar l'Anna quan va arribar a l'estació. A Alba la esperó Ana cuando llegó a la estación. 'Alba, Anna waited for her when (she) arrived at the station.'
	Overt	A l'Alba la va esperar l'Anna quan ella va arribar a l'estació. A Alba la esperó Ana cuando ella llegó a la estación. 'Alba, Anna waited for her when she arrived at the station.'

³³ We manipulated the information structure of the experimental sentences to be able to disentangle the role of the antecedents' syntactic function (subject vs. object), information status (topic vs. focus), and surface position (first vs. second mention) on the interpretation of null and overt anaphoric subject pronouns.

Information Structure	Pronoun	Example
Focused subject (subject cleft)	Null	Va ser l'Anna qui va esperar l'Alba quan va arribar a l'estació. Fue Ana quien esperó a Alba cuando llegó a la estación. 'It was Anna who waited for Alba when (she) arrived at the station.'
	Overt	Va ser l'Anna qui va esperar l'Alba quan ella va arribar a l'estació. Fue Ana quien esperó a Alba cuando ella llegó a la estación. 'It was Anna who waited for Alba when she arrived at the station.'
Focused object (object cleft)	Null	Va ser a l'Alba a qui va esperar l'Anna quan va arribar a l'estació. Fue a Alba a quien esperó Ana cuando llegó a la estación. 'It was Alba whom Anna waited for when (she) arrived at the station.'
	Overt	Va ser a l'Alba a qui va esperar l'Anna quan ella va arribar a l'estació. Fue a Alba a quien esperó Ana cuando ella llegó a la estación. 'It was Alba whom Anna waited for when she arrived at the station.'
Question:		Qui va arribar a l'estació? ¿Quién llegó a la estación? 'Who arrived at the station?'
Possible answers:		Anna / Alba Ana / Alba

4.3.2.2 Design of the experimental items

In the following paragraphs, we argue the decisions behind some criteria in the design of experimental items that deserve special attention, such as the reason why main-subordinate clause order was chosen, or why and how the item's global ambiguity was controlled for.

Clause order: Although more evident effects of the Position of the Antecedent Hypothesis (PAH; Carminati, 2002) arise in subordinate-main sequences (e.g., de Rocafiguera & Bel, 2022), the stimuli designed for the present study follow a main-subordinate clause order. In fact, in addition to testing the predictions of the PAH, the main aim of the present study is to investigate how marked information structures—and the manipulation of the information status and order of mention of the antecedents—affects pronoun resolution in null subject languages. In this regard, we are interested in

testing marked information structures that do not allow the subordinate-main clause order, such as subject and object clefts, as shown in (4.2) and (4.3).

- (4.2) a. *Quan va ser l'Anna qui va esperar l'Alba, va arribar a l'estació. (Catalan)
b. *Cuando fue Ana quien esperó a Alba, llegó a la estación. (Spanish)
- (4.3) a. *Quan va ser a l'Alba a qui va esperar l'Anna, va arribar a l'estació.
b. *Cuando fue a Alba a quien esperó Ana, llegó a la estación.

Given that subordinate-main word order is impossible in some of the marked information structures under study, for the purpose of coherence and to allow for comparability, unmarked sentences and sentences with CLLD structures also followed this clause order.

Ambiguity: Another decision that deserves further explanations is the use of globally ambiguous sentences. All the experimental items were totally ambiguous to observe the bare referential biases that subject pronouns inherently have. It is when there are no disambiguating cues and both antecedents are equally possible that structural preferences affecting the interpretation of null and overt pronouns clearly arise (Carminati, 2002). To create ambiguous items, we controlled for semantico-pragmatic factors by using verbs with neutral implicit causality biases and temporal subordinate clauses, as explained below.

Implicit causality: To ensure that no semantic cues such as implicit causality biases of the verbs in the main clause influenced the interpretation of the ambiguous pronoun in the subordinate clause (e.g., Cozijn et al., 2011; Garvey & Caramazza, 1974; Järvikivi et al., 2017), the implicit causality of all of the main verbs was as neutral as possible. Based on Goikoetxea et al. (2008), who analyzed the implicit causality biases of 100 interpersonal verbs in Spanish, sixteen verbs with a bias higher than 37.5% or lower than 62.5% were chosen (García-Alcaraz, 2015; Järvikivi et al., 2005, 2017). The chosen verbs appear in Table 4.10, as well as their reported bias in Goikoetxea et al. (2008). Each of these verbs was used three times in the main clause of the experimental sentences, in different contexts for each item³⁴. We assumed that the effects of implicit causality are similar across languages (Hartshorne et al., 2013; Rudolph & Försterling, 1997), particularly in closely related languages such as Catalan and Spanish. For this reason, the translated equivalents of the verbs in Spanish were used for Catalan.

³⁴ The verbs in the subordinate clause were only used once and maintained the global semantic ambiguity of the sentences.

Table 4.10*Implicit causality of the verbs in Spanish used in the main clause of the experimental sentences*

Verb	Subject bias (%)	Verb	Subject bias (%)
<i>evitar</i> ('to avoid')	59.3	<i>sorprender</i> ('to surprise')	52.5
<i>desmentir</i> ('to deny')	58.3	<i>alegrar</i> ('to cheer up')	52.0
<i>formar</i> ('to prepare')	56.9	<i>seguir</i> ('to follow')	51.1
<i>asustar</i> ('to scare')	54.7	<i>abandonar</i> ('to abandon')	48.1
<i>dejar</i> ('to leave')	54.7	<i>investigar</i> ('to investigate')	47.1
<i>saludar</i> ('to greet')	54.3	<i>ver</i> ('to see')	46.9
<i>interrumpir</i> ('to interrupt')	53.7	<i>esperar</i> ('to wait')	38.3
<i>soportar</i> ('to stand')	52.9	<i>recoger</i> ('to pick up')	37.5

Note. From Goikoetxea et al. (2008).

Connective: The conjunction that introduced the adjunct subordinate clause was a temporal connective, to favor ambiguity and following previous studies (e.g., Carminati, 2002; de la Fuente, 2015; García-Alcaraz, 2015; Torregrossa et al., 2020; Tsimpli et al., 2004). Different from other studies mixing sentences using *mentre/mientras* ('while') and *quan/cuando* ('when'), which have been found to slightly differ (Martín-Villena et al., 2021), all the experimental sentences used the same connective—*quan/cuando* ('when')—to avoid variability and introducing noise. On top of that, using only *quan/cuando* allowed us to maintain the verbs' grammatical aspect homogeneous across items. They were all in past tense and perfective aspect, both the verbs in the main clause and in the subordinate clause (e.g., Ferretti et al., 2009; Grüter et al., 2017). Also, a temporal connective was chosen to avoid the underlying coherence relations signaled by other types of connectives, which affect pronoun resolution patterns—e.g., causal, consequential, or concessive connectives (e.g., Filiaci, 2011; Godoy et al., 2018; Kehler et al., 2008; Kehler & Rohde, 2013; Koornneef & Sanders, 2013; Mayol, 2018; Stevenson et al., 2000). Finally, Colonna et al. (2015, p. 1318; see also de la Fuente, 2015) suggest that causal or concessive subordinate clauses, different from temporal subordinate clauses, may separate discourse units in a way that they do not correspond to sentence boundaries. Temporal connectives ensure that we are dealing with pronoun resolution in intrasentential contexts (Carminati, 2002) and within discourse units (for the distinction between sentence and discourse unit, see de la Fuente, 2015; Patterson & Felser, 2020). Overall, however, the effects of connectives or coherence relations on null and overt pronoun resolution may require further investigation.

4.3.2.3 Norming study

To select the experimental items that would be included in the task, 8 volunteer native speakers of Catalan and Spanish that did not participate in the experiment were asked to assess the degree of ambiguity of several sentences in Catalan (4 volunteer participants) and in Spanish (4 volunteer participants). With this purpose, a short norming task was designed using Google Forms. It included 64 ambiguous sentences with the 16 abovementioned implicit causality verbs, so four items were tested with each verb. The norming task was restricted to informatively unmarked canonical sentences and null subjects. Participants had to indicate how ambiguous they found each sentence in a Likert scale from 1 to 5 (1 being not ambiguous and 5 being very ambiguous). For each of the implicit causality verbs, the three sentences that had been rated as the most ambiguous were the ones included in the experimental task (range: 2.75-4.60), and the sentence that had been rated as less ambiguous was discarded (range: 1.52-4.04). Since we wanted the sentences in the Catalan and the Spanish task to be equivalent and as comparable as possible, the ratings in Catalan and in Spanish were assessed together. Also, some of the discarded sentences were rated as more ambiguous than some of the sentences we kept, because we discarded the less ambiguous sentence for each verb.

4.3.3 Practice items, fillers, and distractors

In addition to the experimental items, 6 practice items and 72 non-critical items (60 distractors and 12 fillers) were included in the task (Keating & Jegerski, 2015). At the beginning of the task, participants were presented with 6 practice items different from the experimental items to become familiar with the task. Half of these items resembled the experimental items, they were presented with two characters in a main clause and an ambiguous subject pronoun in a subordinate clause and participants were asked to interpret this pronoun (e.g., *La chica ayudó a Nora cuando ganó la lotería. ¿Quién ganó la lotería?*, 'The girl helped Nora when (she) won the lottery. Who won the lottery?'). The implicit causality verb in the main clause was semantically biased and the characters corresponded to a definite noun phrase and a proper name. The other half of practice items followed the structure of distractors.

The distractors were designed for two different experiments that were not related to the present study. There were 48 distractors that assessed the acceptability of differential object marking (e.g., *Los policías buscan a Carmen para detenerla de inmediato*, 'The police are looking for Carmen to arrest her immediately'), and they also appeared in eight different conditions. In this case, participants had to answer the question *¿Cómo te suena esta frase?* ('How does this sentence sound to you?') using the slider bar to determine acceptability. The ends represented the options *muy bien* ('very good') or *muy mal* ('very bad'). The remaining 12 distractors presented locative copular verbs in two different conditions (e.g., *El encargado recuerda que el reloj está en la joyería*, 'The manager remembers that the watch is in the jeweler's'). Similar to the other distractors, they were followed by the question *¿Cómo te suena esta frase?* ('How does this sentence sound to you?').

The 12 filler items contained an ambiguous relative clause to assess low/high attachment preferences in two different conditions (e.g., *El cardiólogo llamó al hijo del abuelo que escribe libros*, 'The cardiologist phoned the son of the old man who writes books'). Similar to the experimental items, participants were asked to interpret the ambiguous relative subject pronoun introducing the relative clause (e.g., *¿Quién escribe libros?*, 'Who writes books?'). The design was alike in both tasks (Catalan and Spanish).

4.4 Procedure

Prior to the experimental sessions, bilingual participants were sent a link to complete the language background questionnaire, which was set up as a web-based questionnaire using Google Forms. In the case of Spanish monolingual participants, they had to fill in some background questions as well before participating in the experiment. The experiments for bilingual participants were mainly conducted in a controlled lab environment, whereas Spanish monolingual participants completed the task online due to the Covid-19 pandemic restrictions at the time of data collection.

Once participants had completed the background questionnaire, the experiment consisted of two sessions in which participants were asked to complete two equivalent tasks: one in Catalan and one in Spanish (except from Spanish monolinguals, who only completed the task in Spanish). The experimental tasks were implemented as Qualtrics Survey questionnaires and took around 40 minutes to complete each. A break in the middle of the task was included to allow participants to rest if needed. The second

session took place 15 days after the first session (if it was not possible to keep this calendar, a minimum of 7 days and a maximum of 1 month separated the two sessions). Presentation by language was counter-balanced: half of the participants started with the task in Catalan in the first session, and the other half completed the task in Spanish first.

As for the instructions of the experimental tasks (included in Appendix D), participants were asked to read the sentences and answer the questions by moving the slider bar toward their preferred answer, depending on how certain they were of their choice. Participants were told that they could take their time to respond each question accurately, but that they should not spend much time thinking about the answer. The experimenter also reminded participants that this task was not based on grammar but on their intuitions, and that there were no correct or incorrect answers.

The 120 sentences of each experimental task were counterbalanced and presented in a pseudorandomized order across eight presentation lists. 48 experimental items were designed for the tasks, so that each list contained six items per condition ($k = 6$). Each item appeared in eight versions or conditions, one for each level of the independent variables. The experimental items were lexically matched in all their versions (i.e., they used exactly the same words and only differed in the information structure of the main clause and the subject pronoun form in the subordinate clause).

The eight lists were constructed following a Latin Square design so that participants would never see the same item in more than one experimental condition (i.e., in more than one version) and items of the same condition would never appear consecutively, nor items using the same neutral implicit causality verb in the main clause. No more than two experimental items were presented in succession. Only one item was presented at a time, and it was not possible to go back to the previous screen. The number of words referring to subject and object antecedents that appeared in the right and the left side of the slider bar (as shown in previous Figure 4.6) were also counterbalanced. Numerical values in the continuous scale were not visible, so that participants just relied on spatial reasoning.

4.5 Data analyses

4.5.1 Data preparation

4.5.1.1 Data cleaning

Data collected from the Qualtrics Survey questionnaires were tidied and cleaned before conducting statistical analyses. Prior to data cleaning, the complete dataset consisted of 11,184 observations. However, some of these observations were removed after an analysis of participants and items, as will be carefully reported in the following paragraphs. Some answers were also removed based on response time.

Four participants, as already mentioned in §4.2.1 were excluded from the analyses for two reasons³⁵. Three of these participants were removed because most of their answers were in the middle of the slider scale (97.92%, 85.42% and 79.17% of the answers they gave were between 45 and 55), indicating that they mainly considered both antecedents as equally possible across conditions. The fourth participant was removed because they showed a persistent preference for object antecedents. More specifically, the mean of their responses for subject or object antecedents across all conditions in both languages was 92.70, being it highly biased toward the object (the interpretation of null and overt pronouns was measured on a 0-to-100 scale, where 0 stood for subject and 100 for object bias)³⁶. According to these criteria, the 288 observations corresponding to these four participants were removed, representing 2.58% of the data. The boxplots used for visual assessment in the participants' analyses are included in Appendix E.

A detailed item analysis was also conducted, and no items were found to be persistently biased toward the same antecedent neither in Catalan, nor in Spanish. It is worth saying, however, that some of the items showed little variation across the eight conditions, as observed in a visual inspection of raw data. Three items could be said to rather favor object interpretations in both languages (two items using the verb *seguir*, 'to follow', and one item using *formar*, 'to prepare'), one item seemed to favor subject interpretations in both languages (using the verb *abandonar*, 'to abandon'), and two items seemed to favor subject interpretations as well, but only in Spanish (both items using the verb *sorprender*, 'to surprise')³⁷. The mean bias of these mentioned items was higher than 70 when biased

³⁵ These participants correspond to the following IDs: BC024, BC035, M059 and M026.

³⁶ The mean bias of all responses for each participant ranged from 31.52 to 76.44.

³⁷ In the materials, these items correspond to the following sentence IDs: FS13, FS23, FS33, and FS45 (affecting both languages), FS31 and FS47 (only affecting Spanish).

toward object antecedents (range: 71.92-78.93), or lower than 30 when biased toward subject antecedents (range: 26.40- 29.19). We ran the same models we report in the results in a dataset in which we had removed these items, but no qualitative differences affecting the answers to our research questions were observed. This is the reason why we concluded that the models accounted for the possible variation within items and decided to keep all the items in the dataset to avoid losing more data. The boxplots used for visual assessment in the items' analyses are included in Appendix F.

In addition to the visual inspection of the data, concerning item analyses, we further explored the data to make sure that thematic role or agentivity were not influencing the results of the analyses. Even though the verbs in the main clause were carefully selected to show neutral implicit causality biases, they took several thematic role structures, and thematic role, particularly agentivity, has been demonstrated to influence anaphora resolution (Schumacher et al., 2016, 2017). Following Goikoetxea et al.'s (2008) classification, which was based on the four-way taxonomy proposed by Rudolph and Försterling (1997), a third of the items in the present study took an agent-patient structure, another third an agent-evocator structure, and another third a stimulus-experiencer structure (representing 31.22% of the items each). In all these structures, the subject antecedent is mapped to an agent-like thematic role, and the object antecedent to a patient-like (i.e., theme) thematic role (see Rissman & Majid, 2019). The psychological verb *suportar/soportar* ('to stand'), was the only one included in the task that took an experiencer-stimulus structure (used in 6.28% of the items, three sentences). In this specific case, the grammatical subject of the sentences is not mapped to an agent-like, but to a patient-like (i.e., experiencer) thematic role. When exploring the data during statistical analyses, the different thematic role structure of the items using *suportar/soportar* was found to be qualitatively affecting the results both in Catalan and in Spanish³⁸. For this reason, these items were removed from the dataset, representing 681 observations and 6.25% of the remaining data³⁹.

Finally, the Qualtrics Survey interface measured the time that participants took to give a response for each item in the task. This information was used to clean the data. All items that had been answered faster than 3 seconds were excluded, since we considered that

³⁸ Goikoetxea et al. (2008) found agent-patient and stimulus-experiencer verbs to overall favor attributions to the subject, and agent-evocator and experiencer-stimulus verbs attributions to the object. In the present data, the thematic role structure of the chosen verbs—with neutral implicit causality biases—does not seem to affect the results.

³⁹ The items using *suportar/soportar* as the main verb correspond to FS14, FS30 and FS46.

this amount of time was not enough to carefully read the sentences and provide an answer accordingly. The observations removed from the dataset were 238, representing an additional 2.32% of the data.

In sum, data removed before conducting statistical analyses represented a total of 1,207 observations and a 10.79% of the amount of data collected (389 observations in Catalan, 8.64% of the data collected in this language, and 809 observations in Spanish, 12.30% of the data in Spanish). The dataset after data cleaning contained 9,977 observations. It is worth mentioning that we checked that a similar number of answers had been removed for each condition in each language, to ensure that data cleaning had not especially penalized any condition.

As will be explained in the following section (§4.5.1.1), undecided answers between 45 and 55 were removed from the dataset to discretize the response variable. The main reason to explain this decision was that participants mainly gave categorical answers when resolving the task. In this context, we opted to remove answers in the middle of the slider scale to avoid attributing them a categorical value, since we interpreted them as undecided. These answers were not associated to any specific condition. They corresponded to 101 observations and represented a 1.01% of the clean dataset. Therefore, the final dataset used for analyzing the results, after data cleaning and after binarizing the responses, consisted of 9,876 observations (a total data loss of 11.70%).

4.5.1.2 Binarization of the responses

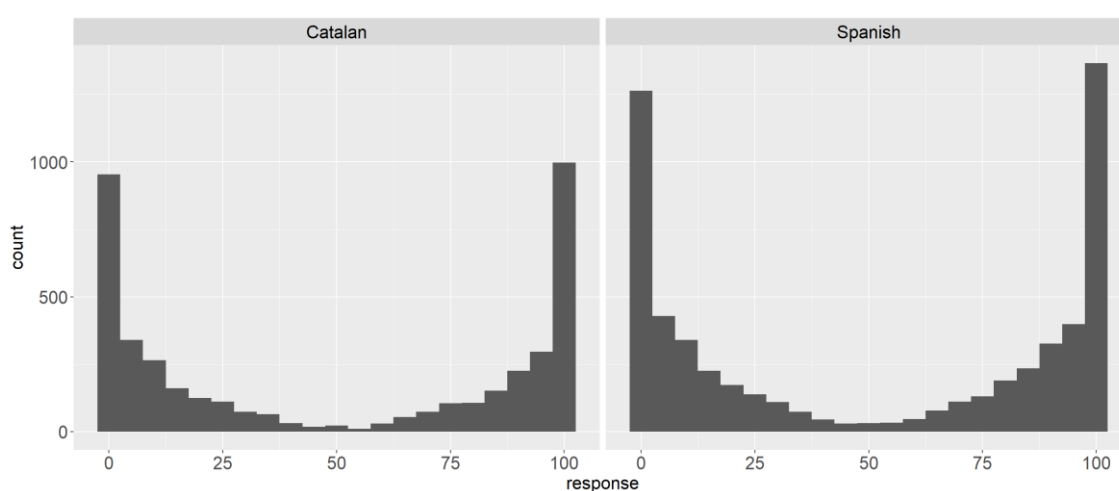
The preference for subject or object antecedents, which was the dependent variable in both Catalan and Spanish tasks, was initially measured on a 0-to-100 visual analogue scale where 0 stood for subject bias and 100 for object bias. Participants were asked to choose a word on the slider displayed in the screen representing a subject or an object antecedent and to show their degree of certainty in a slider scale. The slider scale captured participants' answers as continuous values, since the results were intended to be interpreted as showing gradience in the degree of certainty, or the strength, of interpretation choices (see previous §4.3.1).

However, the answers from participants seem to indicate that they preferred to make rather dichotomous decisions. As shown in Figure 4.7, the histogram representing the distribution of the responses displays a U-shaped or bimodal distribution that reflects a tendency of participants to give dichotomous or categorical answers. More specifically,

in the complete dataset (including answers for both languages), 41.83% of the participants' responses were values between 0 and 20 in the slider scale, that is, very clearly biased toward the subject, and 41.82% of the responses were values between 80 and 100, that is, very clearly biased toward the object. A 7.64% of the responses ranged between 21 and 44 and a 7.71% between 56 and 79. Only a 1.01% of the responses in the clean dataset were comprised between 45 and 55.

Figure 4.7

Histogram of the responses showing the preference toward subject or object antecedents in Catalan and Spanish in a 0-to-100 visual analogue scale

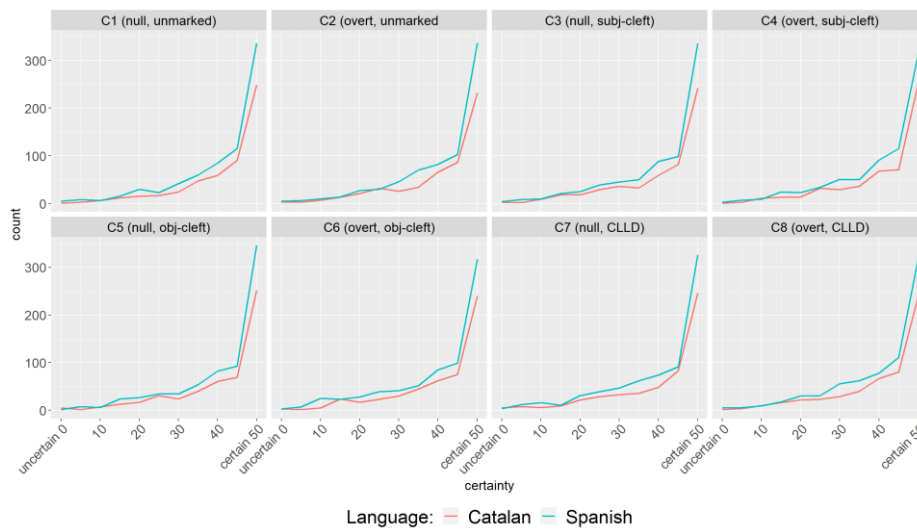


Note. In the response scale, '0' corresponds to subject and '100' to object interpretations. The number of answers in Spanish is higher than Catalan because more participants completed the task in Spanish (Spanish monolinguals).

Figure 4.8 shows that there was not a specific condition favoring more certainty or uncertainty in the responses, or stronger or weaker preferences. On the contrary, it seems that there was a general tendency to provide dichotomous or polarized answers. Participants seem to have responded to the task as if it had been a two-alternative forced choice task.

Figure 4.8

Degree of certainty of the responses to the experimental task per condition



Note. In this figure, the responses in the 0-to-100 scale have been transformed so that '0' in the x-axis, labelled as "uncertain", corresponds to responses at the middle of the scale (undecided answers), and '50', labelled as "certain", corresponds to answers at the edges of the scale (certain answers, irrespectively of whether they refer to subject or object interpretations).

Taking the previous considerations into account, we decided to discretize the response values and treat the dependent variable as binary for further analyses (see Zhang & Davidson, 2021 for a similar design and a similar decision). In this way, the participants' preference for subject or object antecedents' responses was converted to a binary variable in which '1' corresponds to 'subject' choices and '0' to 'object' choices. Undecided responses, with values between 45 and 55 in the slider scale, were removed from the dataset, as anticipated in the previous section (§4.5.1.1).

Finally, it is worth mentioning that continuous responses from the answers provided using the slider scale (the original ones) were first analyzed as a continuous dependent variable using linear mixed-effects models. The assumptions for linear mixed-effects models, however, were not met (Winter, 2019). Not surprisingly, residuals were not normally distributed; the residual plot showed a non-linear pattern, indicating a violation of the linearity assumption and homoskedasticity of residuals and pointing to the existence of binary data. This was an additional reason that confirmed the need to binarize the data and to run logistic regression mixed-effects models, a type of generalized linear models. These models allow for binomial distributions resulting from binary outcomes, so they were more appropriate to account for the data in this study.

4.5.2 Statistical analyses

4.5.2.1 Variables

The dependent variable in this task was the preference to interpret an ambiguous pronominal subject in coreference with a subject or an object antecedent; that is, the number of choices of subject and object antecedents. The analyzed responses were binary: '1' corresponded to 'subject' choices and '0' to 'object' choices. The independent variables were four: Pronoun, Information structure, Group, and Language. Pronoun and Information structure were two within-subjects independent variables (see Table 4.9 in §4.3.2.1 for a summary): Pronoun referred to the form of the ambiguous subject pronoun and had two levels (null and overt), and Information structure⁴⁰ referred to the information structure of the clauses in which subject and object antecedents were presented and had four levels (unmarked structures, object topicalization via clitic-left dislocation, subject focusing via it-cleft, and object focusing via it-cleft). Group was a between-subjects independent variable and had three levels in Catalan (Catalan-dominant bilinguals, balanced bilinguals and Spanish-dominant bilinguals), and four levels in Spanish (Catalan-dominant bilinguals, balanced bilinguals, Spanish-dominant bilinguals, and Spanish monolinguals). Language, which referred to the language in which participants completed the task, had two levels (Catalan and Spanish) and was used as a between-subjects or as a within-subjects variable depending on the aim of the specific analyses (e.g., as a between-subjects factor when comparing Catalan by Catalan-dominant bilinguals vs. Spanish by Spanish-dominant bilinguals, or as a within-subjects factor when comparing the performance in each language by the three bilingual groups, which were tested in both languages).

4.5.2.2 Modeling

4.5.2.2.1 Model fit

Mixed-effects logistic regression models were run in R (v. 4.0.5; R Core Team, 2021) using the *glmer* function of the *lme4* package (v. 1.1.27; Bates, Mächler, Bolker & Walker, 2015) and the *emmeans* package to obtain pairwise contrasts (v. 1.5.5.1; Lenth, 2021). The use of mixed-effects logistic regression models allowed to include fixed effects for Pronoun, Information structure, Group and/or Language, as well as their interactions,

⁴⁰ In the R script, the variable 'Information structure' was labeled as 'Context'.

and random effects for participants and items, when relevant, to adjust their variation (Baayen et al., 2008; Bates et al., 2015; Jaeger, 2008; Winter, 2013, 2019)⁴¹.

Several models were run to analyze the experimental dataset, as summarized in Table 4.11 below. The predictors included in each model were theoretically motivated, based on the research questions that were aimed at answering with each analysis. With respect to the random effects structure, random intercepts by participant and by items were fitted for all models. Regarding random slopes, a series of models were computed including all possible combinations of the relevant random slopes and intercepts. Likelihood ratio tests using the *anova* function of the *stats* package were performed to compare all converging models and to choose the converging model that accounted best for the data. In the reported models, all the assumptions were met: no extreme outliers were identified after standardizing the residuals, and no collinearity was attested. To test collinearity, variance inflation factors (VIF) were used, and their values were always lower than 2.5 (Midi et al., 2010).

Table 4.11

Outline of the mixed-effects logistic regression models computed to account for the preferred interpretation of pronominal subjects

Model	Predictors	Section
1: Effects of marked information structures on pronoun interpretation in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals	<ul style="list-style-type: none"> · Pronoun (null, overt) · Information structure (unmarked, CLLD, subject cleft, object cleft) · Language (Catalan, Spanish) 	Chapter 5 (§5.2, see also Appendix G)
2: Effects of marked information structures on pronoun interpretation in monolingual Spanish	<ul style="list-style-type: none"> · Pronoun (null, overt) · Information structure (unmarked, CLLD, subject cleft, object cleft) 	Chapter 5 (§5.3.1, see also Appendix H)
3: Comparing Catalan by Catalan-dominant bilinguals and Spanish by monolinguals	<ul style="list-style-type: none"> · Pronoun (null, overt) · Information structure (unmarked, CLLD, subject cleft, object cleft) · Group/Language (Catalan-dominant bilinguals in Catalan, monolinguals in Spanish) 	Chapter 5 (§5.3.2, see also Appendix H)

⁴¹ The dataset, the code used to run the analyses in R, and the full outputs of the models are available in the Open Science Framework repository (https://osf.io/nqe5j/?view_only=33edd018f33c424891b5f99507384351).

4: Comparing Spanish by Spanish-dominant bilinguals and monolinguals	<ul style="list-style-type: none"> · Pronoun (null, overt) · Information structure (unmarked, CLLD, subject cleft, object cleft) · Group (Spanish-dominant bilinguals, monolinguals) 	Chapter 5 (§5.3.3, see also Appendix H)
5: Bidirectional analysis of language dominance effects on pronoun interpretation in unmarked and marked information structures	<ul style="list-style-type: none"> · Pronoun (null, overt) · Information structure (unmarked, CLLD, subject cleft, object cleft) · Group (Catalan-dominant, balanced, Spanish-dominant) · Language (Catalan, Spanish) 	Chapter 6 (§6.2 and §6.3, see also Appendix I)

The full model of each analysis is reported in the corresponding chapter and section together with its results. The output of the reported models, as well as the relevant pairwise comparisons, are included in Appendices G, H, and I, as summarized before and as appropriately indicated when they are reported.

4.5.2.2.2 Interpretation of the models' results

The estimates in the results of a logistic regression model are indicated in log-odds (or logits), which express the probability of the second option of two binary responses (e.g., a 1, if the two responses are 0 and 1) on a scale from negative infinity to positive infinity (see Winter, 2019). A log-odds value of 0 corresponds to a probability of 0.5. Given that subject responses in the present study were coded as '1', positive log-odds values will indicate that subject interpretations are more likely to occur than not, and negative log-odds values will indicate that object interpretations are more likely to occur than not (or that subject interpretations are more likely not to occur than to occur).

4.5.2.2.3 Models with no intercept

As we did not only want to compare the effects of each fixed factor on the interpretation of null and overt pronouns, but also to find out whether these pronouns were significantly interpreted as coreferring with subject or object antecedents, the fitted models were additionally run with no intercept. Fitting the model with no intercept forced every category to be compared to zero. Given that in a logistic regression an estimate of zero indicates a fifty-fifty chance, what this model did was comparing each estimate to zero

to let us know which outcomes of each interaction were significantly different from chance. That is, it compared whether subject or object choices—the response variable—in each condition were different from chance.

The following figures (screenshots from Appendix G) aim at illustrating the different outputs obtained by a mixed-effects logistic regression model with a dummy coded category in the intercept level (Figure 4.9), and with no intercept level (Figure 4.10). The intercept in Figure 4.9, which was dummy coded, would correspond to the interpretation of null pronouns in baseline contexts in Catalan. As observed, the positive estimate indicates that the null pronoun represented in the intercept is predicted to be biased toward the subject antecedent ($\beta = 0.960$), and this bias is significantly different from chance ($p < .001$). If we want to know if the overt pronoun in baseline contexts in Catalan is significantly biased toward the object antecedent, we can obtain the value of the estimate from this output (by adding the relevant values to the value of the intercept, in this case -2.232, which would result in an estimate of -1.272). From this output we can also know that the bias of this pronoun is significantly different from the bias of the pronoun in the intercept (a value that can also be obtained using the *emmeans* function), but we would not be able to obtain the p value indicating whether the overt pronoun in baseline contexts in Catalan—or in any other category except from the one in the intercept—is predicted to show a significant bias, different from chance.

Figure 4.9

Screenshot of some of the fixed effects in the output of a mixed-effects logistic regression model with an intercept

```
## Fixed effects:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.96001    0.21874   4.389 1.14e-05 ***
## pronounovert    -2.23180    0.29261  -7.627 2.40e-14 ***
## contextclld-object -0.85463    0.22957  -3.723 0.000197 ***
## contextfocus-object -0.32331    0.23590  -1.371 0.170513
## contextfocus-subject -0.75620    0.23154  -3.266 0.001091 **
```

In the output of the same model run with no intercept, in contrast, the estimates and p -values for each condition are reported. As appreciated in Figure 4.10, null pronouns in baseline contexts in Catalan are significantly biased toward subject antecedents ($\beta = 0.960$, $p < .001$). The values are the same of those in Figure 4.9 because the model is the same, only the information in the output varies. The output of the model with no intercept shows the estimates and p -values for all the other conditions, such as overt

pronouns in baseline contexts in Catalan, which we now certainly know that are significantly biased toward object antecedents ($\beta = -1.272$, $p < .001$).

Figure 4.10

Screenshot of some of the fixed effects in the output of a mixed-effects logistic regression model with no intercept

```
## Fixed effects:
##
## Estimate Std. Error z value Pr(>|z|)
## pronounnull:contextbaseline:languagein_catalan 0.960008 0.218762 4.388 1.14e-05 ***
## pronounovert:contextbaseline:languagein_catalan -1.271784 0.222724 -5.710 1.13e-08 ***
## pronounnull:contextclld-object:languagein_catalan 0.105389 0.206768 0.510 0.61026
## pronounovert:contextclld-object:languagein_catalan 0.069723 0.203242 0.343 0.73156
## pronounnull:contextfocus-object:languagein_catalan 0.636700 0.214737 2.965 0.00303 **
```

In this way, the values in the output of the model with no intercept were used to assess whether null and overt pronouns interpretations were consistently biased toward subject or object antecedents or whether there was not a preferred interpretation different from chance. It is worth highlighting that the formula for these models with no intercept had the exact same structure of both fixed and random effects as the corresponding model⁴². The values obtained through this model with no intercepts would be indeed the values of the intercept's estimate of the reference category in the model with intercepts. These values are the same as those that would be obtained for the intercept by rerunning a model with intercepts as many times as conditions tested changing the reference category of the intercept each time to obtain the values of the estimate for each condition.

4.6 Summary

94 Catalan-Spanish bilinguals participated in the present study (34 Catalan-dominant bilinguals, 31 balanced bilinguals, and 29 Spanish-dominant bilinguals), as well as 40 monolingual speakers of Spanish. To identify the bilingual dominance profile of participants, they had to fill in a language background questionnaire based on the BLP (Birdsong et al., 2012) prior to participating in the experiment. Two equivalent experimental offline tasks were designed in Catalan and in Spanish, and the bilingual

⁴² For instance, the formula for the first reported model (§5.2, see Appendix G) was: `resp2 ~ pronoun * context * language + (1 | sentence_id) + (1 + pronoun | ID)`. The formula used to obtain the estimates of the intercept for each condition was: `resp2 ~ 0 + pronoun:context:language + (1 | sentence_id) + (1 + pronoun | ID)`.

participants completed the two tasks in two separate sessions (Spanish monolinguals only completed the task in Spanish). These tasks tested eight conditions in a 2 × 4 design: pronoun (null, overt), and information structure (unmarked, CLLD, clefted subjects, clefted objects). The experimental task was a two-alternative forced choice plus slider task in which participants had to read sentences including an ambiguous subject pronoun and two plausible antecedents in subject or object positions. After reading the sentences, they had to assign a preferred interpretation to the ambiguous pronoun using a slider scale (with one antecedent at each edge of the scale). Nevertheless, given that the responses were mainly categorical, participants' answers were binarized as if the task had been a two-alternative forced-choice task. The results were analyzed using mixed-effects logistic regressions.

The following chapters report and discuss the results of these tasks. Given the multiple dimensions of this complex design, the results on the impact of marked information structures on pronoun interpretation in Catalan and in Spanish will be first addressed in Chapter 5, where each structure will be separately analyzed to end up integrating the effects of marked information structures in all conditions. In this chapter, Catalan-dominant bilinguals will be taken as the reference group for Catalan, and Spanish-dominant bilinguals as the reference group for Spanish (and they will also be compared to monolingual speakers). The results on the effects of bilingualism on pronoun interpretation in unmarked and marked information structures will be then addressed and discussed in Chapter 6, where all the data by the three bilingual groups in their two languages will be considered.

Chapter 5

Disentangling the role of syntactic, pragmatic, and sequential factors in pronominal anaphora resolution in Catalan and Spanish: results and discussion

5.1 Introduction

The main purpose of this chapter is to investigate how and to what extent syntactic, pragmatic, and sequential factors affect pronoun resolution in Catalan and in Spanish, as well as to compare whether any differences arise between languages. To that end, the interpretation of null and overt pronouns will be assessed in unmarked canonical sentences and in three marked information structures: object topicalization via clitic-left dislocation (CLLD), subject focalization via it-cleft and object focalization via it-cleft. These structures make it possible to disentangle the notions of subjecthood, topicality, and linear position (as well as hierarchical position), which have very often been aligned in the same antecedent in previous research within the framework of the Position of Antecedent Hypothesis (PAH; Carminati, 2002).

Firstly, we aim to describe null and overt pronoun resolution in the language varieties of Catalan and Spanish in Catalonia, as well as how and to what extent syntactic, pragmatic, and sequential (or word order) factors influence these interpretive patterns. To do so, the two extreme poles of language dominance—Catalan-dominant and Spanish-dominant bilinguals—have been taken as a reference to characterize the interpretation of null and overt pronouns in bilingual Catalan and Spanish, respectively. Catalonia is essentially a bilingual society (Idescat & Language Policy Secretariat, 2018; see §3.3.1), where no monolingual speakers are found in either of the two languages (Catalan monolinguals cannot be found outside of Catalonia either). We will therefore be analyzing data from bilinguals. Given that the bilingual Catalan variety in contact with Spanish is described, it is of special interest to contrast it with the Spanish bilingual variety in contact with Catalan. This will allow the comparison of two languages by two groups of bilingual participants that differ in their language dominance profile, but not in their bilingual nature. These two groups (Catalan-dominant and Spanish-dominant bilinguals) are also included in the bilingual spectrum when examining the effects of bilingualism and the role of language dominance in anaphora resolution (in Chapter 6).

Secondly, our aim is also to ascertain whether subject pronoun resolution patterns in (un)marked information structures in the monolingual Spanish variety are similar to those attested in bilingual Spanish by Spanish-dominant bilinguals and in Catalan (by Catalan-dominant bilinguals). With this purpose, monolingual Spanish will also be characterized and then compared to (bilingual) Catalan and bilingual Spanish. In order to identify whether the monolingual and bilingual varieties of Spanish differ from each other, the resolution preferences of these two groups will be compared. Finally, Catalan and monolingual Spanish will also be compared to find out whether there are any differences between the two languages. However, prior to this, the effects of marked information structures in the preference null and overt pronouns for subject and object antecedents are explored in bilingual Catalan and bilingual Spanish.

As such, the present chapter is devoted to analyze in-depth how and to what extent syntactic, pragmatic, and sequential (or word order) factors influence null and overt subject pronoun resolution, and to characterize anaphora resolution in Catalan, bilingual Spanish, and monolingual Spanish, as well as make crosslinguistic comparisons.

5.1.1 Outline of the research questions

With the aforementioned aims in mind, the research questions that guide the present chapter are outlined below. Given the multiple layers that need to be analyzed in order to answer some of the main research questions presented in §1.3, more specific questions have been formulated when relevant to help structure the presentation of the results and discussions.

RQ1 Can the predictions of the Position of Antecedent Hypothesis explain the interpretive biases of null and overt subject pronouns in unmarked structures in bilingual Catalan and bilingual Spanish? (§5.2.3)

RQ2 Are the preferences of null and overt pronouns toward subject and object antecedents in bilingual Catalan and bilingual Spanish affected by the manipulation of information structure? How and to what extent do the syntactic function, information status, and linear position of the antecedent shape null and overt pronoun resolution? (§5.2.4, §5.2.5, and §5.2.6)

- RQ2.1**
- a. How does topicalizing the object via CLLD affect the preference of null and overt pronouns toward subject and object antecedents compared to unmarked structures? (§5.2.4)
 - b. Which are the interpretive biases of null and overt pronouns in CLLD structures? (§5.2.4)
- RQ2.2**
- a. How does focusing the subject or object antecedents via it-cleft structures affect the preference of null and overt pronouns toward subject and object antecedents compared to unmarked structures? (§5.2.5)
 - b. Which are the interpretive biases of null and overt pronouns in subject cleft and object cleft structures? (§5.2.5)
- RQ2.3**
- a. Do null and overt pronouns similarly (dis)prefer subject antecedents conveying focal information in CLLD and subject clefts? (§5.2.6)
 - b. Do null and overt pronouns similarly (dis)prefer first-mentioned object antecedents in CLLD and object clefts? (§5.2.6)
 - c. Do null and overt pronouns similarly (dis)prefer focused subject antecedents in subject clefts and non-focused subject antecedents in object clefts? (§5.2.6)

-
- RQ3** Is microvariation shown in the interpretation of null and overt pronouns in canonical structures in bilingual Catalan and bilingual Spanish? Are the effects of information structure similar in the two languages?
- a. Are null and overt pronouns interpreted similarly in canonical unmarked structures in bilingual Catalan and Spanish? (§5.2.3)
 - b. Are the effects of CLLD different in Catalan and Spanish? (§5.2.4)
 - c. Are the effects of it-clefts different in Catalan and Spanish? (§5.2.5)
-

-
- RQ4** How are null and overt pronouns interpreted in canonical and marked information structures in monolingual Spanish? To what extent do these interpretations differ from (bilingual) Catalan and bilingual Spanish?
- a.** Can the predictions of the Position of Antecedent Hypothesis explain the interpretive biases of null and overt pronouns in unmarked structures in monolingual Spanish? (§5.3.1)
 - b.** Are the preferences of null and overt pronouns toward subject and object antecedents in monolingual Spanish affected by the manipulation of information structure? (§5.3.1)
 - c.** Is microvariation attested in canonical and marked information structures between monolingual Spanish and Catalan? (§5.3.2)
 - d.** Do pronominal anaphora resolution preferences in monolingual Spanish differ from those in bilingual Spanish in contact with Catalan? (§5.3.3)
-

5.1.2 Organization of the present chapter

As previously mentioned, the present chapter is divided into two parts according to its two main purposes. The first part aims to examine the interpretive properties of null and overt pronoun resolution through the manipulation of information structure in the language varieties of Catalan and Spanish in Catalonia (§5.2). In this section, pronoun resolution patterns in each of the scrutinized information structures are addressed separately and finally brought together to analyze how syntactic, pragmatic, and sequential factors shape the interpretation of null and overt pronouns. Once the results in the bilingual varieties of Catalan and Spanish have been presented and discussed, the second part of this chapter aims to characterize monolingual Spanish and compare it to bilingual Catalan and to the bilingual Spanish variety (§5.3). In this section, the biases of null and overt pronouns in (un)marked structures in monolingual Spanish are first described. After identifying resolution patterns in monolinguals, the comparison between monolingual Spanish and Catalan is addressed, as well as the comparison between monolingual Spanish and bilingual Spanish. Overall, RQ1, RQ2, and RQ3 are answered in §5.2, and RQ4 is addressed in §5.3. Finally, a general discussion in §5.4 integrates and discusses the main findings in the present chapter.

5.2 The interpretation of subject pronouns in bilingual Catalan and bilingual Spanish: how do syntactic and pragmatic factors shape null and overt subject pronoun resolution?

In this section we explore how syntactic (i.e., the syntactic function of the antecedents), pragmatic (i.e., the information status of the antecedents), and sequential factors (i.e., the linear surface position of the antecedent, or word order), modulate subject pronoun resolution in bilingual Catalan and bilingual Spanish. To do so, the previous research questions on the interpretation of null and overt pronouns in canonical structures (RQ1) and in marked information structures (RQ2) are addressed, as well as the comparison between the two languages, bilingual Catalan and bilingual Spanish (RQ3). In an attempt to facilitate the interpretation of the answers to these research questions, each structure is addressed separately in a different section.

Firstly, we include a brief reminder of the methodology. We refer to participants (Catalan-dominant bilinguals for Catalan data and Spanish-dominant bilinguals for Spanish data), materials, and we report the only model that was run for the inferential statistical analysis. This model included the answers for all the conditions of the tasks and the two languages (i.e., Pronoun, Information structure and Language as factors). Secondly, an overview of the descriptive results is given, along with the main effects and interactions of the comprehensive statistical model. Thirdly, the results of the study are presented and discussed separately for each structure: 1) canonical unmarked structures, 2) clitic-left dislocation structures (CLLDs), 3) focusing structures (subject focusing via it-cleft and object focusing via it-cleft), and 4) a final section integrates all these findings on marked information contexts by comparing CLLD and focusing structures. Finally, the main findings are summarized at the end of the section.

5.2.1 Method

5.2.1.1 Participants

In order to better understand the linguistic phenomenon under study in each language, Catalan-dominant bilinguals ($N = 34$; mean age: 22.32, range: 18-35) were used as the reference group for Catalan, and Spanish-dominant bilinguals ($N = 29$; mean age: 23.28, range: 19-26) were used as the reference group for Spanish, as explained in the previous section (§5.1). In this way, only Catalan-dominant bilinguals' responses in Catalan and

Spanish-dominant bilinguals' responses in Spanish were included in the analyses to characterize and compare these two languages.

Catalan-dominant bilinguals have been more exposed to Catalan than Spanish and have mainly grown up in Catalan-speaking environments. They all started acquiring Catalan from birth, and Spanish acquisition started either from birth or later (before age 6). At the time of participation in the present study, they used Catalan more often than Spanish across all contexts. They self-rated their proficiency level in Catalan at ceiling in all language skills. They also declared feeling more themselves when speaking Catalan than when speaking Spanish. Spanish-dominant bilinguals show a mirror image. They have been more exposed to Spanish than Catalan and have mainly been raised in Spanish-speaking environments. They all started acquiring Spanish from birth and Catalan either from birth or later (before age 6). They use Spanish to a larger extent than Catalan across all contexts in their day-to-day life. They rate their Spanish proficiency at ceiling and declare feeling more themselves when speaking Spanish. These two groups have been characterized in detail in §4.2 (see also Appendix B).

5.2.1.2 Materials

The experimental tasks completed by participants, a two-alternative forced-choice task in Catalan and an equivalent task in Spanish, have been described in §4.3.2. The two predictors are pronoun (null, overt) and information structure (unmarked structures (5.1), topicalized objects via CLLD (5.2), focused subjects via it-clefts (5.3), and focused objects via it-clefts (5.4)). The results in the present section separately address each information structure. An example for each informational context is provided below.

(5.1) *Unmarked structures*

- a. La Laura va espantar la Maria quan *pro/ella* va entrar a l'habitació. (Catalan)
- b. Laura asustó a María cuando *pro/ella* entró en la habitación. (Spanish)
'Laura scared Maria when she went into the room.'

(5.2) *Clitic-left dislocations*

- a. A la Maria, la va espantar la Laura quan *pro/ella* va entrar a l'habitació.
- b. A María, la asustó Laura cuando *pro/ella* entró en la habitación.
'Maria, Laura scared her when she went into the room.'

(5.3) *Subject clefts*

- a. Va ser la Laura qui va espantar la Maria quan *pro/ella* va entrar a l'habitació.
- b. Fue Laura quien asustó a María cuando *pro/ella* entró en la habitación.
'It was Laura who scared Maria when she went into the room.'

(5.4) *Object clefts*

- a. Va ser a la Maria a qui va espantar la Laura quan *pro/ella* va entrar a l'habitació.
- b. Fue a María a quien asustó Laura cuando *pro/ella* entró en la habitación.
'It was Maria whom Laura scared when she went into the room.'

5.2.1.3 Reported model

We ran a mixed-effects logistic regression with Catalan-dominant bilinguals' responses in Catalan and Spanish-dominant bilinguals' responses in Spanish. The fitted model included Pronoun (null, overt), Information structure (unmarked structures, topicalized objects via CLLD, focused subjects via it-clefts, and focused objects via it-clefts), and Language (Catalan, Spanish) as fixed effects, as well as their interactions. As random effects, varying intercepts for participants and items were added to the model and a by-participant varying slope for the effect of Pronoun. Additional random slopes were tested but they either did not contribute to model fit or led to estimation problems within the models. The summary of this model is provided in Appendix G (Table G.1). The model's total explanatory power was moderate (conditional $R^2 = 0.25$), and the part related to the fixed effects alone (marginal R^2) was 0.09. No multi-collinearity issues were detected (highest VIF = 1.002). This model had a C-index of concordance of 0.77.

The same model with no intercept—with the same fixed effects and random effects' structure—was also computed to obtain the estimates for each condition and to compare whether subject and object antecedent choices of null and overt pronouns differed from chance (see §4.5.2.2). Table G.2 in Appendix G presents the summary of this model. The estimates are given in log odds.

At this point, it is worth mentioning that we are aware that we will be comparing bilingual Catalan and bilingual Spanish from different subjects completing two different tasks in a single model. To control for the variation that having two tasks in two different languages may introduce, it would have been desirable to include a by-item varying slope for the effect of Language in the model. However, including this random slope always led to singular fits. In any case, we meticulously designed the tasks in the two languages to

make them as comparable as possible. The fact that Catalan and Spanish are very close languages, in addition to the fact that the compared structures are parallel in both languages, allowed for the experimental sentences to be translated almost word for word. As such, we consider that the two tasks are similar and comparable enough to conduct between-language comparisons.

5.2.2 General considerations on the results

5.2.2.1 Descriptive results

Table 5.1 summarizes the proportion of subject interpretations of null and overt pronouns given by Catalan-dominants in Catalan and Spanish-dominants in Spanish in the four analyzed contexts (unmarked structures, topicalized objects via CLLD, focused subjects via it-cleft, and focused objects via it-cleft). As can already be observed, the two languages seem to behave similarly in all contexts, although Catalan seems to show more well-defined preferences in unmarked sentences than Spanish. Complementary interpretations of null and overt pronouns seem to be attested only in canonical unmarked sentences; when the information structure of the antecedents' clause is manipulated, the division of labor between null and overt pronouns is attenuated. This is especially the case of structures presenting antecedents in an OVS word order, either in topicalization or in focusing structures; when the object appears in an initial position, null and overt pronouns seem to be in a free distribution or interpreted at chance.

Table 5.1

Proportion of subject interpretations in Catalan (by Catalan-dominant bilinguals) and in Spanish (by Spanish-dominant bilinguals), by information structure and type of pronoun (SD)

	Bilingual Catalan		Bilingual Spanish	
	Null	Overt	Null	Overt
Unmarked	.69 (.46)	.24 (.43)	.60 (.49)	.38 (.49)
Topicalized object	.52 (.50)	.53 (.50)	.50 (.50)	.43 (.50)
Focused subject	.54 (.50)	.33 (.47)	.47 (.50)	.29 (.46)
Focused object	.64 (.48)	.57 (.50)	.63 (.48)	.54 (.50)

5.2.2.2 Main effects and interactions

Several likelihood ratio tests were performed comparing nested models to evaluate the contribution of the predictors and their interactions to the overall fit of the model. The effects revealed by these comparisons are summarized in Appendix G. A model with only random effects was significantly improved by a model with Pronoun as the only fixed effect ($\chi^2(1) = 24.213, p < .001$) and also by a model with Information structure as the only fixed effect ($\chi^2(3) = 54.003, p < .001$). Conversely, Language did not make a significant contribution to the model ($\chi^2(1) = 0.917, p = .338$). The two-way interaction Pronoun \times Information structure also improved a model with no interactions ($\chi^2(3) = 28.299, p < .001$). A model including the three-way interaction Pronoun \times Information structure \times Language was also attested to significantly improve a model including all the two-way interactions ($\chi^2(3) = 10.156, p = .017$). All the reported results of pairwise contrasts in the following sections derive from the three-way interaction between Pronoun, Information structure and Language. The pairwise contrasts can also be found in Tables G.3, G.4, and G.5 in Appendix G.

5.2.2.3 Presentation of the results in the following sections

We present the results separately for each structure to facilitate the presentation, interpretation, and discussion of the results. At the beginning of each section, we reiterate the relevant research questions and we formulate predictions. The relevant pairwise contrasts derived from the three-way interaction Pronoun \times Information structure \times Language are then interpreted. The predicted estimates of subject interpretations are additionally included to identify whether null and overt pronoun's interpretation differs from chance in each structure-language combination. In this way, the results and partial discussions for each context are presented as follows:

- 1) Canonical unmarked structures (§5.2.3): Pairwise contrasts from the perspectives of Pronoun (null vs. overt; RQ1) and Language (Catalan vs. Spanish; RQ3a).
- 2) Clitic-left dislocation structures (§5.2.4): Pairwise contrasts from the perspectives of Information structure (unmarked vs. CLLD structures; RQ2.1a), Pronoun (null vs. overt; RQ2.2a) and Language (Catalan vs. Spanish; RQ3b).
- 3) Focusing via it-cleft structures (§5.2.5, subject clefts and object clefts are presented together): Pairwise contrasts from the perspectives of Information structure (unmarked vs. subject cleft structures, unmarked vs. object cleft

structures, and subject clefts vs. object clefts; RQ2.1b), Pronoun (null vs. overt; RQ2.2b) and Language (Catalan vs. Spanish; RQ3c).

- 4) Integrating findings: marked information structures compared (§5.2.6): Pairwise contrasts from the perspective of Information structure (CLLD vs. subject clefts, RQ2.3a; CLLD vs. object clefts, RQ2.3b).

Finally, the main findings for each structure are brought together in a section that summarizes the main results for each structure, comprehensively relating it to the other analyzed structures, in a summary of our main findings (§5.2.7).

5.2.3 Canonical unmarked structures

5.2.3.1 Aims and predictions

This section aims to identify the interpretive biases of null and overt pronouns in canonical sentences, i.e., with unmarked information structures, illustrated in (5.5), in the Catalan and Spanish bilingual varieties. Hence, it aims to provide an answer to RQ1 and RQ3a, formulated earlier (§5.1.1) and reproduced below. Defining the interpretive preferences of pronouns in canonical sentences is particularly relevant since this condition will be used as a baseline when analyzing the effects of marked information structures on pronoun resolution.

- (5.5) La Laura va espantar la Maria quan *pro/ella* va entrar a l'habitació. (Catalan)
Laura asustó a María cuando *pro/ella* entró en la habitación. (Spanish)
'Laura scared Maria when she went into the room.'

RQ1 Can the predictions of the Position of Antecedent Hypothesis explain the interpretive biases of null and overt subject pronouns in unmarked structures in bilingual Catalan and bilingual Spanish?

Based on previous studies that have tested the validity of the Position of Antecedent Hypothesis (PAH; Carminati, 2002) in intrasentential pronominal anaphora in bilingual Catalan and Spanish, we expect to attest PAH-like biases in both languages (in both Catalan and Spanish: Bel & García-Alcaraz, 2018; in Spanish: Contemori & Di Domenico, 2021; de la Fuente, 2015; de Rocafiguera & Bel, 2022; Filiaci, 2011; García-Alcaraz, 2015; among others). However, some of the aforementioned studies have found

that null pronouns remain unbiased in main-subordinate intrasentential contexts, which are the ones addressed here. According to these studies, the PAH predictions are restricted to subordinate-main anaphora (in both Catalan and Spanish: Bel & García-Alcaraz, 2018; in Spanish: de Rocafiguera & Bel, 2022; Filiaci, 2011). It is therefore possible that null pronouns do not display a clear bias in our data, where main-subordinate anaphora is tested.

RQ3 a. Are null and overt pronouns interpreted similarly in canonical unmarked structures in bilingual Catalan and bilingual Spanish?

The only study on bilingual Catalan and Spanish in contact demonstrated coinciding interpretive behaviors in both languages (Bel & García-Alcaraz, 2018). Parallel biases in the two languages are therefore expected: null pronouns should pick up subject antecedents, and overt pronouns, object antecedents. However, microvariation between null subject languages has also been shown: monolingual Spanish has been found to show less robust patterns than other null subject languages such as Catalan (affecting only overt pronouns: Bel & García-Alcaraz, 2018) or Italian (affecting both pronouns: Contemori & Di Domenico, 2021; affecting only overt pronouns: Filiaci, 2011; Filiaci et al., 2014). If crosslinguistic microvariation is demonstrated in the present study, subject pronouns, and particularly overt pronouns, are predicted to show weaker or milder biases in Spanish than in Catalan.

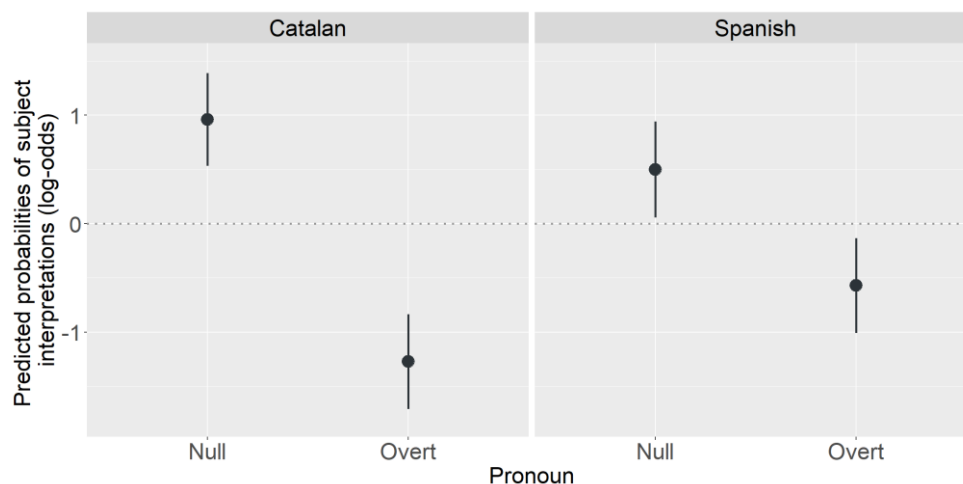
5.2.3.2 Results

The interpretive biases of null and overt pronouns in canonical structures in Catalan and in Spanish are shown in Figure 5.1. This figure, which represents the model estimates, shows the predicted probabilities of interpreting subject pronouns (null or overt) as coreferring with subject antecedents⁴³. As described in the following paragraphs, both languages reproduce the patterns predicted by the PAH: null pronouns show a bias toward subject antecedents, and overt pronouns toward object antecedents.

⁴³ Note that results are presented in log odds. As explained in previous §4.5.2.2, a value of 0 corresponds to chance level (or a 50% of probabilities). If the CI bar illustrated in the graph crosses the line representing chance level, it means that the mean bias represented in the graph (with a point shape) is not reliable or significant. If the predicted probability of subject interpretations for a pronoun is a positive value, it means that this the preferred interpretation of the pronoun is to corefer with subject antecedents. Alternatively, if it is a negative value, its preferred interpretation is to corefer with object antecedents.

