

LANGUAGE ACQUISITION AND ENDOGENOUS GRAMMAR CHANGE: THE RISE OF
GALICIAN COMPLEMENTIZER AGREEMENT

by

BRIAN M. GRAVELY, JR.

(Under the Direction of TIMOTHY GUPTON)

ABSTRACT

The incorporation of empirical and theoretical knowledge from the language acquisition literature into formal diachronic syntax has been largely exiguous despite the growth of the latter as a subfield in formal linguistics over the past thirty years. In this dissertation, it is shown that a greater knowledge of diachronic phenomena, particularly those that are not cross-linguistically attested, may be achieved upon combining what is known about the inner workings of the language acquisition device (LAD) with surfacing studies that show and hypothesize about different deviant structures in child language. In this work, several individual syntactic phenomena, both synchronic and diachronic, are addressed and united in order to explain what is referred to here as the Galician complementizer agreement phenomenon (Galician C-AGR). By reviewing the limited historical data available within the purview of empirical data on cliticization in Western Iberian children and theories regarding cliticization patterns in this language family, it is shown that several of the proposed steps can be accounted for if one assumes that endogenous grammar change seen in child language data is one of the sources of diachronic change.

INDEX WORDS: language acquisition, learning models, clitics, agreement, phi-features,
Romance, Galician, complementizer agreement, grammar change,
language change, morphosyntax

LANGUAGE ACQUISITION AND ENDOGENOUS GRAMMAR CHANGE: THE RISE OF
GALICIAN COMPLEMENTIZER AGREEMENT

by

BRIAN M. GRAVELY, JR.

Bachelor of Arts, Appalachian State University 2014

Master of Arts, Appalachian State University, 2016

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

August 2021

© 2021

BRIAN M. GRAVELY, JR.

All Rights Reserved

LANGUAGE ACQUISITION AND ENDOGENOUS GRAMMAR CHANGE: THE RISE OF
GALICIAN COMPLEMENTIZER AGREEMENT

by

BRIAN M. GRAVELY, JR.

Major Professor: Timothy Gupton
Committee: L. Chadwick Howe
Joshua Bousquette

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
August 2021

DEDICATION

A Fernando, Chus e Paula por me introducirdes á lingua e a xente máis preciosas do mundo e por me apoiardes en todo na época que nos tocou vivir xuntos. Sen vós, non sei onde estaría. Débovos moitísimo máis ca esta tese e endexamais poderei dárvo-lo que merecedes. Sempre vos levarei no meu corazón e sempre serei agradecido por todo o que fixestes por min.

ACKNOWLEDGEMENTS

There are numerous parties that played an important role worthy of gratitude on so many different levels, so I will do my best to include everyone.

To Dr. Gupton, the most influential academic in my life not just during these grueling dissertation years but since I stepped foot in Gilbert Hall in 2016. The amount of work that you put into me as a young, growing professional— from the seven-version conference abstract revisions to the simplest professional development suggestion— there was never a time that you turned away from helping me. You have always been my biggest critic and my biggest supporter, all wrapped in one, and nothing could have been better for me in becoming the academic I am today. You supported me just as much when I did not think I could support myself as when I was on cloud nine with a new idea. You were always just as quick to point out when I nailed it on the first try as you were to come down on me when I skipped crucial details. Over the past five years, you have exaltedly exemplified what it truly means to be a mentor and a thesis director. I will never be able to thank you adequately for all you have done for me, but may the completion of this dissertation serve as the first try.

To Dr. Howe for introducing me to the mesmerizing field of diachronic work. You taught me to connect the dots I had floating around in my head when I lacked the vision to see how they were related. I cannot think of one idea I pitched to you (totally convinced that it was absolute gold, of course) that you did not have three or four immediate criticisms for in order to bring me back down to earth. Most of all, your acceptance of formal syntax and supporting me in exploring how to bridge the gap between generative theories and diachronic ideas awoke a fierce

side in me as a budding academic. I knew I was on the right track the day you said, “This isn’t what the field wants, but it’s what it needs.” You were right.

To Dr. Bousquette for coming in so late in the game and introducing me to the complementizer agreement phenomenon. Without you, there is a strong possibility that I would still be searching for the answer to the dependencies of several phenomena in this dissertation I simply could not understand two years ago. I remember the night I met you at Dr. Gupton’s house—I had barely gotten out “diachronic” and “really strange development” when you looked at me and said, “Yea, I’d love to hop on board.” I wish that I had been able to meet and learn from you before that evening, but I enjoy thinking about how you were the last puzzle piece I needed that slid into place in the nick of time.

A Ramón Freixeiro po-las conversas sobre temas presentados nesta tese e moito máis. A Manuel Ferreiro po-la axuda, por moi pouca que fose no momento, cando lla pedín. A Paula Cousillas Pena e Miguel Giadás Quintela por nunca me deixardes sen respostas por moitas voltas que lle dera a un tema ou outro. Non sabede-la forza da aperta que vos espera. E a Juan Uriagereka por compartires tanta información, relacionada con esta tese ou non, comigo e po-las palabras de ánimo cada vez que falamos. Lembrarei a túa xenerosidade pra sempre.

To Lisa Pearl for always entertaining even the simplest of questions with such enthusiasm and positivity. I could barely string together a coherent sentence during our very first conversation because I was so nervous, but you answered every question I had and walked me through crucial aspects of language acquisition that ended up being invaluable at so many turns of this thesis.

To Kate Bové for being my “thesis hero” and always being there to give me words of encouragement or a swift kick in the ass— whichever was needed at the time. To Amy Rose

Deal for always making time to answer my questions in great detail and for sharing her data and work with me. To Ailís Cournane, Giovanni Cassani, Valentina Colasanti, Jeff Lidz, Monica Irimia, Colin Phillips, Laura Kalin, Gary Baker, and Ed Rubin for the discussions (however brief) that motivated an idea (however small) presented in this thesis.

To all of my family and friends that helped me through the past five years, most notably my dad. Your love for me and your wish to see me succeed in the (sometimes excruciatingly painful) way I chose for myself meant the world to me. Even when I refused to believe it, your positivity and assurance that everything would come together if I continued to put my head down and work hard was spot on every single time, and your constant support in so many ways made vital stepping stones along my journey possible. Words cannot describe how proud I am to be your son.

To Ann, Dave, Ethan, Hannah, and Jennifer— the five of you brought (and continue to bring) a type of joy to my life that I desperately needed during such a trying time the past few years. Although our face-to-face time was limited due to the distance that separated us, the time spent with you was more than just a great time with loved ones. A Jorge, Midons y Marta por todo vuestro amor desde el primer día que nos conocimos en Salamanca. A pesar de la distancia y la locura que ha sido mi vida desde que empecé el doctorado, pienso en vosotros todos los días sin falta.

To Spencer and Ryan for making the last year of my thesis both tolerable and agonizingly insufferable at the same time. I had no idea that I needed daily time away from everything and everyone else, but your consistent reminders of how shit I am at everything I do (linguistics included—“No, that’s semantics, B”) were great reminders not to quit my day job. Dr. B loves you both something fierce.

And last but certainly not least, to Shannon for being everything and more outside of academia. You were patient, understanding, and supportive when I needed it most, even in my darkest moments when I had no idea which way to go. You always respected the time I needed for work, as well as the time I needed away from it. From driving on road trips so I could work or read in the car to gladly taking over mundane responsibilities like dinner or walking Pep—there was nothing that you would not do for me to ensure that I had the time, space, and peace to work. For all of the times that we had plans and canceled them because I said, “Honey, I’m on a roll right now!” and for all of the others when you said, “Shouldn’t you be working right now?” but never pushed it—your omnibenevolent understanding in every and all circumstances far surpassed anything that I could have ever hoped for in someone else. Although the culmination of the research here marks the end of a chapter in both of our lives, so many more await us.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	xii
LIST OF FIGURES	xiii
CHAPTER	
1 Introduction and background	1
1.1 Introduction.....	1
1.2 Syntax and language acquisition.....	5
1.3 The linguistic experience	17
1.4 Syntactic bootstrapping.....	28
1.5 Minimalism and feature-based syntax	36
1.6 The phenomenon investigated in this thesis	41
2 Data and theoretical analyses of Galician complementizer agreement	45
2.1 Introduction.....	45
2.2 Galician data	45
2.3 Previous analyses	56
2.4 Why a diachronic account is necessary	91
2.5 Conclusion	96
3 Reanalysis, category acquisition, and learnability	98
3.1 Introduction.....	98

3.2	The data.....	99
3.3	Enclisis and proclisis in Western Iberian: the learning path.....	111
3.4	An excursus on category learning and lexical categorization.....	139
3.5	The syntax of <i>dizque</i>	145
3.6	The two faces of Step 2.....	151
3.7	Conclusion	157
4	New categories, new grammar.....	159
4.1	Introduction.....	159
4.2	Different processes, the same change	160
4.3	Reanalysis and highly variable input	168
4.4	How arbitrary is endogenous change?	180
4.5	Conclusion	191
5	Clitics, agreement, and complementizer agreement	193
5.1	Introduction.....	193
5.2	Clitic or affix? The longstanding debate.....	194
5.3	A review of Galician C-AGR	199
5.4	Preminger (2009, 2014, 2019)	203
5.5	Nevins (2011).....	208
5.6	Kramer (2014).....	212
5.7	Theoretical underpinnings of Galician C-AGR	215
5.8	Complementizer agreement: How much cross-linguistic variability is there?.....	225
5.9	Theoretical approaches to C-AGR.....	230
5.10	Conclusion	249

6	Conclusions and future considerations	250
6.1	The thesis, reviewed.....	250
6.2	Theoretical importance	255
6.3	Limitations and future directions	257
6.4	Final comments.....	258
	REFERENCES	260

LIST OF TABLES

	Page
Table 1: Clitic paradigms in Cat., Gal., Ast., & Span.....	57
Table 2: Occurrences of Step 1 based on clitic and adverb combinations.....	102
Table 3: Percentage of answers obtained in enclisis contexts	122
Table 4: Percentage of answers obtained in proclisis contexts.....	122
Table 5: Rates of proclisis based on syntactic condition	123
Table 6: Percentage of production systematicity by input group	171
Table 7: Morphosyntactic diagnostics results of Galician C-AGR	218

LIST OF FIGURES

	Page
Figure 1: Grammar transmission in acquisition (Hale 2007:12)	1
Figure 2: Proposed steps for Galician C-AGR	4
Figure 3: The language acquisition device (Pearl 2020a).....	18
Figure 4: Geographical distribution of Galician dialects	83
Figure 5: Geographical distribution of Galician C-AGR (Álvarez 2015:211)	84
Figure 6: Galician C-AGR within the purview of Galician dialects.....	85
Figure 7: Distribution of dialectal variations <i>eí</i> and <i>eiquí</i>	87
Figure 8: Distributions of Galician C-AGR and the dialectal variations <i>eí</i> and <i>eiquí</i>	89
Figure 9: Galician C-AGR paradigm.....	95
Figure 10: West Flemish C-AGR paradigm (Haegeman 1992).....	95
Figure 11: First and second forms of Galician clitics and determiners	104
Figure 12: Example of elicited production for simple clauses (Costa, Fiéis & Lobo 2015:18) ..	120
Figure 13: Example of elicited production task for negative clauses (Costa, Fiéis & Lobo 2015:18)	121

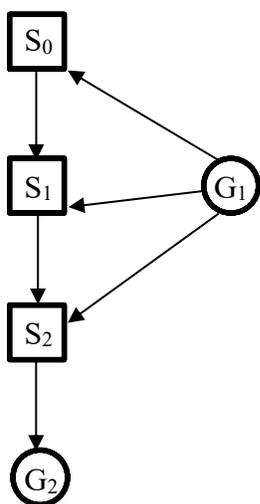
CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1. Introduction

The initial approach to diachronic phenomena on behalf of generative grammarians addressed the question of where one should look for the root of change. Even before Lightfoot's (1979) first generative approach, there was a consensus that the innovations that had been previously noted originated somewhere within the transmission of the grammar of an adult population to young learners. Investigators of the generative framework have maintained the general notion that children are innovators of grammatical sequences that differ from the grammar of previous generations, recently referred to by Cournane (2017, 2019a) as the *Child Innovator Approach*, a term I adopt throughout this thesis.

Figure 1. Grammar transmission in acquisition (Hale 2007:12)



The idea in Figure 1 is that, as the initial grammar (G1) is exposed to the learner producing different states of acquisition along the learning path (S_0 , S_1 , S_2), the resulting grammar (G2) differs in some form or fashion based on the lack of *complete transmission*. From a syntactic point of view, said differences have typically been thought of as the addition or loss of a feature (Roberts & Roussou 2003), although, as I shall show, a less noticeable change such as an additional hypothesis space in a learner may be enough to constitute the birth of a change.

During the advancement of diachronic generative work, the focus on child innovation trailed off and diachronic generative studies followed the synchronic generative literature in focusing on blanket effects that UG is said to have on the learner. Extensive works such as that of Roberts & Roussou (2003) and van Gelderen (2009) opted for a “UG bias” toward simpler structures, ultimately attributing the reanalysis of multi-word structures to single heads or the Merge-over-Move principle to these internal partialities.

Universal Grammar, however, has taken a significant hit with respect to both importance and explanatory value in generative work. A considerable amount of investigation, particularly in computational linguistics, has led to the understanding that the internal computation abilities of the brain can easily serve a learner in order to reduce the strain on UG (Yang 2002, 2016, 2018; Rothman & Chomsky 2018; Chomsky, Gallego & Ott 2019). Furthermore, the assumptions as to what role UG actually plays are often misunderstood. UG is anything that is both innate and specific to human language (Pearl 2020a,b), which makes up a relatively small part of the learning process and most likely attributes even less when one attempts to identify causes of diachronic change. It is, therefore, untenable to consider UG as a cause for change.

In this thesis, I address several phenomena related to grammar change which include cliticization and the clitic-affix debate, fixed sentential position of functional items, and the way

in which learnability and parsing of certain strings affect what come to be recognized as innovative diachronic developments. Specifically, I address these phenomena with respect to the development of what I deem Galician complementizer agreement (C-AGR) (1).

- (1) a. *Velaí o vai Xan!*
 behold _{CL_M.SG} go.PRS.3SG Xan
 ‘There goes Xan!’
- b. *Velaquí as veñen as nenas!*
 behold _{CL_{FEM}.PL} come.PRS.3PL the girls
 ‘Here come the girls!’
- c. *Velaquí os estamos nós!*
 behold _{CL_M.PL} be.PRS.1PL we
 ‘Here we are!’

In (1), we see the presence of a morphologically accusative clitic paired with an intransitive verbal predicate. This clitic must not be an argument of the verb, which raises the question: How may we account for this licit construction? Furthermore, how may we describe the relation between the clitic and the postverbal subject, which seem to agree in [NUMBER] and [GENDER]? Let us observe the data in (2).

- (2) a. **Velaquí as*
 behold _{CL_{FEM}.PL}
 Intended: ‘There they are!’
- b. **Veñen-as / *As veñen*
 come.PRS.3PL-CL_F.PL _{CL_F.PL} come.PRS.3PL
 Intended: ‘Here they come!’

In addition to the questions above, the ungrammatical examples in (2) force us to ask: What structural relation must be accounted for in (2) with respect to the presence of the clitic being illicit when either *velai/velaqui* ('behold') or the verb is elided?

In this investigation, I claim that these questions regarding the present-day data may only be understood within the purview of diachrony. Therefore, I propose the following steps in accounting for the development of the construction in (1).

Step 1: Biclausal, prosodically divided structure

Ve-lo	aí	vai	Xan
see.PRS.2SG-CL _{ACC.M.SG}	there	go.PRS.3SG	Xan

'You see him, there goes Xan.'

Step 2: Biclausal, prosodically unified structure

Ve-lo	aí	vai	Xan
see.PRS.2SG-CL _{ACC.M.SG}	there	go.PRS.3SG	Xan

'You see him there, Xan is going.'

Step 3: Monoclausal structure; reanalyzed verbal structure

Vel(o)aí	vai	Xan
behold	go.PRS.3SG	Xan

'There goes Xan.'

Step 4: Monoclausal, main clause complementizer agreement structure

Vel(o)aí	o	vai	Xan
behold	CL _{ACC.M.SG}	go.PRS.3SG	Xan

'There goes Xan.'

Figure 2. Proposed steps for Galician C-AGR

Unlike in example (1), Step 1 and Step 2 show the presence of an accusative clitic that is clearly the argument of the verb *ver* ('to see'). With the reanalysis of the main clause verb string in Step 2 as a functional element with no argument structure (2a) in Step 3, the clitic that appears in Step 4 must be accounted for by a different structural relation than that of the argument structure of the verb. I claim that the structural relation of this innovative structure is that of complementizer agreement.