Figure 5.1

Predicted subject interpretations of null and overt pronouns in unmarked contexts in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI)



5.2.3.2.1 In Catalan

In informatively neutral contexts in Catalan, as shown in Figure 5.1, the statistical values for the estimates of null and overt pronouns indicated a significant probability of null pronouns to corefer with subject antecedents ($\beta = 0.960$, $p < .001$) and overt pronouns with object antecedents ($\beta = -1.272$, $p < .001$).

Pairwise contrasts in the three-way interaction (Pronoun \times Information structure \times Language)⁴⁴ from the Pronoun's perspective confirmed a significant division of labor between the two pronominal forms in Catalan ($\beta = 2.232$, $p < .001$). In this way, null pronouns were more associated with subject antecedents than overt pronouns, and overt pronouns were more associated with object antecedents than null pronouns.

5.2.3.2.2 In Spanish

In unmarked structures in Spanish, illustrated in Figure 5.1, null and overt pronouns also tended to be resolved following the PAH. The interpretation of null pronouns was biased toward subject antecedents ($\beta = 0.499$, $p = .027$), and the interpretation of overt pronouns was biased toward object antecedents ($\beta = -0.572$, $p < .001$).

⁴⁴ As argued in previous §5.2.2.3, all the results derive from a comprehensive model including all the factors that we are interested in in the present chapter. In the interest of clarity, the results of this general model are being presented by structure and by language.

As corroborated by pairwise comparisons in the three-way interaction from the perspective of Pronoun, the division of labor between null and overt pronouns was also significant ($\beta = 1.070$, $p < .001$).

5.2.3.2.3 Catalan and Spanish compared

The comparison of the two languages under scrutiny also comes from pairwise contrasts in the three-way interaction from the perspective of Language. As can be clearly observed in Figure 5.1 and the previously reported estimates, both pronouns showed more polarized biases in Catalan than in Spanish. Regarding null pronouns, the difference between languages did not reach significance ($\beta = 0.461$, $p = .105$). However, the bias of overt pronouns toward object antecedents was statistically stronger in Catalan than in Spanish ($\beta = -0.700$, $p = .014$). Generally speaking, pronominal resolution biases were more pronounced in Catalan than in Spanish, although statistically significant evidence was only found in favor of overt pronouns.

5.2.3.2.4 Main findings on canonical sentences

Results in both bilingual Catalan and Spanish show that in canonical sentences, null pronouns are preferably interpreted as coreferring with subject antecedents (5.6) and overt pronouns as coreferring with object antecedents (5.7). For overt pronouns, their object bias is stronger in Catalan (by Catalan-dominant bilinguals) than in Spanish (by Spanish-dominant bilinguals).

- (5.6) a. **La Laura_j** va espantar la **Maria_k** quan **pro_j** va entrar a l'habitació.
b. **Laura_j** asustó a **María_k** cuando **pro_j** entró en la habitación.
'Laura scared Maria when (she) went into the room.'
- (5.7) a. La **Laura_j** va espantar la **Maria_k** quan **ella_k** va entrar a l'habitació.
b. **Laura_j** asustó a **María_k** cuando **ella_k** entró en la habitación.
'Laura scared Maria when she went into the room.'

5.2.3.3 Discussion

In both bilingual Catalan and bilingual Spanish, null and overt pronoun resolution in canonical sentences is sensitive to syntactic cues, as predicted by the Position of Antecedent Hypothesis (PAH; Carminati, 2002) (RQ1). Similar interpretive biases are

found in both languages: whereas null pronouns tend to be interpreted as coreferring with antecedents in subject (or Spec IP) position, overt pronouns are prone to be resolved toward antecedents in object or non-subject (or non-Spec IP) position. These preferences are illustrated in (5.8). In line with Bel and García-Alcaraz (2018), these findings confirm that both bilingual Catalan and bilingual Spanish show a clear division of labor between null and overt pronouns.

- (5.8) La Laura_j va espantar la Maria_k quan *pro*/*ella*_k va entrar a l'habitació. (Catalan)
Laura_j asustó a Maria_k cuando *pro*/*ella*_k entró en la habitación. (Spanish)
'Laura scared Maria when she went into the room.'

However, in contrast to Bel and García-Alcaraz (2018), who only found the null pronoun bias in subordinate-main clause order, we found a clear bias of null pronouns toward subject antecedents in main-subordinate contexts. This difference could be related to a task effect (de Rocafiguera & Bel, 2022). While Bel and García-Alcaraz used an acceptability judgment task, other studies using forced-choice tasks have also shown this clear bias of null pronouns in main-subordinate contexts (Contemori & Di Domenico, 2021; de la Fuente, 2015; García-Alcaraz, 2015). Forced-choice tasks tend to increase statistical power to detect differences between conditions and do not seem to capture subtle nuances to the same degree as acceptability judgments (Schütze & Sprouse, 2014). Clause order effects may therefore remain imperceptible in main-subordinate contexts if a forced-choice task is used, as in the present study.

Overt pronouns exhibit a clear interpretive bias toward object antecedents in both Catalan and bilingual Spanish. These results differ from previous studies on monolingual Spanish that had shown overt pronouns to be flexible in their preferences (for Peninsular Spanish: Bel & García-Alcaraz, 2018; Filiaci et al., 2014; for Mexican Spanish: Keating et al., 2016). Bel and García-Alcaraz (2018) found more well-definite PAH-like biases in bilingual Spanish compared to monolingual Spanish and hypothesized that crosslinguistic influence of Catalan on Spanish in the bilingual contact variety could explain this difference. To what extent the attested pattern is a consequence of Catalan influence on bilingual Spanish—or represents a genuine tendency in Spanish—will be discerned further on. It is initially explored in §5.3.3, where monolingual and bilingual data in Spanish are compared directly. It is then analyzed in more detail in Chapter 6, where bilingualism effects are carefully examined. For the moment, as detailed in the next paragraph, overt pronouns have been found to be significantly more biased toward object antecedents in bilingual Catalan than in bilingual Spanish.

Even though PAH-like pronominal biases are attested in both languages, Catalan seems to show more polarized resolution preferences than Spanish in the interpretation of subject pronouns, especially concerning overt pronouns (RQ3a). Therefore, these two null subject languages diverge in the strength of the interpretive patterns, rather than in their resolution biases. Results show that the interpretation of null pronouns in the two languages does not differ significantly, although their subject bias seems to be stronger in Catalan. The overt pronoun preference toward object antecedents is significantly stronger in Catalan than in Spanish. This contrast between Catalan and Spanish is similar to that attested by Contemori and Di Domenico (2021) between Italian and Spanish through a three-alternative forced-choice task. In their study, the interpretive biases of both null and overt pronouns were found to be weaker in Mexican Spanish than in Italian, although the two languages complied with the PAH. Giannakou and Sitaridou (2020), in an offline comprehension questionnaire, found both null and overt pronouns to have more flexible coreference patterns in Chilean Spanish (which did not comply with the PAH) than in Greek (in which well-defined PAH-like biases were attested). Torregrossa et al. (2020), using an acceptability judgment task, found that pronouns in Italian are also more constrained in their reference possibilities compared to Greek. In this sense, Catalan seems to be closer to Italian as far as the PAH is concerned. The present study on Catalan and Spanish expands previous evidence of crosslinguistic microvariation between typologically similar null subject languages. In this regard, null and overt pronouns seem to display different degrees of flexibility in their distribution depending on the language. We will explore this idea further in the general discussion.

5.2.4 Topicalization structures via clitic-left dislocation

5.2.4.1 Aims and predictions

After analyzing unmarked canonical sentences, a question that has been recurrent in the literature on anaphora resolution remains open: are the preferences of null and overt pronouns guided by structural or configurational syntactic constraints, by pragmatic constraints, or by both syntactic and pragmatic constraints? That is, are null pronouns biased toward subject antecedents and overt pronouns toward object antecedents, or rather are they specialized in conveying topic continuity and topic shift (or pragmatic/informational related concepts), respectively? Pragmatically unmarked sentences, even if they appear out-of-the-blue, also have an information structure; the subject is foreseeably read as conveying topical information, and the predicate as

conveying focal information (see §2.4.1). Hence, in canonical sentences, the subject and the sentence topic coincide in the initial surface position (at least in fragmented experimental sentences such as those used here). Therefore, three main factors that have been proposed to influence anaphora resolution overlap: subjecthood, topicality, and order of mention. This section aims to explore the extent to which each of these cues influence pronoun resolution, and/or whether there is a more heavily weighted factor that guides their interpretation.

For this purpose, topicalization structures via clitic-left dislocation (CLLD), as in (5.9), will first be analyzed and compared to canonical sentences as a baseline condition. As detailed in §2.4.1.2, CLLDs allow the notion of subject to be extricated from that of sentence topic, and also from that of initial position. In these contexts, the object is syntactically marked as a topic and appears first-mentioned, in an initial preverbal position. At the same time, the subject conveys focal information and appears second-mentioned, in a postverbal position.

- (5.9) a. A la Maria la va espantar la Laura quan *pro/ella* va entrar a l'habitació.
b. A María la asustó Laura cuando *pro/ella* entró en la habitación.
'Maria, Laura scared her when she went into the room.'

In this section, we will address the research questions outlined below (see §5.1.1). To answer RQ2.1a, the preference of null and overt pronouns for subject and object antecedents in CLLD and canonical sentences will be compared. To answer RQ2.1b, we will look at whether null and overt pronouns show clear interpretive biases in CLLD contexts. To answer RQ3b, we will compare Catalan and Spanish data on CLLDs.

RQ2.1 a. How does topicalizing the object via CLLD affect the preference of null and overt pronouns toward subject and object antecedents compared to unmarked structures?

A number of proposals have highlighted discourse factors such as the information status of the plausible antecedents as being crucial for pronoun resolution (e.g., de la Fuente, 2015; Ellert, 2013; Papadopoulou et al., 2015). Based on these studies, the prediction is that the preference of null and overt pronouns for subject and object antecedents will be sensitive to information structure. As has been generally assumed across null subject languages (e.g., García-Alcaraz & Bel, 2019; Lozano, 2009; Sorace et al., 2009), we expect null pronouns to be more prone to convey topic maintenance. In CLLDs,

topicalized object antecedents should become more prominent for null pronouns, compared to unmarked objects in canonical sentences. We also expect overt pronouns to be more prone to convey topic shift, and therefore topicalized objects to become less preferred for overt pronouns compared to canonical objects.

RQ2.1 b. Which are the interpretive biases of null and overt pronouns in CLLD structures?

In CLLD contexts, subjecthood and topicality are not aligned. If null pronouns are specialized in topic maintenance and overt pronouns in topic shift, and information-structural properties are more determining than syntactic properties, null pronouns should prefer topical object antecedents and overt pronouns should prefer non-topical subject antecedents. However, based on Carminati's (2002) predictions, the preference for topicalized objects (in Spec,TopP) is predicted to compete with—but not to override—the preference for subject antecedents (in Spec,IP). Manipulating the information status of the antecedents may affect the intrinsic biases of null and overt pronouns in CLLDs compared to unmarked sentences, but it should not reverse them: topics and subjects should be similarly (un)accessible. These results would be in line with experimental findings such as those in Colonna et al. (2012) for German.

RQ3 b. Are the effects of CLLD different in Catalan and Spanish?

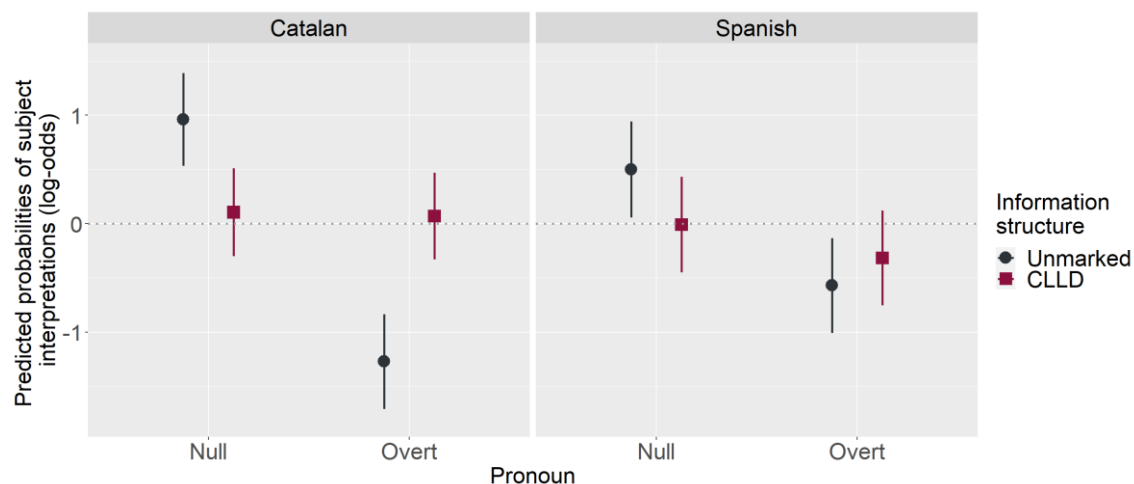
CLLD constructions are not predicted to affect Catalan and Spanish differently. Information structure categories are taken to be universal (although not their linguistic encoding or syntactic reflexes in the grammatical systems) and the syntactic structure of CLLD in the two languages is essentially the same.

5.2.4.2 Results

In CLLDs, object antecedents are first-mentioned and topicalized, whereas subject antecedents appear second-mentioned and convey focus information. The results for null and overt pronouns in this context, together with the baseline condition (unmarked canonical structures), are shown in Figure 5.2, both in Catalan and in Spanish.

Figure 5.2

Predicted subject interpretations of null and overt pronouns in unmarked and CLLD structures, in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI)



5.2.4.2.1 In Catalan

In CLLD contexts, pairwise contrasts of the three-way interaction from the perspective of Information structure revealed that the preference of null pronouns toward topicalized objects significantly increased as compared to objects in unmarked sentences ($\beta = 0.855$, $p = .001$). However, objects conveying a topical information status were still not preferred over subject antecedents. Figure 5.2 shows null pronouns were interpreted at chance level ($\beta = 0.105$, $p = .610$), not showing any bias.

Opposite to null pronouns, the interpretation of overt pronouns as coreferring with object antecedents significantly decreased in CLLD compared to baseline ($\beta = -1.342$, $p < .001$). At the same time, the preference of overt pronouns for topicalized objects increased, and their preference for non-topical subjects decreased. This strong weakening of their object bias drove overt pronouns to remain unbiased. As shown in Figure 5.2, overt pronouns were also interpreted at chance level ($\beta = 0.070$, $p = .732$).

Null and overt pronouns showed no interpretive biases in CLLD contexts: they freely, at chance, picked up subject and object antecedents. Pairwise contrasts from the Pronoun's perspective revealed no differences between the interpretations of the two pronouns, confirming that they were very similarly interpreted ($\beta = 0.038$, $p = .894$).

5.2.4.2.2 *In Spanish*

As in Catalan, as can be observed in Figure 5.2, the interpretive choices of null and overt pronouns in CLLDs in Spanish did not differ from chance (null pronouns: $\beta = -0.009$, $p = .970$; overt pronouns: $\beta = -0.317$, $p = .154$). However, regarding the impact of CLLD constructions in pairwise contrasts from the perspective of Information structure, no significant differences emerged as compared to canonical sentences. Subject antecedents became less accessible for null pronouns ($\beta = 0.507$, $p = .164$) and more accessible for overt pronouns ($\beta = -0.255$, $p = .741$). These tendencies are similar to those reported for Catalan, but the difference between the estimates is smaller and non-significant.

From the Pronoun's perspective, no differences were attested when comparing the interpretation of both pronouns ($\beta = 0.309$, $p = .300$). In CLLD contexts, null and overt pronouns similarly show no clear specialization.

5.2.4.2.3 *Catalan and Spanish compared*

We have seen that, in Catalan, CLLD structures significantly impacted pronominal resolution as compared to the baseline condition. In Spanish, differences between unmarked and CLLD contexts did not reach statistical significance, although null and overt pronouns showed indeterminacy in their final interpretations, similarly to Catalan. When comparing the effects of CLLD in Catalan and Spanish in pairwise comparisons from the perspective of Language, no significant differences between languages were attested (null pronouns: $\beta = 0.114$, $p = .679$; overt pronouns: $\beta = 0.387$, $p = .152$). CLLD constructions seem to have similar effects on the interpretation subject pronouns in the two languages.

5.2.4.2.4 *Main findings on clitic-left dislocation*

Results show that topicalizing the object antecedent via CLLD has a different impact on null and overt pronouns. Compared to canonical sentences, CLLD makes the preference for topicalized object antecedents increase for null pronouns and decrease for overt pronouns. In these structures, object antecedents convey topical information and appear first-mentioned, altering the canonical SVO word order. Subject antecedents, oppositely, convey focal information and appear second-mentioned (see §2.4.1.2 and §2.4.6).

In this way, the preference of null pronouns for subject antecedents seems to be weakened when the subject conveys focal (i.e., non-topical) information and is not first-mentioned. As shown in (5.10), null pronouns show no interpretive biases.

- (5.10) a. A la Maria_k la va espantar la Laura_j quan **pro**_{jk} va entrar a l'habitació.
b. A María_k la asustó Laura_j cuando **pro**_{jk} entró en la habitación.
'Maria, Laura scared her when (she) went into the room.'

On the other hand, the preference of overt pronouns for object antecedents is weakened when the object is topicalized and first-mentioned. As shown in (5.11), overt pronouns show no interpretive biases.

- (5.11) a. A la Maria_k la va espantar la Laura_j quan **ella**_{jk} va entrar a l'habitació.
b. A María_k la asustó Laura_j cuando **ella**_{jk} entró en la habitación.
'Maria, Laura scared her when she went into the room.'

As discussed in the following paragraphs, the increase of the preference of null pronouns for topical objects fits with their discourse nature to mainly express topic continuity, and the decrease of the preference of overt pronouns for topical objects fits with their nature to express mainly topic shift. We do not know whether these results are primarily attributable to the pragmatic function of each pronoun—and the information status of the antecedents—or to the surface position of the antecedents. The subject position was disentangled from the topic and initial surface positions, but topicality and order of mention still overlap in CLLD structures. Results are similar in Catalan and Spanish, but only non-significant effects were found in Spanish, so Catalan may show a stronger sensitivity to manipulated information structure.

5.2.4.3 Discussion

5.2.4.3.1 *The effects of CLLD structures on the preferences of null and overt pronouns compared to canonical sentences*

The findings in Catalan and Spanish show a different impact of the topicalization of object antecedents via CLLD on the interpretation of null and overt pronouns, as predicted (RQ2.1a). In CLLD structures, subject antecedents are not sentence topics and appear second-mentioned (in a postverbal position), whereas object antecedents are topical and first-mentioned. Compared to unmarked contexts, in CLLDs, objects become more preferred for null pronouns, but less preferred for overt pronouns. In line with de la

Fuente's (2015) findings in Spanish, the interpretations of null pronouns as coreferring with object antecedents increased when this antecedent was topicalized. In his study, de la Fuente analyzed hanging topic left dislocation structures, which appear to lead to similar effects to those of CLLD structures. The findings of the present study expand those of de la Fuente by demonstrating the opposite effect on overt pronouns, whose interpretations as coreferring with object antecedents decreased when the object was topicalized.

This different impact of CLLD on the interpretation of null and overt pronouns could be compatible with proposals that consider information structure to be a crucial factor modulating pronominal anaphora resolution. In the present study, null pronouns accept object interpretations more easily if objects are in a topical position, while overt pronouns accept subject interpretations more easily if subjects are in a non-topical position. The (natural) topic continuity and topic shift features that null and overt pronouns seem to have, respectively, may underlie these changes in pronoun interpretation. If we assume that the partition between null and overt pronouns can be projected into the partition between personal and demonstrative pronouns in non-null subject languages, our results resemble those obtained in other studies. For instance, in German, personal pronouns (*er*) have been argued to prefer topical referents, while demonstrative pronouns (*der*) tend to avoid topical referents (Bosch & Hinterwimmer, 2016; Bosch & Umbach, 2007).

It is worth noting, however, that the same effects could also be attributed to order of mention (e.g., Gernsbacher et al., 1989), given that the initial surface position overlaps with the notion of topic. The analysis of focalization structures via it-cleft in §5.2.5 will provide new evidence to disentangle the role of information status and word order factors. As we will see, information-structural factors (i.e., topic maintenance and topic shift features) will be revealed as insufficient for explaining null and overt pronoun coreference patterns.

5.2.4.3.2 The interpretation of null and overt pronouns in CLLD structures

Although both null and overt pronouns are sensitive to the information status of the antecedent (or order of mention linked to topicalization), pragmatic factors do not appear to drive pronoun resolution on their own (RQ2.1b). Null pronouns did not resolve toward topical object antecedents, nor did overt pronouns resolve toward non-topical subject antecedents; their preferences only approached chance level. Therefore, topicalizing the

object does not reverse pronominal interpretive biases in canonical contexts, but rather cancels them. As predicted by Carminati (2002), object antecedents appear to compete with subject antecedents when objects are topicalized and thus occupy a higher specifier syntactic position (Spec,TopP) than subject antecedents (Spec,IP). However, the preference for referents in Spec,TopP does not override the preference for referents in Spec IP. Mirroring these effects on null pronouns, placing the object in higher positions in the syntactic configuration makes this antecedent less preferred for overt pronouns, but they still do not reject the object antecedent in Spec,TopP.

These findings suggest that pronominal anaphora resolution is affected by an interaction of multiple constraints, as widely attested in previous literature. Topicality and/or word order (topical first-mention vs. non-topical second-mention) influence the interpretation of null and overt pronouns in Catalan and Spanish, but the syntactic function of the antecedent (subject vs. object) still plays a role in the interpretive decision. None of the studied factors appear to have a more decisive role over the others: in CLLD structures, information status (or word order, given that the topical antecedent appears in an initial surface position) seems to have a similar weighting to that of syntactic function. This resemblance in weightings results in a mitigation of the interpretive biases of null and overt pronouns when these cues are confronted (rather than aligned), as suggested in other studies addressing competing cues, such as Colonna et al. (2012) or Schumacher et al. (2016).

5.2.4.3.3 No microvariation shown between Catalan and Spanish

Finally, no differences arise between Catalan and Spanish regarding the effects of CLLD on pronoun resolution (RQ3b). In contrast to Catalan, however, the effects of CLLD structures in Spanish do not reach statistical significance. This lack of a clear topicalization effect in Spanish could be related to the fact that pronoun preferences in unmarked contexts also showed a milder polarization compared to Catalan. Although the comparison of CLLD and unmarked structures did not reach significance in Spanish, null and overt pronouns show unbiased interpretations in CLLD contexts in both Catalan and Spanish, and no significant differences between languages are demonstrated. We consider it highly unlikely that there is an effect on one language and not on the other. A better explanation would be to say that the effect size in Catalan for this difference is larger and therefore it is more likely to find a significant impact of CLLD in Catalan than in Spanish.

It is true, however, that a difference in the strength of the effects could also be related to a more extensive use of dislocations in Catalan than in Spanish (Adli, 2011; Vanrell & Fernández-Soriano, 2013; Villalba, 2007). As hypothesized by Colonna et al. (2012), the frequency of a specific construction could affect its impact on pronoun resolution. In their study, Colonna et al. found a clear effect of topicalization enhancing the accessibility of an antecedent in German, but not in French, which they related to the fact that the construction they were testing (hanging topicalization) was not very frequent in French. In any case, little is known about the possible effect of frequency and its scope, making us rather cautious in this regard.

5.2.5 Focusing subjects and objects via it-cleft structures

5.2.5.1 Aims and predictions

The main purpose of testing subject and object it-cleft focus constructions is that these structures will allow to disentangle the role of pragmatic factors and word order factors (in addition to syntactic factors) in pronominal anaphora resolution. In clitic-left dislocation (CLLD) structures, analyzed in the previous section (§5.2.4), topicality was linked to first-mention, as in canonical sentences. In contrast to canonical sentences and CLLDs, in focus structures via it-clefts, the first-mentioned antecedent does not align with the topic of the sentence, but rather with the focus.

In subject clefts, as in (5.12), the subject antecedent appears in initial position, as in canonical sentences, but it is not the sentence topic (i.e., it is the focus). The object antecedent appears second-mentioned, in a postverbal position, and is part of the background in a focus-background configuration, thus conveying non-focal (i.e., topical) information.

In object clefts, as in (5.13), the object is the first-mentioned antecedent, as in CLLD structures, but it conveys focal information. Clefted objects can be considered to be similar in informative terms to objects in unmarked structures (i.e., they both convey focal information), but different in structural or configurational terms. As in CLLD structures, in object clefts, the subject occupies a postverbal position, but in this case, it is part of the background and holds a topical information status (see §2.4.1.2 and §2.4.6).

(5.12) Subject clefts

- a. *Va ser la Laura qui va espantar la Maria quan pro/ella va entrar a l'habitació.*

b. Fue Laura quien asustó a María cuando *pro/ella* entró en la habitación.

'It was Laura who scared Maria when she went into the room.'

(5.13) *Object clefts*

a. Va ser a la Maria a qui va espantar la Laura quan *pro/ella* va entrar a l'habitació.

b. Fue a María a quien asustó Laura cuando *pro/ella* entró en la habitación.

'It was Maria whom Laura scared when she went into the room.'

Building on previous results regarding CLLDs, the analyses of cleft structures will shed some light on the weight of discourse factors and sequential factors. In this section, we will address the following research questions, from §5.1.1 above, and the corresponding predictions. To answer RQ2.2a, the preference of null and overt pronouns for subject and object antecedents in subject clefts and object clefts will be compared to their preference in unmarked structures. To answer RQ2.2b, we will examine whether null and overt pronouns show biased interpretations in each it-cleft structure.

RQ2.2 a. How does focusing the subject or the object antecedents via it-cleft structures affect the preference of null and overt pronouns toward subject and object antecedents compared to unmarked structures?

In light of studies such as de la Fuente (2015), Colonna et al. (2012, 2015) or Patterson et al. (2017), it-cleft structures in intrasentential contexts should lead to an anti-focus effect, affecting, at least, null pronouns. As such, null pronouns are expected to disprefer focused antecedents. Overt pronouns have never been studied before in these contexts, but we would expect them to display a complementary function to that of null pronouns. Overt pronouns are thus predicted to prefer clefted antecedents, maybe to a different extent depending on whether they are subjects or objects. Their preference for clefted subject antecedents may increase with respect to canonical subjects, and their preference for clefted object antecedents may be similar or even stronger than for canonical objects, given the alignment between object and focus features—or syntactic and pragmatic cues.

Another interesting comparison for answering this research question is the contrast between subject and object clefts. If an anti-focus effect emerges that affects null pronouns, and clefted antecedents are complementarily enhanced for overt pronouns, no differences in the preference for clefted antecedents (subjects or objects) should be found. In other words, both subject and object antecedents in cleft positions should be similarly dispreferred for null pronouns and similarly preferred for overt pronouns.

RQ2.2 b. Which are the interpretive biases of null and overt pronouns in subject cleft and object cleft structures?

As in CLLDs, pragmatic and syntactic factors may interact. Null and overt pronouns may not show clear preferences when encountering confronting cues (e.g., Blything et al., 2021; Colonna et al., 2012), and may show enhanced preferences when encountering aligned cues. If it-cleft structures generate an anti-focus effect on null pronouns, they are expected to remain unbiased or even to disprefer focused subject antecedents in subject clefts (confronting cues). In object clefts, null pronouns should maintain their subject bias (aligned cues). If it-cleft structures enhance the preference of overt pronouns for focused antecedents, overt pronouns are expected to remain unbiased or even to prefer focused subject antecedents in subject clefts (confronting cues). In object clefts, the preference of overt pronouns for objects should remain the same or may even be enhanced (aligned cues).

RQ3 c. Are the effects of it-clefts different in Catalan and Spanish?

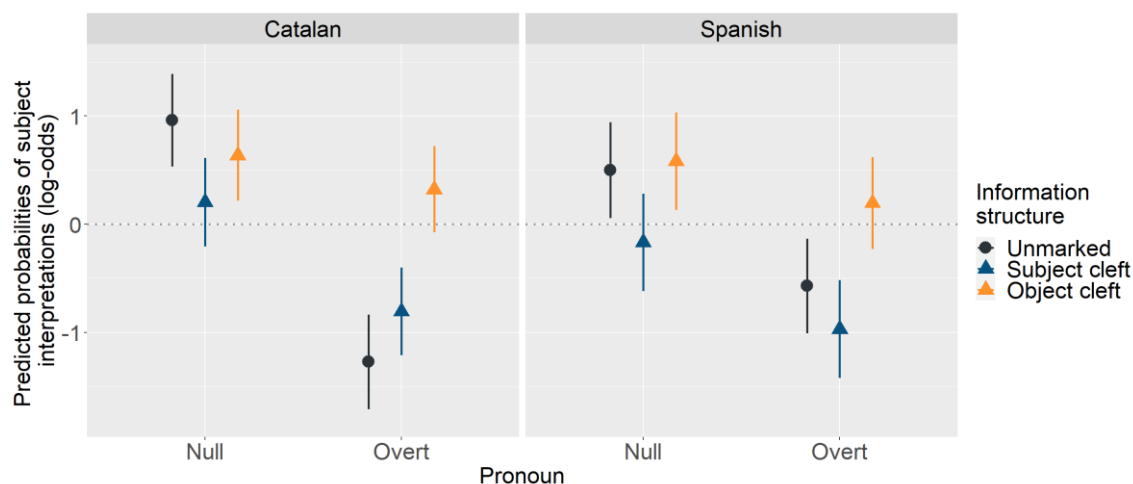
Given that the syntactic structure behind clefts in Catalan and in Spanish is essentially the same, no differences are expected between the two languages. Taking into account that the two languages differed in the strength of the interpretive preferences of pronouns in canonical contexts, we might expect stronger effects of it-cleft structures in Catalan than in Spanish, as found in CLLD structures.

5.2.5.2 Results

The results on the interpretation of null and overt pronouns in subject and object clefts, together with the baseline condition (unmarked canonical structures) are shown in Figure 5.3, both in Catalan and in Spanish. In subject clefts, as already mentioned, subject antecedents hold a focal information status and appear first-mentioned and in a preverbal position. In contrast, object antecedents convey non-focal information and appear second-mentioned in a postverbal position (SVO). In object clefts, object antecedents are focal and first-mentioned or preverbal, whereas subject antecedents are non-focal and second-mentioned or postverbal (OVS).

Figure 5.3

Predicted subject interpretations of null and overt pronouns in unmarked, subject cleft and object cleft structures, in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals (± 95 CI)



5.2.5.2.1 In Catalan

Regarding the effects of it-cleft structures on null pronouns, when the subject was focused, subject antecedents became significantly less preferred compared to unmarked structures ($\beta = 0.756$, $p = .006$). This decrease in the preference of null pronouns for subject antecedents, as illustrated in Figure 5.3, made them lose their subject bias and be interpreted at chance level ($\beta = 0.204$, $p = .329$). When the object was focused, the bias of null pronouns did not differ from unmarked structures ($\beta = 0.323$, $p = .518$) and they kept a clear preference for subject antecedents ($\beta = 0.637$, $p = .003$). The difference between the two focalization structures was not significant ($\beta = -0.433$, $p = .227$), although the point estimate for null pronouns' subject interpretations was higher in object clefts (i.e., when the subject was not focused) than in subject clefts.

The effects of it-cleft structures on overt pronouns were contrary to the formulated predictions. The preference of overt pronouns for subject antecedents in the baseline condition did not increase when the subject was focused. When the subject was clefted, no significant effects were attested compared to unmarked structures ($\beta = -0.465$, $p = .221$) and overt pronouns maintained their preference for object antecedents ($\beta = -0.806$, $p < .001$). When the object was focused, the interpretations of overt pronouns as coreferring with object antecedents significantly decreased compared to unmarked

structures ($\beta = -1.593$, $p < .001$). Thus, focusing object antecedents resulted in unbiased interpretations ($\beta = 0.321$, $p = .114$), even showing a non-significant trend toward subject antecedents. Comparisons of the two focus structures revealed significant differences for overt pronouns: object antecedents were significantly more accessible for overt pronouns when they were not focused and second-mentioned (i.e., in subject clefts) than when they were focused and first-mentioned (i.e., in object clefts) ($\beta = 1.128$, $p < .001$).

Looking at the results of pairwise comparisons of the three-way interaction from the perspective of Pronoun, they indicated that the complementary distribution of null and overt pronouns was only significant when the subject was focused in clefted structures ($\beta = 1.011$, $p < .001$), similarly to unmarked structures (§5.2.3.2). Subject clefts and unmarked structures display an SVO word order (the subject is preverbal and first-mentioned). This division of labor was not attested when the object was clefted ($\beta = 0.315$, $p = .251$), a structure displaying an OVS word order (the subject is postverbal and second-mentioned).

5.2.5.2.2 *In Spanish*

In Spanish, results on how focusing an antecedent can affect pronoun resolution were very similar to those of Catalan, as shown in previous Figure 5.3. In what refers to null pronouns, focused subjects became significantly less preferred than canonical subjects ($\beta = 0.666$, $p = .039$). Thus, when the subject was clefted, null pronouns were interpreted at chance ($\beta = -0.167$, $p = .466$). On the other hand, focusing the object antecedent did not affect the interpretation of null pronouns ($\beta = -0.084$, $p = .987$). Null pronouns remained biased toward subject antecedents ($\beta = 0.970$, $p < .001$), as in canonical sentences (and this bias was even strengthened, although not significantly). When comparing the two focusing constructions, second-mentioned non-focal subject antecedents in object clefts were significantly more preferred for null pronouns than first-mentioned focal subject antecedents in subject clefts ($\beta = -0.750$, $p = .016$).

Concerning overt pronouns, their preference for object antecedents in unmarked constructions was not affected by subject clefts ($\beta = 0.398$, $p = .410$). When the subject was focused, overt pronouns kept their clear-cut bias for object antecedents ($\beta = -0.970$, $p < .001$). When the object was focused, however, the preference of overt pronouns for object antecedents significantly decreased compared to unmarked structures ($\beta = -0.766$, $p = .010$) and they lost their object bias ($\beta = 0.195$, $p = .369$). When comparing

subject clefts and object clefts, object antecedents were found to be significantly more preferred for overt pronouns in postverbal and non-focal positions (i.e., in subject clefts) than in initial and focal positions (i.e., in object clefts) ($\beta = -1.165, p < .001$).

Regarding pairwise contrasts of the three-way interaction from the Pronoun's perspective, as in Catalan, the preferences of null and overt pronouns only displayed a clear division of labor in SVO orders: in subject clefts ($\beta = 0.803, p = .009$), and in canonical sentences ($\beta = 1.070, p < .001$). Significant differences between the two pronouns were not attested in OVS object clefts ($\beta = 0.388, p = .192$).

5.2.5.2.3 Catalan and Spanish compared

As shown by the results for Catalan and Spanish, we see similar effects on pronominal resolution in the two languages for subject and object focalization structures. Pairwise contrasts from the perspective of Language on the impact of cleft structures did not reveal any significant differences between Catalan and Spanish, neither on null pronouns (subject clefts: $\beta = 0.371, p = .186$; object clefts: $\beta = 0.054, p = .851$) nor on overt pronouns (subject clefts: $\beta = 0.163, p = .559$; object clefts: $\beta = 0.127, p = .633$).

5.2.5.2.4 Main findings on subject clefts and object clefts

Firstly, focusing a subject antecedent via it-cleft makes it convey focal information and appear first-mentioned. As for the object antecedent, it conveys background (i.e., non-focal, presupposed) information and appears second-mentioned in its canonical postverbal position.

Results showed that clefting a subject antecedent made it less preferred for null pronouns but did not affect the preferences of overt pronouns. Consequently, as shown in (5.14), null pronouns were interpreted as unbiased (the information status of the subject antecedent is focal). Overt pronouns kept their object bias, as shown in (5.15), preferring coreference with non-focal object antecedents and dispreferring focused subject antecedents.

- (5.14) a. Va ser la Laura_j qui va espantar la Maria_k quan *pro*_{j/k} va entrar a l'habitació.
b. Fue Laura_j quien asustó a María_k cuando *pro*_{j/k} entró en la habitación.
'It was Laura who scared Maria when (she) went into the room.'

- (5.15) a. Va ser la Laura_j qui va espantar **la Maria**_k quan **ella**_k va entrar a l'habitació.
 b. Fue Laura_j quien asustó a **María**_k cuando **ella**_k entró en la habitación.
 'It was Laura who scared Maria when she went into the room.'

Secondly, focusing an object antecedent via it-cleft makes the subject antecedent convey background (i.e., non-focal) information and appear second-mentioned. Object antecedents, conversely, convey focal information and appear first-mentioned.

In this context, as illustrated in (5.16), null pronouns show a clear bias toward non-focal subject antecedents (no conflicting cues are associated with subjects between syntactic and discourse factors). Overt pronouns, however, as in (5.17), remain unbiased in object-clefts, showing that even if the object conveys focus, it is not preferred over the non-focal subject antecedent.

- (5.16) a. Va ser a la Maria_k a qui va espantar **la Laura**_j quan **pro**_j va entrar a l'habitació.
 b. Fue a María_k a quien asustó **Laura**_j cuando **pro**_j entró en la habitación.
 'It was Maria whom Laura scared when (she) went into the room.'

- (5.17) a. Va ser a la Maria_k a qui va espantar la Laura_j quan ella_{j/k} va entrar a l'habitació.
 b. Fue a María_k a quien asustó Laura_j cuando ella_{j/k} entró en la habitación.
 'It was Maria whom Laura scared when she went into the room.'

Overall, the results show that focusing an antecedent via it-cleft does not seem to enhance its "prominence" for either null or overt pronouns but rather decreases it. The effects of it-cleft structures were not significantly different in Catalan and Spanish.

5.2.5.3 Discussion

5.2.5.3.1 *The effects of it-cleft structures on the preferences of null and overt pronouns compared to canonical sentences*

As a response to RQ2.2a, the predictions made concerning the impact of focusing an antecedent via it-cleft structures on the resolution preferences of subject pronouns are not completely borne out. The effects of clefting are the same in Catalan and Spanish. On the one hand, regarding null pronouns, the predicted anti-focus effect is attested: subject antecedents become less preferred for null pronouns when they are focused in subject clefts, as in (5.14) above. In object clefts, the focused object does not become less preferred compared to unmarked structures, but is clearly dispreferred by null pronouns (5.16), which show a well-defined subject bias in object clefts, as in canonical

contexts. On the other hand, the predicted enhanced preference of overt pronouns for focused antecedents is not observed. On the contrary, overt pronouns also appear to be sensitive to an anti-focus effect: object antecedents become less preferred for these pronouns when they are focused in object clefts (5.17). In subject clefts, the focused subject does not become less preferred either compared to unmarked structures, but is clearly dispreferred by overt pronouns (5.15). Overt pronouns show a well-defined object bias in both canonical contexts and subject clefts. In conclusion, the anti-focus effect impacts both null and overt pronouns: clefted antecedents (both subjects and objects) do not appear to be good candidates to be picked up by a subsequent pronominal subject (regardless of pronoun type), at least in intrasentential contexts.

In line with previous studies on intrasentential anaphora resolution, focusing an antecedent via it-cleft structures does not boost the preference for the clefted antecedent in the final interpretation of a pronoun, but does give rise to an anti-focus effect (Colonna et al., 2012, 2015 for German; de la Fuente, 2015 for null pronouns in Spanish; Patterson et al., 2017 for German and Russian; Patterson & Felser, 2020 for German). These findings were expected for null pronouns, since de la Fuente (2015) had already observed this decrease of the null-subject bias in subject clefts in Spanish. However, we did not expect that similar anti-focus effect to the one attested for null pronouns also arose for overt pronouns in Catalan and Spanish. This finding suggests that null and overt pronouns do not exhibit complementary behaviors, or a division of labor, based on pragmatic factors (i.e., [\pm topic shift]). This claim conflicts with the previous reasoning in light of the results on CLLD (described in §5.2.4), which were compatible with a specialization of null pronouns on topic maintenance and overt pronouns on topic shift.

Considering the results on CLLD, the similar anti-focus effect on null and overt pronouns might be explained by different factors, specific for each pronoun, in relation to a form-specific approach (in line with Kaiser & Trueswell, 2008). The anti-focus effect on null pronouns could be guided by pragmatic factors, and the anti-focus effect on overt pronouns could be explained by word order factors. Whereas null pronouns would disprefer antecedents conveying focal information, overt pronouns would disprefer first-mentioned antecedents (also the antecedents in the highest hierarchical position in the syntactic configuration). This possibility is only outlined roughly here and deserves to be explored and developed further. We will examine it in more detail by bringing findings on CLLD and it-clefts together in the following section (§5.2.6). Comparisons between the three marked information constructions (CLLD, subject clefts, and object clefts), which

have not yet been addressed, will provide new evidence to enrich this discussion and a more complete picture of the interaction between syntactic, pragmatic, and sequential factors in the interpretation of null and overt pronouns.

From a different perspective, in terms of "saliency", "prominence", or "accessibility", it-cleft structures did not increase the preference for a focused referent, neither for null nor for overt pronouns. This finding could be interpreted as contrary to previous studies that have argued for a higher activation of focused antecedents and the consequent prominence increase for pronoun resolution (Cowles et al., 2007; Foraker & McElree, 2007). However, in light of Patterson and Felser's (2020) results, it is possible that focused antecedents are perceived as more prominent during initial processing stages and that the anti-focus effect emerges late in comprehension stages. Therefore, even if focused referents are more readily available or more accessible in memory, null and overt pronouns may ultimately be interpreted as coreferring with the less accessible referent, at least in intrasentential contexts⁴⁵. This is what we found in the offline results of the present study. If clefted antecedents are considered to be more salient, or prominent, this raises a very interesting question regarding whether null pronouns are really more associated to more prominent antecedents, as suggested by theories such as the Accessibility Hierarchy (Ariel, 1990, 2001). As argued by Colonna et al. (2012), relying on salience or prominence as an intervening effect integrating a variety of factors is not sufficient and may not be "useful" to explain pronoun resolution. Although topicalized and focused antecedents may be similarly "salient", "prominent", or "accessible", the preference of null and overt pronouns for topicalized and focused antecedents is not the same.

5.2.5.3.2 The interpretation of null and overt pronouns in it-cleft structures

Regarding RQ2.2b, the anti-focus effect was not strong enough to reverse the preferences of each pronoun when their preferred antecedent was clefted: null pronouns did not prefer object antecedents over focused subjects in subject clefts, and overt pronouns did not prefer subject antecedents over focused objects in object clefts. Null pronouns were interpreted as unbiased in subject clefts, and overt pronouns were

⁴⁵ Cowles et al. (2007) and Foraker and McElree (2007) tested it-cleft constructions in intersentential contexts and found a preference for clefted antecedents, whereas we analyzed intrasentential contexts and attested an anti-focus effect. Sentence relationship may intervene in the final pronominal resolution preferences (see Colonna et al., 2012, 2015), but it is not obvious how should it condition the notions of salience/prominence/accessibility.

interpreted as unbiased in object clefts. In the other conditions, null pronouns maintained their subject bias in object clefts, and overt pronouns their object bias in subject clefts.

These findings suggest, as observed in CLLD structures, that syntactic factors and pragmatic factors interact in the process of anaphora resolution. The syntactic function of the antecedent is not more determining than its information status (or than the anti-focus effect), or vice versa. Instead, these two factors appear to have similar weights: when conflicting cues are found in it-clefts (i.e., when the preferred antecedent of a pronoun is clefted), neither null nor overt pronouns show well-defined interpretations, no cue overrides the other cues, and they are interpreted at chance.

5.2.5.3.3 No microvariation attested between Catalan and Spanish

Finally, as previously mentioned, no differences regarding the effects of it-cleft structures were observed between Catalan and Spanish (RQ3c). The effect size of the impact of subject clefts on null pronouns is very similar in both languages, and the anti-focus effect of overt pronouns in object clefts is slightly stronger in Catalan than in Spanish, but is not significantly different.

5.2.6 Integrating findings: topicalization and focusing structures compared

5.2.6.1 Aims and predictions

After reporting and discussing the results separately for each structure in the previous sections, it is also of great interest for the present investigation to overtly contrast topicalization and focusing effects. In this section we directly compare CLLD structures to subject and object cleft structures—the only comparisons that have not been presented yet. This contrast will allow us to finally be able to integrate the main findings by relating the effects of all the tested conditions. Overall, the main aims of this section are 1) to compare CLLD and it-cleft structures and 2) to integrate findings of the four analyzed contexts to achieve a more complete picture of how syntactic, pragmatic, and word order factors interact in pronoun resolution. In the previous sections, we have seen that Catalan and bilingual Spanish show very similar behaviors regarding the role of information structure (in both CLLD and it-cleft structures). To avoid repetition, we do not report the results for each language separately here.

Firstly, CLLD and subject clefts are compared. In these two structures, the subject conveys focal information, but it appears in two different positions: second-mentioned in CLLDs (in a postverbal position, OVS) and first-mentioned in subject clefts (in a preverbal position, SVO). Overall, this comparison can provide further evidence on the effects of focus marking on the interpretation of null and overt pronouns. RQ2.3a was formulated for this purpose.

Secondly, CLLDs and object clefts are compared. In these two structures, the object is first-mentioned and appears in the left periphery. However, it holds a different information status in each structure: topical in CLLDs and focal in object clefts. This comparison will allow the effects derived from pragmatic factors (i.e., object in topic vs. focus position) to be dissociated from those derived from altering the canonical word order (i.e., the object appears first-mentioned). RQ2.3b was formulated for this purpose.

Finally, subject clefts and object clefts are compared. In these two structures, the subject is focused and first-mentioned in subject clefts, and non-focused and second-mentioned in object clefts. Thus, this comparison allows the notions of topic and first-mention and focus and second-mention to be teased apart. In subject clefts, the subject is focused and first-mentioned, and in object clefts, the object is focused and first-mentioned. RQ2.3c was formulated for this purpose.

After answering RQ2.3a, RQ2.3b and RQ2.3c, we will have all of the necessary information in order to answer RQ2 in the general discussion at the end of this chapter (see §5.1.1: "Are the preferences of null and overt pronouns toward subject and object antecedents in bilingual Catalan and bilingual Spanish affected by the manipulation of information structure? How and to what extent do the syntactic function, information status, and linear position of the antecedent shape null and overt pronoun resolution?").

RQ2.3 a. Do null and overt pronouns similarly (dis)prefer subject antecedents conveying focal information in CLLD and subject clefts?

If null pronouns prefer non-focal (i.e., topical) antecedents, we expect CLLDs and subject clefts to make the preference of null pronouns for subject antecedents decrease to a similar extent. Given that subject antecedents convey focal material in both conditions, no differences should be found regarding null pronouns when comparing these two structures. It is true, however, that foci in these two constructions may be of different types (e.g., informative or contrastive; see §2.4.1). Based on de la Fuente (2015), who

compared informative and contrastive focus and found no differences between these two conditions, we do not expect to find any differences between the preference for focal antecedents in CLLDs and subject clefts attributable to the type of focus they convey.

Regarding overt pronouns, if they prefer focal (i.e., non-topical) antecedents, we expect CLLD and subject clefts to similarly enhance the preference of overt pronouns for focal subject antecedents. However, in light of the results in the previous sections, this prediction is not likely to be confirmed. The postverbal subject in CLLDs—which can be interpreted to convey focal information (see §2.4.1.2)—became more preferred for overt pronouns compared to canonical structures, but the focal preverbal subject in subject clefts was clearly dispreferred for overt pronouns. Based on these findings, we hypothesize that overt pronouns may prefer second-mentioned antecedents rather than focal antecedents. If that is the case, postverbal focal subjects in CLLD should be preferred over preverbal/first-mentioned focal subjects in subject clefts.

RQ2.3 b. Do null and overt pronouns similarly (dis)prefer first-mentioned object antecedents in CLLD and object clefts?

Concerning null pronouns, if they prefer non-focal (i.e., topical) antecedents, topical objects in CLLD are expected to be more preferred than focal objects in object clefts. From the perspective of subject preference, this would be equivalent to saying that the preference of null pronouns for subject antecedents should be greater in object clefts (where the subject is non-focal and can thus be interpreted to convey topical information) than in CLLDs (where the subject conveys focal information).

In terms of overt pronouns, if they prefer focal (i.e., non-topical) antecedents, focused objects in object clefts are expected to be preferred over topical objects in CLLDs. Alternatively, if they are prone to prefer second-mentioned antecedents regardless of information status, as suggested at the end of the previous section (see §5.2.5.3), no differences should be found between CLLD and object clefts as the object appears first-mentioned in both constructions.

RQ2.3 c. Do null and overt pronouns similarly (dis)prefer focused subject antecedents in subject clefts and non-focused subject antecedents in object clefts?

If null pronouns prefer non-focused (i.e., topical) antecedents over focused antecedents, they should show a stronger preference for subject antecedents in object clefts compared to subject clefts.

If overt pronouns prefer focal (i.e., non-topical) antecedents, they should show a higher preference for subject antecedents in subject clefts compared to object clefts. If, alternatively, they are more sensitive to word order factors than pragmatic factors, they should show a higher preference for subject antecedents when they are postverbal (i.e., in object clefts) than when they are preverbal (i.e., in subject clefts).

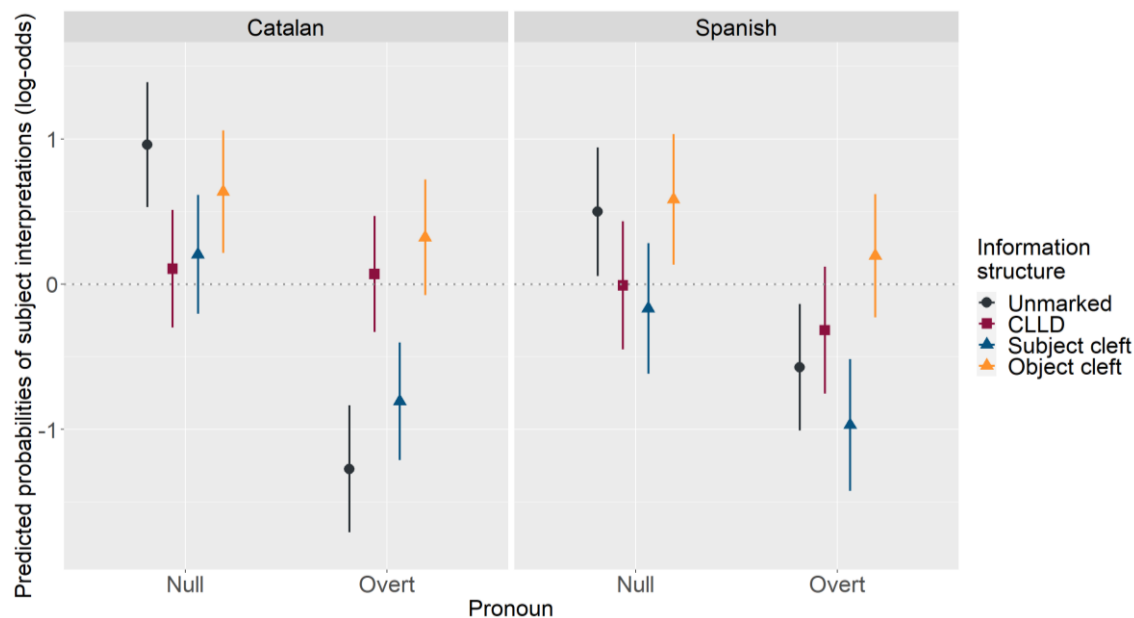
5.2.6.2 Results

As we have previously mentioned, we will not report the results for Catalan and Spanish separately in this section. As reported in §5.2.4 and §5.2.5, no significant differences were attested between the two languages regarding the effects of marked information structures. This can also be appreciated in Figure 5.4 below, which shows the predicted interpretation of null and overt pronouns (in bilingual Catalan, and in bilingual Spanish) in the four scrutinized contexts. In the first place, we refer to pairwise contrasts of the three-way interaction from the Information structure perspective, paying special attention to the contrasts between marked CLLD and it-cleft contexts, which we have not addressed yet. Later, we look at the results from the perspective of Pronoun to summarize the main findings and to identify in which contexts a division of labor between the two pronouns emerges. This way, we will have a complete picture of how null and overt pronouns are interpreted in each informationally marked context.

Firstly, we compare CLLD and subject cleft structures, two contexts in which the subject antecedent conveys focal information and the object antecedent topical or presupposed (i.e., non-focal) information. In these contexts, the preference of null pronouns for subject antecedents in CLLD compared to subject clefts is almost identical (in Catalan: $\beta = -0.098$, $p = .970$; in Spanish: $\beta = -0.159$, $p = .920$). Hence, the decrease of null pronouns' subject interpretations in CLLD and subject clefts compared to canonical sentences (see §5.2.4 and §5.2.5), is similar in both structures. Subjects conveying focus in CLLD or it-clefts were equally (dis)preferred for null pronouns. This does not mean that object antecedents were preferred in these contexts but that null pronouns were interpreted as unbiased.

Figure 5.4

Predicted subject interpretations of null and overt pronouns in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals, by information structure (± 95 CI)



Overt pronouns were also unbiased in CLLD structures, but they were clearly interpreted as coreferring with object antecedents in subject clefts. Postverbal subjects in CLLD were significantly more preferred than first-mentioned subjects in it-cleft structures ($\beta = 0.876$, $p < .001$; in Spanish: $\beta = 0.653$, $p = .056$). Thus, the interpretation of overt pronouns in CLLD structures, in which the object appears first-mentioned, differs from that in subject clefts, in which the object appears second-mentioned.

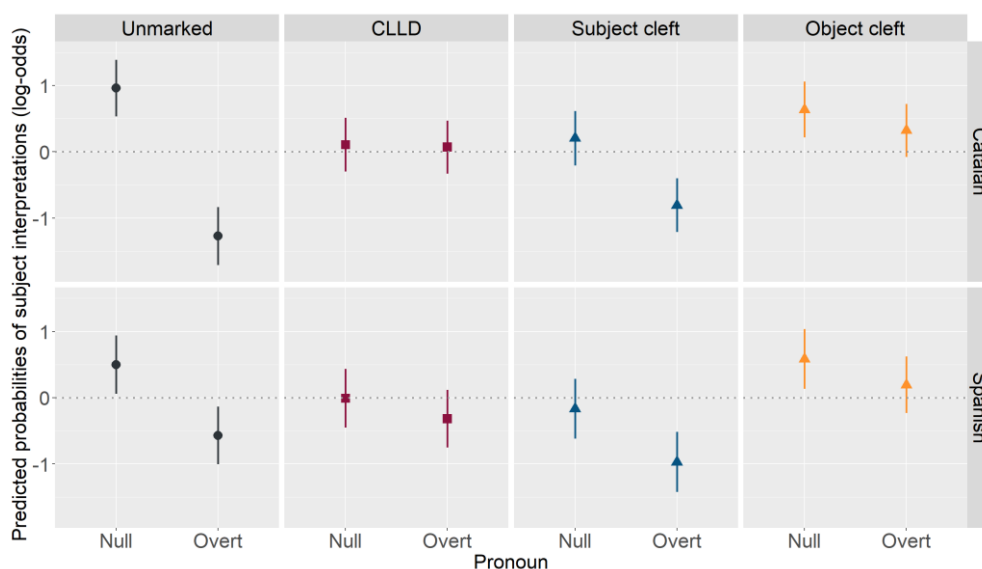
Secondly, we compare CLLD and object cleft structures, the two contexts where object antecedents appear in initial position. Regarding null pronouns, the proportion of subject interpretations was higher in object clefts than in CLLD, which increased the prominence of the topical object. This higher preference for non-focal (i.e., topical) subjects in object clefts compared to focal subjects in CLLD structures only approached statistical significance (in Catalan: $\beta = -0.531$, $p = .087$; in Spanish: $\beta = -0.592$, $p = .082$). The difference between the two constructions reveals that null pronouns were interpreted differently in CLLD structures, where the subject conveyed focus, and in object clefts, where the subject was presupposed and therefore conveyed topical information.

In the case of overt pronouns, no significant differences were attested in their preference for initial dislocated and focused objects (in Catalan: $\beta = -0.252$, $p = .671$; in Spanish: $\beta = -0.511$, $p = .161$). Thus, they were similarly interpreted when the subject was second-

mentioned—postverbal—and the object antecedent was first-mentioned in CLLD structures and object clefts. After the increase of preference for subject antecedents in these contexts (see §5.2.4 and §5.2.5), overt pronouns in CLLD and object clefts were interpreted as unbiased, in contrast to canonical structures (in which overt pronouns are clearly biased toward object antecedents). Altering the canonical SVO word order via CLLD or object clefts also altered the 'canonical' object bias of overt pronouns to a similar extent.

Figure 5.5

Predicted subject interpretations of null and overt pronouns in Catalan by Catalan-dominant bilinguals and in Spanish by Spanish-dominant bilinguals, by pronoun, information structure, and language (± 95 CI)



Regarding the results derived from pairwise contrasts from the perspective of Pronoun, they indicate that null and overt pronoun are not displaying complementary biases in all contexts. Figure 5.5 shows the same results already presented from another perspective to facilitate its interpretation. As can be observed, when the subject antecedent is first-mentioned (i.e., in canonical and subject-cleft structures), null pronouns pick up more subject antecedents than overt pronouns, and overt pronouns pick up more object antecedents than subject pronouns. When the object is topicalized via CLLD or focused via it-cleft—changing the order of mention of subject and object antecedents and displaying an OVS order—no differences between null and overt pronouns are found. As the effect sizes for marked information structures were larger for overt pronouns than for null pronouns, it is not surprising that when focusing the subject antecedent the division of labor between null and overt pronouns is still significant. This structure does not

significantly affect the strong interpretation of overt pronouns toward object antecedents. In structures where the choices of overt pronouns are affected, the division of labor between the two pronominal forms becomes diluted.

5.2.6.3 Discussion

5.2.6.3.1 Comparing the impact of topicalization and focusing structures on null and overt pronouns

The comparison between marked information structures in both Catalan and Spanish provides evidence to support a form-specific multiple-constraint approach (Kaiser & Trueswell, 2008). The preferences of null and overt pronouns are sensitive to different extents to syntactic, pragmatic, and sequential factors (i.e., the syntactic function, the information status, and the order of mention of the antecedents). Overlapping with sequential factors (i.e., word order), it is important to refer to other structural or configurational factors, related to the syntactic structure. The first-mentioned (and preverbal) antecedent is also the hierarchically highest antecedent in the syntactic configuration of our sentences. The second-mentioned (and postverbal) antecedent, in contrast, coincides with the hierarchically lowest antecedent in the phrase structure. The same effects are found in Catalan and Spanish.

When comparing CLLDs and subject clefts (RQ2.3a), null pronouns similarly (dis)prefer focused (i.e., non-topical) subject antecedents, regardless of whether they appear first-mentioned or second-mentioned. In other words, no differences are found between the preference of null pronouns for subject antecedents that appear in a hierarchically higher or lower position in the syntactic structure (i.e., in a preverbal or postverbal position). Overt pronouns, instead, show a higher preference for subject interpretations in CLLDs than in subject clefts. That is, their preference for subjects is higher when the subject is second-mentioned (i.e., postverbal, and therefore appears in a hierarchically lower position than the object antecedent) than when it is first-mentioned (i.e., preverbal, and in a hierarchically higher position than the object antecedent). While null pronouns do not seem to prefer antecedents in the highest structural position, overt pronouns seem to prefer antecedents in the lowest structural position.

When comparing CLLDs and object clefts (RQ2.3b), null pronouns show a stronger preference for subject antecedents when they do not convey focus (i.e., in object clefts) than when they convey focus (i.e., in CLLDs). No differences emerge regarding overt

pronouns, confirming that they similarly (dis)prefer first-mentioned object antecedents, regardless of whether they are topicalized or focused. Null pronouns thus appear to be more sensitive to pragmatic factors than overt pronouns, and overt pronouns more sensitive to word order than null pronouns.

Finally, when comparing subject clefts and object clefts (RQ2.3c), null pronouns do not show a significantly different preference for focused subjects in subject clefts and non-focused subjects in object clefts. However, the subject bias is significant in the object cleft condition, and not in the subject cleft condition, meaning that while focused objects are clearly dispreferred for null pronouns, they do not reject coreference with focused subjects (in this case, null pronouns remained unbiased). On the other hand, overt pronouns prefer non-focused and postverbal subject antecedents (i.e., in object clefts) over focused and first-mentioned subject antecedents (i.e., in subject clefts).

Overall, these findings indicate that, on the one hand, the interpretation of null pronouns is guided by the interaction between syntactic factors (i.e., the syntactic function of the antecedent) and pragmatic factors (i.e., the information status of the antecedent): null pronouns prefer subject and topical antecedents. On the other hand, the interpretation of overt pronouns is guided by the interaction between syntactic factors (i.e., the syntactic function of the antecedent) and sequential factors (i.e., word order or the hierarchical position in the syntactic configuration): overt pronouns prefer object and postverbal (or hierarchically lower) antecedents. This different sensitivity of null and overt pronouns to the interaction of different factors is compatible with a form-specific multiple-constraint approach (in line with Kaiser & Trueswell, 2008), as mentioned at the beginning of the discussion. In the following paragraphs, we will expand on these ideas by integrating the findings on the four structures that, until now, have been separately addressed (see §5.2.3 for an analysis of unmarked canonical sentences, §5.2.4 for an analysis of CLLD, and §5.2.3 for an analysis of subject clefts and object clefts).

5.2.6.3.2 Synthesizing the role of syntactic, pragmatic, and sequential factors in null and overt pronoun resolution

Table 5.2 summarizes whether null and overt pronouns show a significant bias toward subject or object antecedents in each structure. It also specifies how syntactic function, information status, and word order (i.e., order of mention) characterize each antecedent in each of the analyzed information structures. No distinctions are made between

bilingual Catalan and Spanish as the same results are obtained in this regard (they only differ in the strength of the interpretive biases in canonical sentences).

Table 5.2

Summary of the interpretive biases of pronouns in bilingual Catalan and Spanish and the main features characterizing the antecedents in each information structure

Information structure	Canonical unmarked structure		Topicalized object via clitic-left dislocation		Focused subject via it-cleft		Focused object via it-cleft	
Pronoun bias	Null-Subject		Null-Unbiased		Null-Unbiased		Null-Subject	
	Overt-Object		Overt-Unbiased		Overt-Object		Overt-Unbiased	
Syntactic function	Subject	Object	Object	Subject	Subject	Object	Object	Subject
Information status⁴⁶	Topical	Focal	Topical	Focal	Focal	Topical	Focal	Topical
Order of mention⁴⁷	First	Second	First	Second	First	Second	First	Second

Note. The conditions in which null pronouns show a significant bias toward subject antecedents are shown in light grey, and those in which overt pronouns show a significant bias toward object antecedents are shown in dark grey.

Firstly, as stated in the previous paragraph, null pronouns appear to be sensitive to both syntactic factors, preferring subject antecedents, and to pragmatic factors, preferring non-focal antecedents (i.e., topical antecedents). Null pronouns only show a well-defined subject bias in canonical sentences (5.18) and in object clefts (5.19); two structures in which no conflicting cues are encountered, i.e., subjecthood is aligned with topicality. As can be seen from these two structures, null pronouns do not seem to be that sensitive to

⁴⁶ In an attempt to simplify the features of each antecedent in each of the tested information structures, we use the topical-focal dichotomy in this summary table.

⁴⁷ Remember that first-mentioned antecedents (i.e., preverbal) are also the ones that appear in the hierarchically highest position in the syntactic configuration. In SVO structures (i.e., canonical sentences and subject clefts), the subject antecedent is also the antecedent in the highest hierarchical position (in Spec,IP or in Spec,FocP). In OVS structures (i.e., CLLDs and object clefts), it is the object the antecedent that appears in the highest hierarchical position (in Spec,TopP or in Spec,FocP).

word order or configurational factors. They similarly prefer topical subjects, regardless of whether they appear first-mentioned and in the hierarchically highest position, in SVO sequences (i.e., canonical sentences), or second-mentioned and in the hierarchically lowest position, in OVS sequences (i.e., object clefts).

(5.18) *Null pronouns in unmarked structures*

- a. **La Laura_j** va espantar la Maria_k quan **pro_j** va entrar a l'habitació. (Catalan)
- b. **Laura_j** asustó a Maria_k cuando **pro_j** entró en la habitación. (Spanish)
'Laura scared Maria when (she) went into the room.'

(5.19) *Null pronouns in object cleft structures*

- a. Va ser a la Maria_k a qui va espantar la **Laura_j** quan **pro_j** va entrar a l'habitació.
- b. Fue a Maria_k a quien asustó **Laura_j** cuando **pro_j** entró en la habitación.
'It was Maria whom Laura scared when (she) went into the room.'

In CLLD (5.20) and subject clefts (5.21), where the subject conveys focal information, the presence of conflicting cues appears to unbias the preferences of null pronouns. Once more, the proportion of subject interpretations is similar regardless of whether the focal subject appears first-mentioned or second-mentioned. Moreover, results showed that the subject interpretations of null pronouns significantly decrease in both CLLD and subject cleft structures compared to both unmarked and object cleft structures.

(5.20) *Null pronouns in CLLD structures*

- a. A la Maria_j la va espantar la Laura_k quan **pro_{j/k}** va entrar a l'habitació.
- b. A María_j la asustó Laura_k cuando **pro_{j/k}** entró en la habitación.
'Maria, Laura scared her when (she) went into the room.'

(5.21) *Null pronouns in subject cleft structures*

- a. Va ser la Laura_j qui va espantar **la** Maria_k quan **pro_{j/k}** va entrar a l'habitació.
- b. Fue Laura_j quien asustó a María_k cuando **pro_{j/k}** entró en la habitación.
'It was Laura who scared Maria when (she) went into the room.'

Overt pronouns seem to be sensitive to syntactic factors, preferring object antecedents, and to word order or configurational factors, preferring postverbal constituents (i.e., second-mentioned and in the lowest position in the syntactic configuration). In a way, these are both 'structural' factors: overt pronouns prefer object antecedents and antecedents in the lowest hierarchical position in the syntactic configuration. In fact, overt pronouns only show a clear bias in unmarked structures (5.22) and subject clefts (5.23). What these two structures have in common is that the object is second-mentioned and

appears in its canonical postverbal position (hierarchically lower than that of the subject antecedent). As such, in contrast to null pronouns, overt pronouns are sensitive to word order or configurational factors, as well as to the syntactic position of the antecedent, and not so sensitive to pragmatic factors. In unmarked structures, the object antecedent conveys focal information, whereas in subject clefts it conveys non-focal (i.e., topical) information, and the interpretation of overt pronouns in these two structures is similar. Overt pronouns thus appear to be able to corefer with antecedents that convey different pragmatic features (i.e., topic and focus).

(5.22) *Overt pronouns in unmarked structures*

- a. La Laura_j va espantar **la Maria**_k quan **ella**_k va entrar a l'habitació. (Catalan)
 - b. Laura_j asustó a **María**_k cuando **ella**_k entró en la habitación. (Spanish)
- 'Laura scared Maria when she went into the room.'

(5.23) *Overt pronouns in subject cleft structures*

- a. Va ser la Laura_j qui va espantar **la Maria**_k quan **ella**_k va entrar a l'habitació.
 - b. Fue Laura_j quien asustó a **María**_k cuando **ella**_k entró en la habitación.
- 'It was Laura who scared Maria when she went into the room.'

In the two structures in which the object appears first-mentioned, in a preverbal and hierarchically higher position, the comprehender seems to encounter conflicting cues (the object antecedent is not postverbal, nor the hierarchically lowest antecedent). As a result, overt pronouns remain unbiased in their interpretation. The object interpretations of overt pronouns significantly decrease in both CLLD (5.24) and object cleft structures (5.25) compared to both unmarked and subject cleft structures.

(5.24) *Overt pronouns in CLLD structures*

- a. A la Maria_k la va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació.
 - b. A María_k la asustó Laura_j cuando **ella**_{j/k} entró en la habitación.
- 'Maria, Laura scared her when she went into the room.'

(5.25) *Overt pronouns in object cleft structures*

- a. Va ser a la Maria_k a qui va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació.
 - b. Fue a Maria_k a quien asustó Laura_j cuando **ella**_{j/k} entró en la habitación.
- 'It was Maria whom Laura scared when she went into the room.'

At this point, it is worth referring once more to the overlap between sequential or word order factors and configurational factors or hierarchical height. The object bias of overt pronouns demonstrated in canonical sentences remains stable only in structures

following SVO word order. In SVO structures, the object antecedent appears in its canonical position within the VP (i.e., in canonical sentences and subject clefts). Using Carminati's terms, the object antecedent appears in a syntactic position lower than Spec,IP in the phrase structure, both in unmarked and subject cleft sentences. In contrast, overt pronouns remain unbiased in structures where the object appears in the left periphery (i.e., Spec,TopP in CLLD; Spec,FocP in object clefts). In these contexts, the object antecedent appears in a syntactic position higher than Spec,IP (the position where the canonical subject antecedent appears) in the phrase structure. On the whole, what affects the interpretive preferences of overt subject pronouns is the combination of 1) the order of mention of the antecedents (word order) and 2) the position of the antecedent in the syntactic configurations derived from marked information structures (i.e., in the left periphery vs. within the VP). Conversely, the preferences of null pronouns do not seem to be affected by these word order or configurational factors.

5.2.6.3.3 On the Position of Antecedent Hypothesis

In terms of the Position of Antecedent Hypothesis (PAH; Carminati, 2002), the discussed findings are relevant as they disentangle the notions of subject and object antecedents from that of antecedents in Spec,IP and antecedents in a position lower than Spec,IP. Carminati (2002, p. 181) concluded that "the null pronoun prefers an antecedent in the Spec IP position and the overt pronoun an antecedent in a syntactic position lower in the phrase structure".

Regarding null subject pronouns, our results confirm that they tend to prefer antecedents in the Spec,IP (i.e., subject) position, and not in the highest position in the syntactic configuration. Moreover, we provide an empirical answer to a question that Carminati (2002, p. 184) left open, on "whether focalized or topicalized referents are more prominent than referents in Spec IP, such that, when the option is available, the null pronoun prefers to take the former, and not the latter, as its antecedent". As Carminati already intuitively speculated, focalized antecedents (in Spec,FocP) do not compete with subject antecedents (in Spec,IP) in the interpretation of null pronouns. On the other hand, topicalized antecedents (in Spec,TopP) compete with subject antecedents (in Spec,IP). However, topicalized antecedents do not override the preference of null pronouns for subject antecedents. As observed in CLLD structures, where the object was topicalized and the subject appeared in postverbal position, null pronouns remained unbiased, as

well as in subject clefts, where the subject appeared in a higher position than the object, but it conveyed focal (i.e., non-topical) information.

In terms of overt subject pronouns, our results do not appear to support the claim that they systematically tend to prefer an antecedent in a syntactic position lower than Spec,IP. The position of an antecedent in the syntactic configuration is not more determining than the syntactic function of this antecedent. The preference of overt pronouns for antecedents lower than Spec,IP competes with their preference for object antecedents. As observed in CLLD and object cleft structures, when object antecedents appeared in the left periphery and the subject antecedent occupied the lowest position in the syntactic structure, overt pronouns remained unbiased.

5.2.7 Summary of main findings

In the present study, three main features of the antecedent that act as cues for pronoun resolution have been manipulated and confronted using marked information structures: syntactic function, information status and word order. In a form-specific multiple-constraint approach (Kaiser & Trueswell, 2008), different factors appear to guide the interpretive biases of null and overt pronouns: whereas the interpretation of null pronouns seems to be affected by the combination of syntactic and pragmatic factors (syntactic function and information status of the antecedents), the interpretation of overt pronouns seems to be affected by the interaction between syntactic and sequential factors (syntactic function of the antecedents, and word order and the position of the antecedent in the syntactic configuration). The reported results point toward these different effects of topicalization and focusing structures on null and overt pronouns, as will be further explained in the paragraphs below.

To summarize the main claims from the previous sections, we outline how null and overt pronouns are interpreted in each of the analyzed (un)marked information structures: canonical sentences (unmarked structures), clitic-left dislocations (CLLD; topicalization of the object), subject clefts (focusing of the subject), and object clefts (focusing of the object). Finally, we characterize the interpretation of null and overt pronouns across constructions in a wrap-up summary.

5.2.7.1 Canonical unmarked structures

In canonical sentences in both Catalan and Spanish, null pronouns were preferably interpreted as coreferring with subject antecedents, and overt pronouns as coreferring with object antecedents. These PAH-like biases (in line with Carminati, 2002) in unmarked contexts are illustrated in example (5.26).

- (5.26) a. La Laura_j va espantar la Maria_k quan *pro*_j/ella_k va entrar a l'habitació. (Catalan)
b. Laura_j asustó a Maria_k cuando *pro*_j/ella_k entró en la habitación. (Spanish)
'Laura scared Maria when she went into the room.'

5.2.7.2 Object topicalization via clitic-left dislocation structures

In CLLDs (OVS word order), object antecedents are topicalized and appear first-mentioned, whereas subject antecedents are in a non-topical position and appear second-mentioned. In relation to pragmatic factors, object antecedents are marked as topics, whereas subject antecedents convey non-topical information⁴⁸. In contrast to canonical structures, topicalized object antecedents appear in the left periphery (i.e., Spec,TopP), whereas subject antecedents occupy a postverbal position (i.e., either in Spec,IP or in situ in a VP-internal position⁴⁹). Thus, in CLLDs, the subject antecedent appears in a hierarchically lower position than that of the object antecedent.

In CLLDs, null pronouns do not show any preferences in their interpretation (5.27). This lack of bias can be explained by the presence of conflicting cues: the subject antecedent (*Laura*) is not aligned with the topic of the sentence, which is the topicalized object antecedent (*Maria*).

- (5.27) a. A la Maria_j la va espantar la Laura_k quan *pro*_j/*la* va entrar a l'habitació.
b. A María_j la asustó Laura_k cuando *pro*_j/*ella* entró en la habitación.
'Maria, Laura scared her when (she) went into the room.'

Overt pronouns in CLLD conditions do not show any interpretive preferences either (5.28). In this case, the preverbal position of the object antecedent (*Maria*) is what seems to lead overt pronouns to remain unbiased. From a configurational perspective, the

⁴⁸ According to some authors, postverbal subject antecedents in CLLD structures are syntactically marked as conveying focus (e.g., Vallduví, 1992; see also Domínguez, 2018).

⁴⁹ Depending on the analysis, a postverbal subject appears in Spec,IP/Spec,TP (cf. Olarrea, 1998) or remains in its VP-internal position (cf. Gutiérrez-Bravo, 2003). In any case, it appears in a postverbal position.

object antecedent appears in the left periphery (i.e., Spec,TopP), in a higher position in the phrase structure than that of the subject antecedent.

(5.28) a. A la Maria_k la va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació.

b. A María_k la asustó Laura_j cuando **ella**_{j/k} entró en la habitación.

'Maria, Laura scared her when she went into the room.'

5.2.7.3 Subject focalization via it-cleft structures

In subject clefts (SVO word order), subject antecedents are focused and first-mentioned, whereas object antecedents appear in a non-focal position and second-mentioned. In relation to pragmatic factors, in a focus-presupposition configuration, subject antecedents convey focus, and object antecedents convey presupposed information (i.e., topical information⁵⁰). In the phrase structure, subject antecedents occupy a position in the left periphery (i.e., Spec,FocP), whereas object antecedents occupy their canonical postverbal position. As in canonical sentences, subject antecedents appear in a higher position than object antecedents in the syntactic configuration.

Null pronouns in these contexts do not show any preferences in their interpretation (5.29), as in CLLDs. In subject clefts, the subject antecedent (*Laura*) is focused, so it is not aligned with the topic of the sentence. The comprehender thus encounters conflicting cues when interpreting the null pronoun, and it therefore remains unbiased.

(5.29) a. Va ser la Laura_j qui va espantar la Maria_k quan **pro**_{j/k} va entrar a l'habitació.

b. Fue Laura_j quien asustó a María_k cuando **pro**_{j/k} entró en la habitación.

'It was Laura who scared Maria when (she) went into the room.'

The fact that the superficial SVO word order is maintained in subject clefts, as opposed to previous CLLD conditions, does not seem to affect the resolution of null pronouns. Rather, null pronouns seem to be guided by the interaction between the syntactic function and the information status of the antecedent. However, word order seems to have a role in the interpretation of overt pronouns, as shown below.

Overt pronouns in subject clefts do show a clear-cut bias toward object antecedents (5.30), as in canonical sentences. Given that the object appears in its canonical position (i.e., within the VP), comprehenders do not encounter conflicting cues when interpreting

⁵⁰ Given that presupposed elements express given information, they can be considered to convey topical information, as opposed to "new" or focal information.

overt pronouns. In fact, the object (*Maria*) appears in a lower position in the syntactic configuration than the clefted subject (*Laura*).

- (5.30) a. Va ser la Laura_j qui va espantar **la Maria_k** quan **ella_k** va entrar a l'habitació.
b. Fue Laura_j quien asustó a **María_k** cuando **ella_k** entró en la habitación.
'It was Laura who scared Maria when she went into the room.'

5.2.7.4 Object focalization via it-cleft structures

In object clefts (OVS word order), object antecedents are focused and first-mentioned, whereas subject antecedents appear in a non-focal position and second-mentioned. In relation to pragmatic factors, in a focus-presupposition configuration, object antecedents are marked as conveying focus, and subject antecedents convey presupposed or topical⁵¹ information. Regarding configurational factors, object antecedents appear in the left periphery of the phrase structure (i.e., Spec,FocP), whereas subject antecedents appear in a postverbal position (i.e., either in Spec,IP or *in situ* in a VP-internal position). In contrast to canonical sentences, the superficial word order is altered and object antecedents appear in a higher position in the syntactic configuration than subjects.

Null pronouns in object clefts show a well-defined preference for subject antecedents (5.31), as in canonical structures, despite their postverbal position. In object clefts, the subject antecedent (*Laura*) is not the focused antecedent, so it conveys presupposed topical information. Subjecthood and topicality are aligned in this condition and null pronouns show a clear subject bias. Sequential factors do not intervene.

- (5.31) a. Va ser a la Maria_k a qui va espantar la **Laura_j** quan **pro_j** va entrar a l'habitació.
b. Fue a María_k a quien asustó **Laura_j** cuando **pro_j** entró en la habitación.
'It was Maria whom Laura scared when (she) went into the room.'

Overt pronouns in object clefts, on the contrary, do not show any bias in their interpretation (5.32). Once more, word order is altered in this construction (OVS) and the object appears in the left periphery (i.e., Spec,FocP; higher than the subject antecedent). In the presence of conflicting cues in the interaction between syntactic factors and order of mention—or 'structural factors'—, overt pronouns are interpreted at chance.

⁵¹ See the previous footnote.

- (5.32) a. Va ser a la Maria_k a qui va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació.
b. Fue a María_k a quien asustó Laura_j cuando **ella**_{j/k} entró en la habitación.
'It was Maria whom Laura scared when (she) went into the room.'

5.2.7.5 The interpretation of null and overt pronouns across structures

Overall, null pronouns have been found to show a clear subject bias when the subject antecedent conveys topical (or presupposed) information (i.e., in canonical sentences and object clefts). However, when the subject antecedent conveys focal information, null pronouns remain unbiased (i.e., in CLLDs and in subject clefts). As we have argued, the interpretation of null pronouns appears to be guided by an interaction between syntactic and pragmatic factors; that is, the syntactic function and the information status of the antecedents are two similarly weighted cues in the interpretation of null pronouns. Null subject pronouns tend to prefer subject antecedents and antecedents conveying topical information to a similar extent (no cue overrides the other cue). Sequential factors, on the other hand, do not seem to play a determinant role in the interpretation of null pronouns, and neither does the fact that an antecedent occupies the hierarchically highest or lowest position in the phrase structure.

Overt pronouns, on the other hand, have been argued to show a clear object bias when word order is maintained and the object appears second-mentioned, in its canonical postverbal position within the VP (i.e., in canonical sentences and subject clefts). When the object does not appear second-mentioned and does not occupy a lower position in the syntactic configuration than the competing subject antecedent, overt pronouns remain unbiased (i.e., in CLLDs and in object clefts). In summary, their preferences seem to be guided by the combination of syntactic and word order factors; that is, they are similarly sensitive to the interaction between the syntactic function of the antecedent and its sequential position or its structural position in the syntactic configuration. Being an object antecedent and appearing in a hierarchically lower position than the competing (or second-mentioned) subject antecedent are two similarly-weighted cues in the interpretation of overt pronouns (no cue overrides the other cue). On the other hand, overt pronouns do not appear to be as sensitive to pragmatic factors, being able to corefer with both topical and focal antecedents.

5.3 The interpretation of subject pronouns in monolingual Spanish: does it differ from Catalan and bilingual Spanish?

In this section, we will assess pronoun resolution in (monolingual) Peninsular Spanish. More specifically, we will explore whether the impact of syntactic, pragmatic, and sequential factors on pronominal interpretive biases in monolingual Spanish is similar to that described in §5.2 for bilingual Spanish and bilingual Catalan. In addition to describing and characterizing the interpretation of null and overt pronouns in monolingual Spanish, it will be compared to pronoun resolution in Catalan (by Catalan-dominant bilinguals) to confirm whether there is microvariation between the two languages. Moreover, monolingual Spanish will be compared to bilingual Spanish by Spanish-dominant bilinguals to find out whether a specific Spanish variety in Catalonia exists or, alternatively, whether bilingual Spanish in contact with Catalan resembles monolingual Spanish.

Firstly, pronoun resolution preferences in canonical and marked information structures in monolingual Spanish will be addressed (§5.3.1). Monolingual Spanish will then be compared to (bilingual) Catalan (§5.3.2). Finally, monolingual Spanish will be compared to bilingual Spanish (§5.3.3).

5.3.1 Characterizing pronoun resolution in monolingual Spanish and the effects of marked information structures

5.3.1.1 Aims and predictions

The first aim of §5.3 is to describe the interpretation of subject pronouns in the four tested structures in monolingual Spanish. Firstly, pronoun resolution in canonical sentences will be addressed. Following this, the effects of marked information structures will be analyzed, comparing the preferences of null and overt pronouns in object CLLDs, subject clefts, and object clefts to their baseline preferences (in canonical sentences). To this end, only the Spanish monolingual group will be analyzed.

The following research questions will be addressed: RQ4a and RQ4b (see §5.1.1).

RQ4 a. Can the predictions of the Position of Antecedent Hypothesis explain the interpretive biases of null and overt pronouns in unmarked structures in monolingual Spanish?

In view of previous studies using forced-choice tasks, we expect that monolingual Spanish will display pronoun resolution patterns according to the PAH in canonical sentences (e.g., Contemori & Di Domenico, 2021). However, given that the tested sentences follow a main-subordinate clause order, null pronouns may show unbiased patterns and overt pronouns a preference for object antecedents (e.g., de Rocafiguera & Bel, 2022).

RQ4 b. Are the preferences of null and overt pronouns toward subject and object antecedents in monolingual Spanish affected by the manipulation of information structure?

Concerning the impact of marked information contexts, we have formulated our predictions building on de la Fuente's (2015) results for null pronouns and considering the results obtained for bilingual Spanish and Catalan (see the main findings summarized in §5.2.7). On the one hand, the preference of null pronouns for subject antecedents is expected to be reduced when the notions of subject and sentence topic are not aligned; that is, in CLLDs and subject clefts. In object clefts, null pronouns should be unaffected, preferring non-focused subjects. On the other hand, the preference of overt pronouns for object antecedents is expected to be reduced when the object appears in the left periphery; that is, in CLLDs and object cleft structures. In subject clefts, overt pronouns should be unaffected, preferring postverbal objects.

5.3.1.2 Method

5.3.1.2.1 Participants

To characterize pronoun resolution preferences in monolingual Spanish and the role of syntactic, pragmatic, and sequential factors, a group of Spanish monolinguals ($N = 40$; mean age: 20.12, range: 18-30) completed the task. These participants were mainly from Aragon, and speakers of Peninsular Spanish. They have always lived in Spain and were raised monolingually and in monolingual environments. These participants have been described in §4.2.1.

5.3.1.2.2 *Materials*

The experimental task completed by participants, a two-alternative forced-choice task in Spanish, has been detailly described in §4.3.2. The two conditions of the tasks are pronoun (null, overt) and information structure (unmarked structures, topicalized objects via CLLD, focused subjects via it-clefts, and focused objects via it-clefts). An example for each informational context is provided in §5.2.1.2 (in addition to Chapter 4).

5.3.1.2.3 *Reported model*

We ran a mixed-effects logistic regression on the data from the monolingual group's responses in Spanish. The fitted model included Pronoun (null, overt) and Information structure (unmarked structures, CLLD, subject clefts, object clefts) as fixed effects and their interaction. As random effects, varying intercepts for participants and items were added to the model. The model supported no random slopes. The summary of this model is provided in Appendix H (Table H.1). The model's total explanatory power was moderate (conditional $R^2 = 0.22$) and the part related to the fixed effects alone (marginal R^2) was of 0.04. No multi-collinearity issues were detected (highest VIF value = 1.001). This model had a C-index of concordance of 0.76.

The same model with no intercept—with the same fixed effects and random effects' structure—was computed to compare whether subject and object antecedent choices of null and overt pronouns differed from chance (see §4.5.2.2). Table H.2 in Appendix H presents the summary of this model.

5.3.1.3 **Results**

To begin with, Table 5.3 summarizes the proportion of null and overt pronouns' subject interpretations in monolingual Spanish in the four analyzed contexts: unmarked structures, topicalized objects via CLLD, focused subjects via it-cleft, and focused objects via it-cleft. As can already be observed in this table, both null and overt pronouns in monolingual Spanish show very mild preferences across conditions. Both pronouns remain unbiased in canonical sentences (baseline condition). Well-defined interpretive biases of null pronouns toward subject antecedents only seem to be attested in object clefts. Overt pronouns seem to show a very shy bias toward object antecedents in canonical sentences, but this bias only seems to emerge clearly in subject clefts.

Table 5.3*Proportion of subject interpretations in monolingual Spanish (SD)*

	Null	Overt
Unmarked	.54 (.50)	.44 (.50)
Topicalized object	.55 (.50)	.46 (.50)
Focused subject	.51 (.50)	.35 (.48)
Focused object	.64 (.48)	.56 (.50)

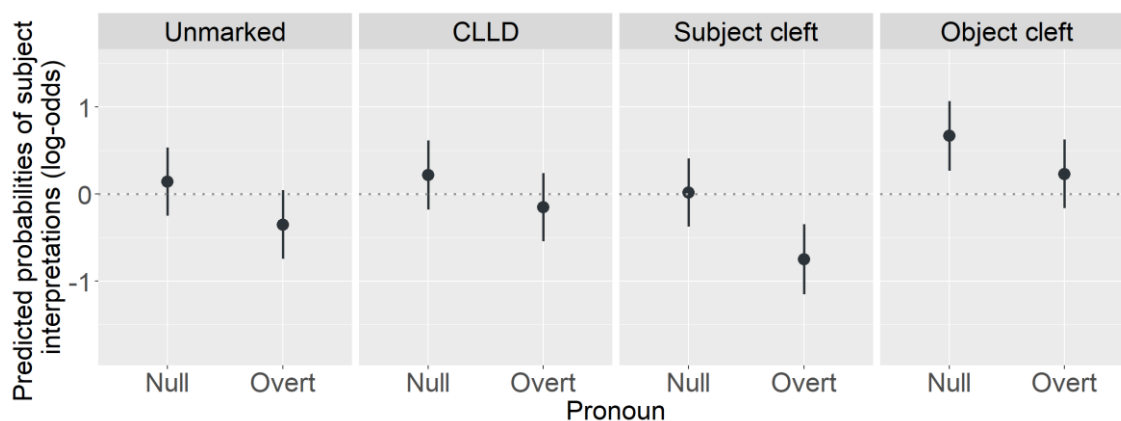
The statistical model showed that the two-way interaction Pronoun \times Information structure was not significant ($\chi^2(3) = 1.986, p = .575$). However, significant effects of both Pronoun ($\chi^2(1) = 22.758, p < .001$) and Information structure ($\chi^2(1) = 29.137, p < .001$) were attested (obtained through likelihood ratio tests, see Appendix H). As significant effects emerged for Pronoun and Information structure, the lack of a significant interaction indicates that the different interpretations of null and overt pronouns may be similarly affected by different structures. This may be related to the very weak interpretive preferences of both pronouns in the baseline condition, which can already be observed in Table 5.3.

Figure 5.6 illustrates the model results and the predicted subject interpretations of null and overt pronouns in monolingual Spanish across contexts. The first finding that needs to be highlighted is that the Position of the Antecedent Hypothesis was not attested in canonical structures in Spanish. Null pronouns were not interpreted differently from chance by Spanish monolinguals ($\beta = 0.144, p = .472$) and overt pronouns were only slightly biased toward object antecedents, which was not statistically significant ($\beta = -0.350, p = .082$).

Regarding marked information structures, results are also shown in the same Figure 5.6. In CLLDs, neither null nor overt pronouns were biased in their interpretation (null pronouns: $\beta = 0.218, p = .282$; overt pronouns: $\beta = -0.147, p = .461$). In subject clefts, null pronouns were also interpreted at chance ($\beta = 0.019, p = .925$) and overt pronouns displayed a significant bias toward object antecedents ($\beta = -0.747, p < .001$). In object clefts, null pronouns were significantly interpreted as coreferring with subject antecedents ($\beta = 0.668, p = .001$) and overt pronouns showed no bias ($\beta = 0.233, p = .249$).

Figure 5.6

Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in Spanish



Null pronouns only displayed a clear subject bias in object clefts. However, the increase of preference for subject antecedents in object clefts compared to unmarked sentences did not reach significance ($\beta = -0.524$, $p = .075$). The other structures, CLLD and subject clefts (in which the subject is not aligned with the topic of the sentence) did not differ in any way either from canonical sentences ($\beta = 0.199$, $p = .790$; $\beta = -0.450$, $p = .169$). Thus, an object antecedent syntactically marked as a focus seems to clarify the preference of null pronouns for subject antecedents, which is not well-defined in canonical contexts. Finally, in object clefts, subject antecedents were more preferred than in subject clefts ($\beta = -0.649$, $p = .015$). This demonstrates that, although there are few significant differences regarding the impact of non-canonical sentences on pronoun resolution, null pronouns are not insensitive to the syntactic and pragmatic position of the antecedents.

Overt pronouns are not insensitive either to changes in the information structure. Subject clefts are the only context in which overt pronouns display a clear object bias. However, this increase in the preference for object antecedents was not significant ($\beta = 0.397$, $p = .257$). Object clefts, on the other hand, increased the preference for subjects—in postverbal position—($\beta = -0.583$, $p = .035$), although overt pronouns were still interpreted as unbiased. Overt pronouns' biases were similar in CLLD and object clefts ($\beta = -0.381$, $p = .285$)—structures in which the object appears first-mentioned or in the left periphery. Differences were attested between CLLD and subject clefts ($\beta = 0.600$, $p = .028$), as well as between subject and object clefts ($\beta = -0.980$, $p < .001$). Significant differences arise when comparing the three marked information structures beyond the baseline condition.

The fact that no significant impact of CLLD or it-clefts is attested compared to unmarked structures may be explained by the lack of biases in the unmarked condition. Different non-canonical structures influence pronoun resolution.

Pairwise contrasts from the Pronoun's perspective reveal that a significant division of labor between null and overt pronouns is attested in three structures: in unmarked structures ($\beta = 0.494$, $p = .020$), in subject clefts ($\beta = 0.766$, $p < .001$) and object clefts ($\beta = 0.435$, $p = .047$). The difference in the interpretive biases of null and overt pronouns in CLLD was smaller ($\beta = 0.365$, $p = .088$).

In summary, the fact that null and overt pronouns do not show significant biases in unmarked contexts is probably masking the effects of marked information structures with respect to canonical sentences. In CLLD structures, both pronouns are similarly interpreted at chance. Subject clefts timidly increase the preference of overt pronouns for object antecedents, and object clefts show a "stronger" impact, increasing the preference of both null and overt pronouns for subject antecedents.

5.3.1.4 Discussion

Firstly, as an answer to RQ4a, null and overt pronouns in monolingual Spanish do not show clear resolution preferences in unmarked sentences: neither null nor overt pronouns show significant biases in their interpretation. Overt pronouns show a tendency to prefer object antecedents, but this did not reach significance. Secondly, as an answer to RQ4b, both null and overt pronouns were found to be sensitive to the manipulation of the information structure. The effects of marked structures are in line with the previous findings on bilingual Catalan and bilingual Spanish, described in §5.2.

5.3.1.4.1 Pronoun resolution in unmarked canonical sentences

Regarding pronoun resolution in unmarked structures in monolingual Spanish (RQ4a), the lack of significant preferences of null and overt pronouns in canonical sentences does not replicate previous studies using similar forced-choice tasks, which showed PAH-like biases in monolingual Spanish (Bel et al., 2016; Contemori & Di Domenico, 2021; de la Fuente, 2015). However, several studies on Spanish have shown that 1) null pronouns may remain unbiased in main-subordinate clause order (Bel & García-Alcaraz, 2018; Chamorro, 2018; de Rocafiguera & Bel, 2022) and that 2) overt pronouns may not

show well-defined preferences in Spanish, unlike other null subject languages (e.g., Filiaci et al., 2014). In contrast to these studies, which found unbiased patterns in the interpretation of either null or overt pronouns, Spanish showed mild biases in the interpretation of both null and overt pronouns in our data.

In terms of null pronouns, a clause order effect could explain their neutral bias in the current study. In main-subordinate sequences, subject and object antecedents have been argued to be similarly accessible in memory. These contexts would favor the retrieval of the object antecedent, weakening the preference of null pronouns for subject antecedents (see de Rocafiguera & Bel, 2022). Carminati (2002) explained this higher accessibility of the object referring to the fact that the temporal subordinate clause is attached to the VP, like the object antecedent, whose position is internal to the VP. The reason why this clause order effect does not similarly emerge in bilingual Catalan and bilingual Spanish (reported earlier in §5.2.3) using the same task and analyses, remains unexplained at this point. We will come back to this idea in the general discussion §5.4. In a less plausible explanation, comprehenders could also be relying on recency (Arnold, 1998, 2010), increasing the preference for the antecedent that has appeared closer to the ambiguous subject pronoun. However, recency was not found to play a role across conditions, and it would also not explain the difference between monolingual Spanish and the previous findings on bilingual Spanish and Catalan.

Regarding overt pronouns in monolingual Spanish, they showed a tendency to corefer with object antecedents, but it did not reach significance ($p = .082$). This non-significant tendency can be related to the flexible behavior that these pronouns seem to have in Spanish compared to other Romance null subject languages (see §2.3.3). Overt pronouns in Spanish seem to display weaker biases than in other null subject languages such as Italian or Catalan (in line with Bel & García-Alcaraz, 2018; Filiaci, 2011; Filiaci et al., 2014). Comparisons between Catalan and monolingual Spanish, as well as between bilingual Spanish and monolingual Spanish, will be further discussed in the next sections (§5.3.2 and §5.3.3).

5.3.1.4.2 Pronoun resolution in marked information structures

With regard to marked information structures (RQ4b) in monolingual Spanish, the results are in line with the previous findings on bilingual Catalan and bilingual Spanish (see

§5.2.7 for a summary). However, some of the effects demonstrated in bilingual Catalan and Spanish did not reach significance in monolingual Spanish data.

Regarding null pronouns, object clefts are the only structure affecting their interpretive preferences significantly. We had predicted that CLLDs and subject clefts would weaken the subject bias of null pronouns, but this decrease in preference was not attested. This is probably related to the fact that null pronouns were found to be unbiased in monolingual Spanish canonical sentences (5.33a). Although CLLD (5.33b) and subject clefts (5.33c) do not have a significant impact on the prominence of subject and object antecedents for null pronouns with respect to canonical sentences, null pronouns are interpreted as unbiased. Regarding object clefts (5.33d), they were predicted to have no effects on null pronouns, but they were found to increase their preference for subject antecedents. In fact, null pronouns are only biased toward subject antecedents in this context. An increase for subject preference in object clefts was also shown in one of de la Fuente's (2015) experiments, in which null pronouns were also interpreted at chance in the baseline condition⁵². As illustrated in example (5.33), null pronouns only clearly prefer the subject antecedent in the main clause (*Laura*) when it appears in a non-focal position (5.33d). It is possible that, despite the weak preferences of pronouns in Spanish, the anti-focus effect attested in object clefts is strong enough to make null pronouns reject clefted object antecedents and show a clear-cut bias toward subject antecedents.

- (5.33) a. *Laura_i asustó a María_j cuando **pro_{i/j}** entró en la habitación. (Canonical)*
'Laura scared Maria when (she) went into the room.'
- b. *A María_j la asustó Laura_i cuando **pro_{i/j}** entró en la habitación. (CLLD)*
'Maria, Laura scared her when (she) went into the room.'
- c. *Fue Laura_i quien asustó a María_j cuando **pro_{i/j}** entró en la habitación. (Subject cleft)*
'It was Laura who scared Maria when (she) went into the room.'
- d. *Fue a María_j a quien asustó **Laura_i** cuando **pro_i** entró en la habitación. (Object cleft)*
'It was Maria whom Laura scared when (she) went into the room.'

These findings do not contradict the idea that null pronouns are particularly sensitive to the syntactic function and the informational status of the antecedent, as demonstrated for bilingual Catalan and bilingual Spanish in §5.2. The fact that they remain unbiased in canonical contexts shows that there may also be other factors intervening in the mixed

⁵² De la Fuente (2015, Experiment 1) attributed the lack of biases of null pronouns in the baseline condition to a task effect: he only tested null pronouns. Using the same sentences in another study, including Pronoun as a condition (null vs. overt), and testing only canonical sentences, he demonstrated complementary null-subject and overt-object biases (Experiment 3).

and non-robust preferences of null pronouns in Spanish (such as word order factors, or the position of the antecedent in the syntactic configuration).

Referring now to overt pronouns, object clefts increase the coreference of overt pronouns with subject antecedents, as predicted. This effect, however, is not significant in CLLDs, in which the subject is also postverbal (OVS word order). In subject clefts, which maintain the surface SVO word order, the preference for object antecedents increases slightly and, only in these constructions, the non-clear-cut bias of overt pronouns becomes well-defined, as shown in example (5.34c).

- (5.34) a. Laura_i asustó a María_j cuando **ella**_{i/j} entró en la habitación. (*Canonical*)
'Laura scared Maria when she went into the room.'
- b. A María_j la asustó Laura_i cuando **ella**_{i/j} entró en la habitación. (*CLLD*)
'Maria, Laura scared her when she went into the room.'
- c. Fue Laura_i quien asustó a **María**_j cuando **ella**_j entró en la habitación. (*Subject cleft*)
'It was Laura who scared Maria when she went into the room.'
- d. Fue a María_j a quien asustó Laura_i cuando **ella**_{i/j} entró en la habitación. (*Object cleft*)
'It was Maria whom Laura scared when she went into the room.'

These findings support the hypothesis formulated following the results in Catalan and bilingual Spanish (§5.2): overt pronouns seem to be especially sensitive to the syntactic function of the antecedents and to their position in the syntactic configuration (which overlaps with word order), but not to their information status. Again, the fact that the impact of CLLD is weaker than that of object clefts may reveal that overt pronouns are not totally insensitive to discourse status. A fronted topicalized object seems to be slightly more accessible than a fronted focused object. It is possible that TopP occurs in a lower position in the syntactic configuration than FocP, so that a topicalized antecedent is structurally more accessible for a subject pronoun than a focused antecedent, as it is slightly closer to the subject pronoun (or to the VP)⁵³.

In short, these findings on marked information structures by Spanish monolinguals are very similar to those demonstrated by Spanish-dominant bilinguals in Spanish and Catalan-dominant bilinguals in Catalan. Thus, they are compatible with the possibility that null pronouns are particularly sensitive to the interaction between syntactic factors

⁵³ Within the cartographic approach, according to Rizzi (1997; see also Rizzi & Bocci, 2017), a recursion of topics is possible, giving rise to different relative orders in relation to FocP.

and the information status of an antecedent, and that overt pronouns are particularly sensitive to the interaction between syntactic factors and word order.

5.3.2 Comparing pronoun resolution in monolingual Spanish and Catalan

5.3.2.1 Aims and predictions

The second aim of §5.3 is to compare the pronominal biases of null and overt pronouns in Catalan and Spanish, crosslinguistically. Given that Catalan monolinguals do not exist, monolingual Spanish will have to be compared to bilingual Catalan by Catalan-dominant bilinguals. This comparison between Catalan and Spanish is necessary to confirm whether the two languages show consistent differences in the polarization of pronoun biases in unmarked structures, as has been suggested in the previous literature (Bel & García-Alcaraz, 2018) and in §5.2.3 above. Therefore, the following research question (RQ4.c) was formulated (see §5.1.1).

RQ4 c. Is microvariation attested in canonical and marked information structures between monolingual Spanish and Catalan?

Microvariation in pronoun resolution between these two null subject languages is expected. Following Bel and García-Alcaraz (2018), Catalan is predicted to show stronger PAH-like biases than monolingual Spanish in canonical sentences, especially regarding overt pronouns. According to Torregrossa et al. (2020), Catalan should show less flexible biases than Spanish, given that Catalan is more restrictive in allowing VSO word order than Spanish. There are no reasons to expect differences between languages regarding the impact of marked information structures (equivalent in both languages).

5.3.2.2 Method

5.3.2.2.1 Participants

The participants that completed the tasks in the present study have been described in §4.2.1. We compared pronoun resolution preferences in monolingual Spanish and in Catalan by modelling the behavior of Spanish monolinguals ($N = 40$; mean age: 20.12, range: 18-30) and Catalan-dominant bilinguals ($N = 34$; mean age: 22.32, range: 18-35).

5.3.2.2.2 *Materials*

See §5.3.1.2 or §4.3.2.

5.3.2.2.3 *Reported model*

We ran a mixed-effects logistic regression with monolingual Spanish responses and Catalan-dominant bilinguals' responses in Catalan. The fitted model included Pronoun (null, overt), Information structure (unmarked structures, CLLD, subject clefts, object clefts), and Language (Catalan, Spanish) as fixed effects, as well as their interactions. As random effects, varying intercepts for participants and items were added to the model, as well as a by-participant varying slope for the effect of Pronoun. Additional random slopes were tested but they either did not contribute to model fit or led to estimation problems within the models. The summary of this model is provided in Appendix H (Table H.5). The model's total explanatory power was moderate (conditional $R^2 = 0.23$) and the part related to the fixed effects alone (marginal R^2) was 0.07. No multi-collinearity issues were detected (highest VIF = 1.001). This model had a C-index of concordance of 0.76.

5.3.2.3 Results

Table 5.4 summarizes the proportion of subject interpretations of null and overt pronouns in monolingual and bilingual Spanish in the four analyzed contexts. These results for monolingual Spanish were already presented in the previous section (§5.3.1.3). Results in Catalan, from the responses given by Catalan-dominant bilinguals, are the same as those in §5.2.

Similar to the microvariation attested between bilingual Catalan by Catalan-dominants and bilingual Spanish by Spanish-dominants (see §5.2.3) in canonical contexts, the comparison between monolingual Spanish and Catalan shown in Table 5.4 seems to confirm crosslinguistic differences between these two languages in unmarked structures. The biases of both null and overt pronouns in Catalan seem much stronger than those of monolingual Spanish. However, the final interpretation of null and overt pronouns in the three marked information conditions appears to be very similar in both languages.

Table 5.4

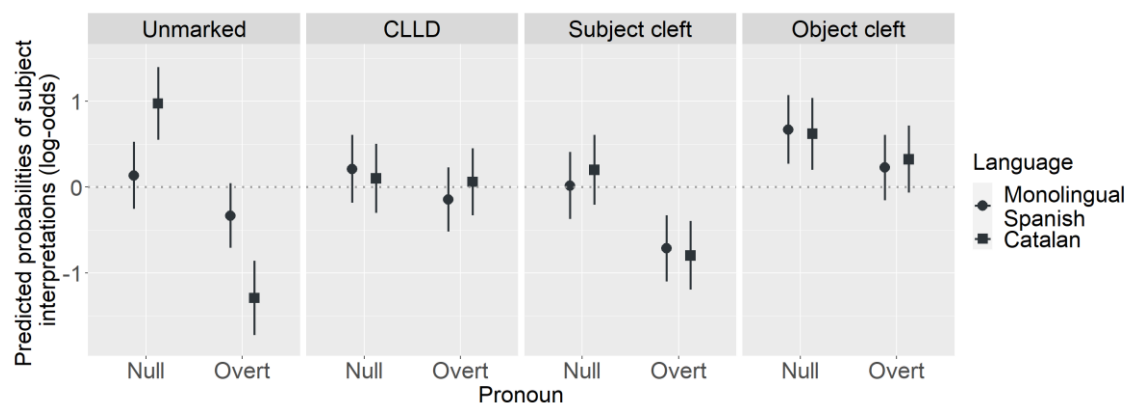
Proportion of subject interpretations in bilingual Catalan (by Catalan-dominant bilinguals), by information structure and type of pronoun (SD)

	Monolingual Spanish		Bilingual Catalan	
	Null	Overt	Null	Overt
Unmarked	.54 (.50)	.44 (.50)	.69 (.46)	.24 (.43)
Topicalized object	.55 (.50)	.46 (.50)	.52 (.50)	.53 (.50)
Focused subject	.51 (.50)	.35 (.48)	.54 (.50)	.33 (.47)
Focused object	.64 (.48)	.56 (.50)	.64 (.48)	.57 (.50)

Statistical analyses revealed no effects of Language alone ($\chi^2(1) = 0.004, p = .950$), but the two-way interaction Pronoun \times Language approached significance ($\chi^2(1) = 3.458, p = .063$) and the three-way interaction Pronoun \times Information structure \times Language was also significant ($\chi^2(3) = 27.200, p < .001$). Pairwise contrasts derived from the significant three-way interaction from the perspective of Language revealed significant differences between Spanish and Catalan only in canonical sentences. Null pronouns were more biased toward subject antecedents in Catalan than in Spanish ($\beta = 0.838, p < .001$), and overt pronouns more biased toward object antecedents in Catalan than in Spanish ($\beta = -0.960, p < .001$). These difference between languages in the unmarked condition is illustrated in Figure 5.7 (black color, round shape). No significant differences involving marked information structures were attested between the two languages.

Figure 5.7

Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in (monolingual) Spanish and in Catalan



5.3.2.4 Discussion

Referring to the comparison between Catalan and monolingual Spanish (RQ4c), results demonstrate that the biases of both null and overt pronouns in canonical sentences are significantly more polarized in Catalan than in Spanish. In marked information structures, on the other hand, no differences between Catalan and monolingual Spanish have been demonstrated in our data and null and overt pronouns show similar interpretive patterns. Crosslinguistic differences are thus only corroborated in unmarked contexts.

These findings provide further evidence on crosslinguistic microvariation between null subject languages (Bel & García-Alcaraz, 2018 for Catalan and Spanish; Contemori & Di Domenico, 2021 for Italian and Spanish; Leonetti-Escandell & Torregrossa, under review for Italian, Greek, and Spanish). Across studies, pronoun resolution in Spanish seems to display more tenuous biases than other null subject languages such as Italian. Regarding Catalan, it seems to behave more like Italian than Spanish. Moreover, as already pointed out by Contemori and Di Domenico (2021), it is not just overt pronouns that show more flexible interpretations in Spanish, but also null pronouns, contradicting the evidence from Filiaci (2011), Filiaci et al. (2014), and Bel and García-Alcaraz (2018). These findings will be further discussed in the general discussion (§5.4.3).

5.3.3 Comparing pronoun resolution in monolingual Spanish and bilingual Spanish in contact with Catalan

5.3.3.1 Aims and predictions

The third aim of §5.3 is to compare monolingual and bilingual Spanish. To do so, the Spanish-dominant bilingual group will be used; this group's results have already been discussed in §5.2. The main purpose of this comparison is to assess whether Spanish-dominant bilinguals present similar interpretive patterns in the tested structures to those of monolinguals. These results will reveal whether there is a bilingual Spanish contact variety in Catalonia that differs from monolingual Spanish. How bilingualism or language contact may be affecting the Spanish of bilingual speakers in Catalonia will be analyzed in more depth in Chapter 6, so it is out of the scope of the purpose of the present section.

This section will address RQ4d (see §5.1.1).

RQ4 d. Do pronominal anaphora resolution preferences in monolingual Spanish differ from those in bilingual Spanish in contact with Catalan? (§5.3.3)

Referring to canonical sentences, if a specific bilingual Spanish variety exists in Catalonia, stronger biases of overt pronouns are expected in bilingual Spanish compared to the monolingual variety, as attested in Bel and García-Alcaraz (2018), due to the potential influence of Catalan (or the convergence with Catalan). Concerning null pronouns, they are predicted to be similarly biased toward subject antecedents in the two Spanish varieties. Regarding marked information structures, there is no evidence indicating that their effects should differ in the two Spanish varieties. However, if bilinguals find these structures involving the syntax-pragmatics interface cognitively challenging, differences between monolingual and bilingual individuals attributable to bilingualism may emerge (Sorace, 2011). Considering our previous findings on bilingual Spanish, differences in this regard are not expected. Spanish-dominant bilinguals were not found to show systematic indeterminacy in their interpretation of overt pronouns.

5.3.3.2 Method

5.3.3.2.1 Participants

The participants that completed the tasks in the present study have been described in §4.2.1. To compare pronoun resolution preferences in monolingual Spanish and in bilingual Spanish, we modelled the behavior of Spanish monolinguals ($N = 40$; mean age: 20.12, range: 18-30) and Spanish-dominant bilinguals ($N = 29$; mean age: 23.28, range: 19-26).

5.3.3.2.2 Materials

See §5.3.1.2 or §4.3.2.

5.3.3.2.3 Reported model

We ran a mixed-effects logistic regression with monolingual Spanish responses and Spanish-dominant bilinguals' responses in Spanish. The fitted model included Pronoun (null, overt), Information structure (unmarked structures, CLLD, subject clefts, object

clefts), and Group (monolinguals, bilinguals) as fixed effects, as well as their interactions. As random effects, varying intercepts for participants and items were added to the model and a by-participant varying slope for the effect of Pronoun. Additional random slopes were tested but they either did not contribute to model fit or led to estimation problems within the models. The summary of this model is provided in Appendix H (Table H.10). The model's total explanatory power was moderate (conditional $R^2 = 0.23$), and the part related to the fixed effects alone (marginal R^2) was 0.05. No multi-collinearity issues were detected (highest VIF = 1.001). This model had a C-index of concordance of 0.76.

5.3.3.3 Results

Table 5.5 summarizes the proportion of subject interpretations of null and overt pronouns in monolingual and bilingual Spanish in the four analyzed information structures. These results for monolingual Spanish were already presented in the previous section (§5.3.1.3). Results in bilingual Spanish, from the responses given by Spanish-dominant bilinguals, are the same as those in §5.2.

Table 5.5

Proportion of subject interpretations in monolingual Spanish and in bilingual Spanish (by Spanish-dominant bilinguals), by information structure and type of pronoun (SD)

	Monolingual Spanish		Bilingual Spanish	
	Null	Overt	Null	Overt
Unmarked	.54 (.50)	.44 (.50)	.60 (.49)	.38 (.49)
Topicalized object	.55 (.50)	.46 (.50)	.50 (.50)	.43 (.50)
Focused subject	.51 (.50)	.35 (.48)	.47 (.50)	.29 (.46)
Focused object	.64 (.48)	.56 (.50)	.63 (.48)	.54 (.50)

Some differences between the varieties of Spanish can be appreciated based on the descriptives. In canonical sentences, the division of labor between null and overt pronouns is larger in the bilingual than in the monolingual variety, and CLLD does not seem to change pronoun preferences in Spanish compared to unmarked contexts. In these constructions, monolinguals do not seem to show interpretive patterns that differ from chance. The results for focusing constructions, however, are almost identical in the

two groups. Bilingual Spanish does not seem to show more optionality compared to monolingual Spanish in non-canonical structures.

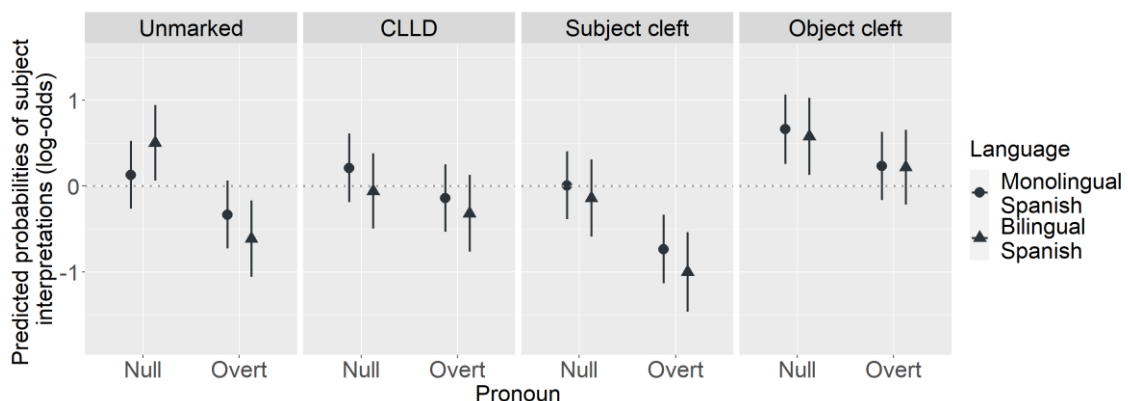
Looking at the results of the model, no significant effect of Group was attested ($\chi^2(1) = 0.792, p = .373$), and the three-way interaction Pronoun \times Information structure \times Group did not improve a model with two-way interactions ($\chi^2(3) = 3.317, p = .345$) or with no interactions ($\chi^2(10) = 12.152, p = .275$). Therefore, the two varieties of Spanish (monolingual vs. bilingual Spanish in contact with Catalan) did not differ statistically when modeled.

However, unlike monolingual speakers, Spanish-dominant bilinguals showed clear-cut interpretive biases of null and overt pronouns in unmarked structures, preferring subject and object antecedents, respectively. As reported in the previous section, neither null nor overt pronouns in monolingual Spanish showed significant biases. When looking at pairwise contrasts from the perspective of Group, in unmarked structures, bilinguals and monolinguals did not significantly differ (null pronouns: $\beta = 0.373, p = .149$; overt pronouns: $\beta = -0.281, p = .279$).

Similar interpretive biases between monolingual and bilingual Spanish were also attested regarding marked information structures. This can be observed in Figure 5.8, in which the preferences of null and overt pronouns in the four analyzed contexts are presented for each group.

Figure 5.8

Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in Spanish by monolinguals and by Spanish-dominant bilinguals



5.3.3.4 Discussion

The comparison of monolingual Spanish and bilingual Spanish (by Spanish-dominant bilinguals) does not prove that the two groups interpret subject pronouns in a different manner in any condition (RQ4d). Like bilinguals, monolinguals interpret null and overt pronouns in marked information structures as unbiased when syntactic, pragmatic, and sequential cues are put into conflict. Null pronouns are interpreted at chance when the subject and the sentence topic do not coincide, and overt pronouns are also interpreted at chance when the object appears in the left periphery, in an external position to the VP.

The only difference between Spanish monolinguals and bilinguals is related to canonical sentences and it was not statistically significant when comparing the two groups overtly. While the biases of monolingual Spanish do not reach significance, bilingual Spanish shows well-defined biases of both null and overt pronouns, closer to the Catalan system. As suggested by Bel and García-Alcaraz (2018), this slight difference between monolingual and bilingual Spanish speakers may be indicative of crosslinguistic influence of Catalan or convergence in language contact situations. In fact, Catalan shows more polarized biases than Spanish in pronoun resolution (see §5.3.2). This comparison between monolingual Spanish and bilingual Spanish by Spanish-dominant bilinguals in Catalonia, however, provides limited evidence to thoroughly examine the emergence of crosslinguistic influence or to assess bilingual effects, if any. Therefore, the aim of the following chapter (Chapter 6) will be to explore bilingual effects and the possible occurrence of crosslinguistic influence in bilingual Spanish or Catalan in relation to language dominance.

Another complementary explanation for the observed differences between bilingual and monolingual Spanish could be related to a bilingual advantage in metalinguistic awareness (e.g., Adesope et al., 2010; Bialystok, 2001) or in statistical learning abilities (see Weiss et al., 2020 and references within). If bilingualism confers an advantage on tracking distributional regularities in a language, as these studies suggest, this may make the two groups perform slightly differently in an essentially metalinguistic task. Bilinguals may outperform monolinguals in uncovering regularities and therefore disclose more categorical and more regular patterns in their responses when interpreting ambiguous anaphoric pronouns. However, we can only point out this possibility, which may be worth further exploration, and we should take into account that an advantage in metalinguistic awareness has not been found systematically across bilingual populations (e.g., Gathercole et al., 2014). Moreover, it is worth keeping in mind that bilingual and monolingual Spanish were not found to diverge significantly.