These data present both commonly attested changes cross-linguistically and language-specific changes, both of which I show must be accounted for through the lens of language acquisition and its relation to intergenerational transmission. Although all cases of grammar change are deviant with respect to the target grammar, I shall show that what comes to be considered change in a diachronic sense is a structure that is internally supported within the language acquisition device based on parsing biases. I also explore how some changes are exogenous and may be linked to parsing biases of the learner, whilst other changes are endogenous and internally driven with little to no evidence of external biases.¹

1.2 Syntax and language acquisition

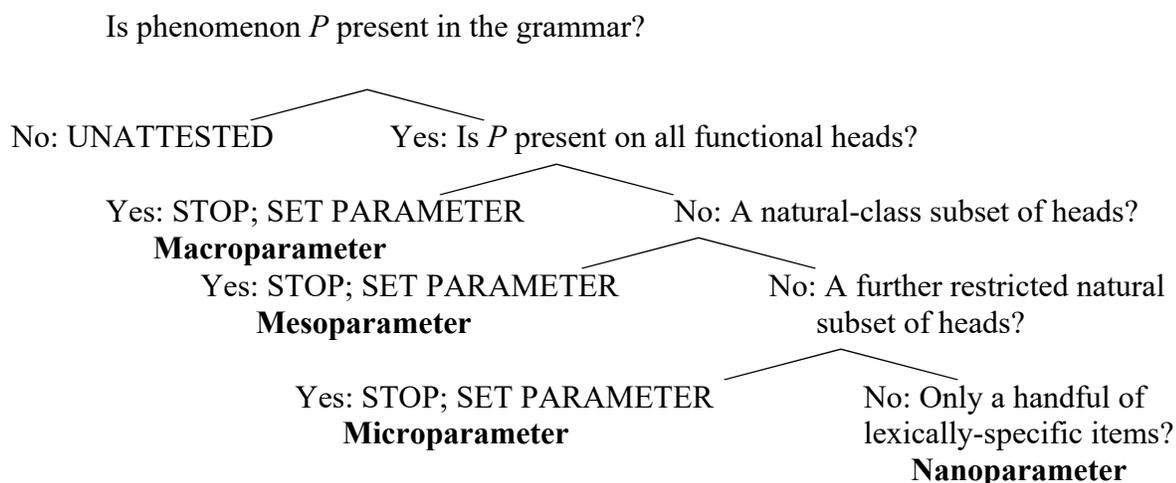
1.2.1 Parameters: Still a viable learning mechanism?

A most consistent mix of generative principles and diachrony have come from researchers of the *Rethinking Comparative Syntax* (ReCos) project headed by Ian Roberts and Theresa Biberauer. In recent years, they have attempted to unify the basic principles of language acquisition and those frequently identified in grammar change by focusing on a set of syntactic parameters unlike those hypothesized during the Principles and Parameters era (Chomsky 1986). With the rise of Minimalism, many investigators called into question the interconnectivity of different phenomena that, on the surface, seemed completely disassociated. Moreover, the ever-

¹ I expand on the notion of *exogenous* and *endogenous* changes in §1.3.3.

increasing hypothesis space of these parameters seemed too much, with the possible combinations of 30 parametric options alone reaching more than 1,073,741,824 options (Roberts 2014:183) (cf. Kayne 2005:14 for a broader view). While this type of learning may be optimal from an acquisition point of view (Pearl & Lidz 2013), the broad-scale linguistic view of parameters is largely believed to be too much to be feasible (Boeckx & Leivada 2013). This forced many research models into a ‘macro- vs. microparameter’ approach of different kinds (e.g. Kayne 2005, Baker 2008, Uriagereka 2007, *i.a.*). In the end, there was one remaining question: Where does this leave us with the investigation of syntactic choices that the learner must make in acquiring her grammar? According to the ReCos models, the choices are based on features, and they must be made on a much finer-grained level in a way that is easily diffusible without the learner falling into a “downward spiral of confusion”.² This is shown in (1).

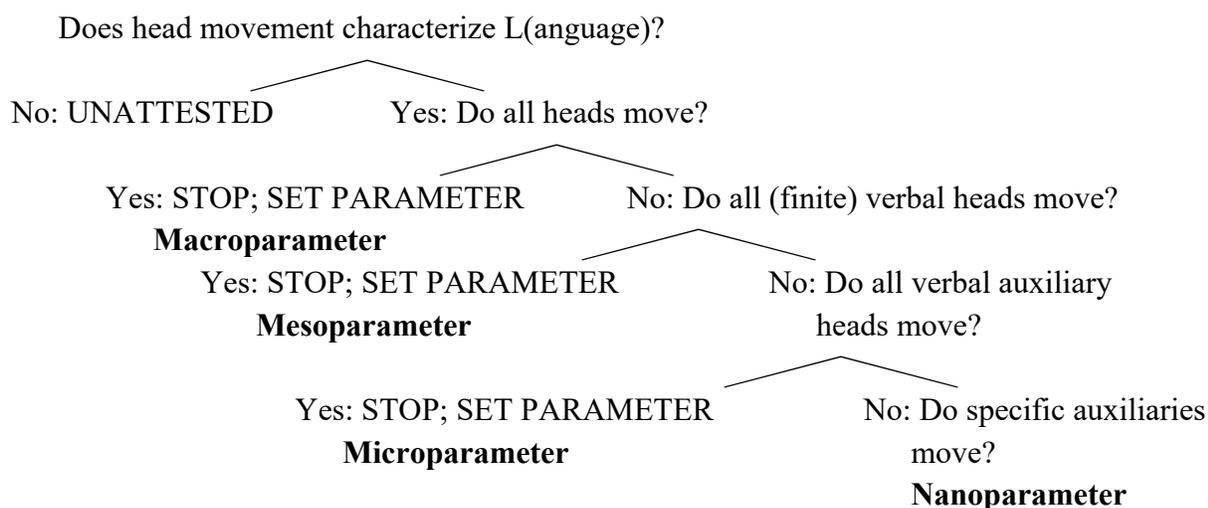
(1) *ReCos Parameter Hierarchies*



² This downward spiral of confusion is what led Berwick (1985) to posit the Subset Condition, a third-factor principle that prevents learners from getting caught in superset traps of potential grammars that may lead them to generate sentences that are not a part of the target grammar. This downward spiral would lead learners to postulate grammars that would be arbitrarily different from those to which they are exposed.

In essence, the learner begins with the option regarding whether said phenomenon is present in her language. If it is not, the parameter is automatically set to ‘No’.³ If it is, further specification must be teased out via the primary linguistic data (PLD): Is this phenomenon present on all heads? If not, is it present on all heads of class H_1 ? If not, is it present on all heads of a smaller class H_2 ? Each sub-specification of a class of lexical items with the same syntactic behavior constitutes a parametric level at which the learner must stop and reassess her analysis of the PLD until all parameters are set. The ‘size’ of any one parameter is in direct relation to another natural class of lexical items that realizes the same phenomenon, whether it be larger or smaller. This is akin to Waddington’s (1977 *apud* Roberts 2012:321) idea of an epigenetic landscape through which the resolutions to the learning problems a learner faces become more and more complex as she follows the path of least resistance. A common parameter such as that of head movement may be modeled as in (3).

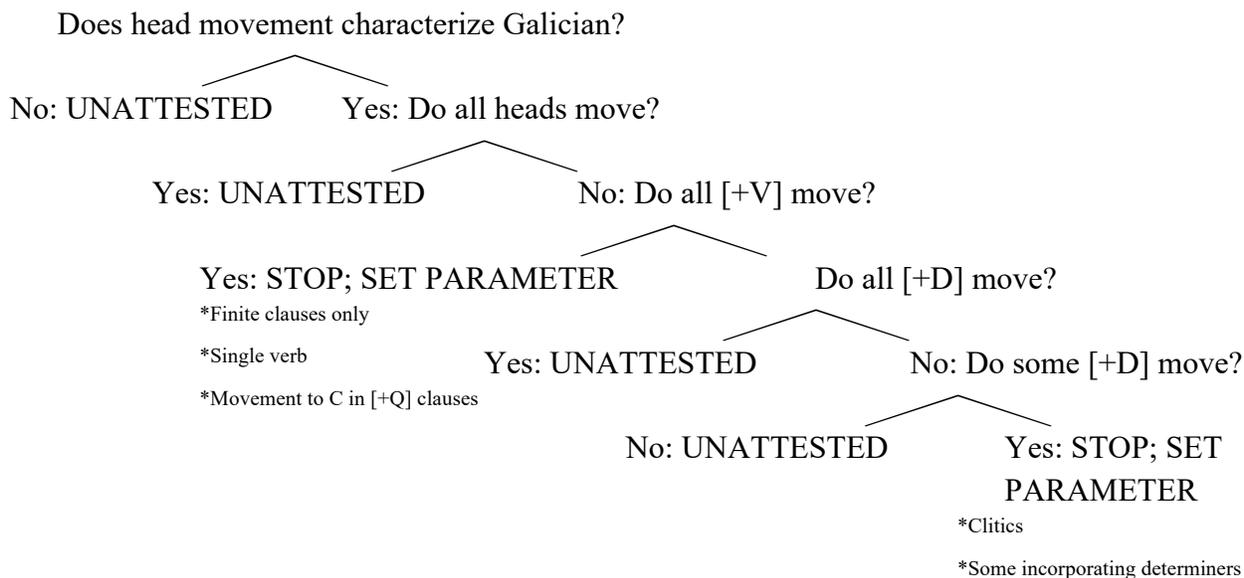
(3) *Head movement hierarchy* (Biberauer & Roberts 2016:260)



³ A more precise way of thinking of this is that the phenomenon in question is never considered, as the child has no evidence for it in order to postulate that it is indeed not a part of her grammar and, therefore, it does not enter into the realm of possible hypothesis spaces. Although the difference may seem irrelevant, I claim that there is indeed a difference, as children do not reject phenomena to which they have had no exposure.

What is important to note is the intent to incorporate single-language parameters into the style of parameter hierarchies typically found in generative syntax (Baker 1996, 2008; Kayne 2005). Attributing a parametric level found in (3) to an individual language shows the descriptive power of parameter hierarchies of this type. However, one quickly understands that a single language may often represent only one of these classifications. For example, learners of a strong polysynthetic language such as Mohawk will stop at the macroparametric setting in (3), while learners of a V2 language such as German will settle on the mesoparametric setting. The microparameter setting is akin to that found in late Middle English (Biberauer & Roberts 2017), and the nanoparametric option is that of modern-day conditional inversion found in English. At no point, however, do we expect a language to fill all tiers at the same time within the same grammar. Should this present a problem for the learner? Not inherently, but what parameter hierarchies have shown us, no matter the fine-grained approach they take, is that there are always exceptions (Pearl 2020a) as we will see below in (4). Were we to imagine that parameters of the type displayed above served a practical purpose to the learner, we would need to consider how these exceptions are accounted for, as these, too, are crucial to the learner converging on the grammar presented to her.

We may draw very conclusive examples from head movement in Romance. In most Romance languages, *v*-to-T movement is well attested. Furthermore, almost all lone [+D] elements incorporate onto a functional head (be that *v* as proposed for Spanish (Gallego 2016) or *f* as in Western Iberian (Raposo & Uriagereka 2005; Gupton 2014; Fernández-Rubiera 2013)). As they both constitute head movement, should these be represented within the same hierarchy? Let us examine head movement in Galician.

(4) *Galician head movement hierarchy*

A Galician child quickly comes to the conclusion that there is head movement in her language, but she does not posit that all heads must move (e.g. no preposition or complementizer movement). What the learner finds, however, is that there are caveats to what heads may move and when. For example, all verbs may undergo *v*-to-T movement, but only in finite clauses where there is only one verb or in *wh*-clauses where the verb may raise to C. In the spirit of keeping her phenomena uniform across all lexical items, she may initially posit an auxiliary construction as biclausal as in (5) based on similar control constructions such as that in (6) (cf. Cournane 2017:20-21 for this approach in English).

(5) [XP Xan_i] [TP [T *está*]] [CP [TP PRO_i [T *a*] [_vP [_v *comer*]]]]

Xan be.PRS.3SG to eat.INF

‘Xan is eating.’

(6) [XP Lúa_i] [TP [T *quere*]] [CP [TP PRO_i [T *comer*]]]

Lúa want.PRS.3SG eat.INF

‘Lúa wants to eat.’

However, after sufficient evidence, she may undergo a repair mechanism in which she now links *está* and *comer* as forming one single clause as in the adult grammar.

- (7) [XP Xani_i] [TP [T *está*]] [PP [P *a*]] [_vP [_v *comer*]]
 Xan be.PRS.3SG to eat.INF
 ‘Xan is eating.’

This realization, in turn, will require her to revisit the initial parameter set that all verbs move, as in this case, *está* does not undergo *v*-to-T movement but is inserted into the derivation on T. Therefore, she must delineate the circumstances in which verb movement takes place: only single, finite verbs move to T.⁴ A Galician learner will also experience a different kind of movement, that of *v*-T-C in interrogatives. This, too, is another asterisk on an overgeneralized parameter setting regarding verb movement.

What about other types of head movement? Some determiners move, as well, but not all of them. The clearest case for a learner will be that of clitic movement, a complicated learning task for a speaker of Galician that I return to in Chapter 3. Not only must she identify that only lone [+D] heads move to meet their verbal host as in (8) but also the possibility of determiner movement when it forms the head of a full DP as in (9).

- (8) Fixemo-lo
 do.PST.1PL-CL_M.SG
 ‘We did it.’
- (9) Fixemo-[lo traballo].
 do.PST.1PL-CL_M.SG work
 ‘We did the work.’

⁴ This itself is an oversimplification, as infinitival forms in complement clauses are said to move to an untensed T node in Romance (cf. Gallego 2009).

As Uriagereka (1996) showed, the type of movement in (9) is morphosyntactically conditioned by the inflection on the verb, presenting an additional learning task and further specification of what may constitute as ‘some incorporating determiners’ in (4). This is a prime example of an exception to a parameter mentioned in the discussion above.

Where does this leave us? Parameters present learners with a compact way of learning, which should be extremely effective upon making sense of the vast information they are faced with. This was the original observation in the P&P framework, as seemingly unrelated phenomena that were tied together underneath the surface provided a learner faced with multiple structural properties as a way to accelerate learning. However, as we have seen, even the simplest parameters have exceptions. Fodor (1998) argues that the first parameter a child sets is that of headedness: either she will see evidence for head-initial or head-final structures. For example, German prepositions precede their complements, but their verbs typically follow their complements. This type of contradictory evidence in the PLD need not be a problem, however. As Pearl (2020a:49-50) notes, the most notable takeaway from the discussion regarding parameters is that they may be seen as emergent overhypotheses based on what a child derives from the incoming input coupled with prior knowledge regarding structured representations.⁵ Based on this view, parameters can be equally as helpful to children based on what does and does not fall in line with the overhypothesis at which children arrive. That is, a learner’s concluding that not every lexical item falls into a particular parameter setting may allow her to adopt future misshapen hypotheses at a quicker rate.

Therefore, we may conclude that parameters, although not flawless, seem to have a realistic capability of assisting learners in acquiring their ambient grammar based on their being rooted in

⁵ An overhypothesis is defined by Kemp, Perfors & Tenenbaum (2007:308) as “any form of abstract knowledge that contributes to a hypothesis space at a less abstract level.”

overhypotheses created via the PLD. As I show in this thesis, certain exceptions that children posit for a particular parameter may bring about change.

1.2.2 Third-factor principles

In addition to the approach to parameter setting detailed above, Biberauer and Roberts propose the presence of two third-factor learning biases in order to account for this style of learning path. These cognitive learning biases, Feature Economy (FE) and Input Generalization (IG), are said to shape the parameter hierarchy in the NO-ALL-SOME (Biberauer 2017, 2019a) structure seen in (3).

(10) *Feature Economy (FE)*

Given two structural representations R and R' for a substring of input text S, R is less marked than R' iff R contains fewer formal features than R'.

(Roberts & Roussou 2003:201)

(11) *Input Generalization (IG)*

If acquirers assign a marked value to H, they will assign the same value to all comparable heads.

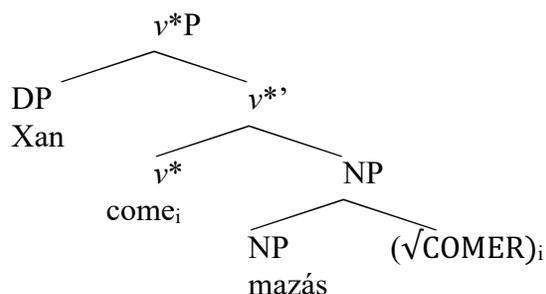
(Roberts 2007:275)

The conditions above are thought of as general cognitive optimization strategies from which parameters may arise, as the interaction of these two learning strategies is a markedness condition (Roberts 2012). Beginning with FE, this is seen as an acquisition bias that forces learners to postulate the fewest features possible from the unambiguous evidence they encounter. This conservative approach by the learner is said to rise from her desire to minimize computation, an idea resonant in Minimalism as a whole (e.g. Chomsky's (1995) original claim that Move (Internal Merge) is more costly than Merge (External Merge)). Based on the individuality of the ReCos style of parameters, feature conservatism is important due to the fact

that more computationally complex options are often found further down the tree (i.e. they are a subset of simpler options applicable to larger classes of lexical items), and, thus, are fewer in number and imaginably more scarce in the PLD. In essence, a learner forced down the parametric tree is a learner faced with greater computational complexity. Turning to IG, this mechanism takes the features identified in the PLD and posited by FE and generalizes them across all relevant lexical items. That is, if a learner postulates verb movement to T based on what she abstracts from the PLD, she will assume that all novel verbs also move to T in the same contexts. This mechanism seems to be a language-specific application of the general cognitive strategy that is the child creating overhypotheses about what she observes around her.

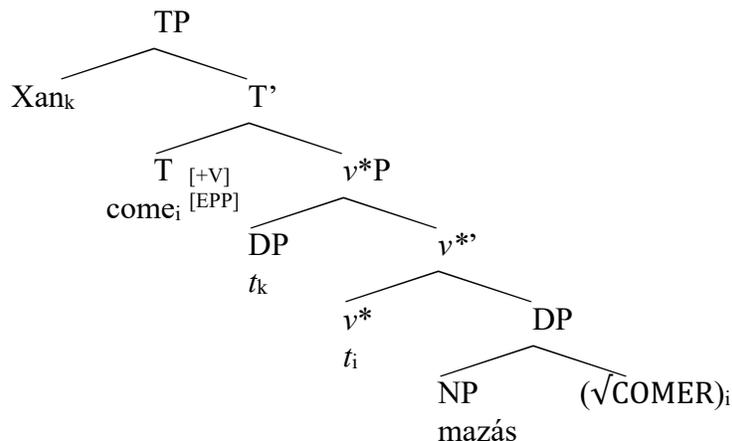
Biberauer & Roberts (2016, 2017) and Biberauer (2017) note that both of these minimal assumption learning strategies are diffusible by the PLD, just as the hierarchical tiers in their parameter structures. That is, if a child initially postulates no features in a string *S* as in (12), this satisfies FE (and indirectly IG). However, we may imagine that a Galician learner, for example, will quickly come to the conclusion that there is verb movement to T (as well as subsequent subject movement to its specifier). Thus, the child will undergo what I shall call *repair*⁶ on this initial hypothesis and posit the necessary features that represent said movement, violating FE (13).

(12) *Xan come mazás* ('Xan eats apples'); in compliance with FE



⁶ I expand on the notion of repair below.

(13) *Xan come mazás* ('Xan eats apples'); violation of FE



The repair of (12) in (13) shows the posit of a categorial [+V] which moves the verb to T and an EPP feature that forces the subject to move, as well. There are a number of aspects that may prove conducive to a structural repair such as this, specifically one dealing with the internal postulation of the necessity for a formal feature, such as morphological verbal inflection (Biberauer & Roberts 2012:271-276). A conclusion such as this immediately violates FE whilst setting a new precedent for IG moving forward. Along these lines, we may predict that children experience this type of repair hundreds of times with dozens of phenomena based on the idea that their initial reading on the PLD is consistently poor with respect to the adult grammar (Crain, Goro & Thornton 2006).

Before moving forward, there are a few issues to address with these learning biases. First, it is explicitly stated throughout the ongoing discussion of the learning path laid out above that the PLD must be unambiguous for FE and IG to aid in learning (Biberauer & Roberts 2017:147; Biberauer, Roberts & Sheehan 2014:53; Roberts & Roussou 2003:203-204). This is in direct contrast with what we find in the child acquisition literature: children must rely on ambiguous data upon parsing in real time due to the fact that unambiguous data is hard to come by for young learners. As Pearl (2020a:23-24) notes, “[i]f a deterministic learner is waiting for perfectly

unambiguous data, she may be waiting forever.” Therefore, ambiguous data must be used when making concluding evidence about input strings. For the interests of thesis, I claim that this inherently sets up learners to make mistakes from which they must later recover (*repair*). However, I claim in section 1.6 that this is precisely where grammar change takes place: the lack of repair in the developing grammar. Therefore, if children must use ambiguous data in order to make immediate decisions regarding their parse of incoming PLD, the syntactic structures they posit for this input must include any and all features and movement. If FE and IG are to play a role in learning, they must utilize the conclusions at which the learner arrives in real time, and these deductions are surely based on (partially) ambiguous data.

Second, there seems to be a divide between the empirical verity of FE and that of IG. Biberauer (2017, 2019a) takes these linguistic specific learning biases to make up part of her cognitive generalization principle *Maximize Minimal Means* (MMM), a third-factor principle by her account. If FE and IG are subcomponents of MMM, they cannot be language specific and, therefore, are not part of UG (first factor) *pace* Roberts & Roussou (2003). In this sense, IG seems to be a linguistic application of the more general strategy of overhypotheses discussed above in that it generalizes conclusions regarding parameter settings made about classes of lexical items. Regardless of whether this is a first- or third-factor principle, there is a clear indication that children do in fact learn this way. In contrast to IG, I wish to argue that FE has little pragmatic use for a learner. As no known language is free of checking procedures or movement, the difference between a “marked” and “unmarked” string is of little importance to a learner attempting to identify consistent and verifiable constituent movement in her intake. For example, children do not fight against a marked hypothesis because of its computational complexity if there is ample reason for positing some feature or diacritic used in order to account

for movement or word order. Furthermore, FE creates further problems for explaining long-standing questions related to change and the timing of change cross-linguistically, such as the *actuation problem*.

(14) *The actuation problem*

What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a particular time, but not in other languages with the same feature, or in the same language at other times?

(Weinreich, Labov & Herzog 1968:102)

Were FE a valid jumping-off point for the learner, one would expect a systematic overriding of similar phenomena based on language typology, contrary to what we find (e.g. V2 in Medieval Romance; Wolfe 2018). In this sense, FE runs against the grain of what we know regarding language acquisition as it relates to change (i.e. Lightfoot's claim that "there is no theory of change to be had independent of theories of grammar and acquisition" (2002:127)).

Returning to the notion of *repair*, I have shown that this learning mechanism allows the learner to recover from a previous conclusion drawn. This repair may be a purely structural one as in (12) and (13), one related to morphosyntactic variation as I will show in (17), or a retreat from a previous conclusion related to features and/or probes with respect to movement and agreement as I show in Chapter 3.⁷ Repair is clearly a linguistic strategy, but I am unsure whether it is found in another biological human system. Therefore, I remain agnostic as to whether it is a first- or third-factor principle.⁸ As for the reason for repair, this may vary

⁷ Repair has also been suggested by Wexler (2014) upon explaining the *Universal Phase Requirement* (Wexler 2004) which states that children take all *v*P and CP heads to be phases until around the age of eight years old. This explains why children perceive passives, unaccusatives, and (subject-to-subject) raising (i.e. 'tough movement') as ungrammatical. Eventually, children learn that these subject DPs may freely move across a *v*P without it being deemed a phase.

⁸ Repair may be a third-factor principle were it able to be connected to theory of the mind, for example.

depending on the stage of development and/or a particular set of data in the PLD that reaches the intake. Although children are thought to be conservative learners in most instances, the literature on child language showing deviance from the corresponding adult grammar abounds. Therefore, repair from a deviant structure may be due to overwhelming evidence in the input or simply a retreat from a structure forged internally for which no evidence was ever given.⁹

In the next section, I outline a detailed theory of the internal workings of the language acquisition device and summarize a theory as to how different subcomponents may play a role in begetting what is considered diachronic grammar change.

1.3 The linguistic experience

1.3.1 The language acquisition device

The general reference to a learner's experience in acquiring language in generative work begins and ends with Universal Grammar. There is, however, a much larger, more encompassing system that may be thought of as the *language acquisition device* (LAD) of which UG is merely a small part. The most comprehensible understanding of not only the innerworkings of the architecture in question, but also the role of the language experience is found in Omaki & Lidz (2015), Lidz & Gagliardi (2015), and Pearl (2020a) as modeled in Figure 3.

⁹ An example of the first case would be a simple reanalysis of infinitival clauses to inflected verbs agreeing with their clausal subjects. The second is the case of work shown by Belletti (2017) mentioned in §1.3.3 and Chapter 5.

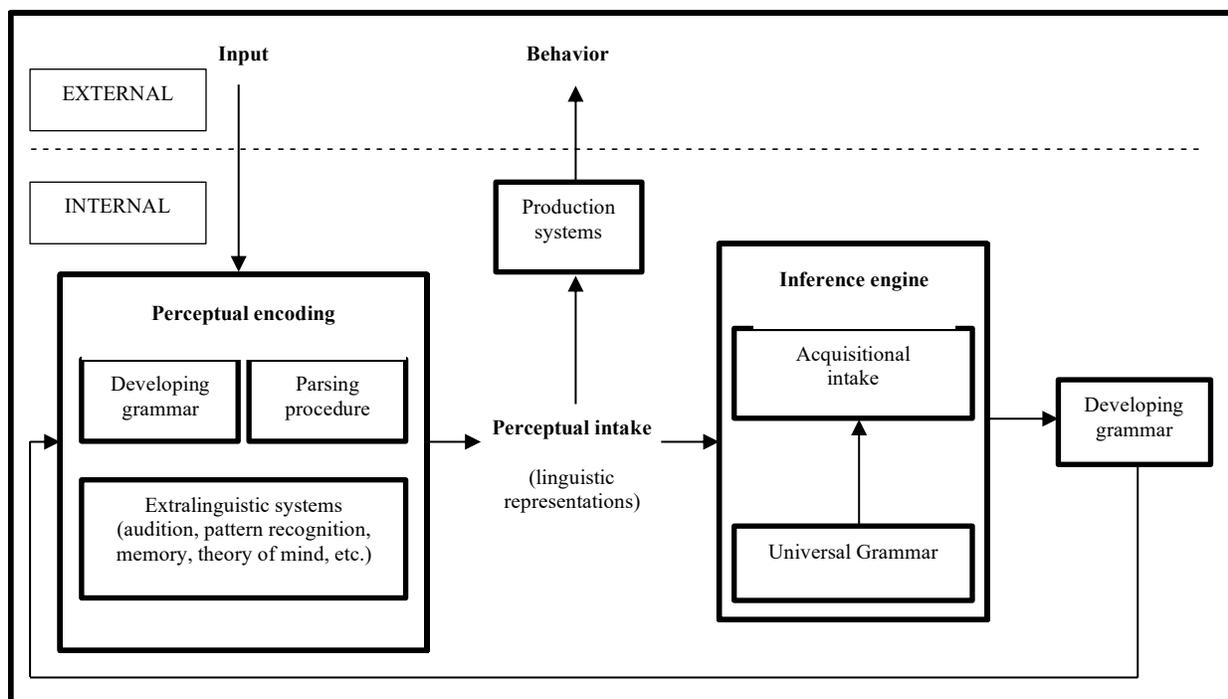


Figure 3. The language acquisition device (Pearl 2020a:4)

Figure 3 represents all of the inner workings of the linguistic motor that is a child's mind. As the child receives external input, several components are at play. First, a representation of what is filtered via the perceptual encoding mechanism is immediately realized based on the child's own developing grammar at the moment, as well as third factor principles dealing with extralinguistic factors that nonetheless play a role in converging on an internal structure for the arriving input (e.g. FE mentioned above, theory of the mind (Wimmer & Perner 1983), etc.). The role that the developing grammar plays is that of assisting in assigning a parse to the input based on the already proposed linguistic experience and, more importantly, the acquisitional intake up to that point. With respect to the syntax, for example, any relevant features acquired and assigned to previous parses will be available for the abstract construction of the surface form provided via the new input, a process that relies on the statistical distribution of the experience that the child has undergone.

The subsequent process is that which is carried out by the perceptual intake. The *perceptual intake* is “the data that are internalized through attention and perception... and is available for further computation in the mind” (Omaki & Lidz 2015:161, fn.1). These data are what receives a parse, which may or may not be accurate. Once this perception has been solidified, it is fed to two distinct systems: the inference engine and the production systems. Beginning with the inference engine, the final parse from the perceptual intake is passed through the constraints placed on the grammar by UG and, in turn, is compared to what is held in the acquisitional intake (i.e. the current developed grammar). The *acquisitional intake* may be thought of as “data that the language acquisition mechanism selectively extracts out of the perceptual intake for the purpose of making inferences about the grammatical structure of the language” (Omaki & Lidz 2015:161, fn. 1). Depending on what the child is trying to learn at a particular stage of acquisition, these data aid in statistical tracking of the minutiae of language (i.e. syntactic behavior, comprehension of theta roles, lexical categorization, phonological alternations, etc.). A key component of the extralinguistic systems is the statistical information recorded for the information that reaches the acquisitional intake. These statistical data go beyond stylistic movement such as topicalization and focus fronting, as discussed in length by Yang (2016).¹⁰ A Spanish-speaking child must learn, for example, that although most adjectives appear postnominally, the adjective *mero* (‘mere’) is never found in a postnominal position. The same goes for an English-speaking child who must learn that the adjectives *responsible* and *available* may fill the predicative, attributive, and postpositive adjectival positions, whereas adjectives

¹⁰ Much of Yang (2016) focuses on statistical learning of irregular forms, such as the irregular past-tense verb forms in English. Although I don’t discuss this phenomenon specifically, I return to some of the commentary on his statistical methods used and its use for children in learning complex phenomena such as that which I present in this thesis.

such as *utter* and *main* may only fill the attributive adjectival slot.¹¹ These idiosyncrasies of fixed-positions within the syntax must be learned and recorded in the acquisitional intake at some point during development, and this seems a prime use for statistical tracking that children employ even before they are able to speak. Therefore, characteristics of statistical acquisition such as where a particular lexical item has appeared in the derivation, in what syntactic context it has been found, and how many times it is found there all undoubtedly have an effect on a child's language experience once she becomes an adult and, therefore, are a vital part of any theory related to grammar change as a byproduct of child language acquisition.

This comparison between what has already been processed and statistically recorded in the inference engine (i.e. what makes up the developing grammar) and incoming strings drives subsequent inferences about the grammatical characteristics corresponding to the new parse under consideration. Based on the linguistic information possessed by the learner and her cognitive maturation, a final derivational structure is processed and, thus, internalized. Once grammatical knowledge is updated, it is added to the developing grammar, which, in the interest of further learning, is fed anew to the perceptual encoding mechanism as new strings of input are taken in.

The perceptual intake mechanism also plays a part in the updating of the systems related to comprehension and the behavior related to the child's response to new incoming sentences, as this is seen as "a function of the developing grammar and the mechanisms that yield a perceptual intake representation for [those] sentence[s]" (Lidz & Gagliardi 2015:337). In the following

¹¹ Predicative adjectives are those that participate in a small clause configuration with copular(-like) verbs such as *be*, *look*, *seem*, etc., as in (i). Attributive adjectives modify nouns in a strict pre-nominal position, as in (ii). Postpositive adjectives are those that solely occur after the noun they modify (iii).

- i) John was *responsible* for the accident.
- ii) A *responsible* student always studies.
- iii) The person *responsible* had no idea of the charges.

sections, I lay out how comprehension, behavior, and the state of the LAD are related to further learning at any given time during the acquisition experience.

1.3.2 General cognition

Before analyzing theoretical models related directly to learning by parsing, a few assumptions are in order. First, we must take into consideration several third-factor principles that affect a learner's ability to parse input strings accurately. The main aspect here is cognitive maturation, which may have several profound effects on interpretation, and thus development, for a child's grammar. Cognitive maturation may have various effects on different parts of the extralinguistic system of the acquisition process, particularly working memory and pattern recognition. Moreover, Gleitman's (1990:12) statement that "an observer who notices *everything* can learn *nothing*" (author's emphasis) is true in part because developing children have a cognitive immaturity that prevents them from learning everything about their language all at once. The flipside to that, however, seems to deal with what children in fact pay attention to and why. In work on audition and sight, this has been referred to as the *Goldilocks effect* (Kidd, Piantadosi, & Aslin 2014). These authors contest that learning is directly dependent on relevant properties of the stimulus in such a way that their focus is on "learning material that is sufficiently novel from—but also sufficiently related to—the infants' existing knowledge" (Kidd et al. 2014: 1796). This focus on intermediate complexity must lie at the heart of a mechanism that attempts to "direct attention to the most important aspects of the world" (Kidd et al. 2014: 1801), which would prove to be a useful strategy to employ as a language learner that is faced with input that is sometimes incomprehensible, sometimes lucid, and sometimes partially blurred. Therefore, it seems only fitting to extend this notion to language learning that goes beyond prosodically structured syllables as analyzed by these authors. If syntactic (or various

components of language learning in more general terms—probably barring lexical word learning) can be shown to follow this path, we see further solidification of language as a biological system (Berwick & Chomsky 2016, Chapter 2; Hauser, Chomsky, & Fitch 2002; Ramus & Fisher 2009; Friederici 2009; *i.a.*), suggesting that the Goldilocks effect is part and parcel of a general learning strategy used in across all developing biological systems.

Applying these observations in other domains of biological maturation of children to language learning lend support to the idea that children not only search for components of their grammar that they have yet to master, but also that children use their assumptions about these elements of language in order to learn. As pointed out by Pearl (2020a), what reaches the acquisitional intake is only a portion of what is sent from the perceptual encoding mechanism. Therefore, I claim that an aspect of a child’s grammar that she is unsure of but interested in learning reaches the acquisitional intake regardless of whether she must later retreat from her initial hypothesis. We may assume that actual learning takes place in the inference engine and, in turn, so does change.

1.3.3 How does change come along, then?

An important point to take away from the system we outlined as the LAD is the difference between *input* and *intake*. For the learner, the input is what is received by the auditory system and may contain information that is unable to be parsed and processed in order to reach the mechanism that internalizes and feeds the developing grammar. The *intake* may be thought of as the mental linguistic representations of the data fed through the language acquisition device for language development. As discussed in §1.3.1, the *intake* is essentially made up of two separate types of data, one leading to the development of the other. For what is to be discussed subsequently, we may think of *intake* as *perceptual intake*, as this is that which affects a learner’s

parsing abilities and influences the structure of a learner's parse. As a child is exposed to more and more of her language, her ability both to parse strings and predict incoming information augments (Lidz & Perkins 2018).

Because the ability to parse an incoming string is vital to learning and hypothesizing about its internal composition (Fodor 1998 2017; Fodor & Sakas 2012, 2017), I begin there. As highlighted by Phillips & Ehrenhofer (2015), three possible outcomes of parsing an incoming string may arise that can affect, either positively or negatively, a developing grammar:

(15) *Three parsing outcomes* (Phillips & Ehrenhofer 2015:416-417)

- i.) A learner may fail to assign a parse to an arriving string, either because it is too complex, or it arrives too quickly;
- ii.) A learner may systematically assign an incorrect parse (per the established adult grammar) to a particular sentence type due to parsing biases or reanalysis failure;
- iii.) A learner that parses an input string successfully so that it may reach the perceptual intake mechanism of the LAD may extract more or less information from said input string depending on her ability to predict the contents of the sentence as it unfolds.