5.3.4 Summary of main findings

In brief, monolingual Spanish was found to display very weak and non-significant biases of both null and overt pronouns in canonical sentences. This lack of biases has not been attested in the previous literature in monolingual Spanish, although mixed findings have been found across studies (e.g., Bel & García-Alcaraz, 2015, 2018; Chamorro et al., 2016; Filiaci et al., 2014; García-Alcaraz, 2015; see §2.3.2). Overall, we have related these mixed findings to a combination of clause order effects (e.g., de Rocafiguera & Bel, 2022) and to the fact that Spanish seems to show more flexible interpretations than other null subject languages (e.g., Contemori & Di Domenico, 2021). Importantly, the biases of null and overt pronouns in marked information structures (CLLD, subject clefts, and object clefts) were similar to those described in §5.2 for bilingual Catalan and bilingual Spanish.

Only in canonical sentences, monolingual Spanish was found to differ from (bilingual) Catalan in the interpretation of both null and overt pronouns, Spanish showing significantly weaker preferences (in line with Bel & García-Alcaraz, 2018). However, although bilingual Spanish showed well-defined interpretive biases in canonical sentences and monolingual Spanish non-significant biases, these two varieties were not demonstrated to differ significantly when they were compared. Overall, these findings provide further evidence on the existence of microvariation among null subject languages (Bel & García-Alcaraz, 2018; Contemori & Di Domenico, 2021; Filiaci et al., 2014; Torregrossa et al., 2020) and do not confirm the existence of a specific variety of bilingual Spanish in contact with Catalan regarding the phenomenon under study.

5.4 General discussion

This chapter has investigated the interpretation of globally ambiguous null and overt subject pronouns in Catalan and Spanish, and how they are affected by syntactic factors (syntactic function of the antecedent), pragmatic factors (information status of the antecedent), and sequential factors (word order or linear position of the antecedents, which overlap with their hierarchic position in the syntactic configuration). To disentangle these factors, the information structure of the clause containing the plausible antecedents for a pronoun was syntactically manipulated. Four different contexts were analyzed: unmarked canonical sentences, object topicalization via clitic-left dislocation structures (CLLD), subject focusing via it-cleft structures (subject clefts), and object focusing via it-

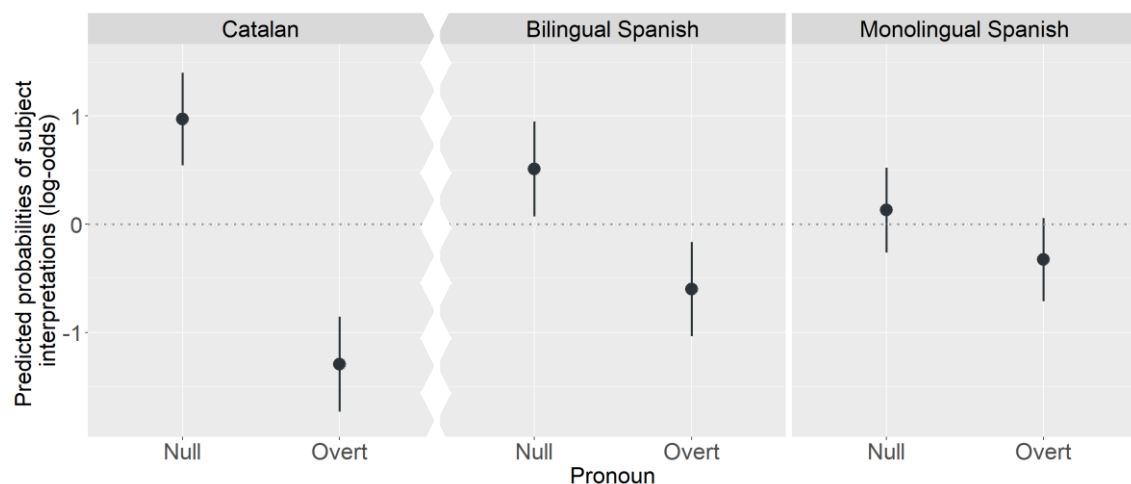
cleft structures (object clefts). At the same time, microvariation between two Romance null subject languages—Catalan and Spanish—was assessed.

5.4.1 The interpretive preferences of null and overt pronouns in canonical sentences in Catalan and Spanish

Firstly, null and overt pronouns in canonical sentences were found to show clear-cut biases toward subject and object antecedents, respectively, in both bilingual Catalan and bilingual Spanish (RQ1). The results therefore confirmed that null and overt pronoun resolution in these two languages can be predicted by the Position of Antecedent Hypothesis (PAH; Carminati, 2002). These well-defined biases, however, were not found in the monolingual Spanish variety, in which both null and overt pronouns were interpreted as unbiased, even though overt pronouns showed a non-significant tendency to corefer with object antecedents (RQ4a). The predicted interpretation of null and overt subject pronouns in Catalan, bilingual Spanish, and monolingual Spanish is graphically illustrated in Figure 5.9 below.

Figure 5.9

Predicted interpretations of null and overt pronouns as coreferring with subject antecedents in canonical sentences in Catalan, bilingual Spanish, and monolingual Spanish



Note. 'Catalan' results come from Catalan-dominant bilinguals completing the task in Catalan, and 'bilingual Spanish' and 'monolingual Spanish' show results from the task in Spanish by Spanish-dominant bilinguals and monolinguals.

In the previous figure, we can observe a progressive attenuation in the polarization of the biases of null and overt pronouns in Catalan, bilingual Spanish, and monolingual Spanish. Catalan differs from both bilingual and monolingual Spanish in pronominal resolution patterns. When it comes to bilingual Spanish, slight differences can be observed in comparison to monolingual Spanish, which could seemingly be attributed to some traces of crosslinguistic influence from Catalan. We will further investigate the effects of bilingualism in Chapter 6, and we will see that the influence of Spanish on Catalan is stronger than that of Catalan on Spanish. However, we can still appreciate how more well-defined interpretive patterns are observed in bilingual Spanish than in the monolingual variety (keeping in mind that bilingual and monolingual Spanish did not differ significantly).

Overall, our results showed that both Catalan and bilingual Spanish display robust PAH-like biases, a finding which seems to be consistent across studies (in line with Mayol & Clark, 2010 for intersentential contexts in Catalan; Bel & García-Alcaraz, 2015, 2018 for subordinate-main intrasentential contexts in bilingual Spanish and Catalan). Although we analyzed main-subordinate sequences, both null and overt pronouns showed well-defined biases, in contrast to Bel and García-Alcaraz's studies, who found null pronouns to remain unbiased in these contexts. These different interpretations of null pronouns in main-subordinate sequences in our study and in Bel and García-Alcaraz (2018) may be attributable to the use of different tasks in the experimental design. When an acceptability judgment task is used (as in Bel & García-Alcaraz, 2015, 2018), participants assess all possible interpretive options for each pronoun, including the less preferred interpretations (i.e., null pronouns as coreferring with object antecedents, and overt pronouns with subject antecedents). These possibilities are not assessed in a forced-choice task (as in the present study). Thus, in main-subordinate contexts, participants may similarly accept the interpretation of null pronouns as coreferring with subject and object antecedents, but may prefer coreference with subject antecedents if they are forced to choose an interpretation. In the case of overt pronouns, they prefer object interpretations and do not seem to accept coreference with subject antecedents. In this regard, acceptability judgement tasks seem to capture more nuances than forced-choice tasks.

However, as we have seen, monolingual Spanish showed very feeble and non-significant biases for both pronouns, even though the same forced-choice task was used. The results did not show any preferences for null pronouns, and only showed a non-

significant tendency of overt pronouns to corefer with object antecedents. These results diverge from previous studies that show a strong bias of null pronouns toward subject antecedents and a more flexible or undefined bias in the case of overt pronouns (Alonso-Ovalle et al., 2002; Bel & García-Alcaraz, 2018; Filiaci, 2011; Filiaci et al., 2014; Jegerski et al., 2011; Keating et al., 2011). In fact, the biases of null pronouns in our data seem to be even weaker than those of overt pronouns, in line with studies demonstrating clause order effects (de Rocafiguera & Bel, 2022), and suggests that the interpretive bias of null pronouns toward subject antecedents may not be as steady as previously assumed. In §5.3.1.4, we attributed the nonsignificant resolution patterns of Spanish monolinguals to structural factors derived from main-subordinate clause order.

There are other factors that may not have favored the null-subject bias. For instance, the subordinate sentences in which the subject pronoun appeared were temporal clauses, instead of, for example, if-clauses. Carminati (2002) showed stronger pronoun resolution preferences in if-clauses than in temporal clauses. She explained this contrast by referring to the fact that if-clauses are more likely to be attached higher (to the IP) than temporal clauses (to the VP), which would favor coreference with antecedents in the same VP and would therefore affect null pronouns in particular. Also, our experimental stimuli used *cuando* as the temporal connector, which does not seem to favor subject coreference as much as *mientras* (Martín-Villena et al., 2021). However, all these factors should have similarly affected the biases of Catalan and bilingual Spanish pronouns, given that the stimuli were the same in Spanish, and equivalent in Catalan, and no differences should be expected between Catalan and Spanish regarding these factors. Therefore, clause order (or the type of clause, or the temporal connector) would not explain why Spanish-dominant bilinguals showed significant biases when interpreting these pronouns in the same context, a pattern that differs from that of Spanish monolinguals, and which was also attested in Catalan.

Evidence from bilingual Spanish and Catalan has provided more arguments to examine the watered-down pronoun resolution patterns of Spanish monolinguals. Taking into account that the intrinsic preferences of Spanish appear to be weaker than those of other null subject languages, it is possible that these already very weak pronominal biases may appear as inexistent when tested in contexts that are globally ambiguous and do not favor coreference with subject antecedents in any manner. In contrast to other studies, the implicit causality of the main verbs was systematically controlled so that the experimental items were kept as ambiguous as possible. It is precisely in ambiguous

contexts where the pronouns should show their intrinsic biases (if any), and in monolingual Spanish null pronouns' intrinsic biases appear as virtually inexistent.

Taken together, the lack of bias of null pronouns in monolingual Spanish shows that null pronouns in monolingual Spanish do not have as well-defined and steady biases as previously assumed, and that both null and overt pronouns show weak resolution patterns. As also mentioned by Contemori and Di Domenico (2021, p. 28), mixed results across studies in Spanish may be explained by an attenuated division of labor in pronominal resolution in this language. If the interpretive biases of null and overt pronouns in Spanish are not so robust, they can change more radically across studies due to small differences in the experimental design.

5.4.2 The influence of syntactic, pragmatic and word order factors on null and overt pronoun resolution

Null and overt interpretive biases toward subject and object antecedents in canonical sentences were not maintained when the information status of subject and object antecedents was manipulated using syntactically marked information structures. This means that information structure impacts pronoun resolution and provides information on the interpretive properties of null and overt subject pronouns. Overall, the results are very similar in Catalan and Spanish, and suggest a different sensitivity of null and overt pronouns to the information status and the linear position of the antecedent (or word order), which interact with the syntactic function of the antecedent. The impact of the three sets of factors being studied on the interpretation of null and overt pronouns is summarized in Table 5.6 below. On the one hand, null pronouns only show a clear subject bias when the subject antecedent conveys non-focal (i.e., topical) information, regardless of the surface linear position of the antecedents. On the other, overt pronouns only show a clear object bias when the object antecedent appears in a non-initial position (i.e., in its canonical VP-internal postverbal position), regardless of the information status of the antecedents. Therefore, null pronouns have been found to be sensitive to the syntactic function and the information status of the antecedents, while overt pronouns appear to be rather sensitive to the syntactic function and the linear/hierarchic position of the antecedents.

Table 5.6

Summary of the features that make an antecedent more preferred for a null or an overt subject pronoun

	Null pronouns	Overt pronouns
Syntactic function	subject > object	subject < object
Information status	topical > focal	topical = focal
Order of mention	first = second mention	first < second mention

Another relevant finding is that when conflicting cues were encountered, pronouns were interpreted at chance. This means that when subject antecedents conveyed focal information, null pronouns were interpreted as unbiased, and when object antecedents appeared in an initial surface position, overt pronouns were interpreted as unbiased (see the summary in §5.2.7). This finding confirms that pronoun resolution is governed by multiple constraints that cannot be ranked, i.e., they have similar weights. Hence, there does not appear to be a more determining cue that outranks the others in the final interpretation of either null or overt pronouns in Catalan and Spanish (among the syntactic, pragmatic, and sequential factors under analysis). If the relevant cues are misaligned, pronominal preferences become flexible and indeterminate.

The fact that pronominal preferences are determined by a combination of factors confirms that there is not a single cue that can account for the interpretive biases of null and overt pronouns. Our findings support a multi-factorial approach (Arnold et al., 2000; Blything et al., 2021, 2022; Järvikivi et al., 2005, 2014; Kaiser & Trueswell, 2008; Schumacher et al., 2016, 2017; among others). Furthermore, the different sensitivity of null and overt pronouns to the analyzed cues supports a form-specific multiple constraint approach to pronominal anaphora resolution; different pronominal forms are affected by different—and multiple—factors (Kaiser & Trueswell, 2008; Bader & Portele, 2019 for personal and demonstrative pronouns in German). This form-specific account for pronoun resolution also implies that null and overt pronouns in Catalan and Spanish may not be in complementary distribution, as previously widely assumed. In completely ambiguous contexts, null pronouns may prefer to be interpreted as expressing topic continuity (García-Alcaraz & Bel, 2019; Lozano, 2016), but overt pronouns may not be as specialized on expressing topic shift. Instead, and in contrast to null pronouns, they appear to be governed mainly by syntactic and structural constraints. It is a possibility

that null pronouns are in complementary distribution with lexical subjects or with demonstrative pronouns, instead of overt pronouns (as suggested by Torregrossa et al., 2020; see also Giannakou & Sitaridou, 2020).

Regarding the interpretation of null pronouns, it is worth referring to Torregrossa et al.'s (2020) proposal. Despite recognizing the need for a multi-factorial approach to account for pronoun resolution, these authors give special importance to the hierarchic position of an antecedent in the syntactic structure for the interpretation of null pronouns (see §2.3.1), based on Rizzi (2018). More specifically, Torregrossa et al. (2020, p. 9) claim that an antecedent will be more prominent (and thus more preferred) for a null pronoun if it is hierarchically higher than the competing antecedent. In relation to the PAH, this claim implies that null pronouns will prefer antecedents that occupy a higher position in the phrase structure rather than antecedents in Spec,IP, regardless of their syntactic function (subject vs. object). In contrast to their proposal, the present research demonstrates that an antecedent that appears higher in the syntactic configuration is not necessarily more preferred for a null pronoun than an antecedent in a lower position. In other words, the interpretation of null pronouns does not seem to be determined by the hierarchic position of the antecedent. Antecedents in Spec,IP (i.e., subjects⁵⁴) are preferred over antecedents in the highest configurational position in the interpretation of null pronouns (e.g., in object clefts, the subject antecedent is preferred over the object antecedent, which appears higher in the syntactic structure).

When the object antecedent appears in a higher Spec,TopP position than the subject (i.e., in CLLDs), it does become more preferred for null pronouns—blurring their subject bias. However, in object clefts (e.g., *Fue a Laura_k a quien asustó María_j cuando pro_j entró en la habitación*; see §5.2.5), null pronouns show a clear-cut bias toward (postverbal) subject antecedents. Hence, despite being in a hierarchically higher position (Spec,FocP), object antecedents do not appear to have any feature that makes them compete with subject antecedents when it comes to the interpretation of null pronouns. This asymmetry in the preference of null pronouns for objects in Spec,FocP and in Spec,TopP reveals that, when referring to the interpretation of null pronouns, topicality might be a more determining cue than structural height (overlapping with surface position), as summarized in Table 5.6 above. According to our data, the antecedent in Spec,IP (i.e., the subject) would be more preferred for null pronouns than the antecedent

⁵⁴ Considering that the postverbal subject antecedent in CLLD and object clefts occupies a Spec IP position, assuming Olarrea's (2002) analysis.

in the highest position in the syntactic structure. Our study also provides specific evidence on the weighting of subjecthood: at least in ambiguous contexts, subjecthood is as relevant as topicality.

Regarding the interpretation of overt pronouns, we have found that they may not be governed by pragmatic features, or not as much as previously assumed in the literature. In contrast to null pronouns, overt pronouns seem to be more sensitive to structural factors, in relation to the syntactic function of the antecedent (they prefer to corefer with object antecedents) and its structural position in the syntactic configuration (they prefer postverbal, hierarchically lower antecedents). As predicted by the PAH (Carminati, 2002), overt pronouns shows a preference for the antecedent in the lowest position in the syntactic structure, unless it is a subject (i.e., it appears in Spec,IP). Filiaci (2011) referred to Cardinaletti and Starke's (1999) crosslinguistic typology of deficient forms to explain the differences in the interpretation of overt pronouns between Italian and Spanish (p. 219). In contrast to the Italian overt pronouns *lui/lei*, the Spanish overt pronouns *él/ella* would be weak elements or structurally deficient, similar to *egli/ella* in Italian (Cardinaletti & Starke, 1999). In this way, overt pronouns in Spanish would not be so restricted to corefer with non-topical object antecedents and would allow coreferences with topical or subject antecedents. In a similar vein, Licerias and Alba de la Fuente (2015) propose that Spanish might have two kinds of overt subject pronouns: 1) weak overt pronouns that would behave like a free phonetically realized counterpart of null pronouns and would convey the same pragmatic features, and 2) strong pronouns, conveying pragmatic features such as focus. Although he does not refer to Romance languages, Vallduví (1994, p. 13) refers to the dichotomy between strong and weak pronouns in English and defines weak pronouns to be "inert as far as information packaging is concerned" (see also Vallduví & Engdahl, 1996, p. 476). Overt pronouns in Catalan and Spanish in the sentences that we tested in the present study could be interpreted as pragmatically 'inert' weak pronouns, not showing a special sensitivity to topic and focus features.

5.4.3 Microvariation between Catalan, bilingual Spanish, and monolingual Spanish

Crosslinguistic differences were observed between Catalan and Spanish (both the bilingual and monolingual varieties) in canonical sentences. Firstly, the very clear and categorical PAH-like interpretations of Catalan contrast with the very tenuous and non-

significant biases of pronoun resolution in monolingual Spanish. The interpretation of both null and overt pronouns was significantly different when comparing Catalan by Catalan-dominant bilinguals and Spanish by monolinguals and Spanish-dominant bilinguals. Secondly, the interpretive biases of Spanish-dominant bilinguals in Spanish seem to be between Catalan and monolingual Spanish. In contrast to monolinguals, Spanish-dominant bilinguals display well-defined biases, but no significant differences are found when the two groups are compared. These findings provide novel evidence on microvariation between null subject languages and confirm that there are differences between Catalan and Spanish (in line with Bel & García-Alcaraz, 2018). However, we cannot reliably confirm that the bilingual Spanish variety in contact with Catalan differs from monolingual Spanish (in contrast to Bel & García-Alcaraz, 2018). Importantly, the resolution patterns of null and overt pronouns in marked information structures (i.e., CLLDs, subject clefts, and object clefts) were similar in both Catalan and monolingual and bilingual Spanish.

The crosslinguistic differences between Catalan and Spanish shown in unmarked contexts are similar to those found by Contemori and Di Domenico (2021) and Leonetti-Escandell and Torregrossa (under review) between Spanish and Italian (see also Filiaci et al., 2014, although they only found differences in relation to overt pronouns). Catalan may behave very similarly to Italian, given that the two languages show very polarized biases and differ from Spanish in a similar manner. Parallel differences have also been found in the comparison between Spanish and Greek, with Spanish showing unbiased and significantly weaker interpretations than the well-defined biases of Greek (Giannakou & Sitaridou, 2020 for overt pronouns; Leonetti-Escandell & Torregrossa, under review for both null and overt pronouns). Torregrossa et al. (2020) also demonstrated microvariation between Italian and Greek, with Italian being more categorical than Greek. This evidence seems to point toward a gradient polarization of pronoun resolution biases in null subject languages. Italian, and possibly Catalan as well, show more categorical biases in the interpretation of null and overt pronouns than Greek, and Greek shows more well-defined biases than Spanish (Italian > Catalan > Greek > Spanish).

The reason why pronominal subjects display more categorical and polarized interpretive biases in Catalan than in Spanish, in which only very tenuous biases can be observed, would deserve a more thoughtful investigation, which is out of the scope of the present thesis and needs to be explored in further research. Therefore, a definitive answer will not be provided and our contribution to this debate will be very modest. It has been

suggested that this asymmetry between null subject languages is related to variation in word order constraints in previous studies (Leonetti-Escandell & Torregrossa, under review for Spanish, Greek and Italian; Torregrossa et al., 2020 for Greek and Italian). These authors explained the more flexible interpretations of Greek compared to Italian by referring to word order constraints. While Greek allows for VSO word order, it is not possible in Italian. As a consequence, these authors hypothesize that, when it comes to anaphora resolution, comprehenders rely on the syntactic position of subject and object antecedents to a greater extent in Italian than in Spanish/Greek (see §2.3.3). This view could also explain the contrast between Catalan and Spanish. Similarly to Italian, VSO order is more restricted in Catalan than in Spanish. More specifically, Catalan does not allow for VSO word order in broad focus sentences (e.g., Colomina, 2019; Ordóñez, 1998; Solà Pujols, 1992).

Depending on how restrictive Romance languages are concerning word order, Leonetti (2016, 2017) proposes a distinction between Central Romance languages (French, Italian, Catalan) and Peripheral Romance languages (Spanish, Portuguese, Romanian). On the one hand, the central Romance group is characterized as being more restrictive in word order and as having the need to establish a topic-comment or focus-background informational partition between the fronted constituent and the rest of the sentence. In this way, these languages show a transparent and straightforward mapping of grammar to informational interpretations. On the other hand, the Peripheral Romance group is less restrictive (or more permissive) regarding marked orders. The mapping grammar-information structure is less straightforward, and the informational interpretation is determined to a greater extent by the context than by the grammar. Thus, it is possible that Italian and Catalan, as Central Romance null subject languages, display more categorical (i.e., more restrictive) pronoun resolution biases. Spanish, as a Peripheral Romance null subject language, may display more flexible (or less restrictive) interpretive biases. Cruschina and Mayol (2022), in an experimental comparative study on the syntactic position of focus in Romance languages, suggest that Leonetti's proposal could be improved by formulating the distinctions in a continuum: French > Italian > Catalan > Spanish. French does not allow focus fronting, Italian was found to show the most restrictive preferences, Catalan was found to be more permissive than Italian, and Spanish was found to be the least restrictive language. A similar continuum might explain the strength of the interpretive biases of null and overt subject pronouns in Italian, Catalan, and Spanish.

In Portuguese, however, which also belongs to the peripheral less restrictive group and allows VSO word order, categorical and well-defined PAH-like biases have been attested in previous research (e.g., Castro et al., 2017; Madeira et al., 2021; Rinke & Flores, 2018; Teixeira et al., 2022). This challenges the idea that more restrictive languages in terms of word order show a greater sensitivity to syntactic cues in pronoun resolution and more constrained interpretations. In fact, Madeira et al. (2021) found European Portuguese to display more categorical biases than Italian, especially in the interpretation of null pronouns (in main-subordinate contexts). In future research, it would be worth comparing pronoun resolution between Catalan and Italian, and comparisons with European Portuguese should also be explored. Comparing anaphora resolution patterns across Romance null subject languages using equivalent tasks could provide valuable insights for the study of microvariation in anaphora resolution.

Chapter 6

Bilingualism effects on pronominal anaphora resolution in Catalan and Spanish in contact: results and discussion

6.1 Introduction

In the previous chapter (Chapter 5), microvariation was found between Catalan and Spanish in the interpretation of null and overt pronouns in these two null-subject languages. In canonical sentences, Catalan follows the biases predicted by the Position of the Antecedent Hypothesis (PAH; Carminati, 2002) in a clear-cut manner, as does bilingual Spanish. Null pronouns are preferably interpreted as coreferring with subject antecedents, and overt pronouns with object antecedents. However, these pronominal biases are less pronounced in bilingual Spanish than in bilingual Catalan (see §5.2.3). Spanish in monolingual speakers showed very mitigated and almost undefined biases of both null and overt pronouns, displaying a very different picture from that of Catalan (see §5.3.2). Although differences in the definiteness of the interpretive patterns of monolingual and bilingual Spanish were observed, no significant differences arose when the two groups were directly compared. As such, inconclusive evidence was found on whether there is a specific variety of Spanish in Catalonia that differs from monolingual Spanish (see §5.3.3). Interestingly, no differences between languages or varieties were shown regarding marked information structures.

With these findings, several questions remain unanswered in relation to bilingualism and on how it may (or may not) be affecting pronoun resolution in Catalan and in Spanish by early Catalan-Spanish bilinguals. To this point, limited evidence has been presented in terms of discussing the effects of bilingualism. We have focused only on studying two groups of bilinguals representing two poles of language dominance: we assessed pronoun resolution in Catalan by Catalan-dominant bilinguals, and in Spanish by Spanish-dominant bilinguals. However, do all Catalan-Spanish bilinguals—with different dominance profiles—behave like Catalan-dominant bilinguals in Catalan and like Spanish-dominant bilinguals in Spanish? How do Catalan-dominant and Spanish-dominant bilinguals interpret pronouns in their weaker language? Do balanced bilinguals differ from Catalan-dominant and Spanish-dominant bilinguals in Catalan and/or in Spanish? Taking all of these factors into consideration, a general question arises: (how)

does language dominance—and bilingualism—modulate pronoun interpretation in highly proficient early bilinguals?

Language dominance—understood as a multidimensional, gradient, and dynamic construct—was used as a proxy for bilingualism to classify Catalan-Spanish bilinguals into three groups: Catalan-dominant, balanced, and Spanish-dominant. For this purpose, participants completed the BLP questionnaire (Birdsong et al., 2012; see §4.2.2). In this way, language dominance was conceived as a relative construct composed of estimations of language exposure, language use, proficiency, and attitudes in each of the bilinguals' languages (Gertken et al., 2014). Catalan-Spanish bilinguals who have been raised in Catalonia are highly proficient and educated in both of their languages, they are all early bilinguals, and they have all grown up in an essentially bilingual society (see §3.3.1 for a description of bilingualism and language contact in Catalonia, and §4.2 for a description of participants). In this context, language dominance has been argued to be more reliable than other measures such as language proficiency as a means of identifying different bilingual profiles (Perpiñán, 2017; Puig-Mayenco et al., 2018; Sebastián-Gallés et al., 2005; among others).

The main goal of the present chapter is to investigate the role of language dominance in pronoun interpretation by early bilinguals in canonical sentences and in marked information structures. To do so, we will refer to two meaningful approaches which have been proposed in the framework of bilingual anaphora resolution (see §3.4): 1) the emergence of crosslinguistic influence (CLI), and 2) the emergence of a general bilingual strategy related to bilingualism per se.

On the other hand, pronominal anaphora resolution has been claimed to be a locus of CLI. Despite the typological similarity of the two languages under study, Catalan and Spanish show microvariation with regards to pronoun interpretation in canonical sentences. Recent studies on pronominal subjects in Romance languages have claimed that CLI occurs in bilinguals' referential systems and is the source of variation among bilingual populations (de Prada Pérez, 2019; Romano, 2019). Based on these authors' proposals, which make testable predictions, we will examine the role of CLI between Catalan and Spanish, as well as its directionality and intensity and whether it is conditioned by language dominance. Given that microvariation between Catalan and Spanish has been shown in unmarked structures but not in marked information structures, CLI would only be expected to emerge in canonical sentences.

On the other hand, we will assess the emergence of general effects of bilingualism attributable to bilingualism in itself. Pronominal anaphora resolution requires the integration of information from different linguistic domains, which has been claimed to be challenging for bilingual speakers. These processing inefficiencies related to bilingualism may lead bilinguals to resort to optionality as a cognitive strategy, mainly affecting the interpretation of overt pronouns (Sorace, 2011; Sorace & Filiaci, 2006; among others). In this view, as opposed to the previous one, CLI would not be the only source of discrepancies in the pronominal resolution preferences of the target language. Rather, pronoun interpretation would be affected by bilingualism itself, irrespective of the language pair. In the light of the mentioned studies, we also analyze whether early functional bilinguals behave similarly to other bilingual populations in showing optionality (i.e., indeterminacy of overt pronouns), as well as whether this possible general bilingual behavior is modulated by language dominance. General bilingualism effects would be predicted to emerge not only in canonical sentences, but also in more complex sentences with marked information structures. Anaphora resolution requires the integration of different syntactic and pragmatic cues and is possibly even more cognitively demanding to interpret in marked information structures than in canonical sentences. These structures thus constitute an interesting context to test the integration abilities in pronoun interpretation of highly proficient bilingual speakers with different dominance profiles.

In summary, comparing Catalan-Spanish early functional bilinguals with different language dominance profiles will allow us to thoroughly investigate whether, how, and to what extent bilingualism and language dominance affect the interpretation of pronominal subjects. Moreover, the combination of canonical and non-canonical structures will constitute a rich context in which we can further evaluate the effects of bilingualism, if differences arise. Specifically, the goals of the present chapter are to examine 1) whether bilinguals with different language dominance profiles are similarly sensitive to syntactic, pragmatic, and sequential factors in anaphora resolution (i.e., to identify each groups' interpretive preferences) and 2) whether language dominance effects are observed, to assess whether these effects are attributable to bilingualism in itself or to CLI.

6.1.1 Outline of the research questions

The overarching question of the present chapter is the following: Does language dominance shape pronoun resolution by early bilinguals? If language dominance effects are attested, we will be interested in discerning whether they can be explained by effects related to bilingualism per se or by CLI. According to the stated goals, the more specific research questions that guide this chapter are outlined below.

RQ5 Does language dominance modulate pronoun resolution in canonical sentences by Catalan-Spanish bilinguals? (§6.2)

- a. Do Catalan-Spanish bilinguals show evidence of effects of bilingualism per se when interpreting pronouns in canonical sentences?
- b. Is CLI observed between Catalan and Spanish in canonical sentences? If this is the case, is it modulated by language dominance, and in which direction does it occur?

RQ6 Does language dominance modulate pronoun resolution in marked information contexts by Catalan-Spanish bilinguals? Is there evidence of a general effect related to bilingualism per se? (§6.3)

6.1.2 Organization of the present chapter

The present chapter is divided into two main sections: the first (§6.2) focuses on pronoun resolution patterns in canonical unmarked sentences and aims at answering RQ5, and the second (§6.3) addresses marked information contexts and aims at answering RQ6. In both cases, the main aim is to explore whether null and overt pronouns are interpreted differently in each context depending on the language dominance profile of the bilinguals, as well as how these effects can be explained. When analyzing marked information structures in §6.3, reference is made to the results on canonical sentences already presented in §6.2, and the overall findings for unmarked and marked contexts are discussed together in §6.4 (General discussion).

As will be explained in §6.2.2.3, a single statistical model was used to analyze all canonical and non-canonical contexts. We considered it preferable to analyze all the responses to the different levels of the conditions of the same task in one model.

6.2 How bilingualism and language dominance affect pronoun resolution in canonical sentences

In the present section, the results referring to unmarked sentences are analyzed to address the question of whether null and overt pronouns are interpreted differently depending on the language dominance profile of the bilinguals. In other words, whether the PAH can account for the interpretive patterns of the three bilingual groups in Catalan and in Spanish. On the one hand, we make between-groups comparisons to see whether anaphoric resolution patterns in Catalan and in Spanish vary depending on the dominance profile of the bilingual groups. On the other hand, we make within-group comparisons to see whether the bilinguals in each group adjust pronoun interpretation depending on the language in which they are being tested. These results are discussed in relation to approaches that predict an effect of bilingualism per se on anaphora resolution, and approaches that predict CLI between languages.

6.2.1 Aims and predictions

The present section (§6.2) aims at addressing RQ5, which concerns unmarked canonical contexts. As mentioned in the introduction (§6.1, see also §3.4), we will contrast two meaningful approaches in the framework of bilingual anaphora resolution that make different predictions regards the bilinguals' performance. The first approach relies on the presence of a general bilingualism effect (e.g., Sorace, 2011, 2016), whereas the second approach is based on the emergence of crosslinguistic influence (CLI) (e.g., Romano, 2019).

RQ5 Does language dominance modulate pronoun resolution in canonical sentences by Catalan-Spanish bilinguals?

If language dominance—used as a proxy for bilingualism—has a role in shaping pronoun interpretation by highly proficient early bilinguals, speakers with different language dominance profiles are expected to show differences in their pronoun resolution patterns, either between groups and/or between their two languages. If language dominance is found to have an effect, more specific research questions have been proposed (RQ5a and RQ5b) to assess whether this effect can be related to bilingual strategies or to CLI.

The lack of language dominance effects could be either indicative of 1) each group having similar anaphora resolution preferences, not affected by bilingualism, or of 2) the presence of bilingualism effects irrespective of the bilinguals' profile (see §3.3.2).

RQ5a. Do Catalan-Spanish bilinguals show evidence of effects of bilingualism per se when interpreting pronouns in canonical sentences?

Based on the Interface Hypothesis (IH) (Sorace, 2011; Sorace & Filiaci, 2006), the interpretation of overt pronouns is expected to be vulnerable to language dominance effects, if any. More specifically, indeterminacy—or optionality—in the interpretation of overt pronouns is expected as a general effect of the participants' bilingual condition. This overextension of overt pronouns should be observed in the weaker language of Catalan- and Spanish-dominant bilinguals⁵⁵, irrespective of whether it is Catalan or Spanish. Balanced bilinguals, lacking, in principle, a more dominant language, could feasibly show optionality in both of their languages. Regarding null pronouns, they should not be vulnerable to language dominance effects. Thus, they are expected to show a consistent bias toward subject antecedents across groups and languages.

If these predictions are borne out, differences are expected between the dominant and the non-dominant languages of Catalan- and Spanish-dominant bilinguals (within-groups). That is, Catalan-dominant bilinguals showing PAH-like interpretations in their dominant Catalan but overgeneralizing the interpretation of overt pronouns in their non-dominant Spanish, and Spanish-dominant bilinguals showing PAH-like interpretations in their dominant Spanish but overgeneralizing the interpretation of overt pronouns in their non-dominant Catalan. Consequently, differences between groups should also arise in Catalan: balanced and Spanish-dominant bilinguals should resort to the indeterminacy of overt pronouns when tested in this language, but not Catalan-dominants. The reverse should also be found in Spanish: balanced and Catalan-dominant bilinguals should resort to the indeterminacy of overt pronouns when tested in this language, but not Spanish-dominants. In contrast to overt pronouns, the interpretation of null pronouns is not expected to differ either between languages (i.e., within groups) or between groups.

⁵⁵ As we saw in the results in §5.2.3, both Catalan-dominant and Spanish-dominant bilinguals display clear PAH-like patterns in pronoun resolution in their dominant language. They do not show evidence of undeterminacy in the interpretation of overt pronouns.

RQ5b. Is CLI observed between Catalan and Spanish in canonical sentences? If this is the case, is it modulated by language dominance, and in which direction does it occur?

Contrary to the IH, which predicts that only overt pronouns would be affected, both null and overt pronouns would be affected by CLI. Although both Catalan and Spanish follow the PAH (Carminati, 2002), the two languages differ in the strength of both pronouns' interpretive biases—showing more polarized biases in Catalan than in Spanish (see §5.3.2). Under these circumstances, no indeterminacy should be attested in pronoun resolution: all groups should interpret null pronouns as corefering with subject antecedents and overt pronouns as corefering with object antecedents. However, these biases may be more or less polarized depending on the language and the bilingual in question's language dominance profile. No differences are expected between the bilinguals' two languages (i.e., within-groups), as they would be transferring the patterns of one of their languages to the other. However, differences can emerge between groups if language dominance modulates the directionality and/or the strength of CLI.

Firstly, regarding the directionality of CLI, we have identified three possible scenarios based on three mutually exclusive hypotheses:

H5b.1: CLI will only occur from Spanish toward Catalan, from the language with less categorical preferences toward the language with more categorical preferences (Romano, 2019).

H5b.2: The stronger categorical preferences of Catalan will not be affected by the more variable distributions of Spanish (de Prada Pérez, 2019; Hulk & Müller, 2000). If CLI occurs, Catalan is expected to influence the flexible preferences of Spanish (in line with Bel & García-Alcaraz, 2018), but not vice versa.

H5b.3: Language dominance will determine the directionality of CLI (Yip & Matthews, 2007). Catalan-dominant bilinguals will transfer Catalan patterns into Spanish, and Spanish-dominant bilinguals will transfer Spanish patterns into Catalan. In this case, between-group differences should be attested. Pronoun resolution preferences in Spanish by Catalan-dominant bilinguals should be more categorical than those of at least Spanish-dominant bilinguals. Pronoun resolution preferences in Catalan by Spanish-dominant bilinguals should be more flexible than those of at

least Catalan-dominant bilinguals. Balanced bilinguals may show preferences in-between the other two groups.

Secondly, language dominance may modulate the occurrence and the strength of CLI (see §3.2.3.2) even if CLI is found to occur unidirectionally (H5b.1 and H5b.2). Although Romano (2019) and de Prada Pérez (2019) do not address a possible role of language dominance, it is possible that CLI only emerges from the dominant language toward the non-dominant language. If this is the case, between-group differences could be expected in the influenced language. Spanish-dominant bilinguals should show a stronger Spanish influence than Catalan-dominant bilinguals when resolving anaphora resolution in Catalan (if CLI occurs from Spanish toward Catalan). Catalan-dominant bilinguals should show a stronger Catalan influence than Spanish-dominant bilinguals when resolving anaphora resolution in Spanish (if CLI occurs from Catalan toward Spanish). As previously mentioned, between-group differences may be observed in the two languages if CLI occurs bidirectionally (H5b.3) (Yip & Matthews, 2007).

6.2.2 Method

6.2.2.1 Participants

For the present section, we take into consideration all the bilingual participants that took part in the study: Catalan-dominant bilinguals ($N = 34$), balanced bilinguals ($N = 31$), and Spanish-dominant bilinguals ($N = 29$). As summarized in Table 6.1, these participants have been comprehensively described in §4.2. These three groups of bilinguals differ in several aspects of the Bilingual Language Profile (BLP) and their language experience.

Overall, in relation to language history, Catalan-dominant and Spanish-dominant bilinguals have been more exposed to their dominant language. The exposure of balanced bilinguals to Catalan was higher than that of Spanish-dominant bilinguals, but lower than that of Catalan-dominant bilinguals. However, balanced bilinguals were similarly exposed to Spanish to Spanish-dominant bilinguals. Language use more visibly discriminates the three identified groups, although balanced bilinguals are less homogeneous than the dominant groups. Even though slight variations distinguished the three bilingual groups in terms of language proficiency, their self-rated skills were high in all cases and for both languages. Finally, language attitudes also discriminated between the three groups, though the differences are tiny. As argued in §4.2.4, the language use

module was the one that more clearly distinguished the three bilingual groups (see the remarks in §4.2.5 for a summary; see also Table 6.1 below).

Table 6.1

Module scores in the Bilingual Language Profile questionnaire, by group

	Catalan-dominant bilinguals (N = 34)				Balanced bilinguals (N = 31)				Spanish-dominant bilinguals (N = 29)			
	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
History	-19.3	8.3	-36.3	-0.9	-2.7	7.6	-22.3	15.0	6.6	6.5	-5.9	20.0
Use	-43.7	7.7	-54.5	-22.9	0.5	13.7	-22.9	19.6	42.2	9.5	13.1	54.5
Proficiency	-3.5	3.6	-13.6	6.8	0.6	2.2	-4.5	4.5	6.1	6.0	0.0	25.0
Attitudes	-16.4	7.0	-29.5	-2.3	-4.0	7.6	-22.7	13.6	7.4	7.3	-4.5	25.0

Note. The maximum score per module was ± 54.5 (positive values are indicative of Spanish dominance, negative values indicative of Catalan dominance, and scores close to 0 indicative of balanced bilingualism).

6.2.2.2 Materials

The experimental task completed by participants, a two-alternative forced-choice task, has been detailly described in §4.3. In the present section, we will only gather at two of the conditions in the task: null pronouns in unmarked contexts, as in (6.1), and overt pronouns in unmarked contexts, as in (6.2).

- (6.1) a. La Laura va espantar la Maria quan *pro* va entrar a l'habitació. (Catalan)
 b. Laura asustó a María cuando *pro* entró en la habitación. (Spanish)
 'Laura scared Maria when (she) went into the room.'

- (6.2) a. La Laura va espantar la Maria quan ella va entrar a l'habitació.
 b. Laura asustó a María cuando ella entró en la habitación.
 'Laura scared Maria when she went into the room.'

6.2.2.3 Reported model

We ran a logistic generalized linear mixed-effects regression to predict subject interpretations with Pronoun, Information structure (Context), Language and Group as fixed effects, with all interactions included. The model included varying intercepts for item and participant and a by-participant varying slope for the effect of Pronoun. Additional random slopes were tested, but they either did not contribute to model fit or led to estimation problems within the models. The summary of this model is provided in Appendix I (Table I.1). The model's total explanatory power was moderate (conditional $R^2 = 0.23$), and the part related to the fixed effects alone (marginal R^2) was 0.07. No multi-collinearity issues were detected (highest VIF = 1.001). This model had a C-index of concordance of 0.75.

The same model with no intercept—with the same fixed effects and random effects structure—was also computed to compare whether subject and object antecedent choices of null and overt pronouns differed from chance (see §4.5.2.2). Table I.2 in Appendix I presents the summary of this model with no intercept.

For the present section, which addresses pronoun resolution in canonical sentences, the results referring to the three marked information structures will not be needed. Hence, we could have done without the non-canonical structures and have run a model for unmarked structures only, without including Information structure as a predictor. However, we would have been separating conditions tested in the same task. To avoid that, we opted for analyzing all the structures together in a sole model. In the present section, only results referring to unmarked canonical sentences are reported and discussed. Results from this same model referring to marked information contexts are addressed later in §6.3.

6.2.3 Results

As previously mentioned, the results reported in the present section only refer to canonical sentences. The results for non-canonical structures (i.e., CLLD, subject clefts and object clefts) are addressed in §6.3. In both sections, the results will be described following the same structure. Firstly, the model's relevant main effects and interactions will be presented together with the descriptive results (§6.2.3.1). Secondly, each group's pronoun interpretation patterns in Catalan and in Spanish will be identified, and within-group comparisons will be addressed, assessing whether bilinguals differ in their

interpretations in the two languages (§6.2.3.2). Thirdly, the three groups will be compared, reporting the results of the between-group comparison in Catalan and Spanish (§6.2.3.3). Before closing the results section, a summary of findings will be included (§6.2.3.4).

6.2.3.1 Overall results

Descriptive results showing the proportion of subject interpretations for null and overt subject pronouns in unmarked canonical sentences are shown in Table 6.2. No striking differences between groups can be observed. However, in Catalan, Catalan-dominant bilinguals seem to display more polarized resolution patterns than the other two groups, especially referring to overt pronouns. In Spanish, no remarkable differences seem to be observed between groups. Differences between the two languages of the bilinguals only seem to arise in the case of Catalan-dominant bilinguals.

Table 6.2

Proportion of subject interpretations for null and overt subject pronouns in canonical sentences (unmarked information structures) by each group of bilingual speakers (SD)

	Catalan		Spanish	
	Null	Overt	Null	Overt
Catalan-dominants	.69 (.46)	.24 (.43)	.66 (.47)	.37 (.48)
Balanced	.65 (.48)	.43 (.50)	.63 (.48)	.42 (.49)
Spanish-dominants	.56 (.50)	.41 (.49)	.60 (.49)	.38 (.49)

Regarding the contribution of each predictor and their interactions to explain pronoun interpretive preferences in unmarked structures in the reported model⁵⁶, Pronoun had a significant effect ($F(1) = 99.463$, $p < .001$), but Language and Group alone did not reach significance. The three-way interaction Pronoun × Group × Language, however, was

⁵⁶ To obtain the values of these effects for unmarked structures alone, the *joint_tests* function of the *emmeans* package was used. This function generates a Type-III-ANOVA-like table of the interaction contrasts for all effects in the model and allows for obtaining separate ANOVA tables based on a certain variable (Lenth, 2018; Lenth et al., 2021). Since in the present section we are not interested in the interactions involving Information structure, we used the *joint_tests* function to separate the main effects and interactions by information structure. The values obtained through likelihood ratio test comparisons using the *anova* function are very similar to those obtained using *joint_tests* (there are no differences in the significance or non-significance of the effects and interactions).

significant ($F(2) = 3.839, p = .022$). Whereas the two-way interaction Pronoun \times Language was not significant, Pronoun \times Group reached significance ($F(2) = 6.516, p = .002$). The dominance profile of bilingual speakers could thus be influencing the interpretation of null and overt pronouns depending on the language, Catalan or Spanish. The main effects and interactions in the reported model are summarized in Appendix I: those referring only to unmarked contexts are shown in Table I.4, and those referring to all the conditions are shown in Table I.3.

For now, we are only interested in unmarked contexts. However, the pairwise contrasts reported in the present section are derived from the four-way interaction between Pronoun \times Information structure \times Language \times Group. This is because the reported model includes all the conditions for Information structure (see §6.2.2.3). The reported contrasts in the present section, in any case, only refer to canonical unmarked sentences, from the perspectives of Group, Language and Pronoun.

To describe the results obtained in the statistical model for canonical sentences, we first address between-group comparisons in Catalan and Spanish to investigate whether language dominance modulates anaphora resolution by early Catalan-Spanish bilinguals in each language (§6.2.3.2). Afterward, we refer to within-group comparisons, scrutinizing each group to identify whether bilinguals display significantly different patterns depending on the language in which they are being tested (§6.2.3.3).

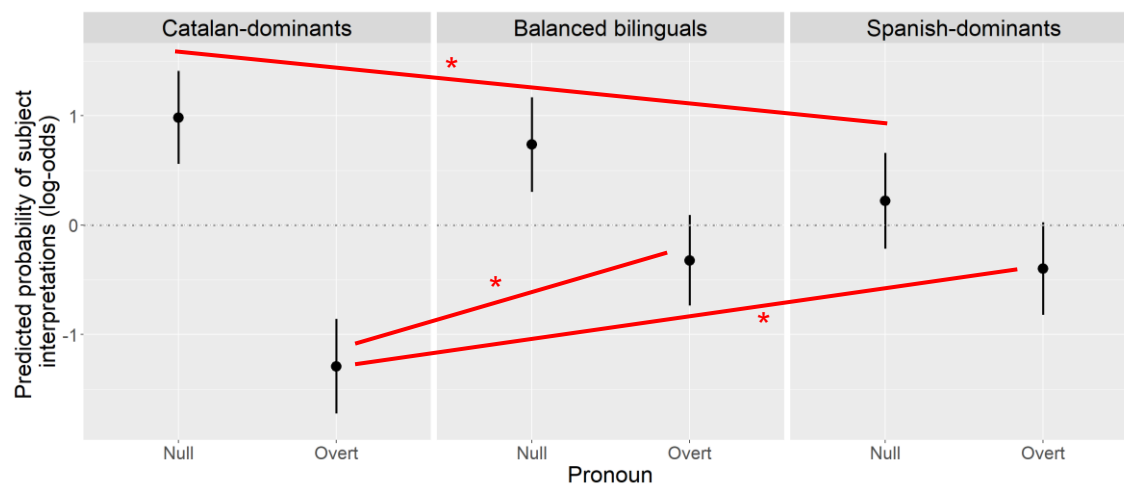
6.2.3.2 Comparing bilinguals with different language dominance profiles (between-group comparisons)

6.2.3.2.1 In Catalan

The interpretive preferences of null and overt pronouns in canonical unmarked contexts in Catalan by Catalan-dominant, balanced and Spanish-dominant bilinguals are illustrated in Figure 6.1. As can easily be observed, language dominance seems to play a relevant role in Catalan anaphora resolution. While Catalan-dominants show strongly polarized interpretive patterns, balanced and especially Spanish-dominant bilinguals show more mitigated biases.

Figure 6.1

Preferred interpretation of null and overt pronouns in canonical unmarked sentences in Catalan, by group of bilinguals ($\pm 95\%$ CI)



As has been already reported in previous §5.2.3, Catalan-dominant bilinguals show categorical biases of null pronouns toward subject antecedents ($\beta = 0.984$, $p < .001$) and of overt pronouns toward object antecedents ($\beta = -1.292$, $p < .001$). From the Pronouns' perspective, the division of labor between null and overt pronouns is shown to be very polarized ($\beta = 2.276$, $p < .001$).

Referring to the behavior balanced bilinguals in Catalan, they show a significant bias of null pronouns ($\beta = 0.736$, $p = .001$) but a non-significant bias of overt pronouns ($\beta = -0.323$, $p = .126$). From the Pronoun's perspective, they display a clear division of labor between the two pronominal forms ($\beta = 1.059$, $p < .001$).

Finally, Spanish-dominant bilinguals show not only a non-significant bias of overt pronouns toward object antecedents ($\beta = -0.400$, $p = .064$), but also a non-significant bias of null pronouns toward subject antecedents ($\beta = 0.221$, $p = .321$). Although their biases do not reach significance, they interpret null and overt pronouns differently, which show a division of labor ($\beta = 0.621$, $p = .024$).

Overall, there seems to be a gradient tendency in pronoun resolution in Catalan: the polarization in the biases of both null and overt pronouns decrease with language dominance. If we address between-group comparisons, from the Group's perspective, only Catalan-dominant bilinguals significantly differ from the other groups. As shown by the asterisks in previous Figure 6.1, Catalan-dominants show a stronger bias of null pronouns compared to Spanish-dominants ($\beta = 0.763$, $p = .018$), and a more well-defined

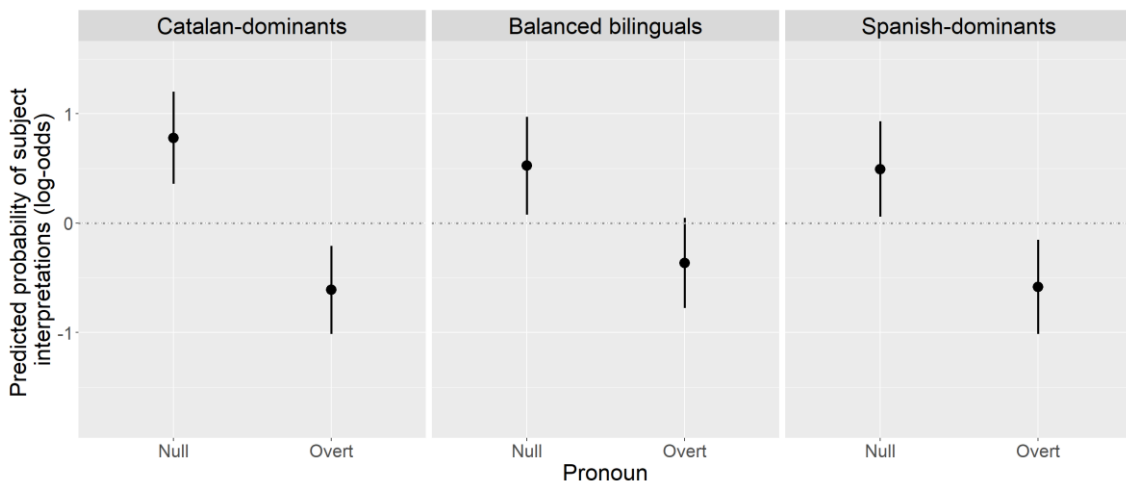
interpretation of overt pronouns with respect to Spanish-dominants ($\beta = -0.893, p = .004$) and also balanced bilinguals ($\beta = -0.969, p = .001$). Spanish-dominant and balanced bilinguals do not differ neither in the interpretation of null pronouns ($\beta = 0.515, p = .163$), nor in the interpretation of overt pronouns ($\beta = -0.077, p = .956$).

6.2.3.2.2 In Spanish

The Spanish data gives rise to a very different picture from the Catalan data. In Spanish, similar patterns are displayed by the three groups of bilinguals. Figure 6.2 shows the interpretive preferences of null and overt pronouns in canonical unmarked contexts in Spanish by Catalan-dominant, balanced and Spanish-dominant bilinguals.

Figure 6.2

Preferred interpretation of null and overt pronouns in canonical unmarked sentences in Spanish, by group of bilinguals ($\pm 95\%$ CI)



Catalan-dominant bilinguals show again a clear-cut preference of null pronouns for subject antecedents ($\beta = 0.779, p < .001$) and of overt pronouns for object antecedents ($\beta = -0.611, p < .001$) in Spanish, their non-dominant language. From the Pronouns' perspective, the division of labor between null and overt pronouns is also significant, they do not prefer subject antecedents to the same extent ($\beta = 1.390, p < .001$).

As to balanced bilinguals, they also show a clear-cut preference of null pronouns for subject antecedents ($\beta = 0.524, p = .022$), and a non-significant tendency of overt pronouns to corefer with object antecedents ($\beta = -0.365, p = .083$). The division of labor between the two pronominal forms is also clear and significant ($\beta = 0.889, p = .001$).

Finally, as previously reported in §5.2.3, Spanish-dominant bilinguals in Spanish show a well-defined bias of null pronouns to prefer subject antecedents ($\beta = 0.494, p = .027$) and of overt pronouns to prefer object antecedents ($\beta = -0.585, p = .008$). They also show a significant division of labor when comparing the subject preferences of the two pronouns ($\beta = 1.079, p < .001$).

If we address comparisons from the Group's perspective, no differences are observed between Catalan-dominant and Spanish-dominant bilinguals, neither affecting null pronouns ($\beta = 0.285, p = .563$), nor affecting overt pronouns ($\beta = -0.246, p = .615$). Moreover, even though balanced bilinguals do not show a clear-cut object bias for overt pronouns, they do not significantly differ from the other groups in the interpretation of overt pronouns (compared to Catalan-dominants: $\beta = -0.246, p = .615$, compared to Spanish-dominants: $\beta = 0.220, p = .699$).

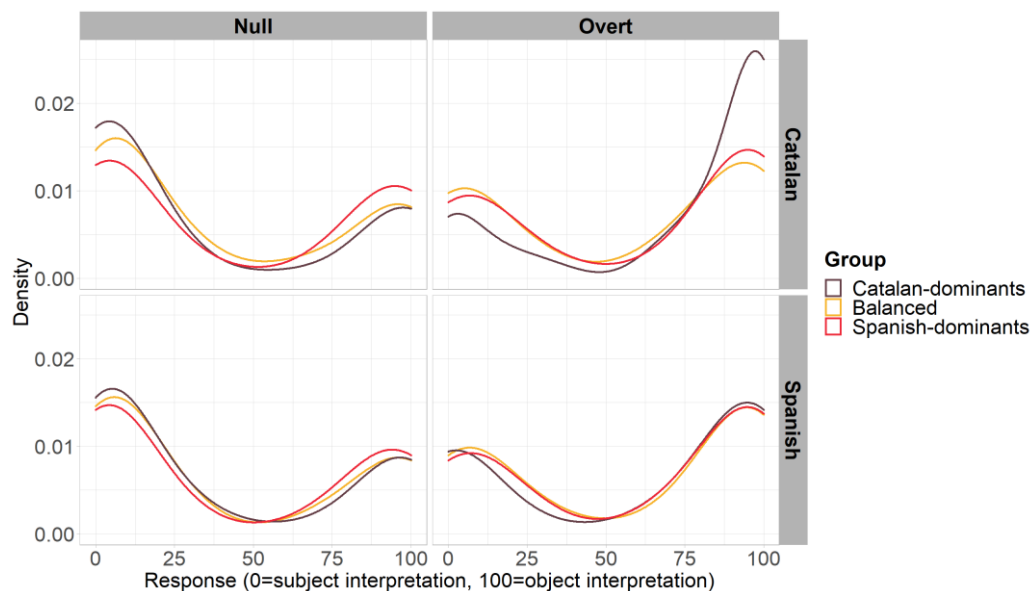
All in all, the three groups do not show reliable differences, so language dominance does not seem to be influencing anaphora resolution in Spanish.

6.2.3.2.3 Looking at the results from a complementary perspective

Before moving on, Figure 6.3 illustrates the same results from a complementary perspective. It shows the distribution of the responses on pronoun interpretation in canonical sentences, by group. Looking at these data allows us to have a complementary representation of the results. Perhaps more remarkable about this graph is that the only density curve that appears singularized is the one corresponding to the interpretation of overt pronouns in Catalan by Catalan-dominant bilinguals. In the case of null pronouns interpretation in Catalan, Catalan-dominant and Spanish-dominant bilinguals also appear distinguished. Conversely, the distribution of each Group's responses practically overlaps in Spanish. In this language, all groups show similar density curves—almost identical in the case of overt pronouns. Similar curves to Spanish are also observed in Catalan by balanced and Spanish-dominant bilinguals.

Figure 6.3

Density plot of raw responses showing the interpretive preferences of null and overt pronouns in canonical sentences in Catalan and Spanish, by group of bilinguals



Note. The variable Response corresponds to the answers (not the predicted answers) that participants gave using the slider scale (i.e., without binarizing the data, see §4.5.1.2).

6.2.3.3 Comparing the two languages of each group of bilinguals (within-group comparisons)

After presenting each group's interpretive preferences in Catalan and Spanish, we report pairwise contrasts from the Language perspective. These comparisons have not yet been addressed and allow investigation into whether bilinguals modulate their preferences depending on the language in which they are being tested, or whether they display similar patterns regardless of language. As shown in Figure 6.4 below, significant differences between languages only emerge in Catalan-dominant bilinguals.

If the performance of Catalan-dominant bilinguals in Catalan and Spanish is compared by looking at pairwise contrasts from the Language's perspective, some differences arise. The apparent decrease in the strength of the null pronouns' bias in Spanish compared to Catalan is not significant ($\beta = 0.204, p = .388$). On the other hand, the overt pronouns' bias is significantly weaker in Spanish than Catalan ($\beta = -0.681, p < .001$). Thus, Catalan-dominant bilinguals attenuate their overt-object bias when they interpret overt pronouns in their non-dominant language, Spanish.

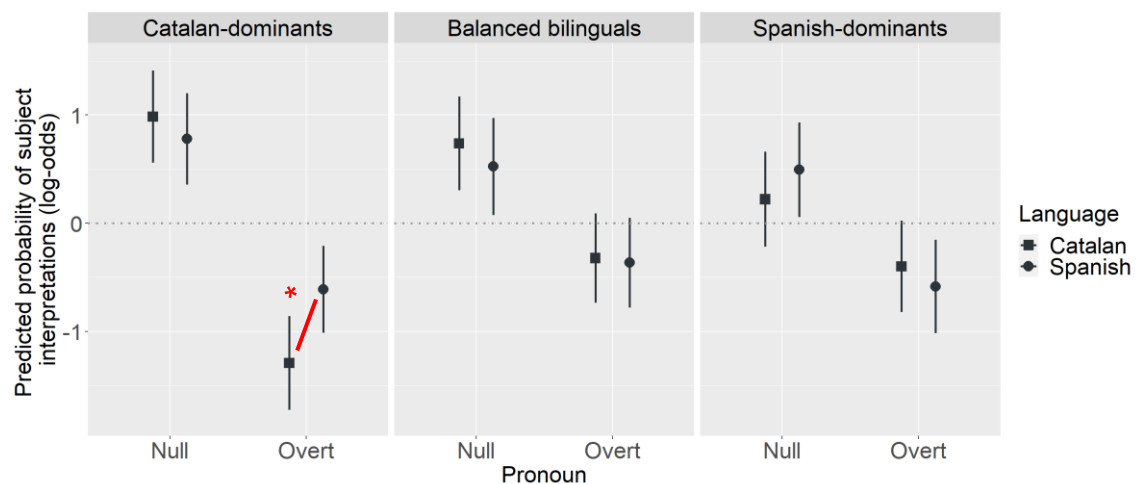
Regarding balanced bilinguals, no differences are attested between their Catalan and their Spanish when interpreting null pronouns ($\beta = 0.213$, $p = .391$) nor overt pronouns ($\beta = 0.042$, $p = .857$).