The first scenario is highly frequent, possibly through and even beyond the two-year stage (Lidz & Perkins 2018). This is not seen as problematic for learning, however, as no information will reach the acquisitional intake and, therefore, nothing is generalized incorrectly. This situation with respect to Figure 3 is best thought of as nothing materializing in the perceptual encoding portion of the LAD. The second scenario has the potential to be highly troublesome if a systematic misparsing of a particular lexical item or string is fed to the acquisitional intake. What is important to note here is that while the input provided to the learner may be both highly informative and optimal from a learning standpoint, the possibility of a learner's intake diverging

from her input is a very real and dangerous problem for learners (Omaki 2010; Omaki & Lidz 2015). It is dangerous in the sense that a child may completely misinterpret and misparse a string when the input is optimal from a communication standpoint. Were this to happen frequently with multiple phenomena, this would be detrimental for both i) the child's development (e.g. she may be the only one of her peers that has misparsed X phenomenon) and ii) the stability of linguistic structures in a speech community. This is the scenario that I shall use as the basis for change in the subsequent chapters, as it is this situation in which a child may be sure of a parse that is fed to the acquisitional intake that, in fact, is not the intended structure and/or meaning of the speaker. The final scenario is what we expect of a learner who has not only acquired the relevant lexical items but their respective syntactic and phonological characteristics, as well. Depending on the level of cognitive maturity and what the child is trying to learn, a learner in this situation will be able to extract more or less information based on what she is looking for and how she is able to predict incoming information based on the context of what in the input string has already unfolded to her in real time. This is not a problematic situation for learning in the sense that it is not divergent from the input and, therefore, does not lead the learner astray.

It follows from these scenarios that systematic misparsing of the input, due to a myriad of reasons (e.g. the underdevelopment of cognitive maturation, an incorrect pragmatic understanding, etc.), has a direct effect on the perceptual encoding mechanism and, in turn, the representational input-divergence that reaches the acquisitional intake. These aspects of language acquisition will provide a solid explanatory base for many of the changes we will observe in Figure 2. Following Cournane (2017, 2019a, 2019b), I claim that these input-divergent strings later become non-divergent structures in the grammar of subsequent generations. Specifically, this thesis does not address how change is spread but rather how change may be internally forged

and how learners may use their own biases to create novel syntactic structures. Historical linguistics has often highlighted the gradualness of change and the fact that change often requires centuries to complete. Hale (1998) was the first to describe historical change as composed of two parts: innovation and diffusion. Strictly speaking, syntactic change has no temporal properties but is the set of differences between, for example, the child grammar and the target grammar (Hale 1998:3). What is perceived as gradual is diffusion, which should not be confused with innovation. Diffusion, as argued by Hale (1998:5), “gives rise to an identity between the innovating grammar and the grammar of the acquirer.” In other words, diffusion is widespread transmission of a phenomenon throughout a specific linguistic community. This is the change often found in the sociolinguistic literature and will not be addressed here.

Returning to the possible divergent result of the analysis that I present outcomes for referenced above, I shall concentrate on the second scenario in (15). Although previous work in diachronic generative work has attributed change to misparsing (Lightfoot 1979), these changes have been primarily one-dimensional in that they assume a simple reanalysis on behalf of the child parser that is predicated on movement of a lexical item to a higher functional head in the clausal spine.¹² Although reanalysis of verbal predicates may occur like this in some instances (Cournane 2014, 2015, 2020), not all change is this simple, however, and it gives little insight as to the broad extensions of grammatical possibilities of a language.

In order to understand the implications for a scenario in which a child may misparse a syntactic string to such a degree that it becomes part of her intake (and, in turn, developing grammar), first we must understand that all grammar change is equal. That is, as originally

¹² Throughout this investigation, I abstain from the term *grammaticalization* and from discussing the potential links to this construction through the eyes of *grammaticalization theory*. First, these notions are thought to be triggered primarily by semantic change, which they leads to syntactic change, the opposite claim that I make here. Second, there is no definition of *grammaticalization* that is accepted by all diachronic investigators as is the case of syntactic reanalysis.

hinted at by Hale (1998:6) and expanded on by Cournane (2019a:132), changes in the developing grammars of children are no different from changes found in diachrony. This view not only shows support for the CIA but holds it at the very epicenter of all changes we see cross-generationally. Conceptually, this makes sense if we consider that the path the child walks along is essentially a string of intermediate grammars leading to an emergent grammar (Biberauer & Roberts 2017; Roberts 2012) on which she converges. Therefore, language acquisition entails grammar change (Crain, Goro, & Thorton 2006).

I follow Cournane (2017, 2019a, 2019b) in that the force behind an acquisition-based model for change derives from the fact that at any one stage of a developing grammar, a child may see distributional regularities that are not actually in the adult grammar. Once this input-divergence has formed part of the child grammar, the acquisition of the target form gives rise to a case of internal syntactic variation. Belletti (2017) gives evidence for this upon citing deviant structures children produced in experiments from Manetti & Belletti (2017). These authors posed a question as in (16a) to children ages 4;1-5;11. As she states, this question welcomes an answer with a clitic-left dislocation (CLLD) structure as in (16b).

- (16) a. Che cosa succede ai miei amici,
 what thing happen.PRS.3SG to-the.PL my.PL friends
 il pinguino e la mucca?
 the penguin and the cow
 ‘What is happening to my friends, the penguin and the cow?’
- b. I pinguino_i, il coniglio lo_i tocca.
 the.PL penguin the rabbit CL_{M.SG} touch.PRS.3SG
 ‘The penguin_i, the rabbit touches him_i.’

Belletti (2017) claims that the children of this age group realized the CLLD structure in a deviant fashion from that which is expected.

- (17) a. Che cosa succede ai miei amici,
 what thing happen.PRS.3SG to-the.PL my.PL friends
 il pinguino e la mucca?
 the penguin and the cow
 ‘What is happening to my friends, the penguin and the cow?’
- b. Il coniglio a i pinguino_i lo_i tocca.
 the rabbit at the.PL penguin CL_{M.SG} touch.PRS.3SG
 ‘The rabbit to the penguin, he touches it.’

(Belletti 2017:2; from Adele age 4;9)

In (17b), the child introduced the dislocated direct object with the preposition *a* (‘at’). In the central variety of Italian (Tuscan), direct objects are never marked with the preposition *a*, making this structure one of endogenous change.¹³

In some cases, the co-existence of deviant and non-deviant structures may produce what has been deemed “grammar competition” (Kroch 1989; Yang 2002). Although I abstain from this notion due to the murkiness regarding its actual implementation within the acquisition device in Figure 3, the possibility of two or more strings being in some kind of internal competition is real, and I expand on it as a potential catalyst for change in Chapter 3.

In the following section, I review the support for *syntactic bootstrapping* as the primary mechanism by which children learn language. I then outline the generative approach to the

¹³ I elaborate more on the specifics of *endogenous* and *exogenous change* in Chapters 4 & 5. For now, *endogenous change* may be thought of one that is created internally, in the mind of the speaker, with no direct evidence for such structure. *Exogenous change*, on the other hand, is typically linked to some sort of change in the input that may be misinterpreted and, in turn, assigned a deviant structure (with respect to the adult grammar) in the learner’s internal grammar.

acquisition of word order and features as proposed by Westergaard (2008, 2009a, 2009b, 2013, 2014) which will serve as my basis for several arguments in Chapters 3 and 4.

1.4 Syntactic bootstrapping

Taking into account the details just reviewed of how incoming information is processed between different parts of the LAD, now we must identify the strategies used by learners in order to arrive at the syntactic conclusions they formulate. What in a learner's input signals to her what she should focus on? How do children learn about what triggers sensitive word order sequences? A learning strategy of this kind should be optimized to focus on salient information in the learner's input, utilizing her conclusions in order to make assumptions about less outstanding words and phrases. This was the original catalyst behind the *Poverty of the Stimulus* (PoVStim) argument on which generative grammar was founded (see Lasnik & Lidz 2017, Pearl 2020b for thorough details on this viewpoint). In the language acquisition literature, the most effective learning strategy concerns a way in which a child may be able to make predictions about incoming information that she is not familiar with based on what she knows about the surrounding words, a concept known as *syntactic bootstrapping* (Gleitman 1990). As summarized in Lidz (2019), referential information regarding the context of a single word is often forgotten; it is the information pertaining to the syntactic context that remains. Syntactic bootstrapping claims that learners acquire knowledge about the distributional patterns they see in their input that ultimately lead to learning the context in which a word is able to occur. It follows from this assumption that semantic information is ultimately parasitic on the syntactic use of a word. If we consider verb learning, it should come as no surprise that a child must receive enough intake of a verb and its uses in multiple syntactic contexts in order to master the fine-grained semantic distinctions that are indicated by the syntax of the adult grammar (Naigles

1996). Building on Harrigan, Hacquard, & Lidz (2016, 2018, 2019), Hacquard & Lidz (2019) show that children use not only the syntactic distribution of verbs in order to learn their meaning, but also their pragmatic context. These authors note that syntactic bootstrapping and pragmatic bootstrapping stand in a mutualistic relationship: as a child learns more about the syntactic distribution of any given word, she in turn learns about its use via the pragmatic inferences she receives based on situational contexts. As I show in Chapter 3, the repetition of the same syntactic context may lead to both a stronger pragmatic inference, as well as a more stable semantic representation in the long run (Cassani, Grimm, Daelemans & Gillis 2018:28).

1.4.1 Syntactic bootstrapping and cue-based learning

There are several approaches used in identifying the need for children to focus on highly frequent words in their input as an aid in both distributional and category learning (Mintz 2003, 2006; Mintz, Newport & Bever 2002; Cassani, Grimm, Gillis, & Daelemans, 2019; Cassani, et al. 2018; St. Clair, Monaghan, & Christiansen 2010). What all of these approaches make clear is that children use functional words such as determiners and verbs in order to anchor their learning. These words appear very frequently, with a plethora of lexical variation, and in a varied local distribution. Learners, therefore, look for cues in their syntax in order to enhance comprehension and learning. In early stages of learning, children may only be able to identify ‘edges’ of functional constituents (Fasanella 2014; Fasanella & Fortuny 2016) based on the headedness of any given language without knowing the actual meaning held by the lexical items within these phrases.

- (18) a. [DP **o** can] [VP **vai**] [PP **por** aí]
 the dog go.PRS.3SG by there
- b. [DP **o** XXX] [VP **vai**] [PP **por** XXX]

the NOUN go.PRS.3SG by PLACE

For a Galician child who has yet to learn the words *can* ‘dog’ and *ai* ‘there’, the interpretation of (17a) as (17b) has been shown to be an optimal strategy cross-linguistically for lexical learning from a computational standpoint (Chemla, Mintz, Bernal & Christophe 2009; Weisleder & Waxman 2010; Bar-Sever & Pearl 2016).

Whilst searching for patterns in which functional words may appear is a favorable manner in which to learn about one’s grammar, there are phenomena that fall outside of this learning strategy such as those related to constituent movement. Cue-based learning in general has much in common with the aforementioned bootstrapping mechanism, as both mechanisms (which may be seen as two sides of the same coin) rely heavily on the input and the variation found in it in order for children to construct a mental model of their grammar. Focusing on a child-based account of acquisition grammar change, Lightfoot (1999, 2006) outlined several key aspects of cue-based learning that should afford children the ability to view salient, target-consistent input as the means for acquiring the grammar of the language around them. He referred to cues as abstract pieces of structure in a child’s I-language, which entails that the input that a child receives is an expression of a cue if the child is able to unambiguously extract it from an input string (2006:78). What is important to note about this approach to learning is that the input itself is not a cue but rather a means by which a cue may be expressed. With respect to a salient phenomenon such as V2 structures in Germanic, a cue can be picked up by fairly simple configurations such as those in (19) and (20).

(19) Cue for OV word order: [_{VP} DP v]

(20) Cue for V2 word order: [_{CP} XP V]

According to Lightfoot, UG is equipped with the cues a child needs to tune into when observing the modeled computation of her input. Therefore, the child's job is simply one of taking note of which cue has been signaled to her by UG, after which said cue will immediately be activated in the child's I-language. In this particular instance, Lightfoot (1999:93) argues that there must be a UG requirement that the verb is obligatorily in C in this configuration.

Lightfoot's attempt to capture how a child is able to identify cues leading to structural hypotheses about her grammar seems promising, but there are several issues that must be addressed. As discussed in §1.2, parameter setting in the P&P approach was seen as affecting major parts of grammar. Lightfoot's approach to cues is similar in that it relies heavily on a fully equipped UG, the cues themselves being provided to the learner from the repertoire of UG. Even more similar to the P&P take on parameters, it also frames learning in such a way that smaller, less salient information is overlooked. Lightfoot's description for cues specifically targets large categories such as V and DP. However, as I shall explain below, a finer distinction must be made.

Children are incredibly sensitive to the distributional possibilities of the linguistic context of words throughout the grammar. Lightfoot's approach seems to be too broad in its attempt to encompass all of the discrepancies children must perceive in the input. As argued by Westergaard (2008, 2009a,b), any variation dealing with information structure would fall outside of the realm of notable cues by which this model would permit children to learn. As we have seen, not only are parameters variable across languages but it has been shown that there is considerable variation within languages, as well. In light of these facts, Westergaard proposes a *micro-cue* model in order to account for complex inter-language distinctions and opts for a finer-grained notion of cue regarding the sensitivity to which a child may react to her input. Moving

forward, I couple the work by Westergaard on micro-cues with several notions of sensitive to learning biases based on work by Fodor (1998, 2017) & Fodor & Sakas (2012, 2017) in order to derive what I consider to be an optimal learning model that is reactive to the syntactic variation I deem prime for change. Below I explain Westergaard's *micro-cue* model and the necessity for such sensitivity in the earliest stages of language acquisition. My application of this model to the acquisition process of cliticization in Galician will be left for Chapter 3.

1.4.2 Westergaard's *micro-cue* approach

Westergaard (2008, 2009a, 2009b, 2013, 2014) proffers a model of what she calls *micro-cue* learning by which a child may accurately arrive at complex decisions related to word order. Her model differs from that of Lightfoot in two important ways. First, it assumes that the cues children follow are not innate I-language abstractions, but emergent results between the interaction of the language acquisition device and the PLD. Second, it allows finer-grained details of information structure, clause type, etc. to be recorded in the developing grammar accurately. She notes that the necessity of distinguishing details finer than those of large categorial labels is key for the accurate interpretation of different word orders in any given language. In order to show this, Westergaard uses the example of V2 in Norwegian to demonstrate that micro-cues may even be specific to particular lexical items, even for mesoparameter types of input (Biberauer & Roberts 2012).

The data for V2 in Norwegian presented in Westergaard (2006, 2008) is that expected of standard V2 languages in that verb-second orders are found with subject-initial (21) and non-subject-initial declaratives (22), despite the presence of negation and adverbs.

- (21) Kristin **liker** ikke norsk vær
 Kristin likes not Norwegian weather
 ‘Kristin doesn’t like Norwegian weather.’

- (22) Italiensk mat **spiser** hun ofte
 Italian food eats she often
 ‘She often eats Italian food.’ (Westergaard 2008:1847)

For Westergaard (2008, 2014), the variation amongst V2 or non-V2 orders is dependent on language-specific preferences based on the XP found in sentence-initial position in main clauses. In Norwegian, she claims that factors such as clause type (declarative vs. *wh*-questions), the morphophonological size of the sentence-initial constituent (phrasal vs. monosyllabic *wh*-words), and information structure (new vs. focused information, particularly for subjects) are all noteworthy components in determining whether the matrix clause will yield a V2 order or not. The cue offered by Lightfoot in (19-20), therefore, would simply not be specific enough for children to accurately arrive at the distinctions mentioned above. Westergaard (2014:36) claims that word order is not the only acquisition goal for children but that the contexts in which a certain word order may appear is highly relevant. This statement matches what I highlighted previously regarding the types of statistics taken on lexical items and word strings. For a Norwegian child, micro-cues such as those in (23-25) are necessary in order to derive the V2 mesoparameter.¹⁴

- (23) V2 with monosyllabic *wh*-words
- a. [IntP [Int [*wh*] [TopP [Top [V ... XP] ...]]]]
- b. **Korsen** bil kjøpte du
 how/which car buy.PST you
 ‘Which car did you buy?’

¹⁴ For a more profound list of V2-related phenomenon triggers, see Westergaard (2009a:60).

(24) V2 with subject shift

- a. [_{TopP} DP_[-FOC] [_{Top} V ...]]
- b. **Han** svarte ikke
 he answer.PRS not
 ‘He didn’t answer.’

(25) V2 in declaratives

- a. [_{DeclP} XP [_{Decl} V ...]]
- b. **da** sitt den så langt unna at den
 then sits.PRS it so far away that it
 ikkje kommer borte
 not comes.PRS touching
 ‘Then it sits so far away that it doesn’t touch it.’

Westergaard’s distinctions for V2 micro-cues in Norwegian is based on an expanded left periphery (Rizzi 1997) with movement of each constituent specified by a specific feature to a particular function head within the C-domain. Her basis for this comes from the data present in Westergaard (2009a) where she shows the accuracy with which children just over two years of age are able to acquire the target V2 and non-V2 orders without any overgeneralization. She assumes an approach resembling the One Feature-One Head (1F1H) architecture (Kayne 2005; see Putnam 2020 for a detailed review) in which children may only map one feature X (attached to a particular lexical item) to the functional category X.¹⁵

¹⁵ As Putnam (2020) notes, the full 1F1H approach has many aspects related to nanosyntax (Starke 2009, 2011) that are not supported by minimalist assumptions. My reason for pointing it out here is the parallel between Westergaard’s claim for a fine-grained acquisition model of complex word order and how children may simplify to extreme measures when learning these phenomena by reducing movement of a specific class of constituents with feature X to functional head X along the clausal spine. In Chapter 3 I show that children may undergo repair in consolidating the features/heads related to certain movement.