Lastly, Spanish-dominant bilinguals also reveal similar interpretations of both null and overt pronouns in their two languages (null pronouns: $\beta = -0.273$, $p = .263$, overt pronouns: $\beta = 0.186$, $p = .450$).

Thus, in the case of balanced and Spanish-dominant bilinguals, they seem to interpret pronouns in Catalan and Spanish following the same patterns. Although Spanish-dominant bilinguals seem to attenuate their preferences in their non-dominant language, Catalan, compared to Spanish, these differences are not significant.

Figure 6.4

Preferred interpretation of null and overt pronouns in canonical unmarked sentences in bilingual Catalan and in bilingual Spanish, by language and group ($\pm 95\%$ CI)



6.2.3.4 Summary of the results

By and large, the three groups generally interpret null and overt pronouns as coreferring with subject and object antecedents, respectively, both in Catalan and Spanish. However, language dominance seems to play a role in pronoun resolution preferences when bilinguals are tested in Catalan. Catalan-dominant bilinguals differ from balanced and Spanish-dominant bilinguals in showing significantly more categorical interpretations in Catalan. In contrast, the three groups of bilinguals show similar behaviors in Spanish.

It should also be highlighted that the object bias of overt pronouns by Catalan-dominant bilinguals is significantly weakened in their non-dominant Spanish compared to Catalan. However, no significant differences are obtained between the two languages of Spanish-dominant bilinguals. They display significant PAH-like biases in Spanish, whereas their interpretations of null *and* overt pronouns in their non-dominant Catalan do not reach significance. In some way, the two groups seem to be attenuating their interpretations in their non-dominant language, although this weakening is not reliably attested in all cases. Finally, balanced bilinguals show the same preferences in their two languages: a clear bias of null pronouns toward subject antecedents and a non-significant tendency of overt pronouns to prefer object antecedents.

6.2.4 Discussion

The main goal of the present section was to examine language dominance effects on pronoun resolution patterns in unmarked contexts in Catalan and in Spanish. We wanted to see whether Catalan-dominant, balanced, and Spanish-dominant bilinguals display similar interpretive biases across groups and between their two languages (RQ5). In this regard, pronoun resolution in Catalan has been found to be affected by language dominance. Differences emerge in the strength of especially overt pronouns' biases toward object antecedents, but also regarding the strength of null pronouns' biases toward subject antecedents. More specifically, in Catalan, while Catalan-dominant bilinguals show very strongly polarized PAH-like interpretations, balanced and particularly Spanish-dominant bilinguals show similar and more attenuated resolution preferences. In Spanish, conversely, similar PAH-like interpretations are observed across groups. Comparisons between the two languages of each group reveal that only Catalan-dominant bilinguals adjust their interpretations depending on the language in which they are tested, showing stronger biases of overt pronouns in Catalan than in Spanish. In discussion of these results, we first refer to general bilingualism effects (RQ5a) and then to CLI (RQ5b).

6.2.4.1 On general effects of bilingualism

The asymmetric role of language dominance in Catalan and Spanish does not seem to support the emergence of a general bilingualism effect (RQ5a). Based on the Interface Hypothesis (IH; Sorace & Filiaci, 2006), we would have expected overt pronouns to

consistently show optionality or indeterminacy 1) in Catalan by balanced and Spanish-dominant bilinguals, and 2) in Spanish by balanced and Catalan-dominant bilinguals. However, a bilingual strategy consisting of the overextension of overt pronouns, as predicted by the IH, is not consistently shown in the interpretive biases of Catalan-Spanish early bilinguals.

On the one hand, Catalan-dominant bilinguals show clear-cut and well-defined PAH-like biases in both Catalan and Spanish. Thus, they are not found to overextend the preferences of overt pronouns in their non-dominant Spanish. On the other hand, Spanish-dominant bilinguals show well-defined PAH-like biases in Spanish, but non-significant preferences for both null and overt pronouns in their non-dominant Catalan. Hence, the two pronominal forms show indeterminacy in the weaker language of Spanish-dominant bilinguals and they cannot be claimed to resort to the overextension of overt pronouns predicted by the IH, either. These findings contradict previous evidence that found no variation in the interpretation of null pronouns and indeterminacy of overt pronouns across bilingual populations, and belie the prediction that residual optionality only affects overt pronouns (Belletti et al., 2007; Chamorro et al., 2016; Filiaci et al., 2014; Sorace & Filiaci, 2006; among others). Null pronouns have also been found to be vulnerable to language dominance (i.e., bilingualism) effects: they showed a strong subject bias in Catalan by Catalan-dominant bilinguals but indeterminacy in Catalan by Spanish-dominant bilinguals. Thus, their bias toward subject antecedents may not be as steady and invulnerable as had been previously assumed (Clements & Domínguez, 2017).

Although our results do not fully support the IH, neither Catalan-dominant nor Spanish-dominant bilinguals show exactly the same interpretive preferences in their dominant and non-dominant languages. Firstly, Catalan-dominant bilinguals significantly attenuate their overt pronouns' bias in their weaker language, Spanish, with respect to Catalan. Secondly, the well-defined PAH-like interpretive biases of Spanish-dominant bilinguals in Spanish become diluted in their weaker language, Catalan, not showing significant preferences for null or overt pronouns in this language. Catalan-dominant and Spanish-dominant bilinguals could be similarly weakening their preferences as regards pronouns in their non-dominant language. It is possible that these two groups of bilinguals were relaxing the interpretive biases in their non-dominant language as a bilingual strategy to alleviate cognitive load. However, the weakening of preferences by Spanish-dominant bilinguals does not even approach significance. Although the final interpretations show different patterns in their Catalan and Spanish, the differences between their two

languages are tiny and, as a consequence, difficult to interpret accurately. The weakening of null pronouns' subject bias in Spanish by Catalan-dominant bilinguals was not significant either. Thus, doubts exist as to whether this 'weakening strategy' is solid enough.

At this point it is necessary to refer to balanced bilinguals. Based on the IH, we had predicted that they would show optionality or indeterminacy in the interpretation of overt pronouns in both of their languages. In this case, the results support our predictions and the IH: the interpretive biases of overt pronouns by balanced bilinguals are not significant, neither in Catalan nor in Spanish. However, the well-defined biases of Catalan-dominant bilinguals in their non-dominant Spanish and the indeterminacy of both null and overt pronouns of Spanish-dominant bilinguals suggest, as argued in the previous paragraph, that the IH alone cannot account for our data. Moreover, whereas the unbiased interpretation of overt pronouns in Catalan by balanced bilinguals differs from the strong object bias of Catalan-dominant bilinguals, no between-group differences are attested in Spanish (neither differences approaching significance). Given that their biases in Spanish do not differ from those of Spanish-dominant bilinguals, we would not claim to have enough evidence to state that balanced bilinguals rely on indeterminacy as a bilingual strategy. Moreover, balanced bilinguals show a tendency to link overt pronouns to object antecedents, especially in Spanish, even if it is not significant ($\beta = -0.365$, $p = .083$). On the other hand, nor can we be sure that the 'weakening strategy' explains balanced bilinguals' pronominal biases. The fact that these bilinguals show similar interpretive patterns in their two languages is in line with their balanced profile. They do not seem to have a weaker language, but they do not show target-like interpretations either.

In summary, our results do not support the emergence of a general bilingualism effect as predicted by the IH that results in the overextension of overt pronouns' interpretation. Instead, unbalanced bilinguals could be relying on a weakening of pronominal biases when interpreting anaphoric pronouns in their non-dominant language as a bilingual strategy to overcome an increased cognitive load. However, we lack strong evidence supporting this proposal. In section §6.3 we will explore whether these weakening effects in the non-dominant language of bilinguals also apply to other contexts (i.e., marked information structures). We also hope that evidence on pronoun interpretation in marked information structures will help us elucidate our findings for balanced bilinguals and either confirm or disconfirm whether they consistently overextend overt pronouns' biases in line with the IH. In the following section, we will discuss the same results from the perspective of CLI, in the interests of determining whether it better accounts for our results.

6.2.4.2 On crosslinguistic influence and its directionality

The language dominance effects attested in Catalan, as well as the lack of language dominance effects in Spanish, could be explained by CLI emerging in pronoun resolution by Catalan-Spanish bilinguals (RQ5b). Firstly, the more flexible biases of balanced and Spanish-dominant bilinguals in Catalan compared to Catalan-dominant bilinguals are compatible with the hypothesis that the less defined preferences of Spanish are influencing the more polarized preferences of Catalan. Secondly, the lack of language dominance effects in Spanish indicates that CLI from Catalan is not attested in Spanish, where all bilinguals display similar PAH-like biases (although weaker than those of Catalan by Catalan-dominant bilinguals). Thus, CLI seems to occur, unidirectionally, from Spanish toward Catalan. These findings are in line with Romano's (2019) approach, according to which CLI takes place from the more flexible language (Spanish) toward the more categorical language (Catalan).

It is important to remember that Catalan by Catalan-dominant bilinguals displays clearly polarized biases of null pronouns toward subject antecedents, and of overt pronouns toward object antecedents. In contrast to Catalan, monolingual Spanish does not show clear-cut biases in pronoun resolution, being more flexible in pronoun interpretation⁵⁷. When analyzing language dominance effects in Catalan, we have found that balanced and Spanish-dominant bilinguals show significantly more flexible biases than Catalan-dominant bilinguals, affecting both null and overt pronouns in the case of Spanish-dominant bilinguals, and only overt pronouns in the case of balanced bilinguals. It is possible that balanced and Spanish-dominant bilinguals interpret pronouns in Catalan with more flexible biases as a result of CLI from Spanish. The differences between Catalan-dominant bilinguals, on the one hand, and balanced and Spanish-dominant bilinguals, on the other hand, also demonstrate that language dominance modulates the occurrence of CLI in Catalan: it does not take place when the source of the influence is the non-dominant language of the bilinguals (i.e., in Catalan-dominant bilinguals). Catalan-dominant bilinguals display clear-cut and very polarized interpretations for both null and overt pronouns and thus seem to be preserved from this influence from Spanish.

Regarding the role of language dominance in the interpretive patterns in Spanish, the three bilingual groups behave alike. Balanced and Spanish-dominant bilinguals do not

⁵⁷ The biases of bilingual Spanish by Spanish-dominant bilinguals do not significantly differ from those of monolingual Spanish and are also weaker than those of Catalan by Catalan-dominant bilinguals, especially regarding overt pronouns (see §5.2.3 and §5.3).

significantly differ in the interpretive preferences of pronouns in their two languages. On the other hand, Catalan-dominant bilinguals seem to adjust their interpretive preferences when tested in Spanish by showing target-like preferences in this language. This may indicate that they master the referential system of their non-dominant language, Spanish, and distinguish it from that of Catalan. Alternatively, it could still be possible that Catalan-dominant bilinguals weaken overt pronouns' biases in Spanish as a strategy to overcome increased processing load, as argued in the previous section. This weakening of pronominal biases would result, in this case, in target-like interpretations.

Another possibility that we laid out as a hypothesis was that CLI could take place from the language with a more categorical distribution (Catalan) toward the language with a more variable distribution (Spanish). Our data, however, disconfirms this hypothesis. The very polarized interpretative biases of Catalan-dominant bilinguals in Catalan are not observed in their Spanish. Hence, the more categorical distribution of Catalan does not seem to lead to more well-defined pronominal resolution biases in bilingual Spanish, neither in Catalan-dominant bilinguals nor the other groups. This suggests that the Vulnerability Hypothesis, proposed by de Prada Pérez (2019) to account for subject expression in Catalan-Spanish bilinguals (see also Giannakou, 2018), cannot be expanded to pronominal subject interpretation. Contrary to what de Prada Pérez's (2019) hypothesis would have predicted in this scenario, the more variable distribution of interpretations in Spanish does not seem to make this language more permeable to CLI.

Bel and García-Alcaraz (2018) speculated that CLI from Catalan could be strengthening pronominal biases in bilingual Spanish compared to monolingual Spanish. Given that we did not find any evidence of Catalan influencing Spanish, our findings do not seem to support this possibility. In the present study, Spanish among Spanish dominant-bilinguals does not differ from monolingual Spanish, nor from Spanish by Catalan-dominant bilinguals. It is true that, in contrast to monolingual Spanish, bilingual Spanish by Spanish-dominant bilinguals showed significant PAH-like biases, and in this regard our findings resemble those of Bel and García-Alcaraz (2018). However, bilinguals' biases were not stronger than monolinguals' biases when the two varieties were compared (§5.3.3). Moreover, Catalan-dominant bilinguals do not show stronger biases than Spanish-dominant bilinguals as a result of a stronger Catalan influence. On the contrary, Catalan-dominant bilinguals significantly attenuate the bias of overt pronouns in Spanish compared to Catalan. All in all, no reliable evidence from CLI occurring from Catalan toward Spanish has been attested in our data.

The third hypothesis that we had contemplated was that CLI would be bidirectional and would take place from the dominant to the non-dominant language (e.g., Bosch & Unsworth, 2021; Foroodi-Nejad & Paradis, 2009; van Dijk et al., 2021; Yip & Matthews, 2007). Our findings also disconfirm this possibility. If CLI had been observed from Catalan-dominant bilinguals' stronger Catalan to their weaker Spanish, they should have shown similarly polarized interpretations in their two languages. Instead, they significantly attenuated overt pronouns' biases in Spanish with respect to Catalan, not differing from Spanish-dominant bilinguals in Spanish. Language dominance, at least under the social conditions of our bilinguals, cannot be said to determine the directionality of CLI, but it does seem to determine its activation. Similarly to what has been found in other linguistic domains such as word processing or phonology coactivation in L2 acquisition studies, CLI seems to be stronger when it comes from the L1 and dominant language of a bilingual (see Lago et al., 2021 and references within).

Overall, in line with Romano (2019), CLI in anaphora resolution appears to be unidirectional: from Spanish to Catalan, and not vice versa. CLI results in Catalan showing more flexible biases by balanced and Spanish-dominant bilinguals and more categorical biases by Catalan-dominant bilinguals. Pronoun resolution in Spanish, on the other hand, does not seem to be affected in any way by CLI from Catalan, so that it remains unaltered by language dominance effects. However, CLI from Spanish is not activated if Spanish is the non-dominant language of the bilinguals. This indicates that CLI is modulated by language dominance: it is only activated with Spanish-dominant and balanced bilinguals and does not emerge among Catalan-dominant bilinguals.

CLI from Spanish to Catalan could also be predicted from Tsimpli et al. (2004). In their study, they found L1 attrition effects in the overextension of overt pronouns by Greek and Italian speakers living in an English-speaking environment. They explained these effects as determined by parametric differences between the L1 and the L2. The more "economical" language (i.e., English) was interpreted to be influencing the least "economical" language (i.e., Greek/Italian). Null pronouns were not affected, whereas overt pronouns were vulnerable to attrition effects. Sorace and Filiaci (2006) extended these predictions to L2 acquisition. In their study, as well as in Sorace (2011), they link the overextension of overt pronouns to a default strategy related to a less efficient processing in situations of processing overload. However, they also claim that these processing strategies may be acting in conjunction with CLI, and they do not exclude the possibility of CLI between null subject languages in the case that microvariation is attested (Sorace & Filiaci, 2006, p.

345). However, by assuming that null subject languages do not differ in pronoun interpretation, they leaned toward attributing their results on bilingual anaphora resolution to processing accounts.

Finally, we would like to draw our attention to the role of language exposure in pronoun interpretation. Anaphora resolution is a linguistic phenomenon that does not belong to the core syntax, but to the interface between syntax and pragmatics, and as such it is particularly sensitive to input (e.g., Sorace et al., 2009; Sorace & Serratrice, 2009; see also Unsworth et al., 2014). Furthermore, Arnold et al. (2018) demonstrated that pronoun comprehension strategies are indeed affected by language exposure—understood as reading experience measured through an Author Recognition Task (i.e., print exposure). This crucial role of language experience is of special interest with respect to the fact that balanced and Spanish-dominant bilinguals differ from Catalan-dominant bilinguals in Catalan, but not among themselves. According to their responses to the BLP (see §4.2.4), balanced and Spanish-dominant bilinguals showed similar scores regarding language history in Spanish, higher than Catalan-dominant bilinguals. Following these observations, we could speculate that our findings on language dominance could be particularly related to language history across one's lifespan, rather than to language use or other variables. This possibility will be further explored in a follow-up analysis in which we analyze language dominance as a continuum—using the numerical BLP global scores instead of categorical groups. This (re)analysis will allow us to 1) assess whether similar language dominance effects on pronoun resolution in canonical sentences in Catalan and in Spanish can be predicted using language dominance as a continuum, and 2) to separately evaluate the impact of language history and language use scores.

6.2.5 Follow-up analysis: Language dominance as a continuous variable

6.2.5.1 Motivation of this follow-up analysis

To analyze the effects of bilingualism on anaphora resolution in Catalan-Spanish bilinguals, we classified participants into three groups based on their language dominance scores in the BLP questionnaire (see §4.2). However, having a continuous measure of language dominance from the BLP makes it worth it to further explore the role of language dominance as a continuous variable. It is well known that perfectly balanced bilinguals do not exist, and several authors have claimed that bilingualism should be conceived essentially as a gradient reality (Silva-Corvalán & Treffers-Daller,

2016). Our group of balanced bilinguals showed a lot of variation, especially regarding language use. Regarding language history, they could not be clearly distinguished from Spanish-dominant bilinguals in Spanish (the two groups showed overlapping scores).

In the present analysis, using language dominance as a continuous variable may provide a complementary perspective to the results previously obtained when comparing groups. Our findings so far suggest that there is a gradation in the polarization of null and overt pronoun interpretive preferences in Catalan, going from more categorical biases by Catalan-dominant bilinguals, to more moderate biases by balanced bilinguals, and very weak biases by Spanish-dominant bilinguals (see Figure 6.1 in §6.2.3.2). In addition to using the BLP global dominance scores, the module-specific scores for language history and language use will also be tested in different models to try to see whether its components history or use can better predict language dominance effects.

6.2.5.2 Method

We ran a logistic generalized linear mixed-effects regression with the same participants and with the same structure described in previous §6.2.2.3, changing only the categorical variable Group for the continuous variable BLP⁵⁸ and taking into account only unmarked structure conditions⁵⁹. Therefore, the fixed effects included to predict subject interpretations were Pronoun, Language and BLP score, and their interactions. The model included varying intercepts for item and participant and a by-participant varying slope for the effect of Pronoun. The model's total explanatory power (conditional R^2) was 0.433, and the part related to the fixed effects alone (marginal R^2) was 0.109. It had a C-index of concordance of 0.864. No multi-collinearity issues were detected (highest VIF = 1.001). The summary of this model is provided in Appendix J (Table J.1).

Afterward, instead of the BLP global score, the BLP history score and the BLP use score were used as predictors in two subsequent models with the same structure as the

⁵⁸ The *scale* function in R was used to standardize the BLP scores.

⁵⁹ The full model, including also Information structure as a fixed effect was also run and lead to very similar results, showing no qualitative differences compared to the model analyzing only canonical sentences. Therefore, for simplification purposes, we will report the results of the model that took into account only unmarked canonical sentences. For the more complex model including Information structure, the total explanatory power (conditional R^2) was of 0.227, and the part related to the fixed effects alone (marginal R^2) was of 0.063. No multi-collinearity issues were detected. This model had a C-index of concordance of .751.

reported model in the previous paragraph. First, we fit a model including Pronoun, Language and BLP history score as fixed effects, their interactions, and varying intercepts for item and participant and a by-participant varying slope for the effect of Pronoun. This model's conditional R^2 was 0.433, the marginal R^2 was 0.111, and the C-index of concordance was 0.864. Second, we fit a model including Pronoun, Language and BLP use score as fixed effects, their interactions, and varying intercepts for item and participant and a by-participant varying slope for the effect of Pronoun. This model had a conditional R^2 of 0.434, a marginal R^2 of 0.108, and a C-index of concordance of 0.864. No multi-collinearity issues were detected in any of the models (highest VIF = 1.0001). The summary of these models is provided in Appendix J (Table J.2 for the one using the BLP history scores, and Table J.3 for the one using the BLP use scores).

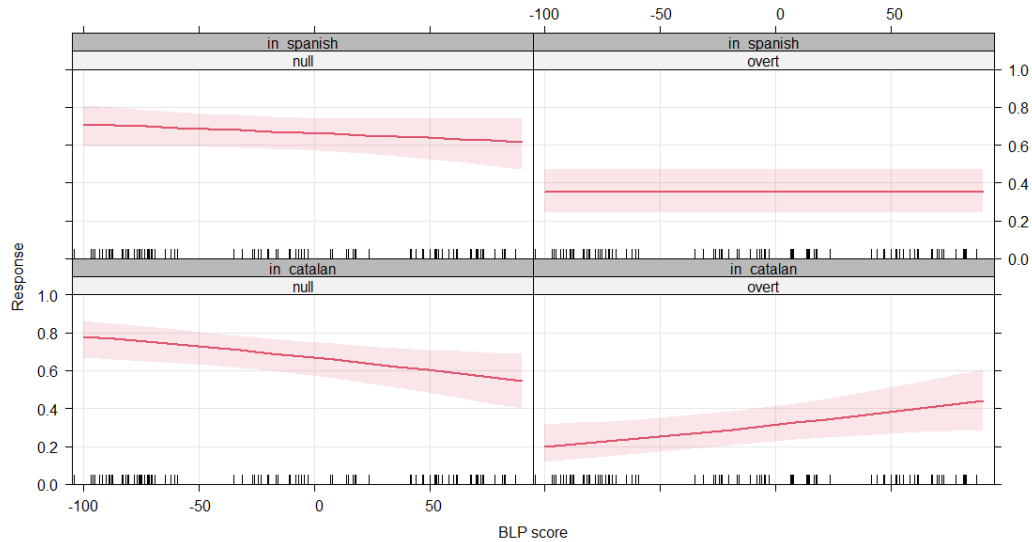
6.2.5.3 Results

The analysis results using the BLP global score as a continuous variable confirm that language dominance effects emerge only in Catalan and affect null pronouns ($\beta = -0.348$, $p = .013$) and overt pronouns ($\beta = 0.373$, $p = .019$). As shown in Figure 6.5 below, the more Catalan-dominant the bilinguals are, the more categorical interpretations they show, clearly associating null pronouns to subject antecedents and overt pronouns to object antecedents. At the other end of the continuum, bilinguals with more dominance in Spanish show more weakened biases, i.e. more flexible interpretations. In Spanish, on the other hand, as also shown in Figure 6.5, language dominance is not found to have an impact on the interpretation of neither null ($\beta = -0.137$, $p = .316$) nor overt pronouns ($\beta = 0.035$, $p = .822$).

Regarding the models using the BLP history module score and the BLP use module score, they lead to very similar results, showing significant effects on both pronouns in Catalan and no effects in Spanish. Looking at the two figures, the scores on language history (Figure 6.6) seem to show a more pronounced effect on pronoun resolution in Catalan than the scores on language use (Figure 6.7). However, the predictions of the two models turn out to be fairly similar when comparing the values of the intercepts of the model assessing language history (null pronouns: $\beta = -0.366$, $p = .009$; overt pronouns: $\beta = -0.402$, $p = .013$) and the model assessing language use (null pronouns: $\beta = -0.326$, $p = .020$; overt pronouns: $\beta = -0.349$, $p = .029$).

Figure 6.5

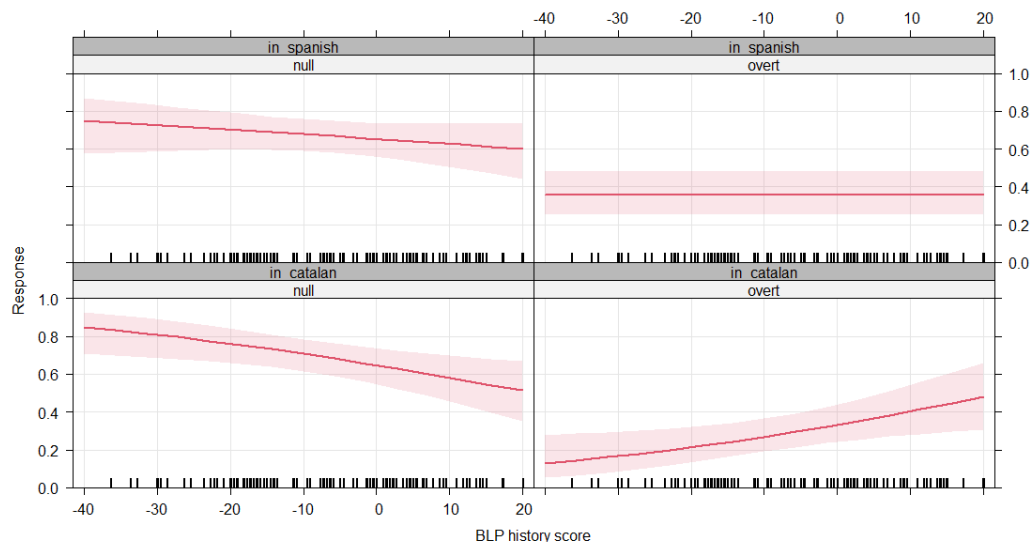
Effects of language dominance (BLP global score) on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals



Note. In the y-axis, the interpretive choice of the participants is represented (1 = subject interpretations, 0 = object interpretations). In the x-axis, the BLP global dominance scores are represented.

Figure 6.6

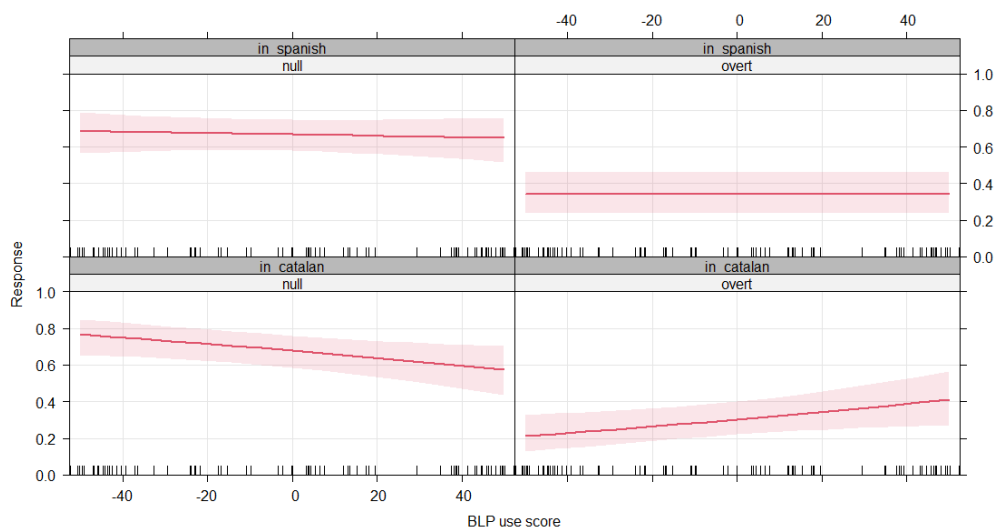
Effects of language history BLP scores on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals



Note. In the y-axis, the interpretive choice of the participants is represented (1 = subject interpretations, 0 = object interpretations). In the x-axis, the BLP scores on the language history module are represented.

Figure 6.7

Effects of language use BLP scores on subject interpretations for null and overt pronouns by Catalan-Spanish bilinguals



Note. In the y-axis, the interpretive choice of the participants is represented (1 = subject interpretations, 0 = object interpretations). In the x-axis, the BLP scores on the language use module are represented.

6.2.5.4 Discussion

These follow-up analyses lend strong support that language dominance predicts differences in pronoun interpretation in Catalan. In Spanish, no evidence of language dominance effects was attested. More specifically, differences within Catalan-Spanish bilinguals related to language dominance lie in how consistently they associate null pronouns with subject antecedents and overt pronouns with object antecedents in Catalan. The more Catalan-dominant the bilinguals are, the more categorical preferences they exhibit. In this way, the trends identified when dividing bilinguals into groups are corroborated by this complementary analysis. On the one hand, they confirm that both null and overt pronouns are affected by CLI. On the other hand, it can be clearly observed that the influence of Spanish into Catalan is modulated by language dominance: it is more intense in speakers that have Spanish as a more dominant language. The less Spanish-dominant bilinguals are, the less strong this influence is, not emerging in the most Catalan-dominant bilinguals. Catalan-dominant bilinguals show very polarized resolution preferences, which are not influenced by Spanish.

As also attested in the group analysis, language dominance has no statistically significant impact on pronoun interpretation in Spanish. As already discussed, this finding could be interpreted as indicating that Spanish shows homogeneous pronoun resolution preferences, not influenced by the more categorical patterns of Catalan (Romano, 2019). Catalan-dominant bilinguals show target-like interpretations in Spanish, their non-dominant language.

Although we speculated that language history could be a better determiner of pronoun resolution than language use, we have found that both components of language dominance make similar predictions. Predictions based on language history are slightly stronger than those of language use, but this small difference does not allow us to make any claims in this regard. As suggested by Arnold et al. (2018), exposure—measured as print exposure and understood in broader terms, as including both language "history" and use—could have a specially relevant role in providing the necessary input to learn about frequent referential patterns and probabilities. This is indeed especially decisive in the case of overt pronouns, which are used/produced to a lesser extent (see corpus studies such as Bel & García-Alcaraz, 2019; Lozano, 2016). Building on Arnold et al. (2018), the effect of language exposure would not be restricted to reading and exposure to books, which is one of the sources of language exposure and in fact was not assessed in the language history module of the BLP questionnaire. However, the global BLP scores and language use scores were also found to predict the same effects. All in all, language history measured in the BLP is not revealed to be more crucial than other factors in predicting the resolution patterns used in pronoun interpretation.

6.2.6 Summary of main findings

Overall, CLI seems to be attested in our results on bilingual pronoun resolution in canonical sentences. Regarding its directionality, CLI appears to take place from the language with more flexible biases (Spanish) toward the language with more categorical interpretations (Catalan), as proposed in Romano (2019; see also Sorace, 2011; Tsimpli et al., 2004). Given that crosslinguistic differences between Catalan and Spanish seem to affect both null and overt pronouns, the two pronominal forms are sensitive to CLI effects in Catalan. The very polarized interpretive biases of Catalan seem to be vulnerable to the influence of less defined interpretive biases of Spanish, resulting in weaker biases of the Catalan by balanced and Spanish-dominant bilinguals compared to that of Catalan-dominant bilinguals. The fact that Catalan by Catalan-dominant

bilinguals displays robust and polarized pronominal biases indicates that the Spanish system does not systematically permeate Catalan, but that CLI activation is modulated by language dominance. Catalan-dominant bilinguals, who have been less exposed to Spanish and use this language to a lesser extent, are preserved from Spanish influence. Moreover, language dominance also seems to modulate the strength of CLI: the more Spanish-dominant bilinguals are, the more flexible interpretations they show in Catalan. Spanish, on the other hand, does not show influence from Catalan, and its steady pronoun resolution preferences across groups do not seem to be affected by bilingualism or language dominance effects.

However, these findings are not incompatible with Catalan-Spanish bilinguals showing effects of bilingualism itself (or by a combination of both CLI and a bilingual strategy). General effects of bilingualism per se would also be modulated by language dominance. It is possible that balanced bilinguals rely on an overextension of overt pronouns, as predicted by the IH (Sorace, 2011, 2012). Notwithstanding, the IH does not seem to account for the results of unbalanced bilinguals. Catalan-dominant bilinguals show well-defined biases of both pronouns in their non-dominant Spanish, whereas Spanish-dominant bilinguals show indeterminacy in both pronoun biases in their non-dominant Catalan. According to our findings, if unbalanced bilinguals relied on a general bilingual strategy when interpreting pronouns in their non-dominant language, this strategy would not consist of the overextension of overt pronouns but in the weakening of the biases of both null and overt pronouns (not necessarily resulting in indeterminacy). Inconclusive evidence has however been found regarding the effects of bilingualism per se, which will need to be further discussed in light of the results on marked information structures.

In the following section, we analyze pronoun resolution patterns in marked information structures, using non-canonical sentences (i.e., CLLDs, subject clefts and object clefts). The corresponding results will shed some light on the findings attested to in canonical sentences, which remain unclear. If bilinguals rely on a general bilingual strategy (in the non-dominant language in the case of unbalanced bilinguals), differences based on language dominance should also emerge in marked information structures, and in a similar manner to unmarked structures. Conversely, if bilingual anaphora resolution can be better explained by CLI effects, no differences should emerge in these contexts because Catalan and Spanish have been demonstrated as not showing microvariation in their resolution biases in marked information structures.

6.3 How bilingualism and language dominance affect pronoun resolution in marked information structures

After investigating language dominance effects in canonical structures, this section analyzes sentences with marked information structures (CLLD, subject clefts, and object clefts) to see whether and how language dominance effects emerge in these more complex contexts. In fact, these sentences constitute an especially interesting context in which the effects of bilingualism per se can be evaluated. Anaphora resolution has been proved to be challenging for bilinguals in previous studies, a fact that has mainly been attributed to this linguistic phenomenon being at the syntax-pragmatics interface, in the light of the Interface Hypothesis (IH; Sorace, 2011, 2012; Sorace & Filiaci, 2006). In the case of marked information structures, which involve the manipulation of the syntactic position and the information status of the antecedent, the predictions of the IH should be more evident.

In the previous section (§6.2), no strong evidence was found as regards early Catalan-Spanish bilinguals systematically overextending the interpretation of overt pronouns in canonical sentences. Contrary to the predictions of the IH, they do not seem to rely on this cognitive strategy in order to deal with the overloaded processing demands attributed to bilingual speakers. However, the interpretation of both null and overt pronouns was found to be vulnerable to language dominance effects. Broadly speaking, we interpreted the results in §6.2 as compatible with both crosslinguistic influence (CLI) occurring from Spanish toward Catalan and/or with a cognitive strategy to alleviate processing costs resulting in the weakening of pronominal biases in the non-dominant language. Given that microvariation between Catalan and Spanish arises in canonical structures, but not in marked information structures (see §5.3.2), CLI is not expected in marked contexts. Any effect of language dominance in non-canonical structures should only be related to bilingualism per se.

6.3.1 Aims and predictions

The present section (§6.3) addresses language dominance effects on pronoun resolution in marked information contexts. As stated above, analyzing these more complex sentences will shed light on whether bilinguals experience any processing demands that make them resort to a bilingual strategy that gives rise to indeterminacy in pronoun

interpretation, particularly in the interpretation of overt pronouns. Interpreting pronouns in marked information contexts would be expected to be even more challenging for bilinguals. Syntactic and pragmatic factors—as well as sequential factors—may present conflicting cues in marked structures regarding the interpretation of null or overt pronominal forms (see Chapter 5). Despite the fact that Catalan and Spanish present microvariation regarding pronoun resolution in baseline sentences, no crosslinguistic differences have been identified in marked information structures. If differences between groups are attested in non-canonical sentences, they will not be attributable to CLI but to the bilingual condition of the participants.

In this context, we address the following research question and two competing hypotheses.

RQ6. Does language dominance modulate pronoun resolution in marked information contexts by Catalan-Spanish bilinguals? Is there evidence of a general effect related to bilingualism *per se*?

H6.1: If bilinguals are less efficient in the integration of information from distinct linguistic domains, as required in anaphora resolution, language dominance effects are expected. Different patterns from 'target-like' interpretations (taking the Catalan from Catalan-dominants and the Spanish from Spanish-dominants as 'target-like') are expected in the less dominant language of Catalan- and Spanish-dominant bilinguals, as well as in balanced bilinguals⁶⁰. To be more precise, we expect to attest differences between Catalan and Spanish within Catalan-dominant bilinguals, and within Spanish-dominant bilinguals. Likewise, we expect to find differences between groups: Catalan-dominant bilinguals differing from balanced and Spanish-dominant bilinguals in Catalan, and Spanish-dominant bilinguals differing from balanced and Catalan-dominant bilinguals in Spanish. Regarding the nature of these differences, we will consider two possibilities:

H6.1a: If a general bilingualism effect results on the overextension of overt pronouns as predicted by the IH (Sorace, 2011), differences according to language dominance should mainly affect overt pronouns. This overextension strategy should be observed in the non-dominant language of unbalanced bilinguals, as well as in balanced bilinguals. Across contexts, and irrespective of the information

⁶⁰ As we have already seen in Chapter 5, Catalan-dominant and Spanish-dominant bilinguals do not present evidence of such inefficiencies in their dominant languages, and they do not differ from monolingual speakers in the effects that manipulating the information status of a specific antecedent has on the resolution of null and overt pronouns.

status of the antecedents, overt pronouns should show undetermined interpretations. Subject clefts are the structures that will allow us to test this hypothesis. In these structures, overt pronouns were shown to display a very clear bias toward object antecedents (see §5.2). According to the IH, bilingualism effects should make overt pronouns display a non-target-like undefined bias in these subject cleft contexts. On the other hand, null pronouns should remain unaffected and maintain their subject bias in object clefts. CLLD structures do not constitute an ideal context in which to observe bilingualism effects, given that both null and overt pronouns have shown unbiased interpretations in these contexts (in bilingual Catalan and Spanish, and also in monolingual Spanish; see Chapter 5).

H6.1b: On the other hand, building on our findings on canonical structures (see §6.2), it could also be possible that null and overt pronouns show weak and indeterminate preferences in the non-dominant language of unbalanced bilinguals, as well as in the two languages of balanced bilinguals. In this case, both pronouns would be subject to this weakening effect: the subject bias of null pronouns in object clefts should be attenuated, as well as the object bias of overt pronouns in subject clefts. Once more, no effects are predicted regarding unbiased interpretations in CLLD structures.

H6.2: As already mentioned, another possible outcome is contemplated: in terms of CLI, no differences between languages or groups are predicted. Since the two languages under study do not seem to diverge with respect to the distribution of the interpretation of pronouns in marked structures, no language dominance effects would be expected if bilingualism effects are driven by CLI. In CLLD, both null and overt pronouns should remain unbiased. In subject clefts, overt pronouns should maintain their clear bias toward object antecedents. In object clefts, null pronouns should show a clear bias toward subject antecedents. The same interpretations are expected in Catalan and in Spanish.

6.3.2 Method

6.3.2.1 Participants

The bilingual participants that will be considered in the following analysis are the same as those reported in §6.2.2.1 (see also §4.2): Catalan-dominant bilinguals (N = 34), balanced bilinguals (N = 31), and Spanish-dominant bilinguals (N = 29).

6.3.2.2 Materials

The experimental task completed by participants was described in §4.3. The conditions involving unmarked contexts have already been analyzed in previous §6.2, and in the present section, we will focus on interpreting null and overt pronouns in marked information contexts. We analyze the effects of bilingualism on null and overt pronoun resolution in non-canonical sentences. In this context, the plausible antecedents for the pronouns appear in a clitic-left dislocation structure, as in (6.3), in a subject cleft, as in (6.4), or in an object cleft, as in (6.5). In these contexts, syntactic factors (i.e., subject vs. object antecedents) interact with discourse factors (i.e., topic vs. focus antecedents) and sequential factors (i.e., first vs. second mentioned antecedents) in the interpretation of each pronominal form (see Chapter 5).

(6.3) *Clitic-left dislocation sentences*

a. A la Maria la va espantar la Laura quan *pro/ella* va entrar a l'habitació.

b. A María la asustó Laura cuando *pro/ella* entró en la habitación.

'Maria, Laura scared her when she went into the room.'

(6.4) *Subject clefted sentences*

a. Va ser la Laura qui va espantar la Maria quan *pro/ella* va entrar a l'habitació.

b. Fue Laura quien asustó a María cuando *pro/ella* entró en la habitación.

'It was Laura who scared Maria when she went into the room.'

(6.5) *Object clefted sentences*

a. Va ser a la Maria a qui va espantar la Laura quan *pro/ella* va entrar a l'habitació.

b. Fue a María a quien asustó Laura cuando *pro/ella* entró en la habitación.

'It was Maria whom Laura scared when she went into the room.'

6.3.2.3 Reported model

The results reported in the present section have been obtained from the model reported in §6.2.2.3. Along the previous section (§6.2), only results referring to canonical unmarked sentences were reported. In the present section (§6.3), we report and discuss the results of the same model that refer to syntactically-informationally marked conditions.

6.3.3 Results

6.3.3.1 Overall results

Results referring to syntactically marked information structures appear in Table 6.3. Unmarked sentences have also been included in this table as a baseline condition (the results are the same as those of Table 6.2 in §6.2.3.1). No remarkable differences between groups or languages seem to emerge from these results: topicalizing object antecedents, focusing subject antecedents, and focusing object antecedents seem to have a similar impact on pronoun resolution by the three groups and in both languages.

Table 6.3

Proportion of subject interpretations for null and overt subject pronouns in unmarked and syntactically marked information structures, by group of bilinguals (SD)

	Catalan		Spanish	
	Null	Overt	Null	Overt
Unmarked				
Catalan-dominants	.69 (.46)	.24 (.43)	.66 (.47)	.37 (.48)
Balanced	.65 (.48)	.43 (.50)	.63 (.48)	.42 (.49)
Spanish-dominants	.56 (.50)	.41 (.49)	.60 (.49)	.38 (.49)
CLLD				
Catalan-dominants	.52 (.50)	.53 (.50)	.46 (.50)	.46 (.50)
Balanced	.48 (.50)	.50 (.50)	.49 (.50)	.41 (.49)
Spanish-dominants	.54 (.50)	.50 (.50)	.50 (.50)	.43 (.50)
Subject clefts				
Catalan-dominants	.54 (.50)	.33 (.47)	.52 (.50)	.30 (.46)
Balanced	.48 (.50)	.40 (.49)	.47 (.50)	.34 (.47)
Spanish-dominants	.48 (.50)	.32 (.47)	.47 (.50)	.29 (.46)
Object clefts				
Catalan-dominants	.64 (.48)	.57 (.50)	.64 (.48)	.54 (.50)
Balanced	.68 (.47)	.60 (.49)	.64 (.48)	.51 (.50)
Spanish-dominants	.58 (.49)	.59 (.49)	.63 (.48)	.54 (.50)

Before reporting the results of the pairwise comparisons of inferential statistical analyses, we refer to the main effects and interactions of the factors included in the model, obtained through likelihood ratio test comparisons, and summarized in Appendix I (Table I.3). Pronoun and Information structure have very strong effects in improving a model with no fixed effects ($\chi^2(1) = 39.822, p < .001$; and $\chi^2(3) = 167.700, p < .001$, respectively), and their two-way interaction is also significant when compared to a model with no interactions

($\chi^2(3) = 77.458, p < .001$). As we saw in the previous chapter (Chapter 5), null and overt pronouns are differently interpreted in unmarked contexts, as well as in certain marked information structures. Indeed, manipulating the information status of an antecedent had a different impact on pronoun interpretation depending on the pronominal form. Language alone also had an effect, although weaker ($\chi^2(1) = 5.166, p = .023$), and the contribution of Group was not significant.

Interestingly, the three-way interaction Pronoun \times Information structure \times Group is significant ($\chi^2(6) = 18.666, p = .004$), whereas Pronoun \times Information structure \times Language is not. It, therefore, seems that the interaction between Pronoun \times Information structure may be affected by Group, rather than Language. It is true that, in the previous Chapter 5, we saw that the same effects of information structure emerged in Catalan and Spanish by the reference groups, so we do not expect to find many differences between the two languages in this respect. However, the effects of information structure on the interpretation of null and overt pronouns may slightly diverge across groups, depending on the dominance profile of the bilinguals. In the present section, we will see whether Group and Language affect pronoun resolution in marked information structures by Catalan-Spanish bilinguals. All the reported pairwise contrasts in the following paragraphs will unfold from the non-significant four-way interaction Pronoun \times Information structure \times Group \times Language (see Appendix I, Tables I.5 to I.8).

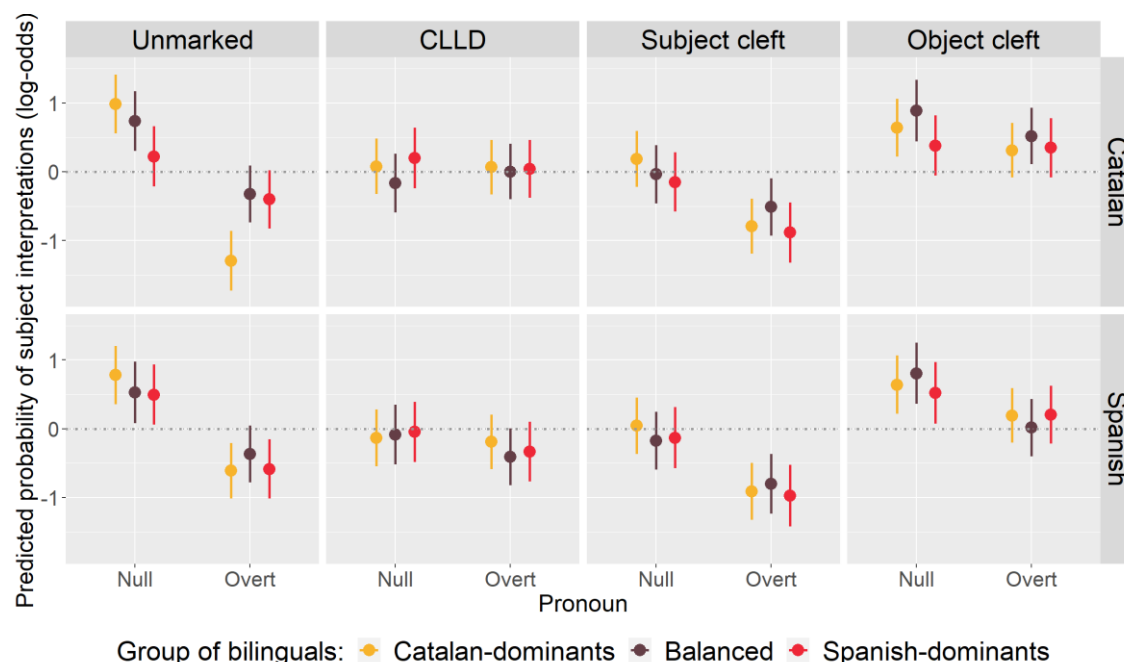
6.3.3.2 The effects of marked information structures per group of bilinguals and language

As thoroughly explained in Chapter 5, based on Catalan by Catalan-dominant bilinguals and Spanish by Spanish-dominant bilinguals, multiple factors guide pronoun interpretation (i.e., syntactic function, information status, or word order). In CLLD, where the object appears in a preverbal position, and the subject has a non-topical status, both null and overt pronouns are interpreted as unbiased. In subject clefts, where the subject is focused, null pronouns are interpreted as unbiased and overt pronouns are biased toward postverbal objects. Finally, in object clefts, where the object appears in preverbal position, overt pronouns are interpreted as unbiased and null pronouns as coreferring with topical subjects. Generally speaking, as can be appreciated in Figure 6.8, Catalan-dominant, balanced, and Spanish-dominant bilinguals display these same coreference patterns in Catalan and Spanish⁶¹.

⁶¹ Given that the three groups seem to show very similar interpretive patterns and in line with the patterns discussed in Chapter 5 and outlined in the previous paragraph, we will only report here

Figure 6.8

Preferred interpretation of subject pronouns in Catalan and in Spanish, by pronoun, information structure, language, and group of bilinguals ($\pm 95\%$ CI)



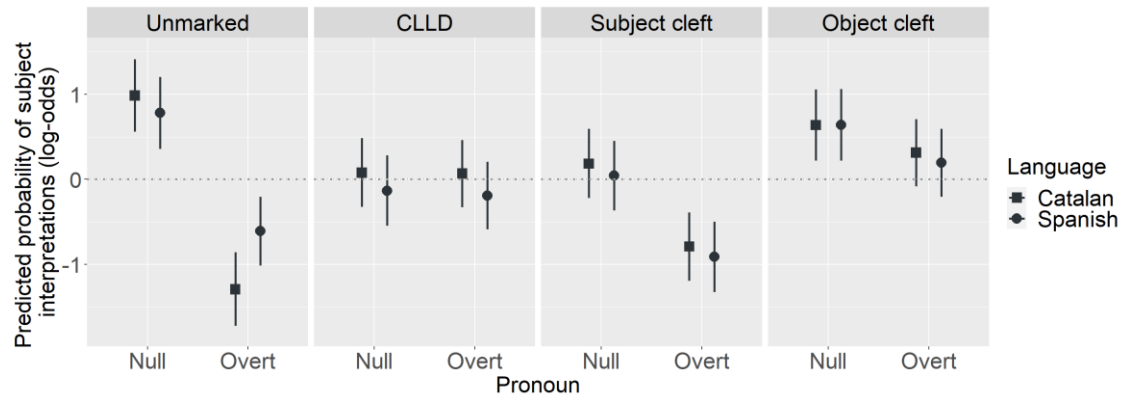
Pairwise comparisons from the Group's perspective confirm that no significant (nor approaching significance) between-group differences were attested (see Table I.5 in Appendix I). Thus, all bilinguals seem to behave similarly in front of the manipulation of syntactic and discourse factors. As illustrated in Figure 6.8, different behaviors did emerge in unmarked canonical contexts, showing language dominance to influence pronoun interpretation (as addressed in §6.2). Conversely, pronoun resolution in marked information structures was reasonably similar across groups. No language dominance effects seem to emerge when interpreting pronouns in these more complex contexts.

Focusing on Catalan-dominant bilinguals, they interpret null and overt pronouns in marked information structures in their non-dominant language, Spanish, as target-like. As illustrated in Figure 6.9, their interpretive preferences in non-canonical conditions are analogous in the two languages. No significant differences were attested when comparing Catalan and Spanish (see Table I.6 in Appendix I).

the interpretations that slightly diverge from these preferences. We will not report all the significance values of the biases of null and overt pronouns in each information context by each group (see Appendix I, Table I.2), nor all the significant pairwise contrasts derived from the statistical model (see Appendix I, Tables I.5 to I.8).

Figure 6.9

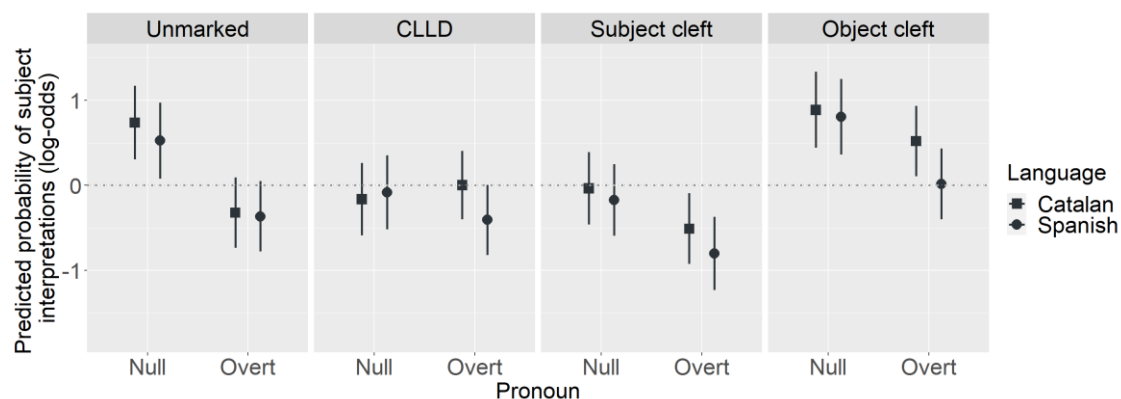
Catalan-dominant bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language, and information structure ($\pm 95\%$ CI)



Balanced bilinguals' interpretation of null and overt pronouns in marked information structures, illustrated in Figure 6.10, mainly display the interpretative patterns described at the beginning of the section. However, two slightly different behaviors from these target-like biases need to be mentioned. In CLLD structures, the preference of overt pronouns for object antecedents approaches significance in Spanish ($\beta = -0.407$, $p = .054$). In object clefts, their interpretation of overt pronouns in Catalan is not unbiased. Instead, they significantly interpret overt pronouns as coreferring with subject antecedents ($\beta = 0.519$, $p = .013$), clearly rejecting coreference with focused objects. In Spanish, however, they interpret overt pronouns in object clefts as unbiased ($\beta = 0.016$, $p = .941$), and they differ in their interpretation of overt pronouns in object clefts in Catalan compared to Spanish ($\beta = 0.519$, $p = .013$).

Figure 6.10

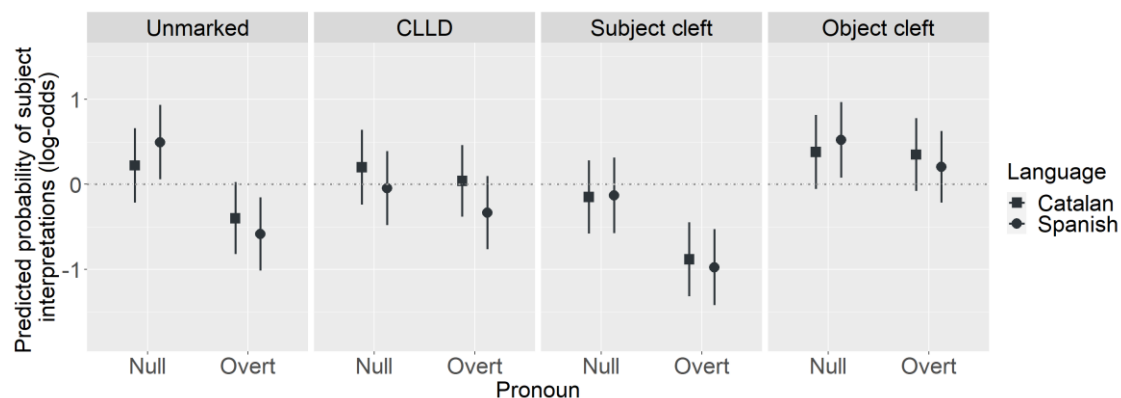
Balanced bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language, and information structure ($\pm 95\%$ CI)



Spanish-dominant bilinguals also show a similar picture to that of Catalan-dominant bilinguals, as shown in Figure 6.11. The interpretation of pronouns in marked information structures follows the previously described patterns in Catalan and Spanish. The only slight difference we can perceive is that, instead of showing a clear subject bias of null pronouns in object clefts, this bias does not reach significance ($\beta = 0.380, p = .087$). This lack of bias is not surprising considering that, in contrast to the other groups, Spanish-dominant bilinguals do not show a clear null-subject bias in canonical sentences either. Thus, object clefts are not modifying the baseline preference of null pronouns ($\beta = 0.371, p = .414$). Like balanced bilinguals, overt pronouns in CLLD are slightly biased toward object antecedents in Spanish, but this bias does not reach significance ($\beta = -0.333, p = .129$). No significant differences emerge when comparing their resolution patterns in Catalan and in Spanish, in any condition (see Table I.6 in Appendix I).

Figure 6.11

Spanish-dominant bilinguals' preferred interpretation of null and overt pronouns in Catalan and in Spanish, by pronoun, language and information structure ($\pm 95\%$ CI)



6.3.3.3 Summary of the results

Overall, no evidence has been found suggesting that language dominance in Catalan-Spanish bilinguals affects the integration of syntactic and discourse cues in pronoun interpretation in marked information structures. Even in these contexts, when a null or an overt pronoun is expected to show a bias toward a specific antecedent, bilinguals do not rely on an overextension strategy resulting in optionality or indeterminacy in their choices. Very few variations in the interpretive patterns of each group of bilinguals is attested and, importantly, no differences emerge when comparing them. Therefore, as illustrated in Figure 6.8, language dominance only seems to affect the pronouns' baseline

preferences in unmarked canonical structures. No relevant differences emerge either when comparing the bilinguals' two languages.

6.3.4 Discussion

The main finding of the present section refers to the absence of language dominance effects on the interpretation of both null and overt pronominal subjects in marked information contexts in Catalan and in Spanish (RQ6). This demonstrates that early Catalan-Spanish bilinguals do not seem to systematically resort to a default strategy in pronoun interpretation. The three groups of bilinguals show very similar resolution patterns in both Catalan and Spanish, so that Catalan- and Spanish-dominant bilinguals show similar interpretations in their dominant and non-dominant languages.

Contrary to the predictions of the Interface Hypothesis (IH; Sorace, 2011, 2012; Sorace & Filiaci, 2006), no evidence of an overextension of overt pronouns is observed in our data on informatively marked structures. Bilinguals consistently interpret overt pronouns, and also null pronouns, with their target-like biases (those described in Chapter 5). Importantly, in subject clefts, where overt pronouns preferably corefer with object antecedents, this overt-subject bias is maintained across all bilingual groups and in both languages. Furthermore, the hypothesized weakening of the interpretation of null and overt pronouns, which was compatible with our findings in canonical structures (see §6.2), cannot explain the results on marked information structures either. In object clefts, where null pronouns corefer with subject antecedents, all bilinguals maintain this null-subject bias, unaltered in both languages.

Therefore, as an answer to RQ6, both H6.1a and H6.1b should be dismissed: no language dominance effects attributable to bilingualism per se are attested in non-canonical contexts. On the other hand, our findings are compatible with H6.2, according to which no differences across bilinguals are expected given the lack of crosslinguistic differences between Catalan and Spanish. All bilinguals show a similar sensitivity to the manipulation of information status, and they successfully integrate cues from different domains in both of their languages and for both null and overt pronouns. Despite the increased processing demands associated with bilingualism, they do not show evidence of relying on a general bilingual strategy. No difficulties are reflected in our data related to a less efficient integration of information from the syntactic and pragmatic interface that can be explained by bilingualism itself. These findings point to a crucial role of the

language combination involved in bilingual anaphora resolution (see also Kraš, 2016; Quesada, 2021; Teixeira et al., 2022). Non-target-like bilingual interpretations as a result of persistent difficulties, or the overextension of overt pronouns, may only be exhibited in contexts where the two bilinguals' languages differ. The fact that two typologically similar null subject languages show microvariation should not be minimized.