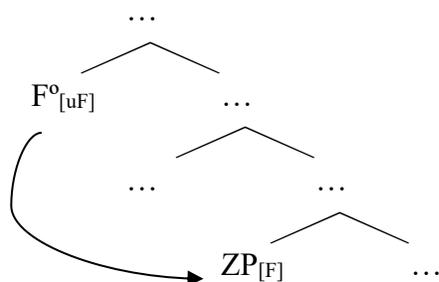
A point of contention in the adult grammar, however, shows that the adverb *kanskje* ('maybe') freely produces both V2 and non-V2 orders in regular declarative sentences with no apparent change in meaning or emphasis, whilst other non-subject-initial constituents require V2 in standard declaratives. Although she does not directly endorse Kroch's (1989) notion of "grammar competition" as he describes it, she indicates that there is clearly a situation of competition between the V2 and non-V2 orders with a sentence-initial *kanskje*. This situation, she explains, is primed for change: when learners are faced with two (or more) strings that show no semantic or pragmatic difference, the choice as to which syntactic option children take forward as the preferred one is truly an endogenous choice and may vary from speaker to speaker across generations. This is the same situation I address for Galician in Chapter 3.

Westergaard's micro-cue model has much in common with the notion of syntactic bootstrapping described above. If children must sort out their perceptual input in such a way that their acquisitional intake provides their developing grammar with subtle distinctions as to when a certain string is used or preferred over another one based on a particular linguistic context, finer-grained notions than major categorial labels is certainly needed. Moreover, certain phenomena may develop sooner rather than later or vice versa depending on a number of factors such as frequency or syntactic complexity (e.g. if a derivation is based on more than one constituent movement or particular feature). Because of this, in Chapter 3 I explore how Westergaard's micro-cue approach is able to explain syntactic complexity based on more than one feature-related movement upon analyzing the learning process of the proclisis-enclisis split in Western Iberian based on findings in Costa, Fiéis & Lobo (2015).

Below, I lay out both the innate and learned aspects of phrase structure building assumed to be available to the child in order to account for the syntactic phenomena discussed subsequently in this investigation.

1.5 Minimalism and feature-based syntax

The departure from the Principles and Parameters notion of Head-Spec agreement (Chomsky 1981) in favor of the prospect of a syntax based on a series of individualized internal operations has proven to be the most profitable advancement in syntactic theory to date. Although more complex in nature in their inception (Chomsky 1995a, 1995b), the push to reduce the compositional complexity of said internal operations has resulted in structure building consisting of two basic steps: *Agree* and *Merge* (cf. Chomsky 2008, et seq.). What has remained a constant since Chomsky (2000, 2001), however, is the notion of how these operations are triggered, namely via the probe-goal model. This model serves as the basis to explain syntactic relations within a derivation, in particular the phenomenon of movement and the need to account for how certain positions come to interact within the syntax. A direct result of this dealt with what came to be known as feature-checking, an operation that was able to account for ϕ -agreement both locally and at a distance. This operation assumes that certain syntactic elements (e.g. DPs) are composed of features that are “attracted” by functional heads (e.g. T) in order to establish a structural relation, often times adjacent. In order to account for the necessity of this operation, features of two kinds were posited: on the probe, *uninterpretable* (uF), and on the goal, *interpretable* (F).

(26) *The probe-goal model*

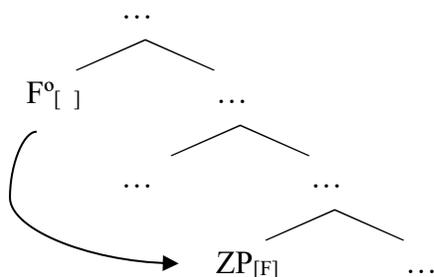
These specifications were brought about in order to account for the fact that certain features (e.g. [PARTICIPANT]) are of relevance for DPs but unmeaningful for verbs although they may be overtly expressed in their morphology (Chomsky 1995b:277-278). Not only did this mechanism force interpretable features to check the uninterpretable features in order for the latter to be deleted before reaching LF, as the failure to do so would result in ungrammaticality, but the features available within the syntax proper were reduced to [GENDER], [NUMBER], and [PERSON].

In recent theoretical work, various aspects of this derivational mechanism have been challenged. Empirical data observed across various language typologies has led to the conclusion that agreement is neither obligatory nor as cut and dried as the simple three-feature bundle. The beginnings of a new model were born in the work of Béjar (2003), Béjar & Rezac (2003), and Béjar & Rezac (2009) regarding PCC effects. These authors showed that person and number features were linked to individual probes, owing to the fact that not all probes were uniform or ever-present. Building on this observation, Preminger (2011, 2014) proposed that the syntactic mechanism not only capitalizes on the fact that not all probes are created equally and that sometimes they may fail without the ungrammatical result posited by Chomsky. Agree is obligatory when possible, thus ungrammaticality only arises when the output string is the result of a derivation not sanctioned by the grammar (Preminger 2014:9). In this thesis, I assume the specificity of Preminger's model and outline its details below.

1.5.1 Feature geometry and relativized probes

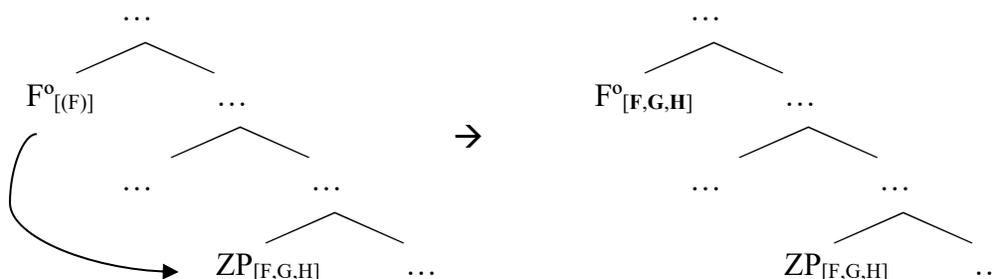
Preminger's (2011, 2014) model, referred to as the *obligatory operations* model, dispenses with the notion of *uninterpretable* features and, instead, assumes that probes may be understood as empty containers for snippets of feature geometry.

(27) *The obligatory operations model*



If F probes its c-command for the feature $[F]$, it is copied from the goal ZP onto probing head (cf. Pesetsky & Torrego 2007). What is appealing about the approach here is the idea that the probe may only search for a subset of the features on the goal, but, as Béjar & Rezac (2003) and Preminger (2014, 2019) show, all of the features borne by the goal may be copied onto the probe.

(28) *One feature probe, full ϕ -set copy*



In (28), although the probe only searches for the feature $[F]$, the entire ϕ -set of the goal ZP is copied. There are instances in which this is not the case, however, as in the *le-for-les*

phenomenon in Spanish (Ausín & Fernández-Rubiera 2017, 2021; Roca 1992, 1996; Fernández-Soriano 1999; *i.a.*).

- (29) a. Les_i enviaron los regalos a los nietos_i
 CL_{DAT.PL} send.PST.3PL the gifts at the grandchildren
- b. Le_i enviaron los regalos a los nietos_i
 CL_{DAT.SG} send.PST.3PL the gifts at the grandchildren
- ‘They sent the gifts to their grandchildren.’

The examples in (29) represent two variations found in all Spanish dialects. In (29a), the dative clitic *les* agrees with its coreferential DP *los nietos* in [NUMBER]. However, (29b) shows a lack of agreement between the clitic and the corresponding dative DP.¹⁶ In Chapter 4, I discuss examples of agreement in which this is the case, as well as others that do not follow this pattern.

Under this approach, probes are specified for certain features related to person, number, animacy, etc. that search for a viable goal to meet their featural specifications. This concept is nothing new; it is essentially the notion of *Relativized Minimality* put forth by Rizzi (1990) but applied to all probes. As Rizzi showed, relativized probing was necessary in order to account for how a probe could skip over a viable goal, as is the case in *wh*-questions.

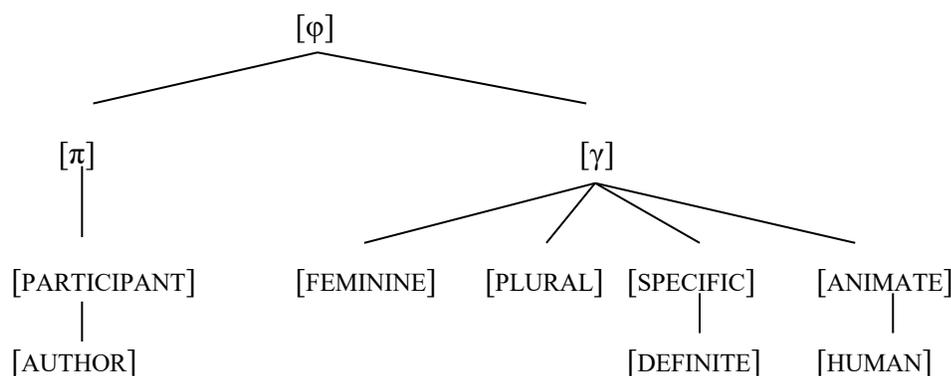
- (30) a. [C_{+Q} [deron [os mestres] [o mellor premio] [a quen]_{+Q}]]
- |-----|
- b. A quen deron os mestres o mellor premio
 to whom give.PST.3PL the teachers the best award
- ‘Whom did the teacher give the best award?’

¹⁶ One could argue, theoretically, that even regular dative clitic doubling is not true doubling considering it does not morphologically represent a full feature exponent of the doubled DP as does, for example, accusative clitic doubling.

In (30a), the *wh*-probe on C° is able to skip over the verb (*deron* ‘they gave’) and the other two postverbal nominal constituents (*os mestres* ‘the teachers’ and *o mellor premio* ‘the best award’) en route to finding a constituent with its featural specification, a [+Q] feature. This type of probing outside of interrogative phrases has been called *omnivorous agreement* (Nevins 2011) in order to account for φ -agreement more generally. However, as Preminger contends, all probes are omnivorous in that they continuously probe for a goal with their featural specifications regardless of the functional head on which they are found. The specified probing feature(s), therefore, ensure syntactic relations via agreement are always borne out.

In addition to this approach to relativized probing, Preminger also assumes, following Béjar & Rezac (2003), that the possible internal structure for goals (and, therefore, the specifications for probes) are found within a feature geometry as proposed by Harley & Ritter (2002), McGinnis (2005), *i.a.*

(31) *Feature geometry*



(adopted and modified from Kalin 2017:18)

The nodes in (31) are privative, i.e. a first-person singular noun would carry [PARTICIPANT] and [AUTHOR], whilst a second-person plural noun would bear [PARTICIPANT] and [PLURAL] but no [AUTHOR] feature. In addition, they are also monovalent, i.e. there are no [-PARTICIPANT] features or the like (*pace* Harbour 2007). Although not discussed by Preminger, viewing features in this

manner allows us to dispense with notions such as “underspecification” in relation to agreement systems. These notions are represented as the roots $[\pi]$, $[\gamma]$, and $[\varphi]$. Therefore, a probe specified for [PLURAL] and [PARTICIPANT] need not be “underspecified” for [FEMININE]; it is simply not part of the feature composition of the probe. This bodes well from a language acquisition standpoint in that learners extract information from positive evidence and do not measure a probe’s specification on what is not salient in the PLD.¹⁷

1.6 The phenomenon investigated in this thesis

As discussed in §1.1, I address the specific diachronic developments proposed for the rise of Galician C-AGR (32) moving forward.

(32) Step 1: Biclausal, prosodically divided structure

Ve-lo	aí	vai	Xan
see.PRS.2SG-CL _{ACC.M.SG}	there	go.PRS.3SG	Xan

‘You see him, there goes Xan.’

Step 2: Biclausal, prosodically unified structure

Ve-lo	aí	vai	Xan
see.PRS.2SG-CL _{ACC.M.SG}	there	go.PRS.3SG	Xan

‘You see him there, Xan is going.’

¹⁷ The lack of specific nodes for [SINGULAR], [THIRD PERSON], etc. are intended to represent cross-linguistic values that are typically seen as “default” generalizations. It is important to note that Harley and Ritter (2002:485) acknowledge that languages use only a portion of their proposed feature geometry which is far more expansive than that which I propose in (27). For some languages, it may be that [SINGULAR] or [THIRD PERSON] must specified overtly, as McGinnis (2005) argues for [DUAL]. I agree that this is a matter of the salience of particular features in the PLD and will vary from language to language.

Step 3: Monoclausal structure; reanalyzed verbal structure

Vel(o)ái vai Xan

behold go.PRS.3SG Xan

‘There goes Xan.’

Step 4: Monoclausal, main clause complementizer agreement structure

Vel(o)ái o vai Xan

behold CL_{ACC.M.SG} go.PRS.3SG Xan

‘There goes Xan.’

In Chapter 2, I present the synchronic data from what I deem Galician complementizer agreement (C-AGR), as well as the distribution of what I claim are main clause complementizers in *velai/velaqui* (‘Behold’). I examine the current approaches to the phenomenon at hand, and I show that the inadequacy of the existing syntactic accounts (Longa, Lorenzo & Rigau 1996, 1998) falls short on multiple theoretical fronts and that the explanation for the previously undiscussed syntactic dependencies of the present-day construction can only be explained within the purview of diachrony. Moreover, I show that the sociohistorical approach (Álvarez 2015) is untenable from the geolinguistic perspective assumed and presents numerous limitations in linking the supposed diachronic developments to the present-day construction.

In Chapter 3, the focus is divided into two main concentrations. First, I discuss the data in Step 1, particularly those related to cliticization. I show the enclisis and proclisis patterns of Western Iberian languages and lay out the theoretical contributions in Raposo & Uriagereka (2005). In order to explain the learning path that a child follows in order to master these cliticization patterns, I discuss the data in Costa, Fiéis & Lobo (2015) and in what aspects these data support the theoretical claims by Raposo & Uriagereka. In the second focus of the chapter, I

discuss the findings in Cassani et al. (2018) and Cassani et al. (2019) related to category learning. I show how a reanalyzed functional item may continue to be used along the clausal spine, contrasting this behavior to what we find with *velai/velaqui*. Finally, I propose that there was microvariation found in Step 2 that led to the reanalysis of *velai/velaqui* as main-clause complementizers. I outline the proposed steps based on well-known learning biases in child acquisition in order to account for the stages of reanalysis I put forward, particularly with respect to the creation of complementizers (Steps 1, 2, and 3). First, I analyze how cliticization is acquired in Western Iberian using the theoretical approach from Raposo & Uriagereka (2005) coupled with the empirical data from Costa, Fiéis, & Lobo (2015). I then explain the importance of this phenomenon for the data I focus on in this thesis. Subsequently, I discuss the details of how a multi-word string may be reanalyzed as one single lexical item from the perspective of language acquisition when children are faced with numerous input strings that are minimally variable (Hudson Kam & Newport 2005, 2009). Furthermore, I discuss the specifics of position-based reanalysis using the example of matrix-clause complementizers in Galician and Romanian (Hill 2012) in combination with work on category acquisition (Cassani, Grimm, Daelemans & Gillis 2018; Cassani, Grimm, Gillis, & Daelemans, in press).

In Chapter 4, I discuss the two proposed types of grammar change in this investigation: *exogenous* and *endogenous* change. First, I show that *exogenous* change is the change seen in cross-linguistic, widespread diachronic phenomena. This type of economical syntactic reanalysis is distinct from *endogenous* change in the sense that the latter is not a simple misparsing of the input a child receives. Instead, this is creation of grammar that cannot be explained by misparsing alone. Reviewing the study on *endogenous* change in Italian detailed by Belletti

(2017), I claim that the transition from Step 3 to Step 4 is an amalgamation of grammatical structures and is an ideal candidate for *endogenous* change.

In Chapter 5, I compare the findings in the literature regarding the differences between syntactic clitics and agreement markers, following the works of Preminger (2009, 2014, 2019), Nevins (2011), and Kramer (2014). I review the literature on complementizer agreement, analyzing the cross-linguistic behavior of this phenomenon. In addition, I consider how the three approaches to complementizer agreement (cliticization, true agreement, post-syntactic insertion) are rationalized. Using the data presented in this thesis from Galician and those from Nez Perce (Deal 2015), I show that neither may be explained via the post-syntactic insertion account in Fuß (2014). Finally, I show that Galician complementizer agreement may be explained via both cliticization and agreement.

CHAPTER 2
DATA AND THEORETICAL ANALYSES OF GALICIAN COMPLEMENTIZER
AGREEMENT

2.1 Introduction

In this chapter I present the data of what I identify as the Galician complementizer agreement (C-AGR) construction. I review the previous analyses of this construction and show the inadequacies of each, either from a synchronic or diachronic perspective. Finally, I present the complications and particularities of this phenomenon that I address in Chapters 3 and 4 regarding this construction being classified as that of complementizer agreement.

2.2 Galician data

The construction in question has several idiosyncrasies that make it difficult to understand without considering diachronic developments. However, I will provide descriptive analyses of the present-day construction in question, beginning with a complete distribution of the lexical items *velaqui* and *velaí*.¹⁸

2.2.1 The complementizers *velaqui* and *velaí*

The lexical items *velaí* and *velaqui* (1a-c) are the result of the reanalysis of the declarative strings *ve-lo aí* ‘You see it there’ and *ve-lo aquí* ‘You see it here’ (Ferreiro 1999).¹⁹ They have a strict distribution that, as we will see, leads me to classify them as matrix clause complementizers. Combining this fact with their presentative nature, meaning that they directly

¹⁸ Chapters 3 and 4 deal with my proposal for the diachronic processes that led to the present-day construction.

¹⁹ There exist as well the variations *veloaí* and *veloaquí* which are simply phonological alternations of *velaí* and *velaqui*. Their presence in the modern grammar and their influence in the diachronic processes that led to this construction will be discussed more thoroughly in Chapter 3.

address attention to something within the immediate speech act, I will gloss them as ‘behold’.

Wood, Horn, Zanuttini & Lindemann (2015:295) claim that presentative elements have five frequently found properties, all of which I show to be true for *velai/velaquí*: i) they often require direct evidence from the senses within the utterance context; ii) they often feature a deictic adverb, demonstrative, or verb; iii) they cannot be negated or questioned and (usually) cannot be embedded; iv) they often co-occur with a nominative or accusative theme; and v) they often imply the presence of an interlocutor (which may or may not be directly signaled in presentative utterance itself; e.g. via a dative morpheme). They always appear in sentence-initial position and have relatively strict selectional properties. They may select transitive (1a), intransitive (1b), and existential predicates (1c). Crucially, none of these may appear under embedding (1a’, 1b’, 1c’).

- (1) a. *Velai te-las chaves*
 behold have.PRS.2SG-CL_F.PL keys
 ‘There you have the keys.’
- a’. **Xan di que velai te-las chaves*
 Xan say.PRS.3SG COMP behold have.PRS.2SG-CL_F.PL keys
 Intended: ‘Xan says there you have the keys.’
- b. *Velaquí están teus pais*
 behold be.PRS.3PL your parents
 ‘Here are your parents.’
- b’. **Creo que velaquí están teus pais*
 believe.PRS.1SG COMP behold be.PRS.3PL your parents
 Intended: ‘I believe here are your parents.’

c. *Velaquí hai un problema.*

behold be-EXT.PRS.3SG a problem

‘Here there is a problem.’

c’. **Acho que velaquí hai un problema*

find.PRS.1SG COMP behold be-EXT.PRS.3SG a problem

Intended: ‘I find that here is a problem.’

In place of (1c), there also exists a common null-copula phrase comprised of *velaquí/velaí* and a DP. Unlike existentials which may only select an indefinite DP (2a) (Rodríguez-Mondoñedo 2007), the nominal they select may be of varying types: either definite (2b) or indefinite (2c) DP, a quantified DP (2d), a DP heading a relative clause (2e), or a predicate nominal of a small clause (2f).

(2) a. *Hai un/*o carteiro*

be.EXT.PRS a/*the mail-carrier

‘There is a/*the mail carrier.’

b. *Velaquí o carteiro*

behold the mail-carrier

‘Here is the mail carrier.’

c. *Velaí un can ceive*

behold a dog free

‘There is an unleashed dog.’

d. *Velaquí tóda-las fotos da avoa*

behold all-CL_{F,PL} pictures of-the grandmother

‘Here are all of grandmother’s pictures.’

- e. *Velaí o que me comentaches onte*
 behold CL_{M.SG} that CL_{DAT.1SG} comment.PST.2SG yesterday
 ‘There is what you mentioned to me yesterday.’
- f. *Velaí a nena contenta*
 behold the girl content
 ‘There is the happy girl.’

They cannot be used, however, in combination with a preverbal subject (3a-b), pre-verbally focused (3c), or interrogative constituent (3d) of any kind. All overt subjects must appear in postverbal position (4).

- (3) a. **Velaquí Xan vén*
 behold Xan come.PRS.3SG
 Intended: ‘Here comes Xan’
- b. **Velaquí, o polbo_i, téñen-o_i*
 behold the octopus have.PRS.3PL-CL_{M.SG}
 Intended: ‘Here they have the octopus.’
- c. **Velaí A CADELA corre (non o cadelo)*
 behold the female.dog run.PRS.3SG NEG the male.dog
 Intended: ‘Here comes the FEMALE DOG (not the male dog).’
- d. **Velaí como vai?*
 behold how go.PRS.3SG
 Intended: ‘How is he there?’

- (4) Velaquí vén Xan
 behold come.PRS.3SG Xan
 ‘Here comes Xan.’

The only possible left-peripheral element that may combine with these lexical items is the complementizer *que*.²⁰

- (5) Velaquí que chegan meus amigos.
 behold COMP arrive.PRS.3PL my friends
 ‘Here come my friends.’

Below I discuss the syntax of the appearance of *velai/velaquí* with the complementizer *que*, as the syntax of these two items co-occurring has direct relevance to the syntax of the C-AGR construction.

2.2.2 Double complementizers

In an extended CP (Rizzi 1997), complementizers in Romance may be hosted in either ForceP or/and Fin(ite)P. As shown by Demonte & Fernández-Soriano (2009, 2014), the appearance of *que* in one head or the other signals different uses at the syntax-pragmatics interface.²¹ Following these authors’ analysis, the most common complementizer, that which marks a declarative subordinate clause (6), is found in ForceP.

- (6) Sabía **que** non había ren.
 know.IMPV.3SG COMP NEG be-EXT.IMPV.3SG nothing
 ‘She knew that there was nothing.’

²⁰ There is some dialectal variation amongst speakers with respect to the combination of *velai/velaquí* and the complementizer *que*. For speakers that do not permit it, *velai/velaquí* are the only lexical items found on the lone C° head. I have yet to find a clear geolinguistic pattern. Therefore, I leave a more detailed study of this data point for further research.

²¹ Although the same line of investigation has not been done on the Force-Finite system and the use of *que* in Galician, I claim that Galician pairs with Spanish in all of the uses highlighted in Demonte & Fernández-Soriano (2009, 2014).

In contrary to the complementizer found in ForceP (38), the *que* found in FinP is said to be related to mood and emphasis in both matrix and subordinate clauses.²² This matrix-clause complementizer is often referred to as the ‘reportative’ or ‘quotative’ *que* (Etxepare 2007, 2010) which is analyzed as marking a speech event. It follows, as Demonte & Fernández-Soriano also claim, that this *que* may or may not be discourse initial and have particular requirements regarding preceding context. Because of this, the *que* that often appears in matrix clauses is said to appear in Fin as opposed to Force.

(7) - Que pasou?

what happen.PST.3SG

- Que tua nai non colle o teléfono

COMP your mother NEG grab.PRS.3SG the telephone

‘What’s going on?’

‘Your mom isn’t picking up the phone.’

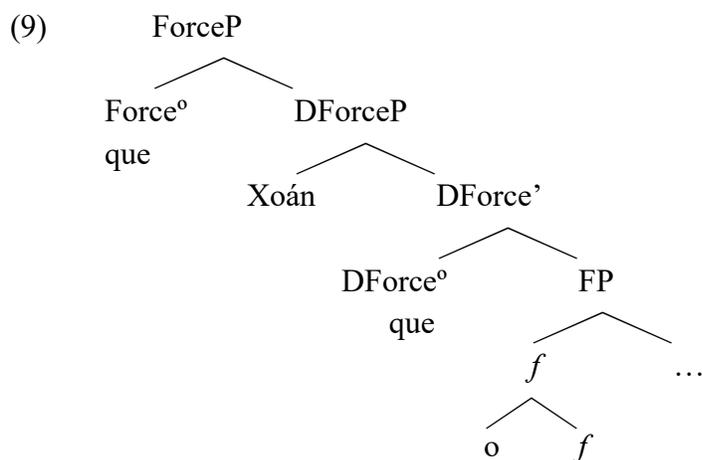
In addition to these facts, there exists the possibility of both heads being filled by *que*, a phenomenon known as *recomplementation*. Gupton (2014:218-222) claims that recomplementation in Galician does not follow the observations in Demonte & Fernández-Soriano (2009) for Spanish, but rather shows closer similarities to the DoubledForceP (DFceP) account as argued for by Martín-González (2002). For Gupton, this claim is supported based on the fact that, unlike Spanish, Galician has contrastive topics that must appear with a resumptive clitic when objects. Adapting the DoubledForceP account to Raposo & Uriagereka’s (2005) observations regarding clitic hosting locality in Western Iberian (WI), Gupton shows that the

²² As *velai/velaquí* may not be paired with the subjunctive, I leave aside the discussion of this *que* and its relation to mood.

resumptive clitic must be in a proclitic position on the verb and hosted prosodically by the second complementizer.^{23,24}

- (8) a. *Dixeron que Xoán que víron-o onte
 say.PST.3PL COMP Xoán COMP see.PST.3PL-CL_{M.SG} yesterday
- b. Dixeron que Xoán que o viron onte
 say.PST.3PL COMP Xoán COMP CL_{M.SG} see.PST.3PL yesterday

‘They said that, Xoán, they saw him yesterday.’



(Gupton 2014:219-220)

Gupton (2014:221) suggests that both contrastive topics and recursive topics in Galician must appear in [Spec,DForceP] which gives explanatory value to the reason why the constituents hosted in these projections may never serve as hosts for proclitics as they are not local enough.

As I showed in (5), the *velai/velaquí* construction permits the selection of a lower complementizer. However, on par with the observations in (3a-c) that preverbal subjects or topics of any kind are impermissible, this construction does not permit recomplementation.

²³ I review Raposo & Uriagereka’s (2005) account of cliticization in Western Iberian, as well as cliticization patterns in Galician more generally, in greater detail in Chapter 3 when I explain its advantages regarding the acquisition of clitics in Galician.

²⁴ Following Gupton (2014), I claim that FinP and FP, the projection that Raposo & Uriagereka argue as the landing site for cliticization in WI, are the same. Therefore, I use these notations interchangeably.

- (10) *Velaquí Uxía que vai
 behold Uxía COMP go.PRS.3SG

Intended: ‘There goes Uxía.’

The construction in (10) should not be confused with a similar biclausal one as in (11).

- (11) Velai Manolo, que ten moito rodaballo
 behold Manolo COMP have.PRS.3SG a-lot turbot

‘There is Manolo, [and] he has got a lot of turbot fish.’

In (1), there is a prosodic break in between *Manolo* and the complementizer *que* ‘that’. Contrary to what we find here, the grammatical examples in (1), (2), (3), and (4) must all be pronounced in one prosodic string. Were (10) pronounced in the same manner as (11), this too would be rendered a grammatical utterance.

- (12) Velaquí Uxía, que vai po-lo monte
 behold Uxía COMP go.PRS.3SG by-CL_{M.SG} hill

‘There is Uxía, [and] she is heading across the hill.’

It is from these observations that (10) and (11) should be considered relative clause constructions, with *velai/velaqui* selecting a lone DP and the subsequent clause beginning in the manner of the reportative or quotative *que* as in (7).

For speakers that permit the combination of *velai/velaqui+que* as in (5), I claim that this represents the exoskeleton of Rizzi’s (1997) extended CP: *velai/velaqui* appear in Force, whilst the second complementizer *que* is located in Fin. This seems to have to do with the selectional properties of these items on the highest functional head, as *velai/velaqui* are not able to select any left-peripheral material aside from a second complementizer *que*. What I wish to point out here is that the double complementizer construction *velai/velaqui+que* construction is in

complementary distribution with structures that may host clitics. That is, *velai/velaqui+que* may not head a structure that hosts a clitic as Gupton showed in (8) and (9) for the recomplementation phenomenon. In this sense, we may imagine that F(in)P in Galician may either a) host a complementizer *que* or b) serve as the functional head to which clitics are attracted, but not both.

The implications of this claim are supported by the data seen above. It seems to be, *pace* Demonte & Fernández-Soriano (2009), that the reportative or quotative *que* in Galician is able to appear on multiple heads: either ForceP or F(in)P.

- (13) Que non o quero!
 COMP NEG CL_{M,SG} want.PRS.1SG
 ‘(I said) I don’t want it!’

Following Gupton’s adaptation of Raposo & Uriagereka’s cliticization account to Martín-González’s DoubleForceP, I claim that instances of matrix-clause complementizers paired with cliticization patterns as in (13) must be instances of *que* in Force and the clitic *o* (and host *non*) in F(in).

- (14)
-
- ```

graph TD
 ForceP --> Force["Force°
que"]
 ForceP --> FP
 FP --> f1["f"]
 FP --> ellipsis["..."]
 f1 --> non["non"]
 f1 --> f2["f"]
 f2 --> o["o"]
 f2 --> f3["f"]

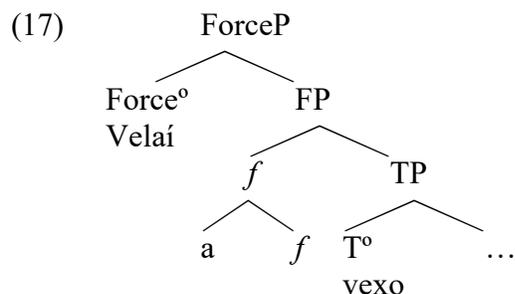
```

When Force is occupied with *velai/velaqui*, however, *que* must appear on a lower functional head, yet it still maintains its discursive properties of emphasis as in (15).

- (15) Velaquí que chega Melba  
 behold COMP arrive.PRS.3SG Melba  
 ‘Here is Melba arriving.’

This construction is illicit in cases of cliticization, however, as *que* and the clitic are found on the same functional head.

- (16) a. \*Velaí que a vexo  
 behold COMP CL<sub>F.SG</sub> see.PRS.1SG  
 Intended: ‘There I see her.’
- b. Velaí a vexo  
 behold CL<sub>F.SG</sub> see.PRS.1SG  
 ‘There I see her.’



The data in (16) and the structure in (17) further validate the claims above that *velai/velaqui* are found in Force and the selection of *que* further down in the structure must be found on F(in). Moreover, as Rizzi (1997:288) notes, the Topic and Focus fields appear on a need-be basis, which inherently implies that not all grammatical structures will license them. I claim this is the case for *velai/velaqui* hosted in C/Force.

Hill (2012) shows evidence for what she refers to as base-generated ‘verby speech act’ particles found in Romanian that she proposes are base generated in the left periphery which may

select the complementizer *că* ('that'). First, she shows that the complementizer *că* ('that') may be found in main (18a) and subordinate (18b) clauses just as I showed above for Galician.

(18) a. Hai **că** nu e bun la nimic  
 PRT COMP NEG be.PRS.3SG good at nothing  
 'Okay (obviously) he's not good at anything.' (Hill 2012:290)

b. A spus **că** vine  
 have.PRS.3SG say.PTCP COMP come.PRS.3SG  
 'She/he has said that she/he is coming.' (Hill 2012:79)

For (18a), Hill claims that these left-peripheral functional elements are not meant for *clause typing* as in (18b). That is, they do not mark a subordinate or complement clause but instead are used for *utterance typing* (assertion versus illocutionary force).

(19) a. \*Maria supne **că hai** **că** Ion nu mai vine  
 Maria say.PRS.3SG COMP PRT COMP Ion NEG more come.PRS.3SG  
 Intended: 'Maria says that Ion doesn't come anymore.' (Hill 2007:2094)

b. \*Maruxa dice **que velaquí** vén Xan  
 Maruxa say.PRS.3SG COMP behold come.PRS.3SG Xan  
 Intended: 'Maruxa says that here comes Xan.'

Much like I have shown for *velai/velaqui* (1a',b',c'; 19b), *hai* cannot be embedded (19a). Thus, I follow the observations made by Hill in proposing that the presentative nature of *velai/velaqui* serves the function of *utterance typing*.<sup>25</sup> Hill compares the c-selection features of verbal predicates that probe for a particular complement to those functional items that mark a particular type of utterance such as exclamatives, vocatives, and ostensives which falls in line with what I

<sup>25</sup> Based on what I show in Chapter 3 and what Hill (2012, 2013) proposes for several phenomena in a detailed cross-linguistic analysis, utterance typing seems to be restricted to main clause phenomena.

have shown of the distributional behavior of *velai/velaqui*.<sup>26</sup> Furthermore, that these functional elements appear in ForceP, above the Topic and Focus fields of the left periphery, is a concomitant of the fact that we may distinguish lexical items used in conversational pragmatics (such as those mentioned above) from those of discourse pragmatics (e.g. topics, focus constituents). This is what I claim for *velai/velaqui*, as well.

Moving forward, I shall assume that *velai/velaqui* are base-generated functional items found in C (Force). This will be crucial for explaining the data found in the subsequent sections.

### 2.3 Previous analyses

In this subsection, I examine two accounts regarding what I have called the Galician complementizer agreement phenomenon. The first, a synchronic syntactic account by Longa, Lorenzo & Rigau (1996, 1998), examines a series of phenomena linked to what they call “clitic recycling.” As we will see, although uniformitarian in principle, their account fails to explain several aspects of the structural relation of the functional elements involved, primarily those of the complementizers previously reviewed above and the appearance of the complementizer agreement morpheme. The second, a socio-historic account by Álvarez (2015), attempts to explain the diachronic pathway by which this construction arose. I shall show that the arguments she makes have no viable links from any sociohistorical, sociolinguistic, or purely dialectal landscape.

#### 2.3.1 Longa, Lorenzo & Rigau (1996, 1998)

The work done by Longa, Lorenzo & Rigau (LLR) is the only generative approach to this construction and centers primarily on the idea that languages ‘recycle’ clitics in their grammar in order to account for ‘missing’ clitics that are specified for a particular function in other

---

<sup>26</sup> In Chapter 3, I return to more elaborate work by Hill (2013) on ostensives, as this class of elements is directly related to the identification of *velai/velaqui* as presentatives.

languages. They claim that languages like Galician, Asturian, and Spanish do this frequently in order to compensate for a smaller clitic repertoire than Catalan and Italian, namely focusing on the substitutions for the locative *hi* and the partitive *en*. LLR (1998:149) propose this formally as in (20).