Referring to bilingualism once more, offline tasks are perhaps not the best method with which to investigate the efficiency of processing resources, given that they involve metalinguistic knowledge. However, they can certainly still provide information on the speakers' processing abilities, as argued by Sorace (2011, p. 20). In fact, differences between monolingual and bilingual pronoun resolution have been widely attested in offline tasks, many of them identifying residual optionality in the interpretation of overt pronouns (e.g., Bel & García-Alcaraz, 2015; Belletti et al., 2007; Giannakou, 2018; Gürel, 2004; Kaltsa et al., 2015; Margaza & Bel, 2006; Sorace & Filiaci, 2006; Tsimpli et al., 2004). Similarly, Slabakova et al. (2017) concur with Sorace in assuming that measures of interpretation can also be indicative of processing costs. In their study, they found lower proficiency L2 speakers to show difficulties in pronoun interpretation with stimuli that implicated a computational cost. Therefore, we can claim that target-like final pronoun biases of early Catalan-Spanish bilinguals do not show evidence of a less efficient processing related to bilingualism.

While we can conclude that bilingualism does not affect bilinguals' final interpretations of null and overt pronouns in non-canonical contexts, we cannot, however, conclude that they do not experience any particular processing difficulties. It is important to keep in mind that the analyzed data was collected using an offline task that was designed to evaluate pronoun interpretation preferences, not real-time processing. Experimental studies combining online and offline methods such as Chamorro et al. (2016) or Patterson and Felser (2020) have shown differences in the results from the perspective of online unconscious processing and offline decisions that involve explicit reasoning. In the present study, we lack information on how Catalan-Spanish bilinguals process null and overt pronouns in real-time, so we cannot be sure that they do not experience any processing difficulties. Bel and García-Alcaraz (2018) showed no processing difficulties in early Catalan-Spanish bilinguals compared to Spanish monolinguals in a self-paced reading task on canonical sentences. However, they did not control for language dominance, and did not analyze more complex marked information structures. Interestingly, in the present study we registered the time that each participant took to

give an answer to each stimulus of the task. In the next section, we will explore this data in a follow-up analysis in order to find out whether any between-group differences are found in this regard. It is possible that balanced bilinguals and unbalanced bilinguals in their weaker language were slower than the reference groups (Catalan-dominant bilinguals in Catalan and Spanish-dominant bilinguals in Spanish) in the conditions requiring more cognitive load (e.g., overt pronouns vs. null pronouns, and marked information structures vs. canonical sentences). Further research could investigate whether early bilinguals show difficulties in real-time processing of anaphoric dependencies across contexts and bidirectionally in their two languages.

In fact, it is still possible that, in presence of conflicting cues, bilinguals show inconsistent patterns of pronoun resolution, but they do not differ from monolinguals in this regard. When syntactic, discourse, and sequential cues are not aligned, both null and overt pronouns are most of the time interpreted as unbiased, showing optionality or indeterminacy. This is not a pattern that only emerges across bilinguals, regardless of their language dominance profile. The same patterns were also attested for Spanish monolinguals (see §5.3), so they cannot be attributed to a specific bilingual behavior. Therefore, this indeterminacy should not be interpreted as a consequence of bilingualism per se, i.e., of an increased competition for processing resources or a less efficient integration of information. It should be rather attributed either to linguistic factors (see §5.2), or to extra-linguistic processing difficulties derived from a cognitively demanding process that does not only affect bilingual systems but also monolingual systems.

From the perspective of information structure, a slight difference in the interpretation of overt pronouns in object clefts by balanced bilinguals was indeed attested in the results. This contrast is almost negligible and does not provide information on bilingualism effects. However, it can be interpreted as providing nuances in the linguistic factors that affect pronoun interpretation in marked information structures (see §5.2). Balanced bilinguals interpret overt pronouns in object clefts in Catalan as significantly coreferring with subject antecedents (e.g., *Va ser a la Maria a qui va espantar la Laura, quan ella, va entrar a l'habitació*; 'It was Maria whom Laura scared when she went into the room'), whereas the other two groups display a non-significant bias of overt pronouns in object clefts. This subject-biased interpretation of overt pronouns in object clefts should be considered with caution because it is not observed in the Spanish of balanced bilinguals or in any other group. However, this difference may point toward a very strong anti-focus effect on overt pronouns in object cleft constructions, which only reverses the overt

pronouns' biases for balanced bilinguals, but that is similarly observed in the other groups as a non-significant tendency. Overall, the effects of focusing the object via it-cleft on overt pronouns could have a stronger impact than those of dislocating the object via CLLD, as has already been suggested.

Finally, it is worth mentioning that we analyzed the impact of language dominance as a continuum on pronoun interpretation in marked information structures, similar to the analyses run for canonical structures in §6.2.5. The results of these analyses showed a non-significant impact of language dominance on null and overt pronouns in all non-canonical contexts, in Catalan and in Spanish (see Table J.4 in Appendix J). These results confirm that language dominance only affects the interpretation of null and overt pronouns in Catalan and in the presence of crosslinguistic differences between languages (i.e., in canonical sentences, as reported in §6.2). Despite being more syntactically complex and also requiring the integration of information from different linguistic domains, bilingualism and language dominance do not alter pronoun resolution patterns in marked information structures.

In conclusion, the lack of language dominance effects in our data indicates that early Catalan-Spanish bilinguals do not show evidence of encountering difficulties in the integration of information from different linguistic domains in pronoun interpretation. Anaphora resolution biases were not found to be altered as a result of bilingualism itself. Our results, rather, are compatible with the possibility that bilinguals show target-like pronoun interpretations in contexts where their two languages do not differ crosslinguistically, such as in marked information structures. The differences that were attested in canonical sentences (see §6.2) could indeed also be predicted by CLI. We will come back to the implications of these findings in the general discussion (§6.4). Prior to this, another follow-up analysis was conducted to evaluate whether any differences attributable to language dominance effects can be observed in the response times that had also been collected in the task and used in the data cleaning process.

6.3.5 Follow-up analysis: Response times

6.3.5.1 Motivation

Before closing the present section, we want to analyze another type of data that we have available: response times. Using the Qualtrics Survey platform allowed for measuring the time each participant took to choose an antecedent for the ambiguous pronoun in

each sentence. Measuring response time was first of all used to clean the data. All the responses that had been given in a too short time to have carefully read each sentence were excluded from the analyses, as reported in §4.5.1.1. However, we think it is also worth exploring these response times to get further information on the behavior of each group of bilinguals. In case any bilingual group experiences an exceptionally high cognitive load when interpreting the experimental sentences in any of their languages, we would expect them to show slower response times compared to the other groups (the Catalan-dominant group is taken as the reference group for Catalan, and the Spanish-dominant group as the reference group for Spanish).

6.3.5.2 Method

We analyzed response times by fitting a linear mixed-effects model with log-transformed Response time⁶² as the dependent variable. Prior to data analysis, outliers were removed in a by-item and by-participant basis. Similar to previous models, it included Pronoun, Information structure, Group, and Language as fixed effects and their interactions. The model included varying intercepts for item and participant and a by-participant varying slope for the effect of Pronoun. The model's total explanatory power (conditional R^2) was 0.265, and the part related to the fixed effects alone (marginal R^2) was 0.030. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The summary of the model's output, as well as pairwise contrasts, can be found in Appendix K.

6.3.5.3 Results and discussion

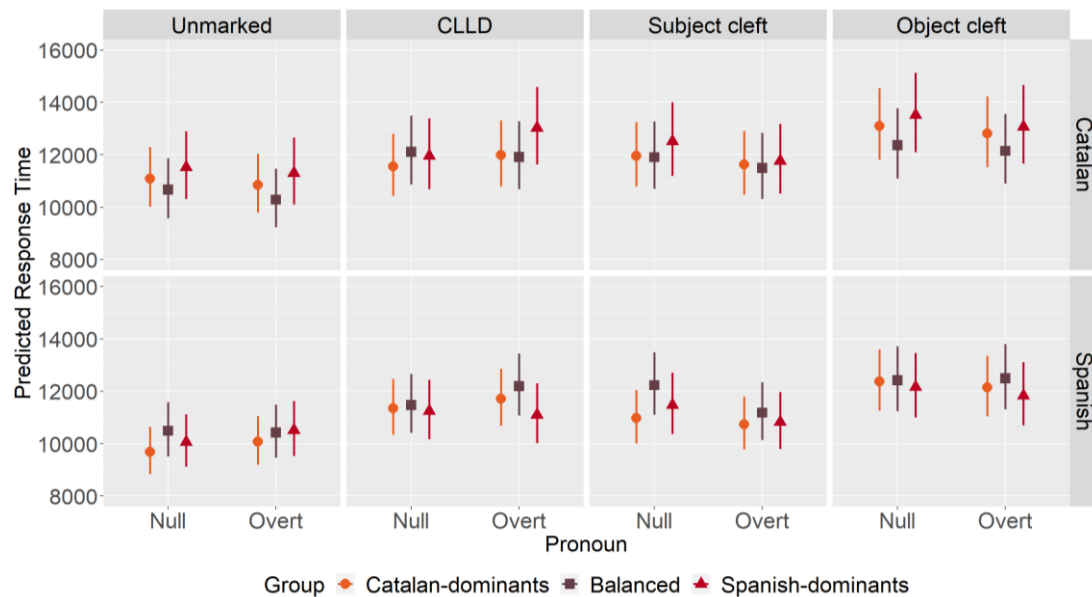
As for the results, illustrated in Figure 6.12, no significant differences are revealed when comparing the three groups of participants in either language. It takes participants a similar amount of time to read each sentence and choose an interpretation for its ambiguous subject pronoun. No evidence suggests that Catalan-Spanish bilinguals' processing is more or less efficient depending on their language dominance profile (see Sorace, 2011). The present task was not designed to measure real-time processing, so we cannot provide conclusive evidence regarding varying processing costs in bilingual

⁶² More specifically, it refers to the time (in milliseconds) that participants took from the moment when the experimental item appeared on screen until they submitted an answer to pass to the next item in the task.

processing of pronominal anaphora. In case they exist, they do not seem to affect the final interpretation of the pronouns or the time it takes them to read and interpret sentences with null and overt ambiguous pronominal subjects.

Figure 6.12

Predicted response times in the interpretation of null and overt pronouns in Catalan and in Spanish, by group of bilinguals



If we further explore the results of this statistical model and look at pairwise contrasts from the Pronoun's perspective, no significant differences emerged in the interpretation between null and overt pronouns (all $p > .05$). Our measure was probably not fine-grained enough to capture these possible contrasts. We also explored a model with Antecedent as a factor to compare null-subject vs. null-object interpretations, or overt-object vs. overt-subject interpretations (i.e., PAH-like vs. non-PAH-like interpretations for each pronoun). No meaningful differences were attested, so we decided not to include Antecedent in the model because it unnecessarily complexified the interpretation of the results.

We found some significant differences in the pairwise contrasts from the perspective of Information structure. Overall, the most relevant finding is that object cleft conditions were responded to slower than baseline conditions, as can be observed in Figure 6.9. This was the case for both types of pronouns by the three groups of bilinguals and in the two languages (all $p < .01$). In Catalan, balanced bilinguals took more time to answer all non-canonical conditions compared to the canonical condition both with null pronouns

(compared to CLLD: $\beta = -0.140$, $p = .004$; subject cleft: $\beta = -0.112$, $p = .026$; object cleft: $\beta = -0.138$, $p = .004$) and with overt pronouns (compared to CLLD: $\beta = -0.144$, $p = .002$; subject cleft: $\beta = -0.114$, $p = .024$; object cleft: $\beta = -0.165$, $p < .001$). Regarding Catalan- and Spanish-dominant bilinguals in Catalan, in the null pronoun condition they only took a longer time to answer object clefts (compared to baseline, Catalan-dominants: $\beta = -0.141$, $p = .002$; Spanish-dominants: $\beta = -0.160$, $p = .001$), not subject clefts neither CLLDs. In the overt pronoun condition, slower response times for object clefts and CLLDs compared to unmarked structures were attested for both Catalan-dominant bilinguals (object clefts: $\beta = -0.143$, $p = .002$; CLLD: $\beta = -0.093$, $p = .080$) and for Spanish-dominant bilinguals (object clefts: $\beta = -0.143$, $p = .002$; CLLD: $\beta = -0.093$, $p = .080$). In Spanish, on the other hand, all groups were generally faster in providing an answer for the baseline canonical condition as compared to all non-canonical structures in null pronoun conditions. In overt pronoun conditions, Spanish-dominant bilinguals were only slower in object clefts, and balanced and Catalan-dominant bilinguals were slower in object clefts and CLLDs, but not in subject clefts (see Table K.3 in Appendix K).

Not surprisingly, the most challenging conditions to answer were object clefts. They are consistently answered slower than canonical sentences across languages and groups, both with null and overt pronouns. CLLDs also appear to be challenging, especially combined with the overt pronoun condition. On the other hand, subject clefts are rarely answered slower than canonical sentences. These results are interesting because they indicate that OVS word order takes longer response times (in CLLD and object clefts).

It is relevant to highlight that no between-group differences were attested, in any condition. Regarding within-group comparisons, no differences were attested either; Catalan- and Spanish-dominant bilinguals were not slower in their non-dominant language compared to their dominant language. All groups were faster when resolving anaphora in Spanish, something for which we do not have an explanation.

Remember, in any case, that the purpose of measuring response times was not to analyze online (processing) data but to clean our dataset. It is interesting to explore these data, but they should be read cautiously as this type of analysis was not a priority when the experiments were designed.

6.3.6 Summary of main findings

Overall, our results on bilingual pronoun resolution in non-canonical sentences show no effects of language dominance. Early Catalan-Spanish bilinguals with different language dominance profiles have been found to show target-like pronominal biases across various contexts (CLLDs, subject clefts, and object clefts) and in both languages, Catalan and Spanish. Therefore, a general bilingualism effect resulting in the overextension of overt pronouns, as predicted by the IH (Sorace, 2011, 2012), does not explain the lack of language dominance effects and the efficient integration of information from syntax and pragmatics domains in bilingual anaphora resolution. However, we also highlighted that our results do not exclude the possibility that bilinguals experience difficulties during online processing. In any case, if such difficulties exist, they do not affect bilinguals' final interpretations of null and overt pronouns in non-canonical contexts, even though these contexts should be more cognitively taxing and thus expected to show evidence of bilingual strategies if they exist.

An account based on CLI, on the other hand, would explain this lack of bilingualism and language dominance effects in pronoun resolution across constructions. In the absence of differences between the two languages under study, CLI does not occur, and null and overt pronouns are interpreted target-like across groups and in both languages. In these contexts, where Catalan and Spanish show parallel patterns, bilingualism does not seem to have an effect. These findings point to a crucial role of the language combination involved in bilingual anaphora resolution (see also Teixeira et al., 2022). Non-target-like bilingual interpretations may only be exhibited in contexts in which the two bilinguals' languages differ, and the fact that two typologically similar null subject languages show microvariation should not be minimized.

Interestingly, we also ran an exploratory analysis of the participants' response times for each condition. Object cleft and CLLD structures, the two contexts altering the canonical SVO word order, generally showed slower response times compared to the baseline condition (especially object clefts, in both null and overt pronouns conditions, and also CLLDs, more often in the overt pronoun condition). These results confirm that marked information structures, particularly those displaying OVS word orders, were more challenging to interpret than canonical sentences. Despite this increased complexity, no reliable differences between groups or languages were attested, in line with the lack of bilingualism and language dominance effects in the final offline interpretations of the pronouns.

6.4 General discussion

This chapter analyzed the impact of language dominance as a proxy for bilingualism on pronoun resolution in canonical sentences and marked information structures in Catalan and in Spanish by early Catalan-Spanish bilinguals, in a bidirectional design. More specifically, we aimed at exploring whether evidence was found suggesting the emergence of a) general bilingualism effects, or b) CLI. Taking into account the interpretive preferences of bilinguals in canonical and marked information structures, our findings are more compatible with the emergence of CLI than with general effects of bilingualism *per se*, as will be further developed in the following paragraphs.

We have provided evidence of language dominance effects on the interpretation of both null and overt pronouns, but only in unmarked canonical sentences and mainly affecting Catalan. In unmarked contexts in Catalan the interpretive preferences of Catalan-dominant bilinguals were significantly more well-defined than those of balanced and Spanish-dominant bilinguals, especially regarding overt pronouns' bias toward object antecedents, but also affecting the null-subject bias in the case of Spanish-dominant bilinguals. In Spanish, the three groups interpreted pronouns in a similar way—in line with the PAH but showing weaker biases than those of Catalan by Catalan-dominant bilinguals. In marked information structures (i.e., topicalized objects via CLLD, focused subjects via *it*-clefts, and focused objects via *it*-clefts), all bilinguals showed similar target-like performance, and in both languages—in line with the effects reported in Chapter 5. We have attributed to CLI the presence and absence of language dominance effects (in unmarked and marked constructions, respectively). The invulnerability of the interpretation of pronouns to bilingualism effects in marked information structures, and not in unmarked contexts, points to interpretive preferences being affected by CLI in the presence of differences between languages⁶³.

More specifically, in canonical sentences, the more flexible biases of Spanish seem to be influencing the more categorical biases of Catalan, in line with Romano (2019; see also Sorace, 2011; Tsimpli et al., 2004). In the absence of between-language differences between Catalan and Spanish (as in non-canonical contexts), no language dominance or bilingualism effects emerge across groups or across languages. Still in relation to CLI,

⁶³ Remember that, in canonical contexts, Catalan and Spanish were found to differ in the extent to which they adhere to the PAH (see §5.2.3 and §5.3.2), whereas in non-canonical contexts the two languages show similar interpretive biases (see §5.2.4, §5.2.5, and §5.3.2).

another finding should be highlighted. Language dominance seems to modulate its occurrence and strength. Catalan-dominant bilinguals display more categorical interpretations—not affected by the weaker biases of Spanish—and balanced and Spanish-dominant bilinguals display more mitigated biases, influenced by the more flexible biases of Spanish. As observed in the analysis taking language dominance as a continuous variable, the intensity of influence from Spanish increases gradually with a higher dominance of this language (resulting in more flexible, even indeterminate interpretations). CLI from Spanish is not activated if it is the non-dominant language of the bilinguals (i.e., in Catalan-dominant bilinguals), and the more dominance bilinguals show in Spanish, the more evident the influence of Spanish becomes (see §6.2).

As far as general effects of bilingualism per se are concerned, early Catalan-Spanish bilinguals do not seem to rely on a general bilingual default strategy that results in the overextension of overt pronouns. Bilinguals have not been found to show optionality in the interpretation of overt pronouns as a strategy to overcome the high cognitive load required by interface phenomena such as anaphora resolution, as predicted by the IH (Sorace, 2011, 2012; Sorace & Filiaci, 2006). If this were the case, we would have expected optionality or indeterminacy in overt pronouns' interpretation at least in the non-dominant language of Catalan- and Spanish-dominant bilinguals. However, Catalan-dominant bilinguals showed well-defined PAH-like biases in Spanish, and Spanish-dominant bilinguals showed optionality in the interpretation of not only overt pronouns, but also in the interpretation of null pronouns in Catalan. This finding showed that both null and overt pronouns were vulnerable to bilingualism effects (Clements & Domínguez, 2017). Also, if bilinguals had relied on a default strategy, the overextension of overt pronouns would have been expected in marked information structures as well, even more-so than in canonical sentences. Contrary to this prediction, in non-canonical sentences both null and overt pronouns were found to show target-like interpretations across groups and in both bilinguals' languages. That is, all bilinguals showed a similar and target-like sensitivity to the interaction between syntactic, pragmatic, and sequential cues, in both languages (see §6.3).

After studying anaphora resolution in adult early bilinguals that have been raised in an officially bilingual society and are highly proficient in both languages (even having different bilingual profiles), we have seen that the predictions of the IH cannot be extended to bilingual (simultaneous or sequential) first language acquisition (see Sorace, 2011), at least for speakers of two Romance null subject languages showing

microvariation in the structures under scrutiny. Most studies on early bilingual acquisition providing evidence of instability in the interpretation of overt pronouns mainly analyzed children populations (Argyri & Sorace, 2007; Iraola, 2014; Sorace et al., 2009). However, pronoun resolution by adult early bilinguals—either simultaneous bilinguals (2L1) or child second language acquirers (cL2)—has been largely unexplored. In line with our findings, evidence of an overextension of overt pronouns was not attested in the scarce studies that have targeted these adult populations (Bel & García-Alcaraz, 2015, 2018). We know that the acquisition of subject pronouns' referential properties is late (Bel, 2001; Bel et al., 2010; Bel & Albert, 2016; Shin & Cairns, 2012) and, in light of the aforementioned studies on bilingual children, it seems to be more challenging for bilinguals than it is for monolinguals (e.g., Sorace et al., 2009). In this regard, it would be interesting to assess pronoun interpretation in Catalan-Spanish bilingual children, in a bidirectional design, to see whether they also show instability in the interpretation of overt pronouns in their weaker language until late stages of acquisition. Adult early bilinguals do not seem to rely on a default strategy in pronoun resolution, contrary to the IH and the proposal referred to as *processing resources account* in Sorace (2011).

We have argued that our results do not totally support the *processing resources account*, but they cannot be claimed to totally support the *representational account* either (Sorace, 2011). Firstly, no between-group and between-language differences in the pronouns' interpretive preferences in marked information structures have been found, but between-group and between-language differences have been attested in unmarked contexts. In opposition to the *processing resources account* (Sorace, 2011), these results suggest that a bilingualism effect per se does not impact pronoun resolution by early bilinguals. Secondly, despite the emergence of CLI in contexts where the two language systems of the bilinguals differ, the *representational account* (based on Tsimpli et al., 2004) would not explain why the interpretation of both null and overt pronouns differs in contexts of microvariation, between language pairings that do not diverge in their parametric choices (see Giannakou 2018 for different findings). From a representational perspective, our proposal that CLI accounts for bilingualism effects in anaphora resolution is not problem-free. As argued in Sorace (2011; see also Sorace & Serratrice, 2009), anaphoric dependencies of subject pronouns are not part of narrow syntax, or core grammar, but they involve contextual information external to grammar. Thus, variability in anaphora resolution should not only be related to grammatical representations, but to factors related to the interaction between syntactic and pragmatic information of this interface

phenomenon. As we have argued, 1) the two languages do not differ from each other and, more importantly, 2) anaphoric dependencies are beyond (grammatical) representations.

A different explanation to account for CLI in bilingual pronoun resolution involving language combinations of two null subject languages could refer to parsing strategies, rather than grammatical representations. Sorace and Serratrice (2009, p. 198) refer to less efficient processing and to the quantity of input as possible explanations for an overgeneralization of overt pronominal forms by bilinguals, but they also consider the possibility that CLI occurs in representations and/or in parsing strategies. We therefore suggest that CLI in anaphora resolution may not occur at the level of grammatical representations but at the level of processing strategies. Similar to attachment preferences in relative clauses (e.g., Dussias, 2003, 2004; Dussias & Sagarra, 2007; Fernández, 2002; Jegerski, 2018; Jegerski et al., 2016; Papadopoulou & Clahsen, 2003), interpretive preferences in anaphora resolution may reflect the use of different strategies that result in different resolution preferences. Bilinguals could be using similar strategies for their two languages to solve syntactic ambiguity, the strategies associated with their dominant language (see Fernández, 2002). This reasoning, however, would not completely explain why Catalan-dominant bilinguals—and possibly Spanish-dominant bilinguals, also—mitigate their preference of overt pronouns for object antecedents in their non-dominant language. Catalan-dominant bilinguals would not be transferring their polarized strategy of Catalan into Spanish, but they would still be showing well-defined patterns. It is possible that a combination of CLI and bilingualism effects resulting in the weakening of pronominal biases in the non-dominant language of bilinguals explains our findings. In the absence of between-language differences, however, no bilingualism effects emerge.

Considering language dominance effects in Catalan in greater depth, CLI has been found to be modulated by language dominance. This finding shows that Spanish influence has not yet systematically permeated the Catalan system in the sense that all Catalan speakers share the same processing strategies when interpreting null and overt pronouns. In our results, two different Catalan referential systems could be identified: 1) a variety represented by Catalan-dominant bilinguals in which null and overt pronouns show very categorical and polarized PAH-like biases, and 2) a variety represented by Spanish-dominant and balanced bilinguals that shows more tolerance to optionality or non-PAH-like interpretations as a result of persistent CLI. As such, a steady and homogeneous Catalan referential system does not seem to exist among our bilinguals,

nor, feasibly, the population our bilinguals represent. In the absence of a reference monolingual Catalan variety, the variety represented by Catalan-dominant bilinguals represents the Catalan system.

Similarly to previous findings on early Catalan-Spanish bilinguals in adulthood, Catalan-dominant bilinguals have been found to differ from the non-Catalan-dominant groups (Boix-Fuster & Sanz, 2008; Perpiñán & Soto-Corominas, 2021). In the case that a homogeneous Spanish referential system exists—as it seems to, given that all groups present parallel patterns of pronoun resolution in Spanish—it permeates the Catalan system more easily in bilinguals who are not Catalan-dominant. A finding that deserves to be highlighted is that the Catalan biases of balanced bilinguals also show CLI from Spanish and correspond with those of Spanish-dominant bilinguals rather than with those of Catalan-dominant bilinguals. As also suggested by Perpiñán (2018) and Perpiñán and Soto-Corominas (2021), it could be indicating a change in progress. In this regard, we can speculate that CLI from Spanish has been incorporated into the Catalan system gradually, from the Spanish-dominant speakers' system toward more balanced systems, as also shown in the analysis that uses the dominance continuum. This influence may hypothetically affect the system of Catalan-dominant speakers as well.

It has been claimed that syntactic predictability is learned from experience of language over time, so language exposure may be determinant in the acquisition of pronominal interpretive patterns (Arnold et al., 2018; Langlois & Arnold, 2020; Williams, 2020). A higher exposure to Spanish, in contexts of use as well as in terms of quantity and quality of input, could facilitate influence of Spanish on Catalan; in §6.2.5 we saw that language history did not make better predictions for our results compared to language use, as measured in the BLP. It is true that Catalan-dominant bilinguals, in contrast to the other two groups, have been less exposed to Spanish and use this language to a much lesser extent. If their exposure to and use of Spanish increased and was not as low as it was for our participants, CLI could probably more easily permeate in their anaphora resolution preferences. Given that balanced bilinguals have mostly been raised in the metropolitan area of Barcelona, we also speculate as to whether this influence would also be attested in balanced bilinguals from other areas in which Spanish is not the environmental majority language. However, given that their dominance in Spanish will be higher than that of Catalan-dominant bilinguals, we do not expect them to be preserved from Spanish influence.

Spanish, in contrast to Catalan, shows homogeneous preferences in pronoun resolution. In this language, generally well-defined PAH-like biases—while less pronounced than those of Catalan—are observed across bilinguals. Influence from the strongly polarized Catalan system does not seem to be attested in any group, nor modulated by language dominance. In a sense, the impassiveness and soundness of the Spanish pronominal system compared to that of Catalan suggests that it is the system that will probably increase in use among Catalan-Spanish bilinguals, through CLI. In line with other studies on bilingual Spanish in contact with Catalan, although they are scarce, bilingual Spanish does not seem to be affected by CLI from Catalan to the same extent as bilingual Catalan is by CLI from Spanish (see also Cuza & Guijarro-Fuentes, 2018; Jiménez-Gaspar et al., 2020; Perpiñán & Soto-Corominas, 2021).

Refining the assertions in Bel and García-Alcaraz (2018), we have not found solid evidence of bilingual Spanish in contact with Catalan to be influenced by Catalan. However, it is still theoretically possible that Spanish does not show such sound and impermeable interpretive biases as we state in the previous paragraph. Despite not showing significant differences when overtly compared, the bilingual and the monolingual Spanish varieties give rise to two slightly different pictures. Bilingual Spanish shows weak but well-defined and significant interpretive preferences, whereas these biases in monolingual Spanish do not reach statistical significance (see §5.3.3)⁶⁴. In light of the findings of the present chapter, evidence showing Catalan influence is not observed when comparing groups. All groups of bilinguals show parallel behaviors and Catalan-Spanish bilinguals do not seem to differ from Spanish monolinguals, regardless of their dominance profile. Based on the different patterns that were attested between the bilingual and monolingual varieties (comparing Spanish-dominant bilinguals and monolinguals), we should not entirely discard the possibility that the coexistence of Catalan and Spanish at a societal level has favored the creation of a steady and homogeneous contact variety (Silva-Corvalán, 2008).

According to this view—the possible existence of a specific bilingual variety of Spanish in contact with Catalan—the lack of language dominance effects in bilingual Spanish

⁶⁴ Significant differences were not attested when comparing bilingual and monolingual Spanish, so we cannot straightforwardly argue that Catalan has influenced bilingual Spanish when comparing it to monolingual Spanish. However, it is not straightforward either to conclude that the lack of significant differences between the two groups should be interpreted as bilingual Spanish showing the same resolution patterns as monolingual Spanish (see Westergaard, 2021 for a similar reasoning).

can be interpreted as suggesting that Catalan influence has already permeated the bilingual system. However, to state that Catalan, the more categorical language, has influenced Spanish, the language with a more flexible distribution, would be at odds with our previous reasonings (and with Romano's 2019 approach). Despite the fact that, generally speaking, bidirectional CLI can occur (e.g., López Otero, 2022; Pavlenko & Jarvis, 2002; Puig-Mayenco et al., 2018), we are not sure whether it would be plausible to argue that bidirectional CLI is differently affecting the same phenomenon at distinct levels (i.e., individual and societal). As argued by Silva-Corvalán (2008, p. 215), the bilingual individual is the locus of transfer, which can result in a societal phenomenon if an innovation gradually spreads throughout a community of speakers. Alternatively, the bilingual variety of Spanish could have resulted from convergence between Catalan and Spanish (Sánchez, 2004, 2015; see also §3.2.4). If this was the case, however, we would have expected 'convergence' patterns in the two languages. In Catalan, they could be observed for Spanish-dominant and balanced bilinguals, but not for Catalan-dominant bilinguals. These reasonings, however, are merely speculative. With our results, we are able only to conclude that we have not found enough evidence to claim that bilingual Spanish differs from monolingual Spanish as a result of contact with Catalan.

Lastly, it should be noticed that only relying on offline measures—where metalinguistic knowledge is involved—may provide us with an incomplete perspective of the phenomenon of bilingual anaphora resolution (see Chamorro et al., 2016). This is a limitation of the present study, which could be interestingly complemented by an online study. Online measures would allow us to better discern whether self-evident bilingualism effects attributable to a less efficient real-time integration of information intervene in Catalan-Spanish bilinguals' processing of pronouns in unmarked and marked information contexts. These bilingualism effects would be expected to interact with language dominance, as in offline data. Also, online data could add information of divergences in the processing of pronouns in the three marked constructions and at different points of time, by comparing earlier and later processing stages (see Patterson & Felser, 2020). In any case, offline data can readily provide us with valuable information (see also the discussion in previous §6.3). While our study cannot conclude that no processing difficulties are experienced by early Catalan-Spanish bilinguals, it does not find any evidence of them in Catalan-Spanish bilinguals pronoun resolution preferences.

On the whole, bilingualism effects in pronoun resolution by Catalan-Spanish bilinguals were only attested if there is microvariation between the two languages in contact. Thus,

the language combination seems to be relevant when studying the impact of bilingualism on anaphora resolution. Language dominance effects suggest a more determinant role of CLI, rather than of bilingualism per se, although a combination of both accounts may account for the vulnerability of anaphora resolution attested across bilingual populations. Importantly, bilingualism effects on anaphora resolution, modulated by language dominance, have also been found in early Catalan-Spanish bilinguals, who live and have been raised in an essentially bilingual society. Regarding the directionality of CLI, it seems to occur from Spanish, the language with less defined biases, toward Catalan, the language with more restrictive biases (Romano, 2019; see also Sorace, 2011). The occurrence and strength of CLI has been found to be modulated by language dominance, thus shaping pronoun resolution preferences depending on the bilinguals' profile. Catalan-dominant bilinguals seem to be preserved from the influence of their non-dominant language, and more weakened pronominal biases are attested with a higher dominance of Spanish.

Chapter 7

Conclusions

7.1 Summary of the study and conclusions

The present thesis has provided evidence on how the interpretation of null and overt subject pronouns 1) is driven by the interaction of multiple language-internal factors such as syntactic, pragmatic, and sequential factors (i.e., the syntactic function, information status, and linear or hierarchical position of the potential antecedents), and 2) is vulnerable to language-external factors such as bilingualism, whose effects are modulated by language dominance. On the one hand, we analyzed the interpretive biases of ambiguous null and overt subject pronouns in syntactically and pragmatically (un)marked information structures in Catalan and Spanish (canonical sentences, topicalization of the object via clitic-left dislocation, focalization of the subject via it-cleft, and focalization of the object via it-cleft). On the other hand, we compared these resolution preferences crosslinguistically (in Catalan and Spanish) among Catalan-Spanish bilinguals differing in their language dominance profile (Catalan-dominant, balanced, and Spanish-dominant bilinguals), also considering Spanish monolinguals.

Regarding language-internal factors, the syntactic function of the antecedents was shown to interact in anaphora resolution with other factors such as the information status and the linear (and hierarchic) position of the antecedents. Importantly, none of the linguistic factors under analysis was found to act as an overriding factor. These findings confirm the need of multi-factorial approaches to account for pronoun interpretation (see also Bader & Portele, 2019; de la Fuente, 2015; Kaiser, 2011; Kaiser & Trueswell, 2008; Schumacher et al., 2017; among others). Moreover, null and overt subject pronouns in null subject languages such as Catalan and Spanish appeared to be sensitive in different ways to these different constraints, supporting a form-specific and multiple-constraint approach (Kaiser & Trueswell, 2008). On the one hand, null pronouns were found to be sensitive to the combination of syntactic and pragmatic factors, but not to sequential factors. They showed a preference for subject and topical antecedents, regardless of their linear (and hierarchical) position. On the other hand, overt pronouns were found to be sensitive to the combination of syntactic and sequential factors, but not to pragmatic

factors. They showed a preference for object and second-mentioned (hierarchically lower) antecedents, regardless of their information status.

In canonical sentences, pronoun resolution biases in bilingual Catalan and bilingual Spanish⁶⁵ were found to be consistent with the predictions of the Position of Antecedent Hypothesis (PAH; Carminati, 2002). In both languages, ambiguous null and overt subject pronouns were preferably interpreted as coreferring with subject and object antecedents, respectively. In these pragmatically unmarked contexts, however, the subject antecedent is also topical, and the object antecedent is also second-mentioned and appears in a hierarchically lower position. Through the manipulation of information structure, we found that when these cues did not coincide in the same antecedent and were misaligned, the biases of null and overt pronouns remained undefined. In clitic-left dislocation structures (CLLD), neither null nor overt pronouns showed a preference for topical objects in sentence-initial (and hierarchically higher) position. In cleft structures, focused subjects and focused objects (first-mentioned, and hierarchically higher) were not found to be the preferred antecedents for a pronoun (null or overt) either. On the other hand, topical (or non-focal) subjects in object clefts were clearly preferred for null pronouns, and second-mentioned (and hierarchically lower) objects in subject clefts were also clearly preferred for overt pronouns. These findings suggest that it is a combination of factors that drives pronoun resolution, and that there is no overarching or more heavily weighted factor that can explain anaphoric dependencies alone.

On the one hand, the preference of null pronouns for topical and subject antecedents demonstrates that their interpretive properties must be necessarily defined in the interplay between syntax and pragmatics. This finding does not exclude the possibility that null pronouns are interpreted as coreferring with non-topical antecedents. In fact, evidence from corpus or semi-spontaneous production data have shown that null pronouns are not uniquely specialized in conveying topic maintenance, but show a wider variety of functions in the pragmatically appropriate contexts (e.g., García-Alcaraz & Bel, 2019; Giannakou & Sitaridou, 2022; Lozano, 2016). It is possible that the interpretation of null pronouns is guided by the coherence of the discourse and therefore prefers the antecedent that makes the most relevant contribution in the incoming discourse, somewhat regardless of the (non-)topical status and (non-)subject function of the antecedent (e.g., Leonetti, 2021).

⁶⁵ Remember that Catalan-dominant bilinguals were taken as the reference group for Catalan, and Spanish-dominant bilinguals were taken as the reference group for Spanish.

On the other hand, the preference of overt pronouns for object and second-mentioned antecedents, which appear lower in the syntactic configuration, suggests that these pronouns may be rather sensitive to syntactic-structural cues, in line with the Position of Antecedent Hypothesis (Carminati, 2002). The lack of apparent sensitivity of these pronouns to pragmatic cues in the structures under analysis is a particularly relevant—and unexpected—finding of the present thesis. Firstly, it calls into question the commonly assumed statement that overt pronouns are specialized in conveying topic shift (they did not reject coreference with topicalized object antecedents). Secondly, it suggests that null and overt pronouns do not necessarily show a division of labor in their interpretations if we not only take syntactic factors into account, but also cues from other linguistic levels.

Regarding between-language comparisons, the impact of information structure on null and overt subject pronoun resolution has been proven to be similar in (bilingual) Catalan, bilingual Spanish, and monolingual Spanish. Therefore, the statements that have been made until this point on the interaction between syntactic, pragmatic, and sequential factors hold for the two languages and the two language varieties. However, microparametric variation was shown between Catalan and Spanish in unmarked sentences. In monolingual Spanish, in fact, well-defined PAH-like biases of null and overt subject pronouns were not observed; both pronouns showed blurred preferences and were not interpreted differently from chance. Comparisons between Catalan and Spanish revealed that Catalan shows stronger and more polarized preferences than both bilingual Spanish (only regarding overt pronouns) and monolingual Spanish (regarding both null and overt pronouns) (see Bel & García-Alcaraz, 2018 for similar findings). Bilingual Spanish in contact with Catalan and monolingual Peninsular Spanish, on the other hand, were not found to differ statistically, although bilingual Spanish showed clear-cut significant biases and monolingual Spanish did not.

Our findings provide further evidence regarding the existence of microparametric differences among Romance and null subject languages (see Bel & García-Alcaraz, 2018; Contemori & Di Domenico, 2021; Filiaci et al., 2014; Giannakou & Sitaridou, 2020; Torregrossa et al., 2020). Taking into account these studies involving crosslinguistic comparisons between Catalan, Spanish, Italian, and Greek, a gradation in the strength of both null and more especially overt pronouns may exist, from the very polarized biases of Italian and possibly Catalan, to the weaker, but well-defined biases of Greek, to the very weak and sometimes non-significant biases of Spanish (Italian > Catalan > Greek > Spanish). In line with Torregrossa et al. (2020), these differences in the strength of

subject pronouns interpretive preferences could be related to different word-order restrictions in each language (i.e., Catalan being more restrictive than Spanish).

The lack of significant biases in monolingual Spanish deserves further attention given that most previous studies testing the PAH have found either a well-defined bias of overt pronouns in main-subordinate contexts or of null pronouns in subordinate-main contexts (see de Rocafiguera & Bel, 2022). In our data on main-subordinate contexts, only a non-significant tendency of overt pronouns was observed, so both null and overt pronouns were interpreted as unbiased (see Giannakou & Sitaridou, 2020 for similar findings in Chilean Spanish). We have interpreted this lack of biases in monolingual Spanish (in contrast to the well-defined biases of bilingual Spanish and Catalan) as indicating that Spanish shows weaker resolution patterns than other null subject languages. These very weak biases may make Spanish pronouns more susceptible to differences in the stimuli of different studies (related to clause order or implicit causality, among others) and this may have led to mixed findings in previous studies (e.g., Bel & García-Alcaraz, 2018; Contemori & Di Domenico, 2021; Filiaci et al., 2014; Giannakou & Sitaridou, 2020). Together with the fact that both null and overt pronouns were found to be affected by microvariation, it is reasonable to conclude that the interpretive patterns of null pronouns may not be as steady when establishing coreference as previously assumed.

Regarding language-external factors, or individual factors, bilingualism and language dominance effects were found in relation to crosslinguistic influence (CLI). In canonical sentences, where microvariation in the interpretation of null and overt pronouns was shown, CLI from Spanish toward Catalan was observed. In Catalan, Spanish-dominant and balanced bilinguals demonstrated significantly weaker biases of null pronouns (only in the case of Spanish-dominant bilinguals) and overt pronouns (for both groups) compared to Catalan-dominant bilinguals. Thus, a gradient in the polarization of pronominal biases was observed, modulated by language dominance. In Spanish, in contrast, homogeneous and similar preferences were found for the two pronouns in the three bilingual groups under study. Referring to marked information structures (i.e., CLLD, subject clefts, and object clefts), where no microvariation had been demonstrated between Catalan and Spanish, the three bilingual groups showed similar interpretive preferences; no effects of language dominance or bilingualism could be identified.

Firstly, our findings showed that both null and overt subject pronouns can be vulnerable to the effects of bilingualism. Microvariation between Catalan and Spanish and CLI was found regarding the two pronominal forms, which demonstrates that the syntactic-

pragmatic constraints that guide anaphora resolution are complex for both null and overt pronouns (in line with Clements & Domínguez, 2017). Secondly, the occurrence of CLI was found to be unidirectional, from the language with more flexible biases, Spanish, toward the language with more categorical biases, Catalan (Romano, 2019; see also Sorace, 2011; Tsimpili et al., 2004). Therefore, language dominance did not determine the directionality of CLI. However, language dominance was found to regulate the strength of CLI: the lower the dominance in Catalan, the higher the influence from Spanish. Thirdly, regarding bilingual Spanish, our findings do not provide reliable evidence in favor of CLI occurring from Catalan toward Spanish. However, PAH-like biases were clearly observed in the bilingual Spanish variety and did not reach significance in the monolingual one. As such, we cannot discard, as suggested by Bel and García-Alcaraz (2018), that traces of Catalan influence or language convergence are observed in bilingual Spanish as a result of a prolonged language contact situation.

Our findings also have implications regarding the Interface Hypothesis (IH; Sorace, 2011, 2012; Sorace & Filiaci, 2006). A general bilingual strategy consisting of the overextension of overt pronouns, as derived from the IH, was not observed in the resolution preferences of Catalan-Spanish bilinguals, not even in complex structures as marked information constructions (rather, both null and overt pronouns were found to be sensitive to CLI in contexts where it occurs). Therefore, even though anaphora resolution—as a linguistic phenomenon at the syntax-pragmatics interface—has been claimed to involve a complex cognitive process that is challenging for bilinguals, regardless of their profile (or proficiency level) and the language combination, no specific difficulties were evidenced for highly functional Catalan-Spanish bilinguals. These observations led us to conclude that 1) a general bilingual strategy as derived from the IH (i.e., from the sole fact of being bilingual) does not affect the performance of Catalan-Spanish bilinguals, and that 2) bilingualism effects appear as CLI, whose occurrence and strength are modulated by language dominance.

7.2 Further research

This study provides valuable evidence on the intrinsic resolution preferences of null and overt subject pronouns in Catalan and Spanish from a novel perspective, by analyzing the relative contribution of syntactic, pragmatic, and sequential (or hierarchic/configurational) factors. In this sense, it has disentangled notions that very often overlapped in previous

studies (those of subject, topic, and first-mention, and those of object, non-topic, and second-mention). However, there are other factors that we did not address and that have also been demonstrated to play a role in anaphora resolution, such as thematic roles or implicit causality. Further studies could disentangle the notions of syntactic function (subject/object) and thematic role (agent/patient) of the antecedent. Thematic roles have indeed been found to be a more determining predictor than subjecthood or topicality for pronoun resolution in German (Patterson & Schumacher, 2021; Schumacher et al., 2016, 2017). In the present study, agentivity and subjecthood totally overlap, so they should be extricated to be able to discern whether null and/or overt pronouns are sensitive to syntactic function or thematic role cues. Regardless of the syntactic function of the antecedent, null pronouns could be preferring agent and topical antecedents, and overt pronouns, patient antecedents and antecedents in the lowest position of the syntactic configuration. Moreover, it would also be interesting to study the relative weight of implicit causality by including verbs with different neutral and non-neutral implicit causality biases to weight the relative role of other semantic factors (see Cristerna-Román, 2020).

Regarding pragmatic factors and the role of information status in anaphora resolution, it would undoubtedly be interesting to see whether the results of the present study could be replicated using contextualized stimuli in which a discourse topic is clearly established. However, it would be challenging to create totally ambiguous contexts for each sentence, and it would also lengthen the task. A way to provide context to the stimuli could be by using a question to introduce the experimental items, as in Mayol (2010). In one of the tasks in de la Fuente (2015; Experiment 4), stimuli were presented with an introductory sentence and a short dialogue in the case of it-cleft sentences, and although he did not overtly contrast these contextualized vs. out-of-the blue sentences (in another experiment, Experiment 1), the two experiments led to similar results. We think that the findings in the present study could be replicated if the same stimuli were tested in discourse contexts, but they may add precision to the pragmatic features envisaged in the present research. Moreover, broader contexts could be used to address to what extent null pronouns are flexible for conveying different functions and contributing to discourse coherence.

Furthermore, it could be interesting to analyze null and overt pronoun resolution in prosodically marked information contexts using oral stimuli to see whether an equivalent information status role is found. In this respect, we would expect the reported findings on the role of discourse factors to be replicated across structures using different

mechanisms to mark the information status of its constituents. Analyzing these contexts would also allow a focus preposing structure different from it-clefts to be used to determine whether the anti-focus effect can be generalized for other focusing structures (as de la Fuente, 2015 demonstrated for focus marking using focus-sensitive particles).

Building on the observation that null and overt subject pronouns may not show complementary distributions or a division of labor at the level of pragmatics, it would also be interesting to contrast their interpretive properties with those of demonstrative pronouns or lexical subjects (Giannakou & Sitaridou, 2020 found strong coreference patterns of the demonstrative *este/esta* toward object antecedents). As observed by Torregrossa et al. (2020), related to the fact that overt subject pronouns are rarely found in (semi)spontaneous production data (see also García-Alcaraz & Bel, 2019; Lozano, 2016; among others), it is possible that null subjects are not in complementary distribution with overt pronouns, but with full DPs.

Beyond the impact of information structure, the present thesis also addressed crosslinguistic comparisons between Catalan and Spanish, expanding the evidence on microvariation among null subject languages. In this regard, however, to gain a more complete picture of language-specific nuances, it would be interesting to overtly address the Italian-Catalan comparison in further research, to see to the extent to which they show similarly polarized patterns, and to contrast the maximum possible number of Romance null subject languages in the same design. Comparisons including European Portuguese or Romanian would be especially relevant to confirm whether they show similar preferences to those of Spanish, as Peripheral Romance languages that allow for VSO, or whether Portuguese more closely resembles Catalan and Italian (as seems to be the case, see Rinke & Flores, 2018), given that it is not as permissive as Spanish and Romanian in terms of constituent order (Leonetti, 2017).

From the perspective of bilingualism, this study has also provided novel insights on the role of language dominance in anaphora resolution on an understudied population, early bilinguals that have been raised in a bilingual society, using a bidirectional design. Given that we found CLI in pronoun resolution to be modulated by language dominance, it would be interesting to investigate the role of language dominance in modulating CLI in the Catalan-Spanish bilingual context from a broader perspective, investigating other linguistic phenomena using similar bidirectional designs (see Benito & Bel, 2022). We also focused on contexts of microvariation between two typologically similar languages, Catalan and Spanish, which could be expanded by studying early bilinguals in bilingual

societies where two typologically distant languages are spoken (e.g., Basque-Spanish; see Iraola Azpiroz, 2014; Iraola Azpiroz et al., 2017), to compare the effects of bilingualism and language dominance in typologically similar and distant languages, also using bidirectional designs.

Lastly, in terms of identifying and characterizing the language dominance profile of the bilinguals, it would also be worth creating a background questionnaire or an improved version of the Bilingual Language Profile questionnaire (Birdsong et al., 2012), which is able to capture the nuances and specificities of bilinguals that are raised in bilingual societies such as Catalonia. Indeed, it would be interesting to further explore the language dominance profiles of Catalan-Spanish bilinguals and to examine the relative contribution to language dominance of different constructs that have been proved to be useful for quantifying the bilingual experience, such as language exposure and use across one's life span, language proficiency, language attitudes, or language mixing practices (see De Cat et al., 2022).

On the whole, the present thesis has contributed to further characterizing the interpretive properties of null and overt pronouns in null subject languages by analyzing marked information structures that manipulated syntactic, pragmatic, and sequential factors. It has shown that these multiple factors interact in shaping the resolution preferences of null and overt pronouns, and that the two pronouns are sensitive to each of these factors in different ways. Besides language-internal factors, however, language-external factors such as bilingualism and language dominance were also found to play a relevant role in anaphora resolution by highly functional Catalan-Spanish bilinguals. Although anaphora resolution, as a phenomenon at the syntax-pragmatics interface, has been proposed to be cognitively challenging for bilingual populations, our participants did not show any difficulties when interpreting pronouns in complex pragmatically and syntactically marked information structures. However, pronoun resolution was sensitive to bilingualism effects in the form of crosslinguistic influence, whose occurrence and strength was modulated by language dominance. Both information structure and language dominance, two understudied factors to date in the widely studied linguistic phenomenon of anaphora resolution, have therefore been shown to play a role in the interpretation of subject pronouns in null subject languages.

References

- Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, 19(4), 689-698. <https://doi.org/10.1017/S1366728916000225>
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207-245. <https://doi.org/10.3102/0034654310368803>
- Adli, A. (2011). A heuristic mathematical approach for modeling constraint cumulativity: Contrastive focus in Spanish and Catalan. *The Linguistic Review*, 28(2), 111-173. <https://doi.org/10.1515/tlir.2011.004>
- Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological review*, 106(4), 748-765. <https://doi.org/10.1037/0033-295X.106.4.748>
- Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass*, 1(1-2), 84-99. <https://doi.org/10.1111/j.1749-818X.2007.00009.x>
- Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., & Clifton, C. (2002). Null vs. Overt pronouns and the topic-focus articulation in Spanish. *Italian Journal of Linguistics*, 14(2), 151-170.
- Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology*, 7: 617. <https://doi.org/10.3389/fpsyg.2016.00617>
- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50(1), 250-263. <https://doi.org/10.3758/s13428-017-0867-9>
- Argenter, J. A. (2020). Languages in contact: A sociocultural approach. In J. A. Argenter & J. Lüdtke (Eds.), *Manual of Catalan Linguistics* (pp. 597-628). De Gruyter. <https://doi.org/10.1515/9783110450408-025>
- Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10(1), 79-99. <https://doi.org/10.1017/S1366728906002835>

- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. Routledge.
- Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics*, 16(5), 443-463. [https://doi.org/10.1016/0378-2166\(91\)90136-L](https://doi.org/10.1016/0378-2166(91)90136-L)
- Ariel, M. (2001). Accessibility theory: An overview. In T. J. M. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (Vol. 8, pp. 29-87). John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.8.04ari>
- Arnaus Gil, L. (2021). Age of onset, language dominance and dialectal variation: Catalan copula selection in locative contexts with (non-)eventive subjects. *International Journal of Multilingualism*, 1-23. <https://doi.org/10.1080/14790718.2021.1892115>
- Arnold, J. E. (1998). *Reference form and discourse patterns* [Doctoral dissertation]. Stanford University.
- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, 31(2), 137-162. https://doi.org/10.1207/S15326950DP3102_02
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4), 187-203. <https://doi.org/10.1111/j.1749-818X.2010.00193.x>
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76(1), B13-B26. [https://doi.org/10.1016/S0010-0277\(00\)00073-1](https://doi.org/10.1016/S0010-0277(00)00073-1)
- Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, 102, 41-54. <https://doi.org/10.1016/j.jml.2018.05.002>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bader, M., & Portele, Y. (2019). The interpretation of German personal pronouns and d-pronouns. *Zeitschrift für Sprachwissenschaft*, 38(2), 155-190. <https://doi.org/10.1515/zfs-2019-2002>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48. <https://doi.org/10.18637/jss.v067.i01>

- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition*, *15*(3), 616-629. <https://doi.org/10.1017/S1366728912000090>
- Bel, A. (2001). Sujetos nulos y sujetos explícitos en las gramáticas iniciales del castellano y el catalán. *Revista Española de Lingüística*, *31*(2).
- Bel, A., & Albert, M. (2016). The development of referential choice in Spanish narratives among school-age children and adolescents. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and Spoken Language Development across the Lifespan: Essays in Honour of Liliana Tolchinsky* (pp. 251-269). Springer International Publishing. https://doi.org/10.1007/978-3-319-21136-7_15
- Bel, A., & García-Alcaraz, E. (2015). Subject pronouns in the L2 Spanish of Moroccan Arabic speakers. In T. Judy & S. Perpiñán (Eds.), *The Acquisition of Spanish in Understudied Language Pairings* (pp. 201-232). John Benjamins Publishing Company. <https://doi.org/10.1075/ihll.3.08bel>
- Bel, A., & García-Alcaraz, E. (2018). Pronoun interpretation and processing in Catalan and Spanish bilingual and monolingual speakers. In A. Cuza & P. Guijarro-Fuentes (Eds.), *Language Acquisition and Contact in the Iberian Peninsula* (pp. 37-62). De Gruyter. <https://doi.org/10.1515/9781501509988-003>
- Bel, A., García-Alcaraz, E., & Rosado, E. (2016). Reference comprehension and production in bilingual Spanish: The view from null subject languages. In A. Alba de la Fuente, E. Valenzuela, & C. Martínez Sanz (Eds.), *Language Acquisition Beyond Parameters: Studies in honour of Juana M. Liceras* (pp. 37-70). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.51>
- Bel, A., Ortells, M., & Morgan, G. (2015). Reference control in the narratives of adult sign language learners. *International Journal of Bilingualism*, *19*(5), 608-624. <https://doi.org/10.1177/1367006914527186>
- Bel, A., Perera, J., & Salas, N. (2010). Anaphoric devices in written and spoken narrative discourse: Data from Catalan. *Written Language & Literacy*, *13*(2), 236-259. <https://doi.org/10.1075/wll.13.2.03bel>
- Bel, A., Sagarra, N., Comínguez, J. P., & García-Alcaraz, E. (2016). Transfer and proficiency effects in L2 processing of subject anaphora. *Lingua*, *184*, 134-159. <https://doi.org/10.1016/j.lingua.2016.07.001>

- Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, 25(4). <https://doi.org/10.1007/s11049-007-9026-9>
- Benito, R. (2020). El marcado diferencial de objeto en catalán: Contextos de aparición y análisis. In C. J. Álvarez López, M. Carrillo Rivas, D. Jiménez Palmero, M. Méndez Orense, A. Moratinos Flórez, & Padilla Herrada (Eds.), *Lingüística prospectiva: Tendencias actuales en estudios de la lengua entre jóvenes investigadores* (pp. 273-286). Editorial Universidad de Sevilla.
- Benito, R., & Bel, A. (2022). *Els efectes de l'estabilitat referencial i la dominança lingüística en el Marcatge Diferencial d'Objecte en català* [Oral presentation]. Jornades de lingüística catalana a Viena, Vienna.
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3-4), 129-181. <https://doi.org/10.1515/tl-2013-0009>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511605963>
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological bulletin*, 143(3), 233-262. <https://doi.org/10.1037/bul0000099>
- Bini, M. (1993). La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro pro-drop. In J. M. Liceras (Ed.), *La lingüística y el análisis de los sistemas no nativos* (pp. 126-139). Dovehouse.
- Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, 35(4), 374-392. <https://doi.org/10.1093/applin/amu031>
- Birdsong, D. (2016). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 85-105). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.005>
- Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. <https://sites.la.utexas.edu/bilingual/>
- Blas Arroyo, J. L. (2007). El contacto de lenguas como factor de retención en procesos de variación y cambio lingüístico: Datos sobre el español en una comunidad bilingüe peninsular. *Spanish in Context*, 4(2), 263-291. <https://doi.org/10.1075/sic.4.2.07bla>

- Blas Arroyo, J. L. (2015). The scope of language contact as a constraint factor in language change: The periphrasis *haber de plus infinitive* in a corpus of language immediacy in modern Spanish. *International Journal of Bilingualism*, 19(5), 499-524. <https://doi.org/10.1177/1367006914524644>
- Bloomfield, L. (1933). *Language*. Holt.
- Blything, L. P., Iraola Azpiroz, M., Allen, S., Hert, R., & Järvikivi, J. (2022). The influence of prominence cues in 7- to 10-year-olds' pronoun resolution: Disentangling order of mention, grammatical role, and semantic role. *Journal of Child Language*, 49(5), 930-958. <https://doi.org/10.1017/S0305000921000349>
- Blything, L. P., Järvikivi, J., Toth, A. G., & Arnhold, A. (2021). The influence of focus marking on pronoun resolution in dialogue context. *Frontiers in Psychology*, 12, 684639. <https://doi.org/10.3389/fpsyg.2021.684639>
- Boix-Fuster, E., & Sanz, C. (2008). Language and identity in Catalonia. In J. Rothman & M. Nino-Murcia (Eds.), *Bilingualism and Identity* (Vol. 37, pp. 87-106). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.37.07boi>
- Bonvin, A., Brugger, L., & Berthele, R. (2021). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2020-0093>
- Bosch, J. E., & Unsworth, S. (2021). Cross-linguistic influence in word order: Effects of age, dominance and surface overlap. *Linguistic Approaches to Bilingualism*, 11(6), 783-816. <https://doi.org/10.1075/lab.18103.bos>
- Bosch, P., & Hinterwimmer, S. (2016). Anaphoric reference by demonstrative pronouns in German: In search of the relevant parameters. In A. Holler & K. Suckow (Eds.), *Empirical Perspectives on Anaphora Resolution* (pp. 193-212). De Gruyter. <https://doi.org/10.1515/9783110464108>
- Bosch, P., & Umbach, C. (2007). Reference determination for demonstrative pronouns. *ZAS Papers in Linguistics*, 48, 39-51. <https://doi.org/10.21248/zaspil.48.2007.353>
- Bullock, B. E., & Gerfen, C. (2004). Phonological convergence in a contracting language variety. *Bilingualism: Language and Cognition*, 7(2), 95-104. <https://doi.org/10.1017/S1366728904001452>
- 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>