(20) *Clitic recycling strategy*

Use the (most) unmarked clitic to fill in gaps in the system.

Due to the fact that Galician, Asturian, and Spanish lost the aforementioned locative and partitive clitics during the sixteenth century (Moraes de Castilho 2005), LLR claim that they must account for these ‘gaps’ in their respective clitic paradigms by utilizing clitic forms that, in principle, are typically used elsewhere. These may be seen in Table 1.

*Table 1. Clitic paradigms in Cat., Gal., Ast., & Span.*

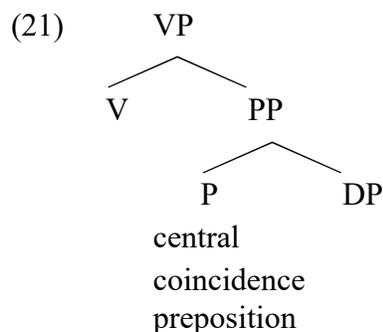
|            | Catalan              | Galician     | Asturian             | Spanish          |
|------------|----------------------|--------------|----------------------|------------------|
| ACCUSATIVE | el, la, ho, els, les | o, a, os, as | lu, la, lo, los, les | lo, la, los, las |
| PARTITIVE  | en                   | -            | -                    | -                |
| LOCATIVE   | hi                   | -            | -                    | -                |

The core argument behind LLR’s theoretical account lies in Galician, Asturian, and Spanish pulling from their accusative clitic paradigms in order to fill in the missing locative and partitive clitics possessed by Catalan.

### 2.3.1.1 <HAVE> and clitic recycling

Their claim to this approach begins with the licensing of different clitics with the existential <HAVE> and the stative <BE>, as these two predicates are derived from the same root.<sup>27</sup>

Following Rigau (1995), they argue that the shared syntactic composition of each verb is as in (21).



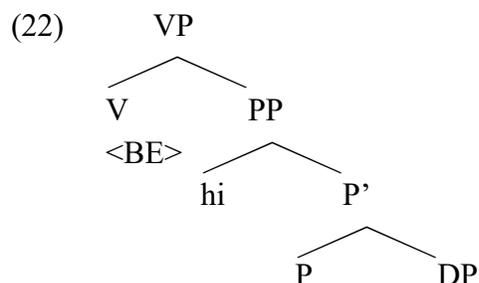
In LLR’s approach, V represents the abstract root. Following Hale & Keyser (1993), they claim that the spell-out of the verb as either <BE> or <HAVE> is the result of the incorporation (or lack thereof) of the P, identified as a central coincidence preposition. Hale & Keyser’s observations come from their view on Lexical Relational Structure (LRS), which is seen as “an unambiguous system of structural relations holding between the head, its categorial projections, and its arguments” (1993:53). For LLR’s model in particular, if there is no incorporation of P to V, then the spell-out is <BE>. If there is, <HAVE> is rendered.

Beginning with <HAVE>, LLR identify differences between Catalan and its neighboring languages Spanish and Galician. The first difference to note is that Catalan exhibits a locative clitic *hi* that is separate from the verb stem.<sup>28</sup> LLR account for this by claiming that this clitic is

<sup>27</sup> Technically, they claim that these verbs derive from “a single abstract verb” (1998:125). However, based on the details of their argument that “V has no relevant meaning” (1998:127) and the general notion of roots in generative theory (Marantz 2013 *i.a.*), this can be equated to the same abstract root selected by different functional heads.

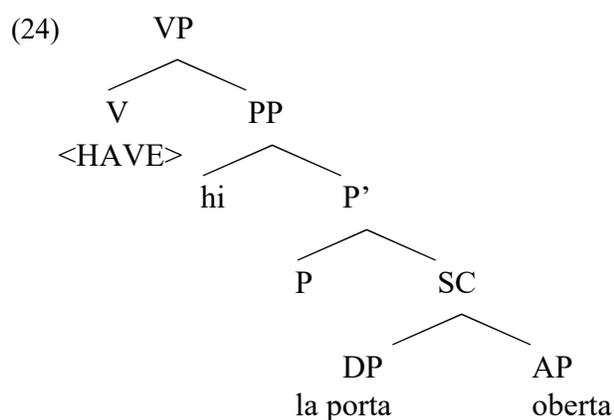
<sup>28</sup> It should be noted that this clitic appears fused onto the present tense existential forms in Galician and Asturian (*hai*) and Spanish (*hay*). However, these languages use the regular conjugation (*había*) of the verb *haber* in the imperfect tense, for example, with no trace of the clitic.

generated in a separate overt head sandwiched in between V and P in the syntactic structure of the existential <HAVE>.



LLR refer *hi* as “the subject of the predication” of the complement DP or small clause, as in (23), claiming that its lack of person features permits the impersonalization of the existential construction *hi ha* ‘there is/are’.

- (23) a. A casa, sempre hi ha [SC la porta oberta]  
 at home always CL<sub>LOC</sub> be-EXT.PRS.3SG the door open  
 ‘At home, the door is always open (there).’ (LLR 1998:130)



LLR claim that Galician, Asturian, and Spanish share a silent version of this clitic, <HI>, base generated in the same position, undergoing the same movement, albeit covertly, to T as *hi* in Catalan.<sup>29,30</sup>

<sup>29</sup> LLR use AGRsP, which I interpret as TP based on arguments made by Hornstein Nunes & Grohmann (2005:161-169).

The second difference between these languages is that <HAVE> in Catalan also bears another distinct clitic, the partitive *en* (25).<sup>31</sup> Galician (26) and Asturian utilize the recycling strategy proposed by LLR and utilize morphologically accusative clitics in order to fill in the gaps in their clitic paradigms. Distinctly from both of these groups, Spanish rejects the use of a clitic altogether (27).

(25) a. No hi ha pa.  
 NEG CL<sub>LOC</sub> be-EXT.PRS.3SG bread

b. No n'hi ha.  
 NEG CL<sub>PART-CL<sub>LOC</sub></sub> be-EXT.PRS.3SG

(26) a. Non hai pan.  
 NEG be-EXT.PRS.3SG bread

b. Non o hai.  
 NEG CL<sub>M.SG</sub> be-EXT.PRS.3SG

(27) a. No hay pan.  
 NEG be-EXT.PRS.3SG bread

---

<sup>30</sup> Although LLR do not specify, the assumption here is that first the verb is moved to T, then *hi* is probed. This would derive the necessary order *hi ha* ('there is/are') as in (31).

<sup>31</sup> LLR fail to mention where the partitive *en* is base generated. Although this is not immediately problematic for the part of their analysis that I argue against, these authors do not make it clear what the structural relation would be between *hi* and *en* in <BE> (i), <HAVE> (ii), or other transitive (iii) constructions.

- i) No n'hi ha  
 NEG CL<sub>PART-CL<sub>LOC</sub></sub> be.EXT.PRS.3SG  
 'There isn't any. (e.g. bread, money, etc.)'
- ii) Hem de fer-n'hi.  
 have.PRS.1PL of do.INF-CL<sub>PART-CL<sub>LOC</sub></sub>  
 'We have to do it there. (e.g. take a trip to London)'
- iii) N'hi lliureu uns quants?  
 CL<sub>PART-CL<sub>LOC</sub></sub> deliver.PRS.2PL ones some  
 'Are you delivering some to them?'

- b. No (\*lo) hay.  
 NEG CL<sub>M,SG</sub> be-EXT.PRS.3SG

In conjunction with these data, LLR show that in Catalan (28a,c) but not in Spanish or Galician (28b,d) <HAVE> may select an indefinite or definite DP.

- (28) a. Hi havia un flequer  
 CL.LOC be-EXT.IMPV.3SG a baker  
 b. Había un panadeiro  
 be-EXT.IMPV.3SG a baker  
 ‘There was a baker.’  
 c. Hi havia el flequer  
 CL.LOC be-EXT.IMPV.3SG the baker  
 d. \*Había o panadeiro.  
 be-EXT.IMPV.3SG the baker  
 ‘The baker was there.’

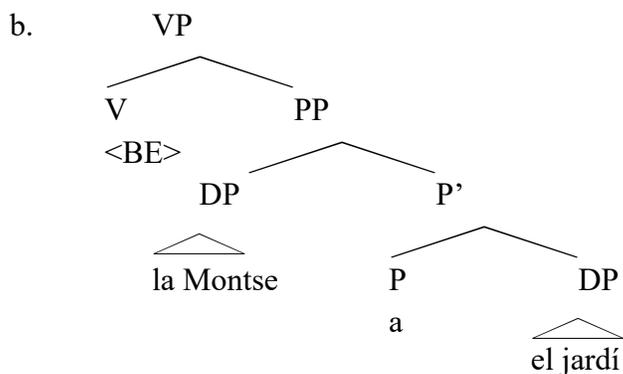
Based on the appearance of the partitive clitic in Catalan coupled with the fact that in Galician, Asturian, and Spanish <HAVE> is unable to select a definite complement, LLR conclude that <HAVE> assigns partitive case. Because Galician, Asturian, and Spanish do not have a clitic solely dedicated for partitive case, LLR claim that these languages utilize their morphological accusative clitic patterns in order to mark this argument.

### 2.3.1.2 <BE> and clitic recycling

Their proposal for <BE> follows a slightly different analysis. As opposed to P incorporating into V as proposed for <HAVE>, the preposition linked to <BE> surfaces independently. For

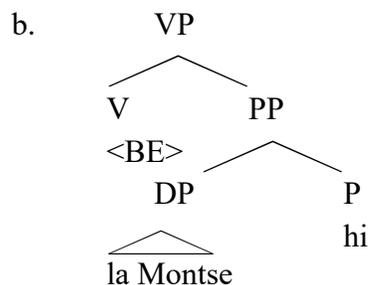
Catalan, the stative verb *ésser* ‘be’ selects a PP complement in which an argument DP is embedded as in (29b).

- (29) a. La Montse és al jardí  
 the Montse be-STAT.PRS.3SG to-the garden  
 ‘Montse is in the garden.’



It is important to note that <BE> is not considered a typical predicate but works as “a pure auxiliary unit” (LLR 1998:145), whilst the subject *la Montse* and the PP *el jardí* are in a predicative relationship. This is a common approach in work on copulas going back to Abney (1987). Unlike <HAVE> which boasts a partitive clitic, here <BE> may license the locative clitic *hi*. In this instance, LLR claim that *hi* heads P and substitutes the prepositional portion of the predication as in (30).

- (30) a. La Montse hi és.  
 the Montse CL<sub>LOC</sub> be-STAT.PRS.3SG  
 ‘Montse is there.’



As Figure 1 shows, Galician, Asturian, and Spanish do not boast a special locative clitic in their paradigms. Instead, LLR argues that Galician and Asturian, but not Spanish, once again recycle their accusative clitic for this function.

- (31) a. Les    muyeres    taben                    xunto    la figal    pero    los paisanos  
           the    women    be-STAT.PRS.3PL    together    the fig-tree    but    the men  
           nun    lo            taben.  
           NEG    CL<sub>NEUT.SG</sub>    be-STAT.PRS.3PL
- b. As    mulleres    estaban                    xunto    á    figueira    pero    os homes  
           the    women    be-STAT.PRS.3PL    together    to-the    fig-tree    but    the men  
           non    o            estaban.  
           NEG    CL<sub>M.SG</sub>    be-STAT.PRS.3PL
- c. Las    mujeres    estaban                    junto    a    la higuera    pero  
           the    women    be-STAT.PRS.3PL    together    to    the fig-tree    but  
           los hombres    no            estaban  
           the men            NEG    be-STAT.PRS.3PL
- ‘The women were beside the fig tree, but the men weren’t.’

(LLR 1998:147)

LLR stress that their clitic recycling strategy defined in (20) plays an important role here, as the clitic used in this construction is the least marked in each respective language that utilizes it. The

most accurate representation of this is in Asturian (31b): this language uses its neuter pronoun *lo* as opposed to the masculine singular *lu*. Galician, which does not have a morphologically independent accusative clitic for neuter marking, utilizes the common masculine singular *o*.

### 2.3.1.3 Modal expressions and clitic recycling

Building on their approach to <BE>, LLR take advantage of their clitic recycling strategy in identifying what I have deemed C-AGR in Galician as yet another instance of locative-related clitic use. These authors claim that this construction is similar to those found in the Northern Italian dialects (NIDs), and, therefore, label this clitic “a nominative-related subject clitic” (LLR 1998:153). They identify a few key components of this construction, the primary being the use of “deixis ad oculos” (Bühler 1934 *apud* LLR 1996, 1998) in which the “subject clitics in this language are only permitted if the sentence has a very special modal orientation: that of an exclamatory deictic (LLR 1998:156). This modal orientation, they claim, deals with the emphasis of the sentence given by the locative elements *aí/aquí* that head this construction and, more importantly, the fact that these adverbs “do not have a locative meaning” (LLR 1998:159). Instead, they deem them “emphasizer operators” in that they introduce deixis on behalf of the utterance as a whole. This idea is supported by the contrast in meaning in (32).

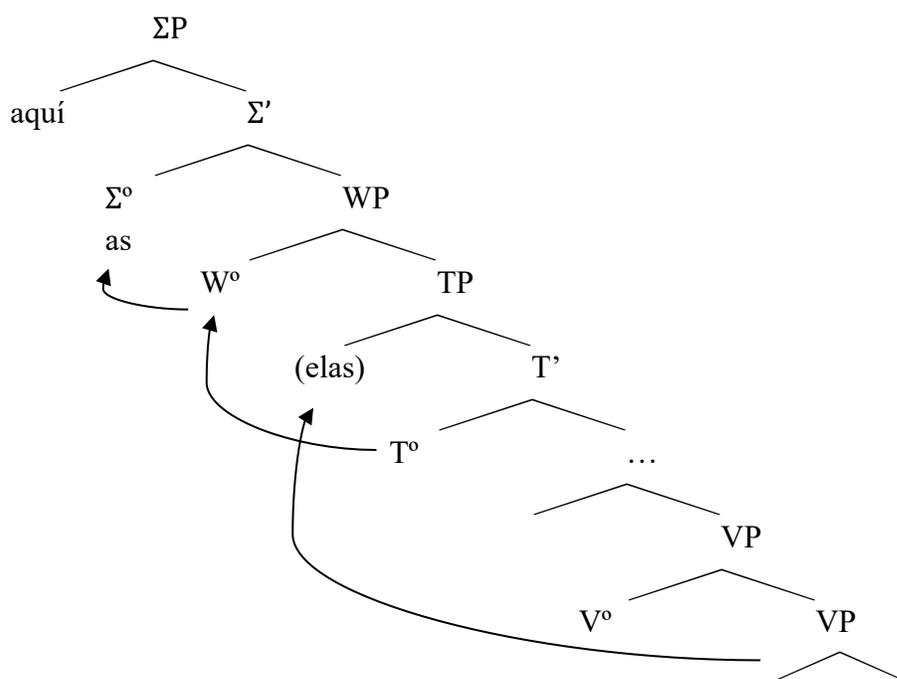
- (32) Aquí as veñen elas ≠ Aquí é onde veñen elas  
 here CL<sub>F.PL</sub> come.PRS.3PL they here be.PRS.3SG where come.PRS.3PL they  
 ‘Here they come!’ ‘Here is where they are coming!’

LLR distinguish the phrase headed by *aquí* with a “subject clitic” from the biclausal construction on the right based on the fact that in the former, but not the latter, the adverbial “is nothing more than a marker of the emphatic-deictic meaning” (LLR 1998:59) of the phrase. The difference, they claim, is that the locative adverbial as an “emphasizer operator” never participates in a

syntactic relation within or around the VP, in contrast to that of normal locatives. Therefore, they conclude that these elements are base-generated high in the structure, in [Spec,  $\Sigma$ P], a projection proposed by Laka (1990) for certain emphatic functions.<sup>32</sup>

LLR identify many of the same properties that I showed in §2.2. Therefore, I will not review them. Their theoretical hypothesis as to the generation of this clitic, however, is related to its function as a nominative marker. They claim that it must generate under T as “a manifestation of the subject agreement features of the sentence” (LLR 1998:159).

(33) *Aquí as veñen elas* (adopted and modified from LLR 1998:160)



These authors propose that the clitic moves from T to a head specified to host clitics, W (for Wackernackel), and then moves to  $\Sigma$  in order to aid in the modality expression alongside *aquí*. The lexical subject then moves to [Spec,TP] at LF, assuming that it must be in a Head-Spec relationship with the verbal predicate.

<sup>32</sup>  $\Sigma$ P is also the projection claimed by Martins (1993, *et seq.*) that attracts clitics in European Portuguese.

Their final puzzle deals with the question of why only verbal predicates with a locative meaning may be used in this construction. Their argument revolves around the notion that only locative verbs may be accompanied by these “emphasizer operators” because they are “the only ones with which a central coincidence expression like *aquí* cannot be naturally interpreted in its original locative sense” (LLR 1998:160). That is, verbs such as *ir* ‘go’ and *andar* ‘walk’ have a locative meaning built into their semantics, which “contradicts the central coincidence meaning of *aquí*.” Therefore, transitive examples as in (34), are illicit because *aquí* is always interpreted as a locative adverb.

- (34) *Aquí* (\*o) ten as mazás (Xan)  
 here CL<sub>M.SG</sub> have.PRS.3SG the apples Xan  
 ‘Here he (Xan) has the apples.’