- Bylund, E., Hyltenstam, K., & Abrahamsson, N. (2021). Age of acquisition – not bilingualism – is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism: Language and Cognition*, 24(1), 18-30.
<https://doi.org/10.1017/S1366728920000188>
- Camacho, J. (2013). *Null subjects*. Cambridge University Press.
- Carbonell, J. M. (2019). The two main challenges to Catalan identity. *American Behavioral Scientist*, 63(7), 789-806.
<https://doi.org/10.1177/0002764218763479>
- Cardinaletti, A., & Starke, M. (1999). The typology of structural deficiency: On the three grammatical classes. *Working Papers in Linguistics*, 4, 41-109.
- Carminati, M. N. (2002). *The processing of Italian subject pronouns* [Doctoral dissertation]. University of Massachusetts Amherst.
- Casielles-Suárez, E. (2004). *The syntax-information structure interface: Evidence from Spanish and English*. Routledge. <https://www.routledge.com/The-Syntax-Information-Structure-Interface-Evidence-from-Spanish-and-English/Casielles-Suarez/p/book/9780415537506>
- Castro, T., Rothman, J., & Westergaard, M. (2017). Comparing anaphora resolution in early and late Brazilian Portuguese-European Portuguese bidialectal bilinguals. *Revista Española de Lingüística Aplicada*, 29(2), 429-461.
<https://doi.org/10.1075/resla.29.2.03cas>
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 25-55). Academic Press.
- Chamorro, G. (2018). Offline interpretation of subject pronouns by native speakers of Spanish. *Glossa: A Journal of General Linguistics*, 3(1).
<https://doi.org/10.5334/gjgl.256>
- Chamorro, G., Sorace, A., & Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition*, 19(3).
<https://doi.org/10.1017/S1366728915000152>
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. MIT Press.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3-42.
<https://doi.org/10.1017/S0142716406060024>

- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693-706.
<https://doi.org/10.1017/S0272263117000250>
- Clements, M., & Domínguez, L. (2017). Reexamining the acquisition of null subject pronouns in a second language: Focus on referential and pragmatic constraints. *Linguistic Approaches to Bilingualism*, 7(1).
<https://doi.org/10.1075/lab.14012.cle>
- Colomina, M. P. (2019). *VSO order in Romance: A distinctness account* [Oral presentation]. 49th Linguistic Symposium on Romance Languages, University of Georgia.
- Colonna, S., Schimke, S., & Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. *Linguistics*, 50(5), 991-1013. <https://doi.org/10.1515/ling-2012-0031>
- Colonna, S., Schimke, S., & Hemforth, B. (2015). Different effects of focus in intra- and inter-sentential pronoun resolution in German. *Language, Cognition and Neuroscience*, 30(10), 1306-1325.
<https://doi.org/10.1080/23273798.2015.1066510>
- Comínguez, J. P., Sagarra, N., Bel, A., & García-Alcaraz, E. (2017). *The Position of the Antecedent Strategy in Spanish: A cross-dialectal comparison* [Oral presentation]. 47th Linguistic Symposium on Romance Languages, Newark, Delaware. <https://sites.udel.edu/lslr147/>
- Consell Superior d'Avaluació del Sistema Educatiu. (2017). *El coneixement de les llengües catalana i castellana de l'alumnat de Catalunya*. Generalitat de Catalunya, Departament d'Ensenyament.
<http://csda.gencat.cat/web/.content/home/arees-actuacio/publicacions/informes-avaluacio/Informes21catala.pdf>
- Consell Superior d'Avaluació del Sistema Educatiu. (2021a). *Estudi sociodemogràfic i lingüístic de l'alumnat de 4t d'ESO a Catalunya 2006-2013-2021*. Generalitat de Catalunya, Departament d'Ensenyament.
<http://csda.gencat.cat/web/.content/home/arees-actuacio/publicacions/informes-avaluacio/29-informes.pdf>
- Consell Superior d'Avaluació del Sistema Educatiu. (2021b). *L'avaluació de quart d'ESO 2021*. Generalitat de Catalunya, Departament d'Ensenyament.
[279](http://csda.gencat.cat/web/.content/home/arees-</p>
</div>
<div data-bbox=)

actuacio/publicacions/quaderns-avaluacio/quaderns-avaluacio-49/quaderns-avaluacio-enllac-49.pdf

- Contemori, C., & Di Domenico, E. (2021). Microvariation in the division of labor between null- and overt-subject pronouns: The case of Italian and Spanish. *Applied Psycholinguistics*, 42(4). <https://doi.org/10.1017/S0142716421000199>
- Contemori, C., & Dussias, P. E. (2020). The processing of subject pronouns in highly proficient L2 speakers of English. *Glossa: A Journal of General Linguistics*, 5(1), 1-19. <https://doi.org/10.5334/gjgl.972>
- Contemori, C., & Ivanova, I. (2021). Bilingual referential choice in cognitively demanding situations. *Bilingualism: Language and Cognition*, 24(1), 83-95. <https://doi.org/10.1017/S1366728920000176>
- Corona, V., Nussbaum, L., & Unamuno, V. (2013). The emergence of new linguistic repertoires among Barcelona's youth of Latin American origin. *International Journal of Bilingual Education and Bilingualism*, 16(2), 182-194. <https://doi.org/10.1080/13670050.2012.720668>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491-511. <https://doi.org/10.1016/j.jml.2004.02.002>
- Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, 15(5). <https://doi.org/10.1038/nrn3709>
- Cowles, H. W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus and pronouns. *Topoi*, 26(1), 3-18. <https://doi.org/10.1007/s11245-006-9004-6>
- Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. M. (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, 64(4), 381-403. <https://doi.org/10.1016/j.jml.2011.01.001>
- Cruschina, S. (2016). Information and discourse structure. In *The Oxford Guide to the Romance Languages*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0034>
- Cruschina, S. (2021a). Topicalization in the Romance languages. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.650>

- Cruschina, S. (2021b). The greater the contrast, the greater the potential: On the effects of focus in syntax. *Glossa: A Journal of General Linguistics*, 6(1).
<https://doi.org/10.5334/gjgl.1100>
- Cruschina, S. (2022). Focus and focus structures in the Romance languages. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.649>
- Cruschina, S., & Mayol, L. (2022). The realization of information focus in Catalan and in Spanish. [Oral presentation]. GLIF Seminar, Universitat Pompeu Fabra.
- Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 20(4).
<https://doi.org/10.1017/S1366728916000675>
- Cuza, A., & Guijarro-Fuentes, P. (2018). The distribution of copulas ser and estar in Spanish/Catalan bilinguals. In A. Cuza & P. Guijarro-Fuentes (Eds.), *The Distribution of Copulas Ser and Estar in Spanish/Catalan Bilinguals* (pp. 63-90). De Gruyter Mouton. <https://doi.org/10.1515/9781501509988-004>
- Davidson, J. (2020). Asymmetry and directionality in Catalan–Spanish contact: Intervocalic fricatives in Barcelona and Valencia. *Languages*, 5(4).
<https://doi.org/10.3390/languages5040060>
- Davidson, J. (2022). On Catalan as a minority language: The case of Catalan laterals in Barcelonan Spanish. *Journal of Sociolinguistics*.
<https://doi.org/10.1111/josl.12545>
- de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, 9(3). <https://doi.org/10.3390/bs9030033>
- de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, 8.
<https://doi.org/10.3389/fpsyg.2017.00522>
- De Cat, C., Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & The Q-BEx Consortium. (2022). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition*, 1-13.
<https://doi.org/10.1017/S1366728922000359>
- De Cesare, A. M. (2017). Cleft constructions. In A. Dufter & E. Stark (Eds.), *Manual of Romance Morphosyntax and Syntax* (pp. 536-5682). De Gruyter.
<https://doi.org/10.1515/9783110377088-015>

- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press.
- De Houwer, A. (2005). Early bilingual acquisition: Focus on morphosyntax and the separate development hypothesis. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford University Press.
- De Houwer, A. (2009). *Bilingual first language acquisition*. Multilingual Matters. <https://doi.org/10.21832/9781847691507>
- de la Fuente, I. (2015). *Putting pronoun resolution in context: The role of syntax, semantics, and pragmatics in pronoun interpretation* [Doctoral dissertation, Université Paris Diderot]. HAL Theses. <https://tel.archives-ouvertes.fr/tel-01535977>
- de Prada Pérez, A. (2009). *Subject expression in Minorcan Spanish: Consequences of contact with Catalan* [Doctoral dissertation, The Pennsylvania State University]. ProQuest. <https://www.proquest.com/openview/3dc0db4582ccc6021352f3206db09360/1?pq-origsite=gscholar&cbl=18750>
- de Prada Pérez, A. (2019). Theoretical implications of research on bilingual subject production: The Vulnerability Hypothesis. *International Journal of Bilingualism*, 23(2), 670-694. <https://doi.org/10.1177/1367006918763141>
- de Rocafiguera, N., & Bel, A. (2022). On the impact of clause order on pronoun resolution: Evidence from Spanish. *Folia Linguistica*, 56(1), 1-24. <https://doi.org/10.1515/flin-2021-2002>
- Del Maschio, N., Sulpizio, S., Toti, M., Caprioglio, C., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2020). Second language use rather than second language knowledge relates to changes in white matter microstructure. *Journal of Cultural Cognitive Science*, 4(2), 165-175. <https://doi.org/10.1007/s41809-019-00039-z>
- DeLuca, V., Segaert, K., Mazaheri, A., & Krott, A. (2020). Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. *Journal of Neurolinguistics*, 56. <https://doi.org/10.1016/j.jneuroling.2020.100930>
- Di Domenico, E., Baroncini, I., & Capotorti, A. (2020). Null and overt subject pronouns in topic continuity and topic shift: An investigation of the narrative productions of Italian Natives, Greek Natives and near-native second language speakers of Italian with Greek as a first language. *Glossa: A Journal of General Linguistics*, 5(1). <https://doi.org/10.5334/gjgl.1009>

- Di Pisa, G., Pereira Soares, S. M., & Rothman, J. (2021). Brain, mind and linguistic processing insights into the dynamic nature of bilingualism and its outcome effects. *Journal of Neurolinguistics*, 58. <https://doi.org/10.1016/j.jneuroling.2020.100965>
- Dijkstra, T. (2005). Bilingual word recognition and lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.
- Domínguez, L. (2013). *Understanding Interfaces: Second language acquisition and first language attrition of Spanish subject realization and word order variation*. John Benjamins Publishing Company.
- Domínguez, L. (2018). Information Structure. In K. L. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics* (pp. 372-391). Cambridge University Press. <https://doi.org/10.1017/9781316779194.018>
- Döpke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25(3), 555-584. <https://doi.org/10.1017/S0305000998003584>
- Dunn, A. L., & Fox-Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition*, 12(3). <https://doi.org/10.1017/S1366728909990113>
- Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25(4), 529-557. <https://doi.org/10.1017/S0272263103000238>
- Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism*, 8(3), 355-371. <https://doi.org/10.1177/13670069040080031001>
- Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 10(1), 101-116. <https://doi.org/10.1017/S1366728906002847>
- Ellert, M. (2013). Information structure affects the resolution of the subject pronouns er and der in spoken German discourse. *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics*, 12. <https://doi.org/10.4000/discours.8756>
- Enrique-Arias, A. (2010). On language contact as an inhibitor of language change. In A. Breitbarth (Ed.), *Continuity and Change in Grammar* (pp. 97-117). John Benjamins Publishing.

- Enrique-Arias, A. (2014). Efectos del contacto de lenguas en el castellano de Mallorca: Una perspectiva histórica. In A. Enrique-Arias, M. J. Gutiérrez, A. Landa, & F. Ocampo (Eds.), *Perspectives in the Study of Spanish Language Variation: Papers in Honor of Carmen Silva-Corvalán* (pp. 271-297). Universidade de Santiago de Compostela.
- Enrique-Arias, A. (2019). Testing contact-induced change in the Spanish of Mallorca: Insights from a historical perspective. In G. Rei-Doval & F. Tejedo-Herrero (Eds.), *Lusophone, Galician, and Hispanic Linguistics: Bridging Frames and Traditions*. Routledge.
- Enrique-Arias, A., & Méndez Guerrero, B. (2020). On the effects of Catalan contact in the variable expression of Spanish future tense: A contrastive study of Alcalá de Henares, Madrid and Palma, Majorca. In L. A. Ortiz López, R. E. Guzzardo, & M. González-Rivera (Eds.), *Issues in Hispanic and Lusophone Linguistics* (Vol. 22, pp. 316-334). John Benjamins Publishing Company.
<https://doi.org/10.1075/ihll.22.13enr>
- Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*. Oxford University Press.
- Fábregas, A. (2016). Information structure and its syntactic manifestation in Spanish: Facts and proposals. *Borealis – An International Journal of Hispanic Linguistics*, 5(2). <https://doi.org/10.7557/1.5.2.3850>
- Fedele, E., & Kaiser, E. (2014). Resolving null and overt pronouns in Italian: An experimental investigation of syntax-semantics interactions. In C. Brown, Q. Gu, C. Loos, J. Mielens, & G. Neveu (Eds.), *Proceedings of the 15th Texas Linguistic Society* (pp. 53-72). <http://tls.ling.utexas.edu/2014tls/TLS15-Proceedings.pdf>
- Feldhausen, I., & Vanrell, M. del M. (2014). Prosody, focus and word order in Catalan and Spanish: An optimality theoretic approach. *Proceedings of the 10th International Seminar on Speech Production*, 122-125.
- Felser, C. (2020). Do processing resource limitations shape heritage language grammars? *Bilingualism: Language and Cognition*, 23(1), 23-24.
<https://doi.org/10.1017/S1366728919000397>
- Fernández, E. M. (2002). Relative clause attachment in bilinguals and monolinguals. In R. R. Heredia & J. Altarriba (Eds.), *Advances in Psychology* (Vol. 134, pp. 187-215). North-Holland. [https://doi.org/10.1016/S0166-4115\(02\)80011-5](https://doi.org/10.1016/S0166-4115(02)80011-5)

- Ferré, P., & Brysbaert, M. (2017). Can Lextale-Esp discriminate between groups of highly proficient Catalan–Spanish bilinguals with different language dominances? *Behavior Research Methods*, *49*(2), 717-723. <https://doi.org/10.3758/s13428-016-0728-y>
- Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language*, *61*(2), 191-205. <https://doi.org/10.1016/j.jml.2009.04.001>
- Filiaci, F. (2011). *Anaphoric preferences of null and overt subjects in Italian and Spanish: A cross-linguistic comparison* [Doctoral dissertation]. The University of Edinburgh.
- Filiaci, F., Sorace, A., & Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, *29*(7). <https://doi.org/10.1080/01690965.2013.801502>
- Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters.
- Flors-Mas, A. (2021). Immigration and languages in Catalonia: From policies to results. *Journal of Iberian and Latin American Research*, *27*(2), 368-385. <https://doi.org/10.1080/13260219.2021.1994712>
- Flors-Mas, A., Sorolla, N., Pradilla, M. À., & Vila, F. X. (2021). The recent evolution of first languages in Catalonia: Between minoritization and language maintenance. *Language Problems and Language Planning*, *45*(1), 31-55. <https://doi.org/10.1075/lplp.20028.flo>
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357-383. <https://doi.org/10.1016/j.jml.2006.07.004>
- Foroodi-Nejad, F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian–English bilinguals. *Bilingualism: Language and Cognition*, *12*(4), 411-427. <https://doi.org/10.1017/S1366728909990241>
- Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro. *Natural Language & Linguistic Theory*, *25*(4), 691-734. <https://doi.org/10.1007/s11049-007-9025-x>
- Fukumura, K., & van Gompel, R. P. G. (2015). Effects of order of mention and grammatical role on anaphor resolution. *Journal of Experimental Psychology:*

- Learning, Memory, and Cognition*, 41(2), 501-525.
<https://doi.org/10.1037/xlm0000041>
- Garachana, M. (2018). Gramáticas en contacto: Inhibición del cambio lingüístico y gramaticalización en la convivencia entre el español y el catalán en Barcelona. *Revista Internacional de Lingüística Iberoamericana*, 32(2), 67-81.
<https://www.jstor.org/stable/48562746>
- Garachana, M. (2021). Contacto lingüístico y cambio gramatical: Convergencia y profundidad histórica en la constitución de la variedad de español de Cataluña. *Studia linguistica romanica*, 6, 192-219. <https://doi.org/10.25364/19.2021.6.8>
- García-Alcaraz, E. (2015). *Comprensión y producción de los pronombres nulos y explícitos de tercera persona en posición de sujeto en la adquisición temprana del español L2* [Doctoral dissertation, Universitat Pompeu Fabra]. TDX.
<http://www.tdx.cat/handle/10803/311971>
- García-Alcaraz, E., & Bel, A. (2019). Does empirical data from bilingual and native Spanish corpora meet linguistic theory? The role of discourse context in variation of subject expression. *Applied Linguistics Review*, 10(4).
<https://doi.org/10.1515/applirev-2017-0101>
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3), 459-464.
- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12(2), 213-237.
<https://doi.org/10.1017/S1366728909004015>
- Gathercole, V., Thomas, E., Kennedy, I., Prys, C., Young, N., Viñas-Guasch, N., Roberts, E., Hughes, E., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00011>
- Gelormini-Lezama, C., & Almor, A. (2011). Repeated names, overt pronouns, and null pronouns in Spanish. *Language and Cognitive Processes*, 26(3).
<https://doi.org/10.1080/01690965.2010.495234>
- Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, 16(1), 161-179. <https://doi.org/10.1017/S0305000900013490>

- Genesee, F., & Nicoladis, E. (2007). Bilingual first language acquisition. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 324-342). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470757833.ch16>
- Georgopoulos, A. (2017). Anaphora resolution in the interlanguage of Greek and English learners of Spanish: A corpus study. *Studies in Greek Linguistics*, 37, 239-252.
- Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6), 699-717. [https://doi.org/10.1016/0749-596X\(88\)90016-2](https://doi.org/10.1016/0749-596X(88)90016-2)
- Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language*, 28(6), 735-755. [https://doi.org/10.1016/0749-596X\(89\)90006-5](https://doi.org/10.1016/0749-596X(89)90006-5)
- Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual Language Profile. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208-225). Multilingual Matters.
- Giannakou, A. (2018). *Spanish and Greek subjects in contact: Greek as a heritage language in Chile* [Doctoral dissertation, University of Cambridge]. Apollo. <https://doi.org/10.17863/CAM.30352>
- Giannakou, A., & Sitaridou, I. (2020). Microparametric variation in the syntax of Spanish and Greek pronominal subjects. *Glossa: A Journal of General Linguistics*, 5(1). <https://doi.org/10.5334/gjgl.960>
- Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-language Study* (pp. 1-42). John Benjamins. <https://doi.org/10.1075/tsl.3>
- Godoy, M. C., Weissheimer, J., & Mafra, M. A. (2018). When grammar meets pragmatics: Subject preference and coherence relations in Brazilian Portuguese pronoun interpretation. *Journal of Portuguese Linguistics*, 17(1). <https://doi.org/10.5334/jpl.197>
- Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods*, 40(3). <https://doi.org/10.3758/BRM.40.3.760>
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT)

- and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 15(3), 594-615.
<https://doi.org/10.1017/S1366728911000332>
- Grant, A. P. (2020). Contact and language convergence. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 113-128). Wiley.
<https://doi.org/10.1002/9781119485094.ch5>
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515-530.
<https://doi.org/10.1080/20445911.2013.796377>
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3-15. [https://doi.org/10.1016/0093-934X\(89\)90048-5](https://doi.org/10.1016/0093-934X(89)90048-5)
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131-149.
<https://doi.org/10.1017/S136672899800025X>
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press.
<https://doi.org/10.4159/9780674056459>
- Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A Framework for Modelling the Local Coherence of Discourse. *IRCS Technical Reports Series*.
https://repository.upenn.edu/ircs_reports/116
- Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, 7(2), 199-229. <https://doi.org/10.1075/lab.15011.gru>
- Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., & Titone, D. (2021). Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics*, 42(2), 245-278.
<https://doi.org/10.1017/S0142716420000521>
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274-307.
<https://doi.org/10.2307/416535>
- Gürel, A. (2004). Selectivity in L2-induced L1 attrition: A psycholinguistic account. *Journal of Neurolinguistics*, 17(1), 53-78. [https://doi.org/10.1016/S0911-6044\(03\)00054-X](https://doi.org/10.1016/S0911-6044(03)00054-X)

- Gutiérrez-Bravo, R. (2003). Subject inversion in Spanish relative clauses. In T. Geerts & H. Jacobs (Eds.), *Romance Languages and Linguistic Theory* (p. 16). John Benjamins.
- Gutiérrez-Bravo, R. (2008). La identificación de los tópicos y los focos. *Nueva Revista de Filología Hispánica*, 56(2), 363-401.
- Halliday, M. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 199-244. <https://doi.org/10.1017/S00222226700016613>
- Hartshorne, J. K., Sudo, Y., & Uruwash, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, 60(3), 179-196. <https://doi.org/10.1027/1618-3169/a000187>
- Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511614132>
- Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00506>
- Hervé, C., Serratrice, L., & Corley, M. (2016). Dislocations in French–English bilingual children: An elicitation study. *Bilingualism: Language and Cognition*, 19(5), 987-1000. <https://doi.org/10.1017/S1366728915000401>
- Himmelman, N. P., & Primus, B. (2015). *Prominence Beyond Prosody—A First Approximation*. In A. De Dominicis (Ed.), *Proceedings of the pS-prominenceS International Conference* (pp. 38-58). DISUCOM Press.
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1-27. <https://doi.org/10.1017/S0305000910000759>
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120(4), 901-931. <https://doi.org/10.1016/j.lingua.2009.06.004>
- Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, 7(1), 96-130. <https://doi.org/10.1075/lab.14027.hop>
- Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, 8(1), 235-256. <https://doi.org/10.1146/annurev-linguistics-030821-054113>
- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), 227-244. <https://doi.org/10.1017/S1366728900000353>

- Idescat, & Language Policy Secretariat. (2018). *Survey on Language Uses of the Population*. <https://www.idescat.cat/pub/?id=eulp&lang=en>
- Institut d'Estudis Catalans. (2016). *Gramàtica de la llengua catalana*. IEC. <https://giec.iec.cat/>
- Iraola Azpiroz, M. (2014). *Anaphora resolution in basque: An experimental study on the interpretation of null and overt subject pronouns by children and adults* [Doctoral dissertation]. Universidad del País Vasco - Euskal Herriko Unibertsitatea.
- Iraola Azpiroz, M., Santesteban, M., Sorace, A., & Ezeizabarrena, M. J. (2017). Pronoun preferences of children in a language without typical third-person pronouns. *First Language, 37*(2), 168-185. <https://doi.org/10.1177/0142723716679799>
- Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica, 35*(1)1.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: Tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience, 29*(7), 877-892. <https://doi.org/10.1080/01690965.2013.804941>
- Järvikivi, J., van Gompel, R. P. G., & Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research, 46*(3), 525-550. <https://doi.org/10.1007/s10936-016-9451-1>
- Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science, 16*(4), 260-264. <https://doi.org/10.1111/j.0956-7976.2005.01525.x>
- Jegerski, J. (2018). Sentence processing in Spanish as a heritage language: A self-paced reading study of relative clause attachment. *Language Learning, 68*(3), 598-634. <https://doi.org/10.1111/lang.12289>
- Jegerski, J., Keating, G. D., & VanPatten, B. (2016). On-line relative clause attachment strategy in heritage speakers of Spanish. *International Journal of Bilingualism, 20*(3), 254-268. <https://doi.org/10.1177/1367006914552288>

- Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, 27(4). <https://doi.org/10.1177/0267658311406033>
- Jiménez-Gaspar, A. (2020). *Bilingualism and language change: The production of the pronominal clitics and the differential object marking in Majorcan Catalan and Spanish* [Doctoral dissertation, Universitat de les Illes Balears]. RepositoriUIB. <http://hdl.handle.net/11201/157726>
- Jiménez-Gaspar, A., Pires, A., & Guijarro-Fuentes, P. (2020a). Transfer and convergence between Catalan and Spanish in a bilingual setting. In L. A. Ortiz López, R. E. Guzzardo Tamargo, & M. González-Rivera (Eds.), *Hispanic Contact Linguistics: Theoretical, methodological and empirical perspectives* (pp. 179-212). John Benjamins. <https://doi.org/10.1075/iHLL.22.08jim>
- Jiménez-Gaspar, A., Pires, A., & Guijarro-Fuentes, P. (2020b). Bilingualism and language change: The case of pronominal clitics in Catalan and Spanish. *International Journal of Bilingual Education and Bilingualism*, 23(2), 113-131. <https://doi.org/10.1080/13670050.2017.1333487>
- Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, 26(10), 1625-1666. <https://doi.org/10.1080/01690965.2010.523082>
- Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5), 709-748. <https://doi.org/10.1080/01690960701771220>
- Káldi, T., & Babarczy, A. (2021). Linguistic focus guides attention during the encoding and refreshing of Working Memory content. *Journal of Memory and Language*, 116. <https://doi.org/10.1016/j.jml.2020.104187>
- Kaltsa, M., Tsimpli, I., & Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua*, 164, 266-288. <https://doi.org/10.1016/j.lingua.2015.06.002>
- Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2022). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition*, 25(1), 29-41. <https://doi.org/10.1017/S1366728921000390>

- Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 37(1). <https://doi.org/10.1017/S0272263114000187>
- Keating, G. D., Jegerski, J., & Vanpatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, 19(1). <https://doi.org/10.1017/S1366728914000418>
- Keating, G. D., VanPatten, B., & Jegerski, J. (2011). Who was walking on the beach?: Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition*, 33(2). <https://doi.org/10.1017/S0272263110000732>
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1). <https://doi.org/10.1093/jos/ffm018>
- Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39(1-2), 1-37. <https://doi.org/10.1515/tl-2013-0001>
- Kiss, K. É. (1998). Identificational focus versus information focus. *Language*, 74(2), 245-273. <https://doi.org/10.1353/lan.1998.0211>
- Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, 28(8), 1169-1206. <https://doi.org/10.1080/01690965.2012.699076>
- Kraš, T. (2008). Anaphora resolution in near-native Italian grammars: Evidence from native speakers of Croatian. *EUROSLA Yearbook*, 8(1), 107-134. <https://doi.org/10.1075/eurosla.8.08kra>
- Kraš, T. (2016). Cross-linguistic influence at the discourse–syntax interface: Insights from anaphora resolution in child second language learners of Italian. *International Journal of Bilingualism*, 20(4), 369-385. <https://doi.org/10.1177/1367006915609239>
- Kremin, L. V., & Byers-Heinlein, K. (2021). Why not both? Rethinking categorical and continuous approaches to bilingualism. *International Journal of Bilingualism*, 25(6), 1560-1575. <https://doi.org/10.1177/13670069211031986>
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4), 243-276. <https://doi.org/10.1556/aling.55.2008.3-4.2>

- Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25(5), 497-514. <https://doi.org/10.1080/20445911.2013.799170>
- Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1(1), 377-394. <https://doi.org/10.1146/annurev-linguist-030514-124937>
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 56, pp. 229-262). Academic Press. <https://doi.org/10.1016/B978-0-12-394393-4.00007-8>
- Kroll, J. F., & Navarro-Torres, C. A. (2018). Bilingualism. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1-29). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119170174.epcn307>
- Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531-553). Oxford University Press.
- Kupisch, T. (2007). Determiners in bilingual German–Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition*, 10(1), 57. <https://doi.org/10.1017/S1366728906002823>
- Kupisch, T. (2012). Specific and generic subjects in the Italian of German–Italian simultaneous bilinguals and L2 learners. *Bilingualism: Language and Cognition*, 15(4), 736-756. <https://doi.org/10.1017/S1366728911000691>
- Kupisch, T. (2014). Adjective placement in simultaneous bilinguals (German–Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition*, 17(1), 222-233. <https://doi.org/10.1017/S1366728913000382>
- Kupisch, T., & van de Weijer, J. (2016). The role of the childhood environment for language dominance: A study of adult simultaneous bilingual speakers of German and French. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 174-194). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.009>

- La Morgia, F. (2015). Assessing the relationship between input and strength of language development: A study on Italian–English bilingual children. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 195-218). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.010>
- Lago, S., Mosca, M., & Stutter Garcia, A. (2021). The role of crosslinguistic influence in multilingual processing: Lexicon versus syntax. *Language Learning*, 71(S1), 163-192. <https://doi.org/10.1111/lang.12412>
- Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620607>
- Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics*, 39(3), 463-516. <https://doi.org/10.1515/ling.2001.021>
- Langlois, V. J., & Arnold, J. E. (2020). Print exposure explains individual differences in using syntactic but not semantic cues for pronoun comprehension. *Cognition*, 197. <https://doi.org/10.1016/j.cognition.2019.104155>
- Langsford, S., Perfors, A., Hendrickson, A. T., Kennedy, L. A., & Navarro, D. J. (2018). Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa: A Journal of General Linguistics*, 3(1). <https://doi.org/10.5334/gjgl.396>
- Leal, T., Destruel, E., & Hoot, B. (2018). The realization of information focus in monolingual and bilingual native Spanish. *Linguistic Approaches to Bilingualism*, 8(2), 217-251. <https://doi.org/10.1075/lab.16009.lea>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2). <https://doi.org/10.3758/s13428-011-0146-0>
- Leonetti, M. (2016). On non-focal fronting in Italian and Spanish. In A.M. De Cesare & D. Garassino (Eds.), *Current Issues in Italian, Romance and Germanic Non-canonical Word Orders: Syntax Information Structure Discourse Organization* (pp. 15-36). Peter Lang.
- Leonetti, M. (2017). Basic constituent orders. In A. Duffer & E. Stark (Eds.), *Manual of Romance Morphosyntax and Syntax* (pp. 887-932). De Gruyter. <https://doi.org/10.1515/9783110377088-024>
- Leonetti, M. (2021). Topics and the interpretation of referential null subjects. In C. Gianollo, K. von Heusinger, & M. Napoli (Eds.), *Determiners and quantifiers: Functions, variation and change* (pp. 94-129). Brill.

- Leonetti-Escandell, V., & Torregrossa, J. (under review). *The interpretation of null and overt subject pronouns in Spanish compared to Greek and Italian: The role of VSO and DOM*.
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, 38(2), 202-210. <https://doi.org/10.3758/BF03192770>
- Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, 17(3), 673-680. <https://doi.org/10.1017/S1366728913000606>
- Liceras, J. M., & Alba de la Fuente, A. (2015). Typological proximity in L2 acquisition: The Spanish non-native grammar of French speakers. In T. Judy & S. Perpiñán (Eds.), *The Acquisition of Spanish in Understudied Language Pairings*. John Benjamins.
- López, L. (2009). *A derivational syntax for information structure*. Oxford University Press.
- López Otero, J. C. (2022). Bidirectional cross-linguistic influence on DOM in Romanian-Spanish bilinguals. *International Journal of Bilingualism*. <https://doi.org/10.1177/13670069211063606>
- Lozano, C. (2006). The development of the syntax-discourse interface: Greek learners of Spanish. In V. Torrens & L. Escobar (Eds.), *The acquisition of syntax in romance languages* (pp. 371-399). John Benjamins. <https://digibug.ugr.es/handle/10481/22166>
- Lozano, C. (2009). Pronominal deficits at the interface: New data from the CEDEL2 corpus. In C. M. Bretones Callejas (Ed.), *Understanding Language and Mind: Applied Linguistics Now* (pp. 213-227). Universidad de Almería.
- Lozano, C. (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus. In M. Alonso Ramos (Ed.), *Spanish Learner Corpus Research: State of the Art and Perspectives* (pp. 236-265). John Benjamins.
- Lubbers-Quesada, M., & Blackwell, S. E. (2009). The L2 acquisition of null and overt Spanish subject pronouns: A pragmatic approach. In J. Collentine (Ed.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 117-130). Cascadilla Proceedings Project.

- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605-621. <https://doi.org/10.1080/20445911.2013.795574>
- Mackey, W. F. (1962). The description of bilingualism. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 7(2), 51-85. <https://doi.org/10.1017/S0008413100019393>
- Madeira, A., Fiéis, A., & Teixeira, J. (2021). Microvariação na resolução de sujeitos pronominais: Português europeu vs. italiano. *Revista da Associação Portuguesa de Linguística*, 8. <https://doi.org/10.26334/2183-9077/rapln8ano2021a11>
- Margaza, P., & Bel, A. (2006). Null subjects at the syntax-pragmatics interface: Evidence from Spanish interlanguage of Greek speakers. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference: The Banff Conference* (pp. 88-97). Cascadilla Proceedings Project. <http://www.lingref.com/cpp/gasla/8/abstract1491.html>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4). [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97-115. <https://doi.org/10.1017/S1366728903001068>
- Martín-Villena, F., & Lozano, C. (2020). Anaphora resolution in topic continuity: Evidence from L1 English–L2 Spanish data in the CEDEL2 corpus. In J. Ryan & P. Crosthwaite (Eds.), *Referring in a Second Language: Studies on Reference to Person in a Multilingual World* (pp. 119-141). Routledge.
- Martín-Villena, F., Lozano, C., & Sorace, A. (2021, mayo). *Offline interpretation of null and overt subject pronouns in L1 Spanish: Investigating the role of temporal subordinating conjunctions*. International Conference on the Acquisition and Processing of Reference and Anaphora Resolution, Brussels.
- Matras, Y. (2009). *Language contact*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809873>

- Mayol, L. (2010). Refining salience and the Position of Antecedent Hypothesis: A study of Catalan pronouns. *University of Pennsylvania Working Papers in Linguistics*, 16(1). <https://repository.upenn.edu/pwpl/vol16/iss1/15>
- Mayol, L. (2018). Asymmetries between interpretation and production in Catalan pronouns. *Dialogue & Discourse*, 9(2), 1-34. <https://doi.org/10.5087/dad.2018.201>
- Mayol, L., & Clark, R. (2010). Pronouns in Catalan: Games of partial information and the use of linguistic resources. *Journal of pragmatics*, 42(3), 781-799. <https://doi.org/10.1016/j.pragma.2009.07.004>
- McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1040-1052. <https://doi.org/10.1037/0278-7393.19.5.1040>
- Meir, N., Walters, J., & Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism*, 7(5), 514-553. <https://doi.org/10.1075/lab.15007.mei>
- Meisel, J. M. (1989). Early differentiation of languages in bilingual children. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss* (pp. 13-40). Cambridge University Press.
- Meisel, J. M. (2001). The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammars. In J. Cenoz & F. Genesee (Eds.), *Trends in Bilingual Acquisition* (pp. 11-41). John Benjamins Publishing Company.
- Meisel, J. M. (2007). The weaker language in early child bilingualism: Acquiring a first language as a second language? *Applied Psycholinguistics*, 28(3), 495-514. <https://doi.org/10.1017/S0142716407070270>
- Meisel, J. M. (2009). Second Language Acquisition in Early Childhood. *Zeitschrift Für Sprachwissenschaft*, 28(1), 5-34. <https://doi.org/10.1515/ZFSW.2009.002>
- Meisel, J. M. (2011). Bilingual language acquisition and theories of diachronic change: Bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition*, 14(2), 121-145. <https://doi.org/10.1017/S1366728910000143>
- Meisel, J. M. (2021). Diversity and divergence in bilingual acquisition. *Zeitschrift Für Sprachwissenschaft*, 40(1), 65-88. <https://doi.org/10.1515/zfs-2021-2025>

- Mishra, R., & Abutalebi, J. (2020). Introduction to the special issue on cognitive consequences of bilingualism. *Journal of Cultural Cognitive Science*, 4(2), 123-125. <https://doi.org/10.1007/s41809-020-00070-5>
- Montalbetti, M. (1984). *After binding. On the interpretation of pronouns*. [Doctoral dissertation]. Massachusetts Institute of Technology.
- Montrul, S. (2016). Dominance and proficiency in early and late bilingualism. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 15-35). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.002>
- Montrul, S., & Rodríguez-Louro, C. (2006). Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In V. Torrens & L. Escobar (Eds.), *The Acquisition of Syntax in Romance Languages* (pp. 401-418). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.41.19mon>
- Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, 16(4), 484-500. <https://doi.org/10.1177/1367006911429518>
- Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4(1), 1-21. <https://doi.org/10.1017/S1366728901000116>
- Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*, 16(4), 709-730. <https://doi.org/10.1017/S1366728912000727>
- Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 912-918. <https://doi.org/10.1037/0096-1523.31.5.912>
- Newman, M., Patiño-Santos, A., & Trenchs-Parera, M. (2013). Linguistic reception of Latin American students in Catalonia and their responses to educational language policies. *International Journal of Bilingual Education and Bilingualism*, 16(2), 195-209. <https://doi.org/10.1080/13670050.2012.720669>
- Nicoladis, E. (2006). Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), 15-32. <https://doi.org/10.1017/S136672890500235X>

- Olarrea, A. (2012). Word order and information structure. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The Handbook of Hispanic Linguistics* (pp. 603-628). John Wiley and Sons. <https://doi.org/10.1002/9781118228098.ch28>
- Ordóñez, F. (1998). Post-verbal asymmetries in Spanish. *Natural Language & Linguistic Theory*, 16(2), 313-345. <https://doi.org/10.1023/A:1006051703562>
- 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.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445-449. <https://doi.org/10.1111/1467-9280.00383>
- Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 25(4), 501-528. <https://doi.org/10.1017/S0272263103000214>
- Papadopoulou, D., Peristeri, E., Plemenou, E., Marinis, T., & Tsimpli, I. (2015). Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua*, 155, 98-120. <https://doi.org/10.1016/j.lingua.2014.09.006>
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213-237. <https://doi.org/10.1075/lab.1.3.01par>
- Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language*, 30(2), 371-393. <https://doi.org/10.1017/S0305000903005609>
- Patterson, C., Esaulova, Y., & Felser, C. (2017). The impact of focus on pronoun resolution in native and non-native sentence comprehension. *Second Language Research*, 33(4), 403-429. <https://doi.org/10.1177/0267658317697786>
- Patterson, C., & Felser, C. (2020). Cleft focus and antecedent accessibility: The emergence of the anti-focus effect. In A. Holler, K. Suckow, & I. de la Fuente (Eds.), *Information Structuring in Discourse* (Vol. 40, pp. 56-85). Brill. https://doi.org/10.1163/9789004436725_004
- Patterson, C., & Schumacher, P. B. (2021). Interpretation preferences in contexts with three antecedents: Examining the role of prominence in German pronouns.

- Applied Psycholinguistics*, 42(6), 1427-1461.
<https://doi.org/10.1017/S0142716421000291>
- Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, 23(2), 190-214. <https://doi.org/10.1093/applin/23.2.190>
- Perpiñán, S. (2017). Catalan-Spanish bilingualism continuum: The expression of non-personal Catalan clitics in the adult grammar of early bilinguals. *Linguistic Approaches to Bilingualism*, 7(5), 477-513.
<https://doi.org/10.1075/lab.15004.per>
- Perpiñán, S. (2018). On convergence, ongoing language change, and crosslinguistic influence in direct object expression in Catalan–Spanish bilingualism. *Languages*, 3(2). <https://doi.org/10.3390/languages3020014>
- Perpiñán, S., & Soto-Corominas, A. (2021). Indirect structural crosslinguistic influence in early Catalan–Spanish bilinguals in adulthood: Predicate selection in Catalan existential constructions. *Applied Psycholinguistics*, 42(6), 1463-1502.
<https://doi.org/10.1017/S0142716421000308>
- Picallo, M. C. (1998). *On the Extended Projection Principle and null expletive subjects*. 10(2), 219-242. <https://doi.org/10.1515/prbs.1998.10.2.219>
- Pliatsikas, C. (2019). Multilingualism and brain plasticity. In J. W. Schwieter & M. Paradis (Eds.), *The Handbook of the Neuroscience of Multilingualism* (pp. 230-251). Wiley. <https://doi.org/10.1002/9781119387725.ch11>
- Polinsky, M. (2008). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal*, 6(1), 40-71.
<https://doi.org/10.46538/hlj>
- Portele, Y., & Bader, M. (2020). Coherence and the interpretation of personal and demonstrative pronouns in German. In A. Holler, K. Suckow, & I. de la Fuente (Eds.), *Information Structuring in Discourse* (pp. 24-55). Brill.
https://doi.org/10.1163/9789004436725_003
- Prentza, A., & Tsimpli, I. (2013). The interpretability of features in second language acquisition: Evidence from null and postverbal subjects in L2 English. *Journal of Greek Linguistics*, 13(2), 323-365. <https://doi.org/10.1163/15699846-13130204>
- Prince, E. (1981). Towards a taxonomy of given/new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-255). Academic Press.
- Puig-Mayenco, E., Cunnings, I., Bayram, F., Miller, D., Tubau, S., & Rothman, J. (2018). Language dominance affects bilingual performance and processing outcomes in adulthood. *Frontiers in Psychology*, 9, 1-16.

- Puig-Mayenco, E., Rothman, J., & Tubau, S. (2020). Language dominance in the previously acquired languages modulates rate of third language (L3) development over time: A longitudinal study. *International Journal of Bilingual Education and Bilingualism*, 1-24.
<https://doi.org/10.1080/13670050.2020.1792408>
- Pujolar, J. (2020). Migration in Catalonia: Language and Diversity in the Global Era. In J. A. Argenter & J. Lüdtke (Eds.), *Manual of Catalan Linguistics* (pp. 723-738). De Gruyter. <https://doi.org/10.1515/9783110450408-032>
- Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, 57, 5-16. <https://doi.org/10.1027/1618-3169/a000002>
- Quesada, T. (2021). *Studies on Anaphora Resolution in L1 Spanish – L2 English & L1 English – L2 Spanish adult learners: Combining corpus and experimental methods* [Doctoral dissertation, Universidad de Granada]. DIGIBUG.
<https://digibug.ugr.es/handle/10481/72052>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramírez, M., & Simonet, M. (2018). Language dominance and the perception of the Majorcan Catalan /ʎ-/ɟ/ contrast: Asymmetrical phonological representations. *International Journal of Bilingualism*, 22(6), 638-652.
<https://doi.org/10.1177/1367006916688334>
- Reinhart, T. (1982). Pragmatics and linguistics: An analysis of sentence topic. *Philosophica*, 27, 53-94.
- Renwick, M. E. L., & Nadeu, M. (2019). A survey of phonological mid vowel intuitions in central Catalan. *Language and Speech*, 62(1), 164-204.
<https://doi.org/10.1177/0023830917749275>
- Rinke, E., & Flores, C. (2018). Another look at the interpretation of overt and null pronominal subjects in bilingual language acquisition: Heritage Portuguese in contact with German and Spanish. *Glossa: A Journal of General Linguistics*, 3(1). <https://doi.org/10.5334/gjgl.535>
- Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, 26(6), 1850-1869.
<https://doi.org/10.3758/s13423-019-01634-5>

- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar: Handbook in Generative Syntax* (pp. 281-337). Springer Netherlands. https://doi.org/10.1007/978-94-011-5420-8_7
- Rizzi, L. (2018). Subjects, topics and the interpretation of pro. In R. Petrosino, P. Cerrone, & H. van der Hulst (Eds.), *From Sounds to Structures*. De Gruyter. <https://doi.org/10.1515/9781501506734-019>
- Rizzi, L. & Bocci, G. 2017. Left periphery of the clause: Primarily illustrated for Italian. In M. Everaert & H. C. van Riemsdijk (Eds.), *The Wiley Blackwell Companion to Syntax*, 1-30. <http://doi.org/10.1002/9781118358733.wbsyncom104>
- Rodríguez-Ordóñez, I., & Sainzmaza-Lecanda, L. (2018). Bilingualism effects in Basque subject pronoun expression: Evidence from L2 Basque. *Linguistic Approaches to Bilingualism*, 8(5), 523-560. <https://doi.org/10.1075/lab.16024.rod>
- Romano, F. B. (2019). Remarks on research of anaphora resolution in situations of language contact: Cross-linguistic influence and the PAS. *International Journal of Bilingualism*, 23(1). <https://doi.org/10.1177/1367006917693410>
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2), 155-163. <https://doi.org/10.1177/1367006909339814>
- Roussou, A., & Tsimpli, I. (2006). On Greek VSO again! *Journal of Linguistics*, 42(2), 317-354. <https://doi.org/10.1017/S0022226706003914>
- Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, 121(2), 192-218. <https://doi.org/10.1037/0033-2909.121.2.192>
- Sánchez, L. (2003). *Quechua-Spanish Bilingualism: Interference and Convergence in Functional Categories*. John Benjamins Publishing Company. <https://doi.org/10.1075/lald.35>
- Sánchez, L. (2004). Functional convergence in the tense, evidentiality and aspectual systems of Quechua Spanish bilinguals. *Bilingualism: Language and Cognition*, 7(2), 147-162. <https://doi.org/10.1017/S136672890400149X>
- Sánchez, L. (2015). Crosslinguistic influences in the mapping of functional features in Quechua-Spanish Bilingualism. In T. Judy & S. Perpiñán (Eds.), *The Acquisition of Spanish in Understudied Language Pairings* (pp. 21-48). John Benjamins. <https://doi.org/10.1075/ihll.3.02san>

- Schimke, S., & Colonna, S. (2016). Native and nonnative interpretation of pronominal forms: Evidence from French and Turkish. *Studies in Second Language Acquisition*, 38(1), 131-162. <https://doi.org/10.1017/S0272263115000303>
- Schimke, S., de la Fuente, I., Hemforth, B., & Colonna, S. (2018). First language influence on second language offline and online ambiguous pronoun resolution. *Language Learning*, 68(3), 744-779. <https://doi.org/10.1111/lang.12293>
- Schmeißer, A., Hager, M., Arnaus Gil, L., Jansen, V., Geveler, J., Eichler, N., Patuto, M., & Müller, N. (2016). Related but different: The two concepts of language dominance and language proficiency. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 36-65). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.003>
- Schumacher, P. B., Dangl, M., & Uzun, E. (2016). Thematic role as prominence cue during pronoun resolution in German. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 213-240). De Gruyter. <https://doi.org/10.1515/9783110464108-011>
- Schumacher, P. B., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua*, 185, 25-41. <https://doi.org/10.1016/j.lingua.2016.07.004>
- Schütze, C. T., & Sprouse, J. (2014). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27-50). Cambridge University Press.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240-255. <https://doi.org/10.1016/j.jml.2004.11.001>
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3(1), 3-25. <https://doi.org/10.1075/lab.3.1.01ser>
- Serratrice, L. (2016). Cross-linguistic influence, cross-linguistic priming and the nature of shared syntactic structures. *Linguistic Approaches to Bilingualism*, 6(6), 822-827. <https://doi.org/10.1075/lab.6.6.15ser>
- Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2009). Bilingual children's sensitivity to specificity and genericity: Evidence from metalinguistic awareness.

- Bilingualism: Language and Cognition*, 12(2), 239-257.
<https://doi.org/10.1017/S1366728909004027>
- Shin, N. L., & Cairns, H. S. (2012). The development of NP selection in school-age children: Reference and Spanish subject pronouns. *Language Acquisition*, 19(1), 3-38. <https://doi.org/10.1080/10489223.2012.633846>
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford University Press.
- Silva-Corvalán, C. (2008). The limits of convergence in language contact. *Journal of Language Contact*, 2(1), 213-224.
<https://doi.org/10.1163/000000008792525246>
- Silva-Corvalán, C., & Treffers-Daller, J. (Eds.). (2016). *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345>
- Simonet, M. (2011). Production of a Catalan-specific vowel contrast by early Spanish-Catalan bilinguals. *Phonetica*, 68(1-2), 88-110.
<https://doi.org/10.1159/000328847>
- Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, 43, 26-37.
<https://doi.org/10.1016/j.wocn.2014.01.004>
- Slabakova, R. (2013). Adult second language acquisition: A selective overview with a focus on the learner linguistic system. *Linguistic Approaches to Bilingualism*, 3(1), 48-72. <https://doi.org/10.1075/lab.3.1.03sla>
- Slabakova, R., White, L., & Brambatti Guzzo, N. (2017). Pronoun interpretation in the second language: Effects of computational complexity. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01236>
- Solà Pujols, J. (1992). *Agreement and subjects* [Doctoral dissertation]. Universitat Autònoma de Barcelona.
- Soler-Carbonell, J., Gallego-Balsà, L., & Corona, V. (2016). Language and education issues in global Catalonia. Questions and debates across scales of time and space. *Language, Culture and Curriculum*.
<https://doi.org/10.1080/07908318.2016.1132647>
- Solsona-Puig, J., Sansó Galiay, C., Rodríguez-Valls, F., & Janés Carulla, J. (2021). Plurilingualism within the global village: A comparative analysis of California and Catalonia attempts for linguistic equity. *Review of Education, Pedagogy, and Cultural Studies*, 1-29. <https://doi.org/10.1080/10714413.2021.1999142>

- Sorace, A. (2000). Syntactic optionality in non-native grammars. *Second Language Research*, 16(2), 93-102. <https://doi.org/10.1191/026765800670666032>
- Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(2). <https://doi.org/10.1017/S1366728904001543>
- Sorace, A. (2005). Syntactic optionality at interfaces. *Syntax and Variation: Reconciling the Biological and the Social*, 46-111.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1). <https://doi.org/10.1075/lab.1.1.01sor>
- Sorace, A. (2012). Pinning down the concept of interface in bilingual development: A reply to peer commentaries. *Linguistic Approaches to Bilingualism*, 2(2), 209-217. <https://doi.org/10.1075/lab.2.2.04sor>
- Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguistic Approaches to Bilingualism*, 6(5), 669-684. <https://doi.org/10.1075/lab.15055.sor>
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339-368. <https://doi.org/10.1191/0267658306sr271oa>
- Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11), 1497-1524. <https://doi.org/10.1016/j.lingua.2004.07.002>
- Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195-210. <https://doi.org/10.1177/1367006909339810>
- Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119(3), 460-477. <https://doi.org/10.1016/j.lingua.2008.09.008>
- Soto-Corominas, A. (2018). *Acquisition of Catalan and Spanish morphosyntax in the Catalan-Spanish bilingual context* [Doctoral dissertation, The University of Western Ontario]. Electronic Thesis and Dissertation Repository. <https://ir.lib.uwo.ca/etd/5441>
- Soto-Corominas, A. (2021). Acquisition of quantified partitivity in Catalan-Spanish bilingualism: Influence from child-level and language-level factors. *Linguistic Approaches to Bilingualism*, 11(6), 873-908. <https://doi.org/10.1075/lab.19016.sot>

- Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics*, 2(1). <https://doi.org/10.5334/gjgl.236>
- Stevenson, R., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9(4), 519-548. <https://doi.org/10.1080/01690969408402130>
- Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3), 225-262. <https://doi.org/10.1080/016909600386048>
- Strubell, M. (1996). Language planning and bilingual education in Catalonia. *Journal of Multilingual and Multicultural Development*, 17(2-4), 262-275. <https://doi.org/10.1080/01434639608666279>
- Strubell, M., & Boix-Fuster, E. (2011). *Democratic policies for language revitalisation: The case of Catalan*. Springer.
- Suïls, J., & Huguet, À. (2001). The Occitan speech community of the Aran valley. In M. T. Turell (Ed.), *Multilingualism in Spain* (pp. 141-164). Multilingual Matters.
- Tarrés Larrègola, M. (2021). *L'adquisició de les propietats morfosintàctiques i discursives dels pronoms clítics en català L2: Un estudi en infants lusòfons i francòfons* [Doctoral dissertation, Universitat Pompeu Fabra]. TDX. <http://www.tdx.cat/handle/10803/671268>
- Teixeira, J. (2020). Gradient optionality in L2 acquisition at the syntax-discourse interface: Evidence from inversion in advanced and near-native English. *Lingua*, 245. <https://doi.org/10.1016/j.lingua.2020.102947>
- Teixeira, J., Fiéis, A., & Madeira, A. (2022). Are non-native speakers sensitive to microvariation in anaphora resolution? The case of Italian learners of European Portuguese. In Y. Gong & F. Kpogo (Eds.), *Proceedings of the 46th annual Boston University Conference on Language Development* (pp. 783-796). Cascadilla Press.
- Thomason, S. G. (2001). *Language Contact: An Introduction*. Edinburgh University Press and Georgetown University Press.
- Thomason, S. G. (2020). Contact explanations in linguistics. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 31-49). Wiley. <https://doi.org/10.1002/9781119485094.ch1>

- Thomason, S. & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
<https://doi.org/10.1525/9780520912793>
- Timmer, K., Costa, A., & Wodniecka, Z. (2021). The source of attention modulations in bilingual language contexts. *Brain and Language*, 223.
<https://doi.org/10.1016/j.bandl.2021.105040>
- Torregrossa, J., Andreou, M., & Bongartz, C. M. (2020). Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa: A Journal of General Linguistics*, 5(1). <https://doi.org/10.5334/gjgl.1011>
- Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. (2021). Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, 24(4), 694-706.
<https://doi.org/10.1017/S1366728920000826>
- Treffers-Daller, J. (2016). Language dominance: The construct, its measurement, and operationalization. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 235-265). Cambridge University Press.
<https://doi.org/10.1017/CBO9781107375345.012>
- Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, 5(1), 375-393. <https://doi.org/10.1146/annurev-linguistics-011817-045554>
- Tsimpli, I., & Mastropavlou, M. (2007). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The Role of Formal Features in Second Language Acquisition*. Lawrence Erlbaum.
- Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257-277.
<https://doi.org/10.1177/13670069040080030601>
- Unsworth, S. (2016a). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 156-173). Cambridge University Press.
<https://doi.org/10.1017/CBO9781107375345.008>

- Unsworth, S. (2016b). Quantity and quality of language input in bilingual language development. In E Nicoladis and S Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103-121). De Gruyter Mouton. <https://doi.org/10.1037/14939-007>
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), 765-805. <https://doi.org/10.1017/S0142716412000574>
- Unsworth, S., Chondrogianni, V., & Skarabela, B. (2018). Experiential Measures Can Be Used as a Proxy for Language Dominance in Bilingual Language Acquisition Research. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01809>
- Vallduví, E. (1992). *The informational component*. Garland.
- Vallduví, E. (1994). The dynamics of information packaging. In E. Engdahl (Ed.), *Integrating information structure into constraint-based and categorial approaches* (pp. 1-26). Institute for Logic, Language and Computation.
- Vallduví, E. (2002). L'oració com a unitat informativa. In J. Solà, R. M. Lloret, J. Mascaró, & M. Pérez-Saldanya (dir.), *Gramàtica del català contemporani* (Vol. 2, pp. 1221-1279). Empúries.
- Vallduví, E., & Engdahl, E. (1996). *The linguistic realization of information packaging*. 34(3), 459-520. <https://doi.org/10.1515/ling.1996.34.3.459>
- Van Coetsem, F. (1988). *Loan Phonology and the Two Transfer Types in Language Contact*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110884869>
- van Dijk, C., van Wonderen, E., Koutamanis, E., Kootstra, G. J., Dijkstra, T., & Unsworth, S. (2021). Cross-linguistic influence in simultaneous and early sequential bilingual children: A meta-analysis. *Journal of Child Language*, 1-33. <https://doi.org/10.1017/S0305000921000337>
- van Dijk, T. A. (1977). Sentence topic and discourse topic. *Papers in Slavic Philology*, 1, 49-61. <https://doi.org/10.1515/9783110826142.177>
- Vanrell, M. del M., & Fernández-Soriano, O. (2018). Language variation at the prosody-syntax interface. In M. García García & M. Uth (Eds.), *Focus Realization in Romance and Beyond* (pp. 33-70). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.201.02van>
- Vanrell, M. del M., & Fernández-Soriano, O. M. (2013). Variation at the Interfaces in Ibero-Romance: Catalan and Spanish prosody and word order. *Catalan Journal of Linguistics*, 12, 253-282.

- Vila, F. X. (2020a). Language Demography. In J. A. Argenter & J. Lüdtke (Eds.), *Manual of Catalan Linguistics* (pp. 629-648). De Gruyter.
<https://doi.org/10.1515/9783110450408-026>
- Vila, F. X. (2020b). Teaching and Learning of Catalan. In J. A. Argenter & J. Lüdtke (Eds.), *Manual of Catalan Linguistics* (pp. 669-682). De Gruyter.
<https://doi.org/10.1515/9783110450408-028>
- Villalba, X. (2007). La dislocació a la dreta en català i castellà: Microvariació en la interfície sintaxi/pragmàtica. *Caplletra. Revista Internacional de Filologia*, 42.
- von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, 154, 117-127.
<https://doi.org/10.1016/j.pragma.2019.07.025>
- Wei, L. (2007). Dimensions of bilingualism. In L. Wei (Ed.) *The Bilingualism Reader* (pp. 2-21). Routledge.
- Weinreich, U. (1953). *Languages in contact, findings and problems*. Linguistic Circle of New York.
- Weiss, D. J., Schwob, N., & Lebkuecher, A. L. (2020). Bilingualism and statistical learning: Lessons from studies using artificial languages. *Bilingualism: Language and Cognition*, 23(1), 92-97.
<https://doi.org/10.1017/S1366728919000579>
- Westergaard, M. (2021). L3 acquisition and crosslinguistic influence as co-activation: Response to commentaries on the keynote 'Microvariation in multilingual situations: The importance of property-by-property acquisition'. *Second Language Research*, 37(3), 501-518.
<https://doi.org/10.1177/02676583211007897>
- Williams, E. D. (2020). *Language experience predicts pronoun comprehension in implicit causality sentences* [Doctoral dissertation, The University of North Carolina at Chapel Hill]. Carolina Digital Repository.
<https://doi.org/10.17615/0x4r-rq30>
- Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications* (arXiv:1308.5499). <https://doi.org/10.48550/arXiv.1308.5499>
- Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge.
<https://doi.org/10.4324/9781315165547>
- Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun resolution. *Language and Cognitive Processes*, 19(6), 665-675.
<https://doi.org/10.1080/01690960444000034>

- Wolna, A., Durlík, J., & Wodniecka, Z. (2022). Pronominal anaphora resolution in Polish: Investigating online sentence interpretation using eye-tracking. *Plos One*, 17(1), e0262459. <https://doi.org/10.1371/journal.pone.0262459>
- Woolard, K. A. (2016). *Singular and plural: Ideologies of linguistic authority in 21st century Catalonia*. Oxford University Press.
- Woolard, K. A., & Frekko, S. E. (2013). Catalan in the twenty-first century: Romantic publics and cosmopolitan communities. *International Journal of Bilingual Education and Bilingualism*, 16(2), 129-137. <https://doi.org/10.1080/13670050.2012.720663>
- Yip, V., & Matthews, S. (2000). Syntactic transfer in a Cantonese–English bilingual child. *Bilingualism: Language and Cognition*, 3(3), 193-208. <https://doi.org/10.1017/S136672890000033X>
- Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length utterance differentials. *Language Assessment Quarterly*, 3(2), 97-116. https://doi.org/10.1207/s15434311laq0302_2
- Yip, V., & Matthews, S. (2007). *The bilingual child: Early development and language contact*. Cambridge University Press.
- Zhang, Y., & Davidson, K. (2021). De re interpretation in belief reports: An experimental investigation. *Experiments in Linguistic Meaning*, 1. <https://doi.org/10.3765/elm.1.4874>
- Zirnstien, M., Bice, K., & Kroll, J. F. (2019). Variation in language experience shapes the consequences of bilingualism. In I. A. Sekerina, L. Spradlin, & V. Valian (Eds.), *Studies in Bilingualism* (Vol. 57, pp. 35-47). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.57.03zir>
- Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. MIT Press.
- Zubizarreta, M. L. (1999). Las funciones informativas: Tema y foco. In I. Bosque & V. Demonte (dir.), *Gramática descriptiva de la lengua española* (Vol. 3, pp. 4215-4244). Espasa Calpe.

Appendices

Appendix A: Language background questionnaire (based on the BLP questionnaire)

I. Información biográfica

Iniciales o código identificador proporcionado por la investigadora: _____

Edad: _____

Lugar de nacimiento: _____ / Lugar de residencia actual: _____

Si no es donde naciste, indica desde cuándo vives en el lugar actual: _____

Nivel más alto de formación académica completada:

- Menos de la escuela secundaria
- Educación secundaria
- Bachillerato
- Universidad (grado, diplomatura, licenciatura)
- Formación profesional
- Máster
- Doctorado
- Otro

¿En qué lengua/s te sientes más cómodo/a?

- Español
- Catalán
- Otra

Indica la/s lengua/s que utilizas habitualmente para hablar con:

	Catalán	Español	Otra	NA
Madre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Padre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hermano(s)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pareja	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

II. Historial lingüístico

1. ¿A qué edad empezaste a aprender las siguientes lenguas?

Español: Desde que nací 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

Catalán: Desde que nací 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

2. ¿A qué edad empezaste a sentirte cómodo usando las siguientes lenguas?

Español: Tan pronto como recuerdo 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

Catalán: Tan pronto como recuerdo 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

3a. ¿Cuántos años de clases (historia, matemáticas, etc., excluyendo las clases de lengua) has tenido en las siguientes lenguas durante la educación infantil y primaria?

Español: 0 1 2 3 4 5 6 7 8 9 10+

Catalán: 0 1 2 3 4 5 6 7 8 9 10+

3b. ¿Cuántos años de clases (historia, matemáticas, etc., excluyendo las clases de lengua) has tenido en las siguientes lenguas en los estudios secundarios y superiores (ESO, bachillerato y estudios superiores)?

Español: 0 1 2 3 4 5 6 7 8 9 10+

Catalán: 0 1 2 3 4 5 6 7 8 9 10+

4. ¿Cuántos años has pasado en un país o región donde se hablan las siguientes lenguas?

Español: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

Catalán: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

5. Cuántos años se han hablado las siguientes lenguas en tu familia?

Español: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

Catalán: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

6. ¿Cuántos años has pasado en un ambiente de trabajo donde se hablan las siguientes lenguas?

Español: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

Catalán: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

III. Uso de las lenguas

7. En una semana normal, ¿qué porcentaje del tiempo usas las siguientes lenguas con tus amigos? (la suma debe dar 100%)

Español: _____ / Catalán: _____ / Otras lenguas: _____

8. En una semana normal, ¿qué porcentaje del tiempo usas las siguientes lenguas en la escuela/en el trabajo? (la suma debe dar 100%)

Español: _____ / Catalán: _____ / Otras lenguas: _____

9. En una semana normal, ¿qué porcentaje del tiempo usas las siguientes lenguas en la escuela/en el trabajo? (la suma debe dar 100%)

Español: _____ / Catalán: _____ / Otras lenguas: _____

10. Cuando te hablas a ti mismo, ¿con qué frecuencia te hablas a ti mismo en las siguientes lenguas? (la suma debe dar 100%)

Español: _____ / Catalán: _____ / Otras lenguas: _____

11. Cuando haces cálculos, ¿con qué frecuencia cuentas en las siguientes lenguas? (la suma debe dar 100%)

Español: _____ / Catalán: _____ / Otras lenguas: _____

IV. Competencia lingüística

12. ¿Qué nivel de comprensión oral tienes en las siguientes lenguas?

(0 = no muy bueno | 6 = muy bueno)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

13. ¿Qué nivel de comprensión escrita tienes en las siguientes lenguas?

(0 = no muy bueno | 6 = muy bueno)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

14. ¿Qué nivel de expresión oral tienes en las siguientes lenguas?

(0 = no muy bueno | 6 = muy bueno)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

15. ¿Qué nivel de expresión escrita tienes en las siguientes lenguas?

(0 = no muy bueno | 6 = muy bueno)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

V. Actitudes

16. Me siento "yo mismo" cuando hablo en...

(0 = no estoy de acuerdo | 6 = estoy de acuerdo)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

17. Me identifico con una cultura...

(0 = no estoy de acuerdo | 6 = estoy de acuerdo)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

18. Es importante para mí hablar de manera adecuada y correcta el...

(0 = no estoy de acuerdo | 6 = estoy de acuerdo)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

19. Quiero que los demás piensen que no tengo acento en...

(0 = no estoy de acuerdo | 6 = estoy de acuerdo)

Español 0 1 2 3 4 5 6

Catalán 0 1 2 3 4 5 6

Appendix B: Results of the language background questionnaire

Table B.1

Results of the history module in the BLP, by language and group of bilinguals

	Catalan				Spanish			
	<i>M</i>	<i>SD</i>	Min	Max	<i>M</i>	<i>SD</i>	Min	Max
<i>Age of onset of acquisition (max. 20+ years)</i>								
Catalan-dominant	0	0	0	0	1.82	2.21	0	6
Balanced	0.77	1.36	0	4	0.1	0.54	0	3
Spanish-dominant	1.41	1.66	0	6	0	0	0	0
<i>Age when started to feel comfortable using each language (max. 20+ years)</i>								
Catalan-dominant	0.35	1.74	0	10	7.74	8.08	0	20
Balanced	1.45	3.49	0	15	1.81	4.46	0	17
Spanish-dominant	4.69	6.61	0	20	0.14	0.74	0	4
<i>Language of schooling during primary school (max. 10 years)</i>								
Catalan-dominant	9.5	0.62	8	10	0.97	2.5	0	10
Balanced	8.9	2.18	3	10	2.68	3.71	0	10
Spanish- dominant	8.45	2.4	0	10	3.14	4.14	0	10
<i>Language of schooling during secondary school and higher education (max. 10 years)</i>								
Catalan-dominant	8.09	2.17	3	10	3.09	2.7	0	10
Balanced	8.84	1.83	3	10	3.87	3.59	0	10
Spanish-dominant	6.93	2.85	0	10	4.07	3.34	0	10
<i>Years spent in a country or region using each language (max. 20+ years)</i>								
Catalan-dominant	19.65	0.6	18	20	19.12	3.08	2	20
Balanced	19.32	1.8	10	20	19.65	0.49	19	20
Spanish-dominant	18.93	3.69	1	20	19.17	3.52	1	20
<i>Years spent in a family using each language (max. 20+ years)</i>								
Catalan-dominant	19.65	0.6	18	20	4.24	7.79	0	20
Balanced	13.97	8.18	0	20	18.65	3.84	0	20
Spanish-dominant	5.21	8.17	0	20	19.83	0.38	19	20
<i>Years spent in a work environment using each language (max. 20+ years)</i>								
Catalan-dominant	6.76	7.33	0	20	3.03	4.66	0	20
Balanced	4.94	5.6	0	19	4.77	6.27	0	19
Spanish-dominant	3.38	5.27	0	20	5.38	6.38	0	20

Table B.2*Results of the use module in the BLP, by language and group of bilinguals*

	Catalan				Spanish			
	<i>M</i>	<i>SD</i>	Min	Max	<i>M</i>	<i>SD</i>	Min	Max
<i>Use with family</i>								
Catalan-dominant	98.53	3.59	90	100	1.47	4.53	0	20
Balanced	38.06	33.16	0	100	44.03	30.15	0	100
Spanish-dominant	4.48	13.32	0	70	95.69	10.83	50	100
<i>Use with friends</i>								
Catalan-dominant	80.29	18.38	30	100	1.03	2.96	0	10
Balanced	44.35	25.94	0	100	61.45	33.37	0	100
Spanish-dominant	13.45	12.75	0	50	95.52	13.32	30	100
<i>Use at school/work</i>								
Catalan-dominant	73.09	18.3	30	100	17.21	16.84	0	70
Balanced	55.65	24.62	15	100	54.52	26.28	0	100
Spanish-dominant	26.55	23.03	0	100	85.52	13.72	50	100
<i>Use when talking to oneself</i>								
Catalan-dominant	96.62	9.11	50	100	23.09	18.46	0	70
Balanced	47.9	26.04	0	100	37.1	22.05	0	80
Spanish-dominant	3.28	7.11	0	35	66.38	24.89	0	100
<i>Use when counting</i>								
Catalan-dominant	97.35	6.99	70	100	2.5	5.54	0	25
Balanced	53.39	29.28	0	100	47.1	24.93	0	90
Spanish-dominant	3.1	7.12	0	30	95.17	7.85	65	100

Table B.3*Results of the proficiency module in the BLP, by language and group of bilinguals*

	Catalan				Spanish			
	<i>M</i>	<i>SD</i>	Min	Max	<i>M</i>	<i>SD</i>	Min	Max
<i>Oral comprehension</i>								
Catalan-dominant	5.97	0.17	5	6	5.85	0.36	5	6
Balanced	5.81	0.4	5	6	5.77	0.43	5	6
Spanish-dominant	5.69	0.66	3	6	6	0	6	6
<i>Written comprehension</i>								
Catalan-dominant	5.88	0.41	4	6	5.74	0.51	4	6
Balanced	5.68	0.65	3	6	5.68	0.6	4	6
Spanish-dominant	5.62	0.62	4	6	5.97	0.19	5	6
<i>Oral expression</i>								
Catalan-dominant	5.85	0.44	4	6	4.88	0.88	3	6
Balanced	5.55	0.68	4	6	5.68	0.6	4	6
Spanish-dominant	4.86	0.95	2	6	6	0	6	6
<i>Written expression</i>								
Catalan-dominant	5.76	0.55	4	6	5.47	0.71	3	6
Balanced	5.39	0.8	3	6	5.55	0.68	4	6
Spanish-dominant	5.03	1.09	2	6	5.93	0.26	5	6

Note. From 0 to 6 (0 = not good at all | 6 = very good)

Table B.4*Results of the attitudes module in the BLP, by language and group of bilinguals*

	Catalan				Spanish			
	<i>M</i>	<i>SD</i>	Min	Max	<i>M</i>	<i>SD</i>	Min	Max
<i>Feel yourself</i>								
Catalan-dominant	6	0	6	6	2.32	1.79	0	6
Balanced	5.19	1.05	3	6	4.97	1.49	0	6
Spanish-dominant	3.03	1.5	0	6	5.9	0.41	4	6
<i>Culture</i>								
Catalan-dominant	5.79	0.59	4	6	1.56	1.44	0	6
Balanced	4.94	1.34	1	6	3.26	1.91	0	6
Spanish-dominant	4.38	1.8	0	6	4.45	1.53	0	6
<i>Correct use</i>								
Catalan-dominant	5.97	0.17	5	6	5.74	0.86	2	6
Balanced	5.61	0.8	3	6	5.39	1.05	2	6
Spanish-dominant	5.72	0.59	4	6	5.93	0.26	5	6
<i>Accent</i>								
Catalan-dominant	2.24	2.66	0	6	3.15	2.6	0	6
Balanced	1.35	1.87	0	6	1.74	2.39	0	6
Spanish-dominant	1.17	1.98	0	6	1.28	2.07	0	6

Note. From 0 to 6 (0 = disagree | 6 = agree)

Appendix C: Materials for the experimental task

Table C.1

Experimental items (in the condition of null pronoun, unmarked information structure) used in the two-alternative forced-choice task in Catalan

ID	Item and question (in Catalan)	Answer A	Answer B
FC01	La Mar va abandonar la Gal·la quan va deixar la feina. Qui va deixar la feina?	Mar	Gal·la
FC02	La Laura va alegrar la Maria quan va venir a la ciutat. Qui va venir a la ciutat?	Maria	Laura
FC03	En David va espantar en Roger quan va entrar a l'habitació. Qui va entrar a l'habitació?	Roger	David
FC04	En Jaume va deixar en Ramon quan es va mudar al Japó. Qui es va mudar al Japó?	Jaume	Ramon
FC05	L'Anna va esperar l'Alba quan va aparcar el cotxe. Qui va aparcar el cotxe?	Anna	Alba
FC06	La Mariona va evitar la Isabel quan va estar al poble. Qui va estar al poble?	Isabel	Mariona
FC07	En Xavier va formar l'Enric quan va cantar a l'òpera de Nova York. Qui va cantar a l'òpera?	Enric	Xavier
FC08	En Roc va desmentir en Nil quan va parlar de l'incendi. Qui va parlar de l'incendi?	Roc	Nil
FC09	En Mohamed va interrompre en Youssef quan va preguntar per les fotos. Qui va preguntar per les fotos?	Mohamed	Youssef
FC10	L'Isaac va investigar l'Oriol quan va treballar per al govern. Qui va treballar per al govern?	Oriol	Isaac
FC11	La Fàtima va recollir la Nàdia quan va decidir anar-se'n. Qui va decidir anar-se'n?	Nàdia	Fàtima
FC12	La Irene va saludar la Gemma quan es va asseure a la terrassa del bar. Qui es va asseure a la terrassa del bar?	Irene	Gemma
FC13	En Ferran va seguir en Kevin quan va marxar del concert. Qui va marxar del concert?	Ferran	Kevin
FC14	En Daniel va suportar en Felip quan va tornar a la festa. Qui va tornar a la festa?	Felip	Daniel

FC15	L'Abril va sorprendre la Sandra quan es va casar. Qui es va casar?	Sandra	Abril
FC16	La Chloe va veure la Lía quan va travessar el carrer. Qui va travessar el carrer?	Chloe	Lía
FC17	En Joan va abandonar en Pere quan va anar al lavabo. Qui va anar al lavabo?	Pere	Joan
FC18	L'Aniol va alegrar en Pau quan va celebrar l'aniversari. Qui va celebrar l'aniversari?	Aniol	Pau
FC19	La Bet va espantar la Tina quan va allunyar-se de la multitud. Qui va allunyar-se de la multitud?	Bet	Tina
FC20	L'Alícia va deixar l'Ariadna quan va tornar del viatge. Qui va tornar del viatge?	Ariadna	Alícia
FC21	En Josep va esperar en Carles quan va arribar a l'estació. Qui va arribar a l'estació?	Carles	Josep
FC22	En Lluç va evitar l'Albert quan va tenir problemes. Qui va tenir problemes?	Lluç	Albert
FC23	La Marta va formar la Laia quan va participar en el projecte. Qui va participar en el projecte?	Marta	Laia
FC24	L'Olga va desmentir l'Eva quan va explicar la historia. Qui va explicar la història?	Eva	Olga
FC25	La Neus va interrompre la Montse quan va començar a cuinar. Qui va començar a cuinar?	Montse	Neus
FC26	L'Àngela va investigar la Sílvia quan va invertir en l'empresa. Qui va invertir en l'empresa?	Àngela	Sílvia
FC27	En Martí va recollir en Guillem quan va sortir de l'oficina. Qui va sortir de l'oficina?	Martí	Guillem
FC28	En Marc va saludar en Gerard quan va baixar de l'escenari. Qui va baixar de l'escenari?	Gerard	Marc
FC29	L'Olívia va seguir la Carolina quan va posar-se en política. Qui va posar-se en política?	Carolina	Olívia
FC30	La Roser va suportar la Clàudia quan va ser secretària del club. Qui va ser secretària del club?	Roser	Clàudia
FC31	L'Eduard va sorprendre en Miquel quan va viatjar a Brasil. Qui va viatjar a Brasil?	Eduard	Miquel

FC32	L'Ernest va veure en Francesc quan va pujar al tren. Qui va pujar al tren?	Francesc	Ernest
FC33	L'Adela va abandonar la Teresa quan va fugir del país. Qui va fugir del país?	Adela	Teresa
FC34	L'Elsa va alegrar la Marina quan va assabentar-se de la notícia. Qui va assabentar-se de la notícia?	Marina	Elsa
FC35	L'Adrià va espantar l'Andreu quan va tancar la porta. Qui va tancar la porta?	Adrià	Andreu
FC36	En Pol va deixar en Pep quan va recuperar-se de l'accident. Qui va recuperar-se de l'accident?	Pep	Pol
FC37	En Quim va esperar l'Arnau quan va agafar l'autobús. Qui va agafar l'autobús?	Quim	Arnau
FC38	En Santi va evitar en Toni quan va demanar disculpes. Qui va demanar disculpes?	Toni	Santi
FC39	L'Èlia va formar la Judit quan va estar a la universitat. Qui va estar a la universitat?	Èlia	Judit
FC40	L'Eulàlia va desmentir l'Agnès quan va descriure la situació. Qui va descriure la situació?	Agnès	Eulàlia
FC41	En Jordi va interrompre en Sergi quan va intervenir a la conversa. Qui va intervenir a la conversa?	Jordi	Sergi
FC42	L'Amir va investigar l'Omar quan va obrir el negoci. Qui va obrir el negoci?	Omar	Amir
FC43	L'Amàlia va recollir l'Helena quan va acabar de dutxar-se. Qui va acabar de dutxar-se?	Amàlia	Helena
FC44	La Samira va saludar la Yasmina quan va passar pel forn de pa. Qui va passar pel forn de pa?	Yasmina	Samira
FC45	L'Héctor va seguir en Teo quan va amagar-se al bosc. Qui va amagar-se al bosc?	Héctor	Teo
FC46	L'Òscar va suportar l'Àxel quan va actuar al teatre. Qui va actuar al teatre?	Àxel	Òscar
FC47	La Simona va sorprendre la Indira quan va publicar els poemes. Qui va publicar els poemes?	Simona	Indira
FC48	La Margarida va veure la Fernanda quan va treure el cap per la finestra. Qui va treure el cap per la finestra?	Fernanda	Margarida

Table C.2

Experimental items (in the condition of null pronoun, unmarked information structure) used in the two-alternative forced-choice task in Spanish

ID	Item (in Spanish)	AnswerA	AnswerB
FS01	Mar abandonó a Gala cuando dejó el trabajo. ¿Quién dejó el trabajo?	Mar	Gala
FS02	Laura alegró a María cuando vino a la ciudad. ¿Quién vino a la ciudad?	María	Laura
FS03	David asustó a Nicolás cuando entró a la habitación. ¿Quién entró a la habitación?	Nicolás	David
FS04	Jaime dejó a Ramón cuando se mudó al Japón. ¿Quién se mudó al Japón?	Jaime	Ramón
FS05	Ana esperó a Alba cuando aparcó el coche. ¿Quién aparcó el coche?	Ana	Alba
FS06	Mariona evitó a Isabel cuando estuvo en el pueblo. ¿Quién estuvo en el pueblo?	Isabel	Mariona
FS07	Javier formó a Enrique cuando cantó en la ópera de Nueva York. ¿Quién cantó en la ópera?	Enrique	Javier
FS08	Leo desmintió a Hugo cuando habló del incendio. ¿Quién habló del incendio?	Leo	Hugo
FS09	Mohamed interrumpió a Youssef cuando preguntó por las fotos. ¿Quién preguntó por las fotos?	Mohamed	Youssef
FS10	Alejandro investigó a Mario cuando trabajó para el gobierno. ¿Quién trabajó para el gobierno?	Mario	Alejandro
FS11	Fátima recogió a Nadia cuando decidió irse. ¿Quién decidió irse?	Nadia	Fátima
FS12	Irene saludó a Gema cuando se sentó en la terraza del bar. ¿Quién se sentó en la terraza del bar?	Irene	Gema
FS13	Fernando siguió a Kevin cuando se marchó del concierto. ¿Quién se marchó del concierto?	Fernando	Kevin
FS14	Daniel soportó a Felipe cuando volvió a la fiesta. ¿Quién volvió a la fiesta?	Felipe	Daniel
FS15	Abril sorprendió a Sandra cuando se casó. ¿Quién se casó?	Sandra	Abril

FS16	Chloe vio a Lía cuando cruzó la calle. ¿Quién cruzó la calle?	Chloe	Lía
FS17	Juan abandonó a Pedro cuando se fue al baño. ¿Quién se fue al baño?	Pedro	Juan
FS18	Álvaro alegró a Pablo cuando celebró el cumpleaños. ¿Quién celebró el cumpleaños?	Álvaro	Pablo
FS19	Alma asustó a Vera cuando se alejó de la multitud. ¿Quién se alejó de la multitud?	Alma	Vera
FS20	Alicia dejó a Adriana cuando regresó del viaje. ¿Quién regresó del viaje?	Adriana	Alicia
FS21	José esperó a Carlos cuando llegó a la estación. ¿Quién llegó a la estación?	Carlos	José
FS22	Lucas evitó a Alberto cuando tuvo problemas. ¿Quién tuvo problemas?	Lucas	Alberto
FS23	Marta formó a Laia cuando participó en el proyecto. ¿Quién participó en el proyecto?	Marta	Laia
FS24	Olga desmintió a Eva cuando contó la historia. ¿Quién contó la historia?	Eva	Olga
FS25	Nieves interrumpió a Montse cuando empezó a cocinar. ¿Quién empezó a cocinar?	Montse	Nieves
FS26	Ángela investigó a Silvia cuando invirtió en la empresa. ¿Quién invirtió en la empresa?	Ángela	Silvia
FS27	Martín recogió a Guillermo cuando salió de la oficina. ¿Quién salió de la oficina?	Martín	Guillermo
FS28	Marcos saludó a Gerardo cuando bajó del escenario. ¿Quién bajó del escenario?	Gerardo	Marcos
FS29	Olivia siguió a Carolina cuando se metió en política. ¿Quién se metió en política?	Carolina	Olivia
FS30	Rocío soportó a Claudia cuando fue secretaria del club. ¿Quién fue secretaria del club?	Rocío	Claudia
FS31	Eduardo sorprendió a Miguel cuando viajó a Brasil. ¿Quién viajó a Brasil?	Eduardo	Miguel
FS32	Ernesto vio a Francisco cuando subió al tren. ¿Quién subió al tren?	Francisco	Ernesto

FS33	Adela abandonó a Teresa cuando huyó del país. ¿Quién huyó del país?	Adela	Teresa
FS34	Elsa alegró a Marina cuando se enteró de la noticia. ¿Quién se enteró de la noticia?	Marina	Elsa
FS35	Adrián asustó a Andrés cuando cerró la puerta. ¿Quién cerró la puerta?	Adrián	Andrés
FS36	Aitor dejó a Iker cuando se recuperó del accidente. ¿Quién se recuperó del accidente?	Iker	Aitor
FS37	Joaquín esperó a Bruno cuando cogió el autobús. ¿Quién cogió el autobús?	Joaquín	Bruno
FS38	Santiago evitó a Antonio cuando pidió disculpas. ¿Quién pidió disculpas?	Antonio	Santiago
FS39	Elia formó a Judit cuando estuvo en la universidad. ¿Quién estuvo en la universidad?	Elia	Judit
FS40	Eulalia desmintió a Inés cuando describió la situación. ¿Quién describió la situación?	Inés	Eulalia
FS41	Jorge interrumpió a Sergio cuando intervino en la conversación. ¿Quién intervino en la conversación?	Jorge	Sergio
FS42	Amir investigó a Omar cuando abrió el negocio. ¿Quién abrió el negocio?	Omar	Amir
FS43	Amalia recogió a Elena cuando acabó de ducharse. ¿Quién acabó de ducharse?	Amalia	Elena
FS44	Samira saludó a Yasmina cuando pasó por la panadería. ¿Quién pasó por la panadería?	Yasmina	Samira
FS45	Héctor siguió a Teo cuando se escondió en el bosque. ¿Quién se escondió en el bosque?	Héctor	Teo
FS46	Óscar soportó a Áxel cuando actuó en el teatro. ¿Quién actuó en el teatro?	Áxel	Óscar
FS47	Simona sorprendió a Indira cuando publicó los poemas. ¿Quién publicó los poemas?	Simona	Indira
FS48	Margarita vio a Fernanda cuando se asomó por la ventana. ¿Quién se asomó por la ventana?	Fernanda	Margarita

Appendix D: Instructions of the experimental task

Instructions at the beginning of the task:

“A continuación leerás unas frases seguidas de una pregunta con dos opciones de respuesta. Mueve la barra deslizante hacia la respuesta que consideres más adecuada, teniendo en cuenta que puedes moverla más o menos según estés más o menos seguro de la respuesta.

Ve a tu ritmo, responde cada pregunta de forma acurada pero tan rápido como te sea posible. Podrás hacer una pausa de descanso en el lugar indicado, cuando llegues a la mitad del experimento.

Ahora te proponemos unas frases de práctica para que te familiarices con la dinámica.

¡Empezamos!”

“A continuació llegiràs unes frases seguides d'una pregunta amb dues opcions de resposta. Mou la barra cap a la resposta que consideris més adequada. Tingues en compte que la pots moure més o menys cap a un dels costats en funció de si estàs més o menys segur de la resposta.

Ves al teu ritme, respon cada pregunta de manera acurada però tan ràpid com et sigui possible. Podràs fer una pausa de descans en el lloc indicat, quan arribis a la meitat de l'experiment.

Ara et proposem unes frases de pràctica perquè et familiaritzis amb la dinàmica.

Comencem!”

Instructions after the six practice items and before the experimental items:

“¿Tienes alguna duda? Recuerda que siempre puedes preguntar a la investigadora y estará encantada de ayudarte.

Es muy importante que tengas en cuenta que las valoraciones son personales y no hay respuestas correctas e incorrectas. ¡Simplemente queremos conocer tu intuición como hablante de lengua española!”

“Tens algun dubte? Recorda que en qualsevol moment pots fer preguntes a la investigadora i estarà encantada d'ajudar-te.

És molt important que tinguis en compte que les valoracions són personals i que no hi ha respostes correctes i incorrectes. Simplement volem veure quines intuïcions tens com a parlant del català!”

Appendix E: Boxplots for participants' analyses

This appendix includes the boxplots that were used for visual inspection of each participants' responses for each condition and in each language (Catalan and Spanish). Note that, in the responses (resp2), '0' corresponds to subject interpretations and '1' to object interpretations. Referring to the conditions, 'C1' corresponds to null pronouns in unmarked contexts, 'C2' to overt pronouns in unmarked contexts, 'C3' to null pronouns in subject clefts, 'C4' to overt pronouns in subject clefts, 'C5' to null pronouns in object clefts, 'C6' to overt pronouns in object clefts, 'C7' to null pronouns in clitic-left dislocation contexts, 'C8' to overt pronouns in clitic-left dislocation contexts.

Figure E.1

Responses of Catalan-dominant bilinguals to each condition in Catalan, by participant

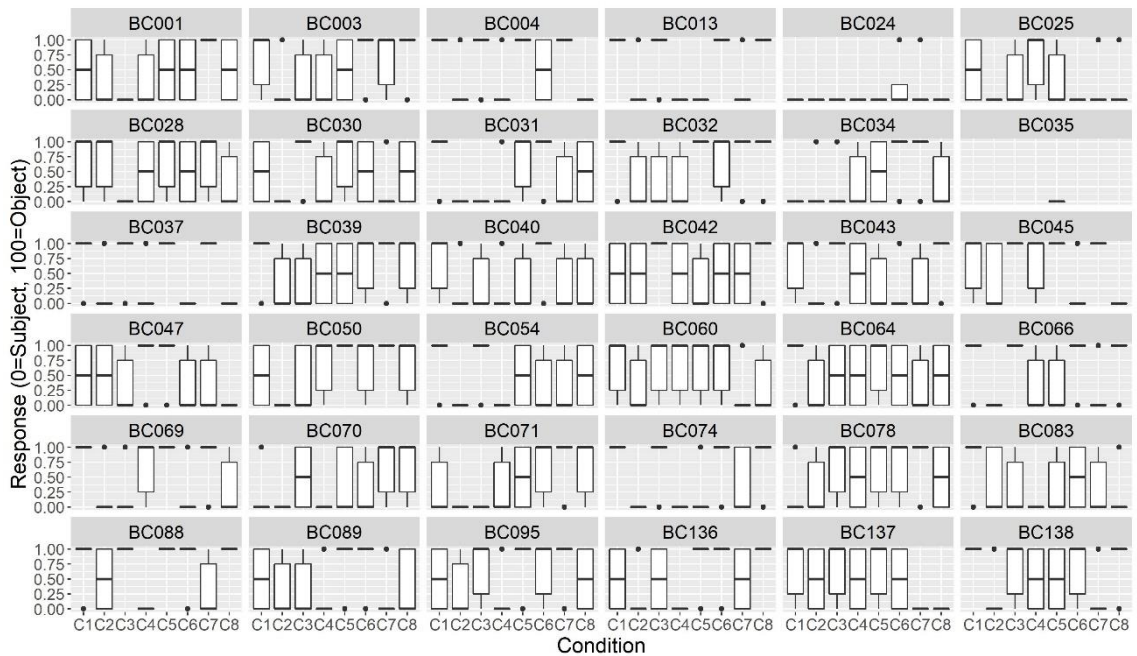


Figure E.2

Responses of balanced bilinguals to each condition in Catalan, by participant

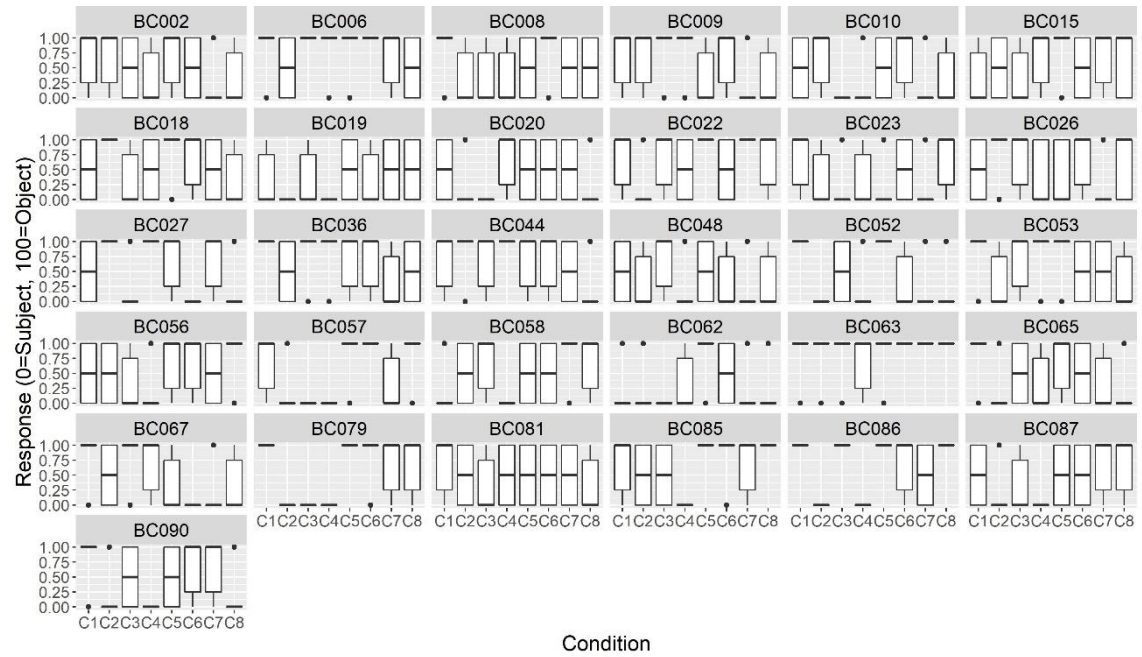


Figure E.3

Responses of Spanish-dominant bilinguals to each condition in Catalan, by participant

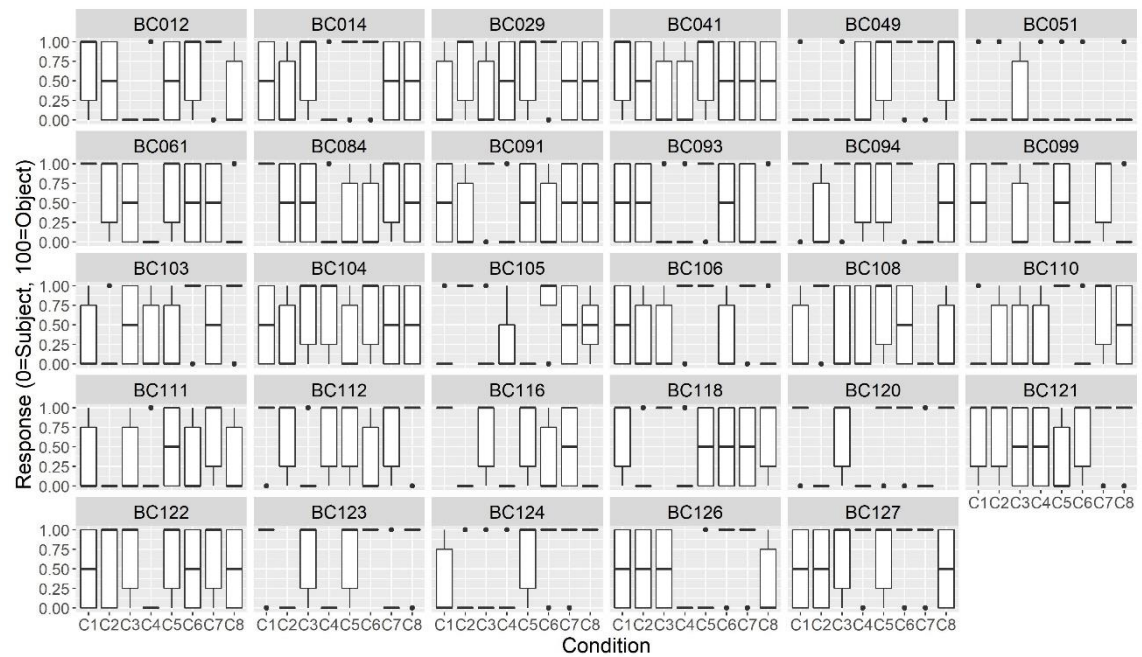


Figure E.4

Responses of Catalan-dominant bilinguals to each condition in Spanish, by participant

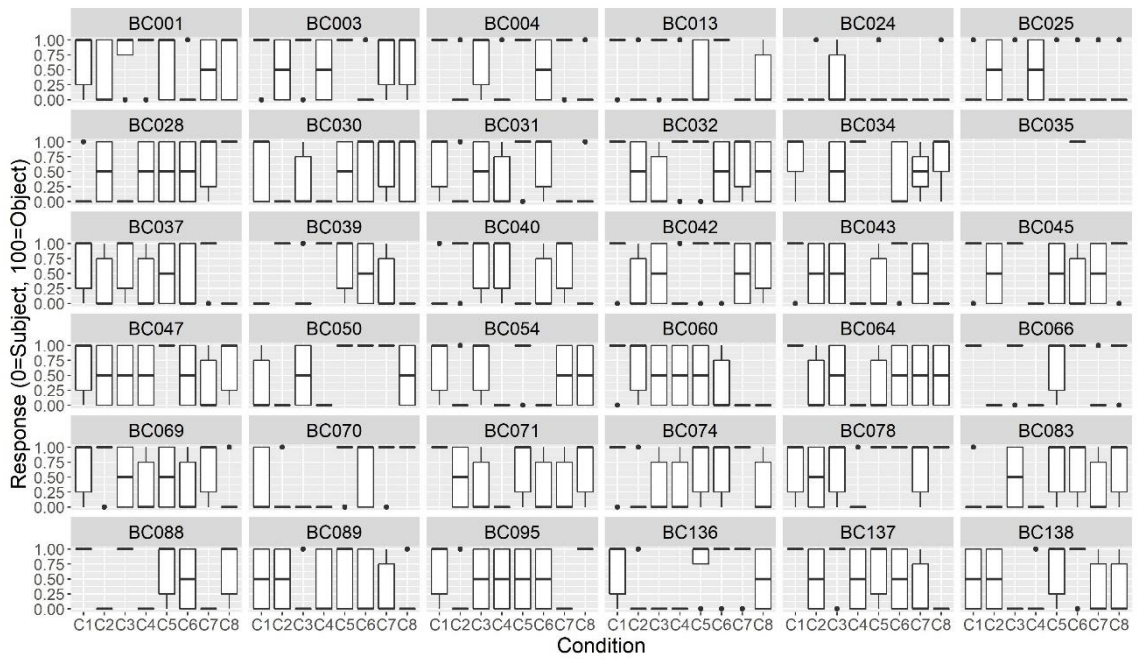


Figure E.5

Responses of balanced bilinguals to each condition in Spanish, by participant

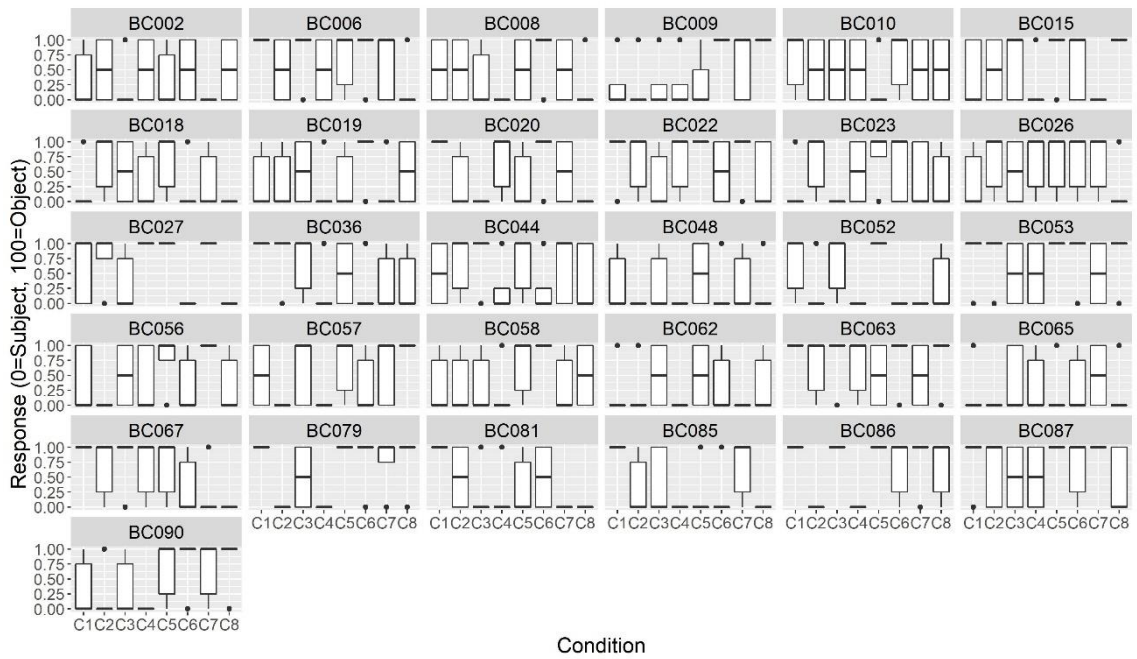


Figure E.6

Responses of Spanish-dominant bilinguals to each condition in Spanish, by participant

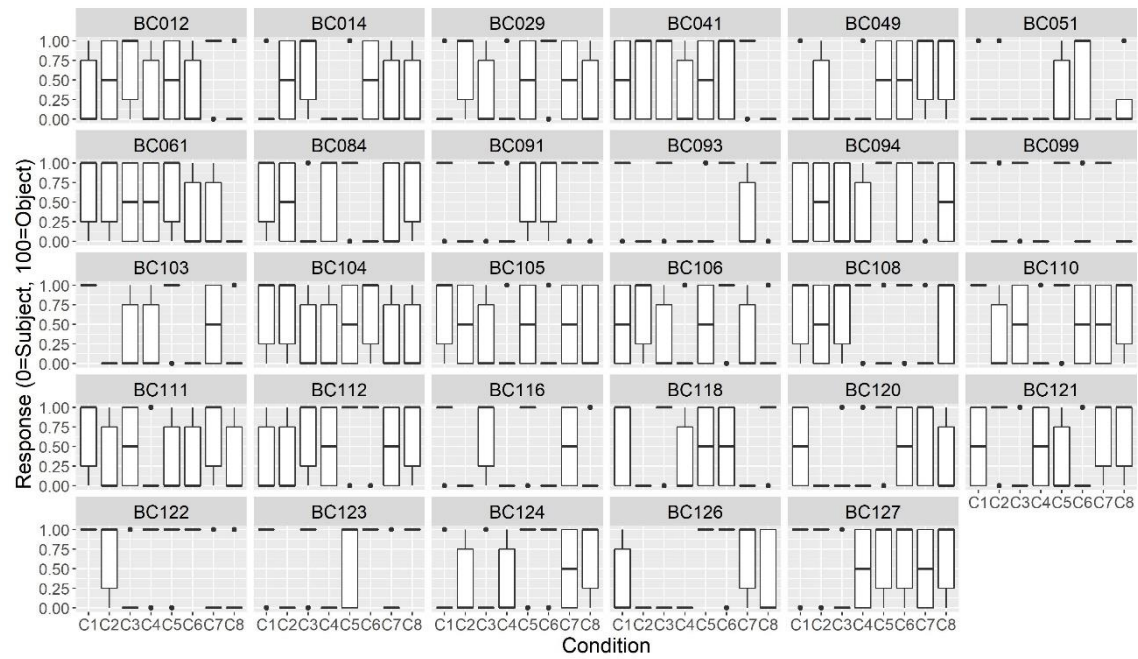
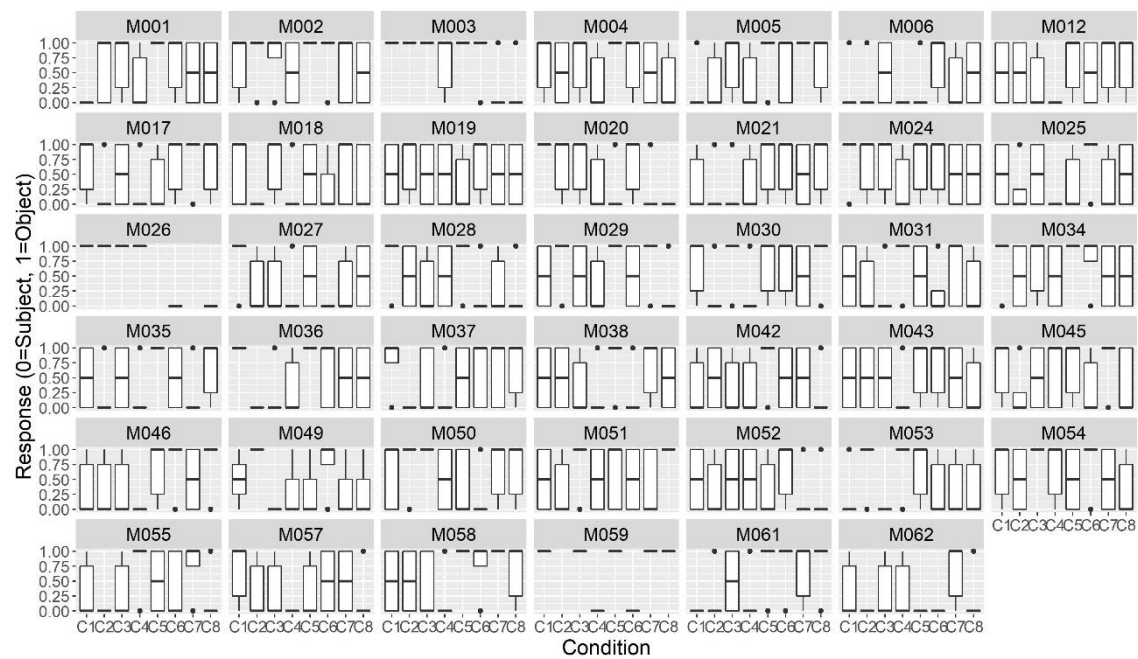


Figure E.7

Responses of Spanish monolinguals to each condition in Spanish, by participant



Appendix F: Boxplots for items' analyses

Figure F.1

Responses of all the participants to each condition in Catalan, by item (I)

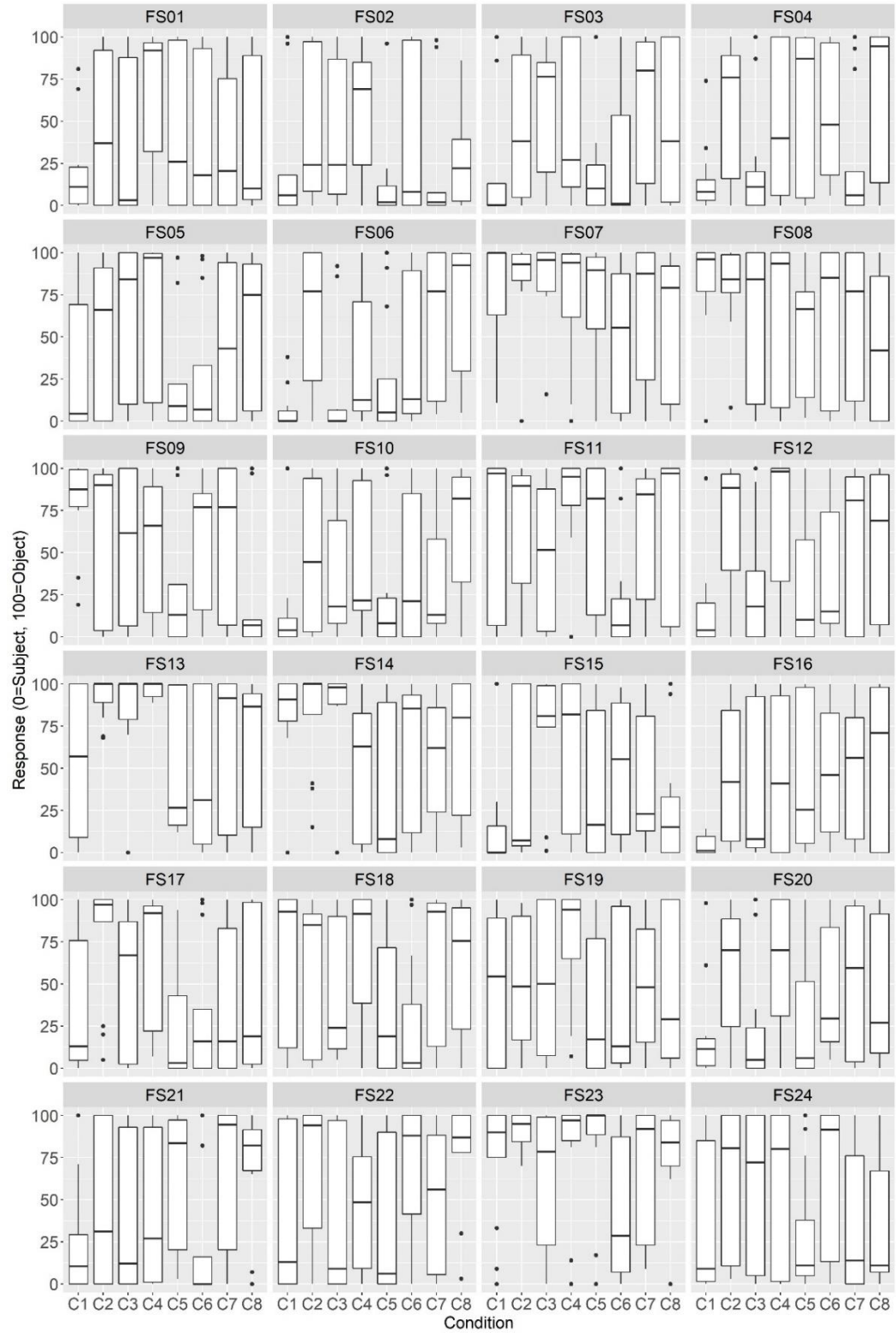


Figure F.2

Responses of all the participants to each condition in Catalan, by item (II)

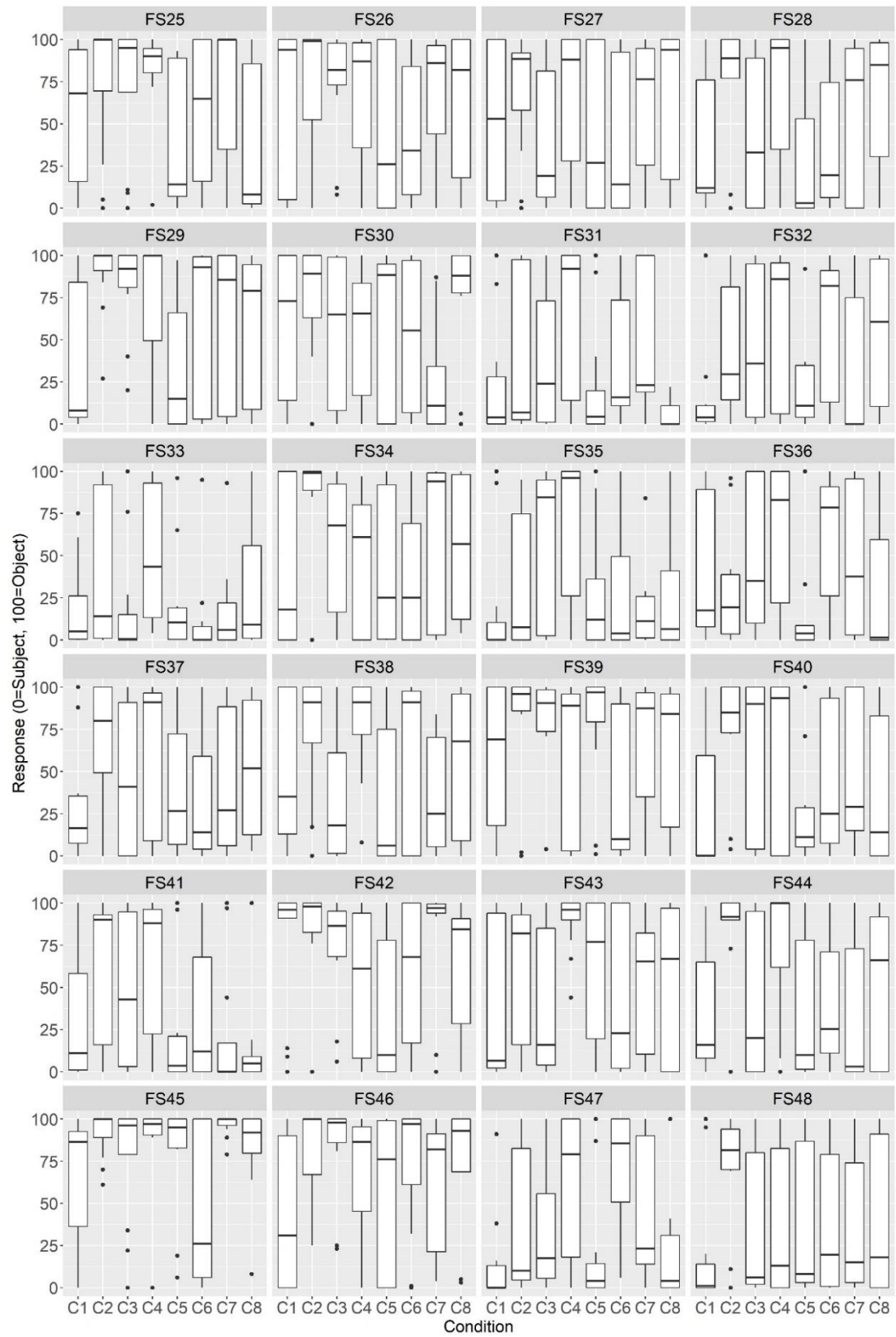


Figure F.3

Responses of all the participants to each condition in Spanish, by item (I)

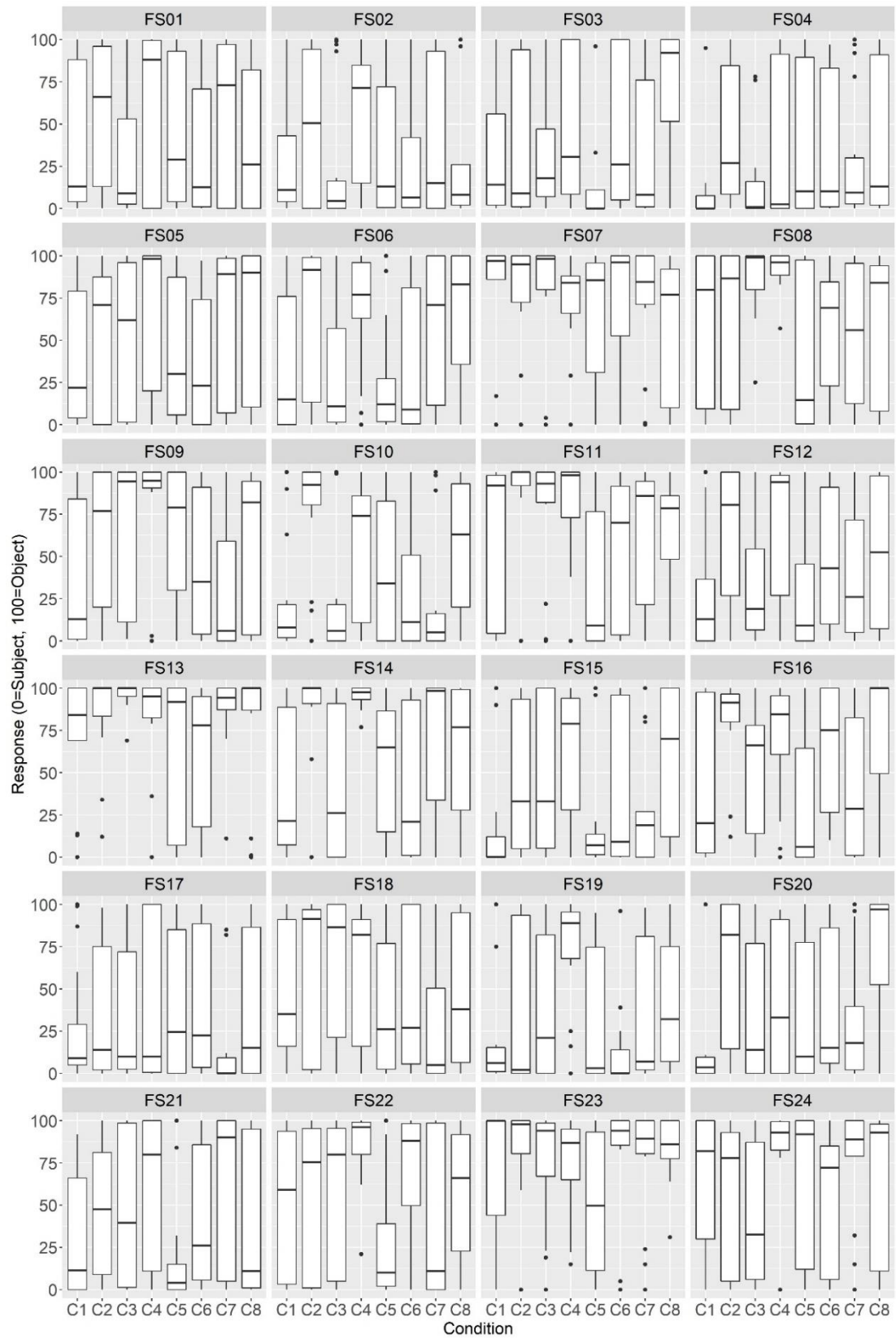
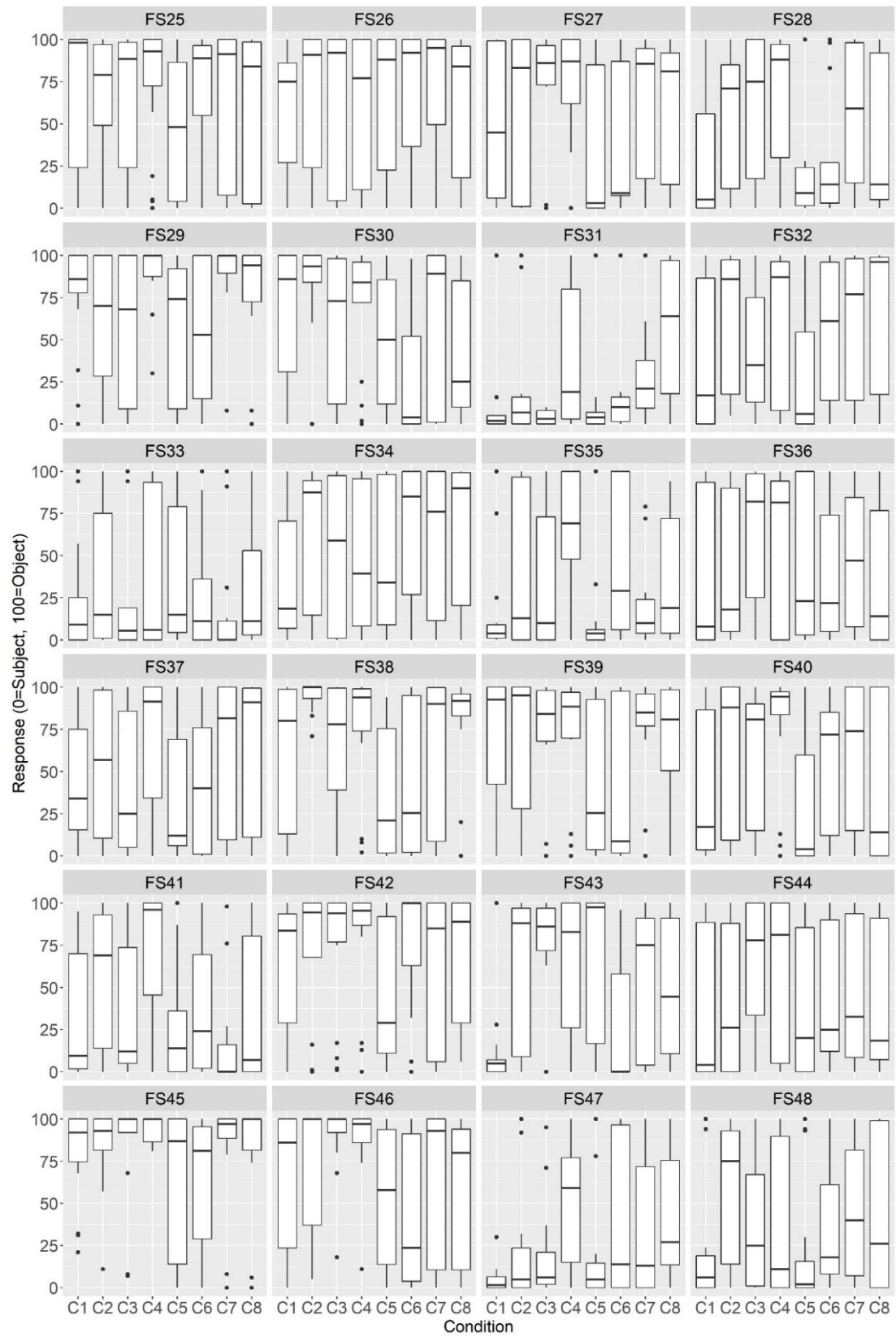


Figure F.4

Responses of all the participants to each condition in Spanish, by item (II)



Appendix G: Model 1, reported in §5.2

Appendix H: Models 2-4, reported in §5.3

Appendix I: Model 5, reported in §6.2 and §6.3

Appendix J: Follow-up analyses, reported in §6.2.5

Appendix K: Follow-up analyses, reported in §6.3.5

This set of appendixes (G, H, I, J and K) are available in the Open Science Framework repository (https://osf.io/nqe5j/?view_only=33edd018f33c424891b5f99507384351). The dataset and the code used to run the analyses in R can also be found in this repository.