Based on the fact that only the “emphasizer operators” are able to head expressions with these “subject clitics” related to the emphatic modality of the utterance, they should never arise in these constructions because *aquí* has a true locative meaning in these instances. They also claim that non-locative elements may be used in this construction:

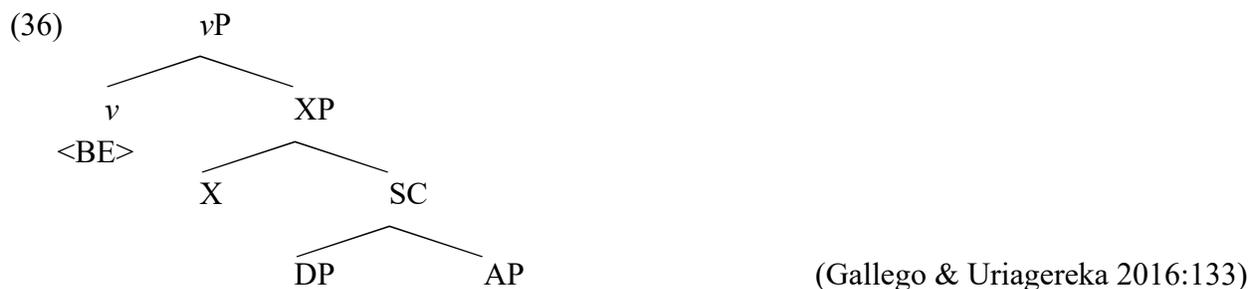
- (35) *Xa* as chegan elas!  
 now CL<sub>F.PL</sub> arrive.PRS.3PL they  
 ‘Now (here) they come!’

Their explanation for (35) is that *xa* ‘now’ acts as an “emphasizer operator” that marks the emphatic-deictic meaning of the utterance as a whole.

#### 2.3.1.4 Where clitic recycling has it wrong

Here I call into question the use of “clitic recycling” for locative purposes as hypothesized by LLR above. Because these authors tie in the use of what I have called the Galician C-AGR

construction with the data in 2.3.1.2, this must be addressed first. The first issue worth examining is their account of the structure <BE> and the supposed recycling of accusative clitics in Galician and Asturian locative constructions. The structure provided in (29b) is better viewed as in Gallego & Uriagereka (2016) in which <BE> selects a small clause complement.



Gallego & Uriagereka's investigation is based on distinguishing the stage-level predicate *estar*, the temporary-locative copula in Galician, Asturian, and Spanish from the individual-level predicate *ser*, the copula used with immutable attributive adjectives. Because LLR do not address this difference, I shall not lay out the specifics regarding this point of investigation.<sup>33</sup> In Gallego & Uriagereka's approach, X incorporates into *v*, yet not in the same way the central coincidence preposition does in LLR's approach. Here, X is an abstract head that changes the status of <BE> from *ser* to *estar* in Galician, Asturian, and Spanish, which has locative and temporary adjectival meanings.<sup>34</sup> The two daughter constituents of the small clause are pair-merged in order to account for the predication relationship they share. What LLR fail to recognize is that predicate substitution in *estar* copular constructions in Galician and Asturian are for transient adjectival arguments only (37a-b) and do not apply to locative phrases (37c-d) as they show for Catalan (38c-d). In the latter instances, no clitic is permitted in Galician, Asturian, or Spanish.

<sup>33</sup> I also leave out differences between predicational and equational clauses in *ser* and *estar*.

<sup>34</sup> It is important that for our examples here Catalan only uses the verb *ésser* for locative functions compared to *estar* in Galician, Asturian, and Spanish. The verb used in transient predication, however, is *estar* for all four languages.

(37) a. Xán está contento (Galician)

Xán be-STAT.PRS.3SG happy

‘Xan is happy.’

b. Estáo.

be-STAT.PRS.3SG-CL<sub>M,SG</sub>

‘He is (that).’

c. Xán está en Foz

Xán be-STAT.PRS.3SG in Foz

‘Xan is in Foz.’

d. Está(\*o).

be-STAT.PRS.3SG-CL<sub>M,SG</sub>

‘He is (there).’

Moreover, LLR fail to mention that Catalan also substitutes predicational arguments of the subject with the neutral accusative clitic *ho* (38b). However, this clitic is used in predicational clauses only and is not interchangeable with the locative clitic *hi* (38b,d).

(38) a. Carlota está malalta (Catalan)

Carlota be-STAT.PRS.3SG ill

‘Carlota is ill.’

b. \*Hi está / Ho está.

CL<sub>LOC</sub> be-STAT.PRS.3SG CL<sub>NEUT,SG</sub> be-STAT.PRS.3SG

‘She is.’

c. Carlota és a Badalona.

Carlota be-STAT.PRS.3SG at Badalona

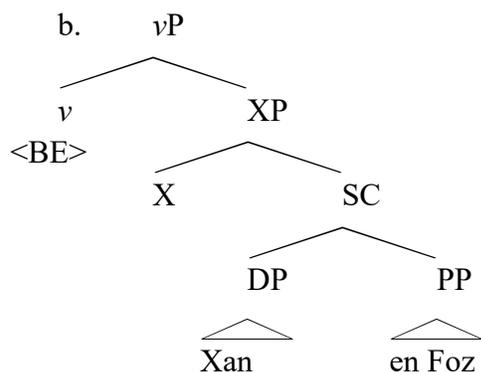
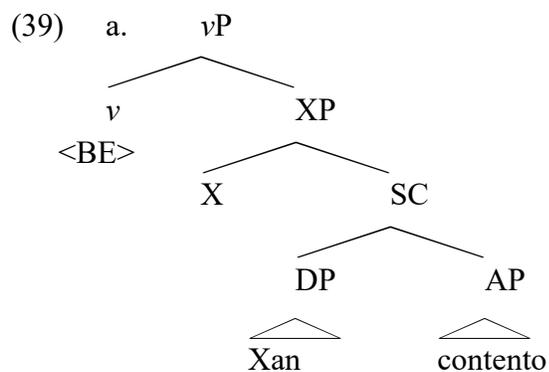
‘Carlota is in Badalona.’

d. Hi és / \*Ho és.

CL<sub>LOC</sub> be-STAT.PRS.3SG CL<sub>NEUT.SG</sub> be-STAT.PRS.3SG

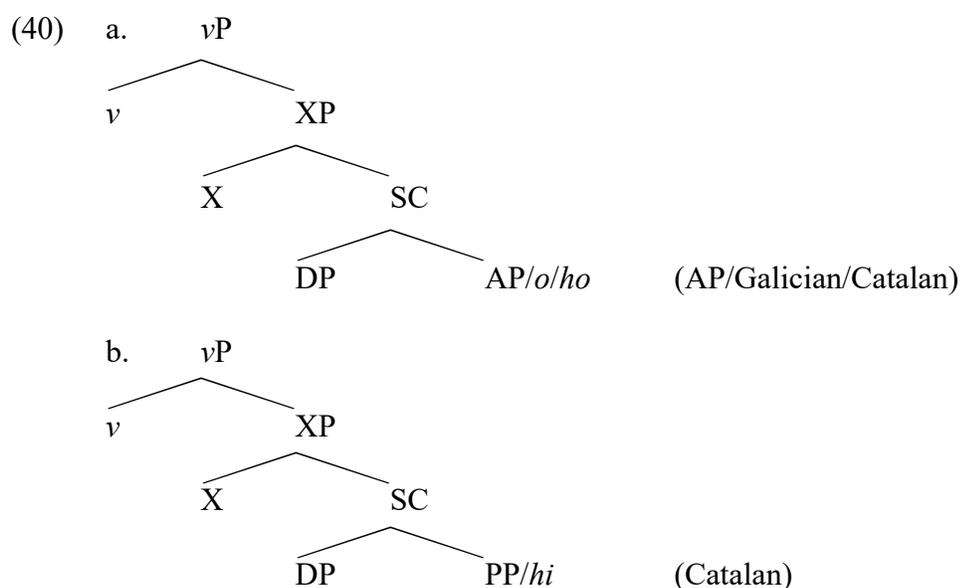
‘She is (there).’

For Galician, Freixeiro (2006:128) claims that the ability for pronoun substitution comes from the arranged attributes used in predicational clauses. Even though both an adjectival phrase (39a) and a prepositional phrase (39b) are sisters of the subject DP within Gallego & Uriagereka’s small clause model, only the adjectival phrase is available for pronoun substitution because of its status as the predicating argument of the subject DP. The examples in (37a) and (37c) are modeled subsequently in (39a) and (39b).



Although there is no syntactic predication between the two daughter constituents of the small clause in (39b), Gallego & Uriagereka claim that this small clause configuration is still necessary in order to derive the correct thematic relation between the subject and the prepositional phrase.

Based on the data in (37) and (38), I conclude that while a form of “clitic recycling” may take place, it does not proceed as explained by LLR. In Galician, Asturian, and Spanish only the predicational configuration in (40a) is available for pronoun substitution. For Catalan, both (40a) and (40b) are acceptable structures, with the latter serving solely for locative phrases.



In (40a), the predicating argument of the subject may appear as a full-fledged AP as in (37a, 38a) or the “recycled” argumental clitic in (37b, 38b). Equally, a non-predicating constituent may either surface as a PP or, in Catalan, as the locative clitic *hi*.

The account in LLR introduces several fundamental problems not only with accounting for the C-AGR construction in Galician but viewing clitic recycling as a whole in the manner described in sections 2.3.1.1-2.3.1.3. As I have shown, LLR’s approach to “clitic recycling” with respect to locative constructions does not support the empirical data I have laid out above and, therefore, is unsuitable as the basis for describing the Galician C-AGR construction.

### 2.3.1.5 Neither subject clitics nor emphasizer operators

Regarding LLR's hypothesis (§2.3.1.3) in which only sentences with a locative meaning may take *ai/aquí* as “emphasizer operators,” I believe that their views on the licensing of this construction overgenerate. Were all that is needed in order to license this construction, an intransitive verb and an “emphasizer operator,” one would expect the matrix-clause *que* (presented in §2.2.2) to be able to appear in this construction, contrary to fact.

- (41) - U-lo                      Brais?  
           where-CL<sub>M.SG</sub>        Brais  
 - \*Que o                    vai                    teu            neno!  
           COMP CL<sub>M.SG</sub>        go.PRS.3SG        your        child  
           ‘Where is Brais?’

Intended: ‘There goes your son!’

As described by Extepare (2007, 2010), the complementizer *que* in (41) belongs in the structure of the entire utterance as opposed to when it serves its more common function of marking relative clauses. This is precisely what LLR propose for the non-locative *xa* in (35): as opposed to modifying the LRS of the verb, it adds emphasis to the entire utterance, which explains its lack of locative function.

First, my informants accept neither *xa* in (35) nor *que* in (41) as a licenser of the C-AGR construction.<sup>35</sup> Although I agree with the emphatic nature of the construction identified by LLR, I do not agree with the idea that any emphatic functional element may license the clitic in question. Moreover, their claim that the “emphasizer operator,” a function word, base generates in a specifier position is an odd claim in minimalism. Base-generated functional elements in the

---

<sup>35</sup> I have discussed the synchronic data presented in this thesis with more than one dozen native Galician speakers that live throughout the area in which this phenomenon is used. All identify Galician as their native language and come from families in which Galician is the primary language spoken.

left periphery show a pattern of always being heads, as they are functional items related to discourse setting and typically have relatively little structure to them. Before addressing the diachrony of how this phenomenon happened in Chapters 3 and 4, it is important to emphasize that the empirical data show only four elements with the ability to license this construction: *velaqui*, *velai*, *aquí* and *ái*.

#### 2.3.1.5.1 Identifying subject clitic paradigms in Romance

The last issue I wish to shed light on is that of the status of these clitics as “subject clitics.” LLR cite Poletto (1993) and Brandi & Cordin (1989) for several Northern Italian dialects (NIDs), but much work over the past decade has revealed a significant amount on this topic that can shed light on reanalyzing subject clitics in Romance. Roberts (2014:196-197) identifies four types of subject clitic paradigms in Romance: fully redundant, null-subject (42), non-null-subject (43), non-redundant null-subject (44), and complementary (45). These categorical types come from Roberts’ focus on both morphophonological distinctions within the subject clitic system as well as that of the verbal paradigm. These paradigms are shown in Roberts (2014:196-197) as such:

|      |              |                   |                                                |
|------|--------------|-------------------|------------------------------------------------|
| (42) | (SG)         | (PL)              | (Florentine (Italian): <i>parlare</i> ‘speak’) |
|      | (E) parlo    | Si parla          |                                                |
|      | Tu parli     | Vu parlate        |                                                |
|      | E/la parla   | E/le parlano      |                                                |
| (43) | (SG)         | (PL)              | (French: <i>dormir</i> ‘sleep’)                |
|      | Je dors      | Nous dormons      |                                                |
|      | Tu dors      | Vous dormez       |                                                |
|      | Il/Elle dort | Els/Elles dorment |                                                |

(44) (SG) (PL) (Como (Italian): *dormir* ‘sleep’)

(-) dərmi (-) dərnum

ta dərmat (-) dərnuḞ

al/la dərma (-) dərmaḡ

(45) (SG) (PL) (Carrara (Italian): *dormir* ‘sleep’)

a dərma a durmiḡ

tə dərma (-) durmitə

i/al dərma i/al dərmaḡ

In (42), there is a two-way syncretism between both the third-person singular and the first-person plural inflections and the third-person singular and plural subject clitics. On the basis of agreement, this paradigm is fully redundant and, thus, is a null-subject variety. The French data in (43) show a fully-fledged subject clitic paradigm but a three-way syncretism in the singular first-, second-, and third-person verbal inflections.<sup>36</sup> Specifically regarding this clitic paradigm, it has been analyzed by Kayne (1991) and Rizzi & Roberts (1989) as one of *phonological* subject clitics. However, its syntax is argued to behave as that of full DPs. The Italian variety Como in (44) shows a non-redundant verbal paradigm capable of licensing null-subjects with only two clitic forms available for the second- and third-person singular. Finally, the Carrara variety (45) shows impoverishment of both the subject clitic and verbal systems. The first has a “three two-way” syncretism (singular and plural first-persons; singular and plural third-person masculine; singular and plural third-person feminine), and the latter has “one three-way” syncretism (first-, second-, and third-person singular). Roberts (2014:197) claims that due to the overlap of the two

<sup>36</sup> Roberts (2014:197) notes that the final /t/ and /s/ are both silent in pronunciation, making these three forms indistinguishable despite the orthographic difference.

paradigms, there are six distinct pairs, which feed the ability to license null subjects in this variety.

Comparing the data to what I have presented for Galician in §1.1, modified and repeated below (46-47) there are numerous reasons to believe that the clitic in the Galician C-AGR construction (in bold) is not a subject clitic but an agreement marker.

- (46) a. Velaquí **\*o/a** está ela!  
 behold  $CL_{M.SG}/CL_{F.SG}$  be.PRS.3SG she  
 ‘There she is!’
- b. ‘Velaquí **\*as/os** andan os rapaces!  
 behold  $CL_{F.PL}/CL_{M.PL}$  walk.PRS.3PL the boys  
 ‘There go the boys!’
- (47) Velaquí **os** estamos nós!  
 behold  $CL_{M.PL}$  be.PRS.1PL we  
 ‘Here we are!’

First, a postverbal subject is permissible with this construction, which is only the case with the phonological subject clitics of French in the paradigms shown in (43). Second, the clitic does not agree with the subject (implicit or explicit) in [PERSON] (i.e., it does not share the same  $\varphi$ -set with the subject). Per Preminger (2011, 2014), doubled clitics must have the same featural composition of the noun phrase they double.

(48) *The coarseness property of clitic doubling*

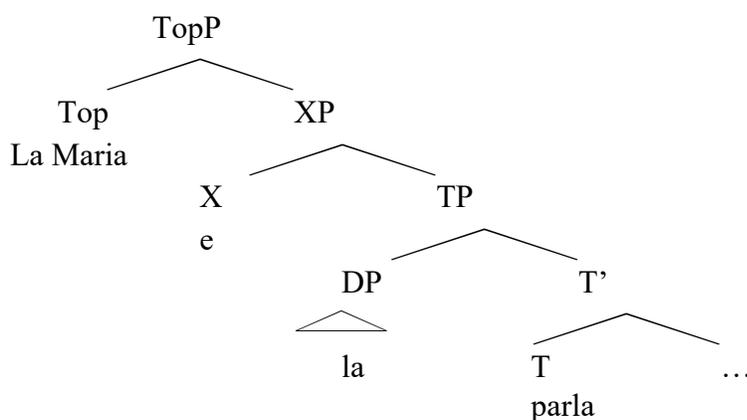
If  $CL^\circ$  is the result of clitic doubling of some noun phrase  $\alpha$ , then  $CL^\circ$  will reflect the full set of  $\varphi$ -features on  $\alpha$ . (Preminger 2014:51)

The constraint in (48) forces us to assume that because the clitic and the subject in the Galician C-AGR construction do not share the same  $\phi$ -set, this cannot be an instance of clitic doubling.<sup>37</sup> Based on these observations, the clitic in the Galician C-AGR construction would seem to fall in between a subject clitic and the topic agreement marker in Fiorentino mentioned in Brandi & Cordin (1989:114).

- (49) a. *Te, e tu parli troppo.*  
 you AGR you talk.PRS.2SG too-much  
 ‘(As for you) you talk to much.’
- b. *La Maria, e la parla troppo.*  
 the Maria AGR she talk.PRS.3SG too-much  
 ‘(As for Maria) she talks too much.’

Brandi & Cordin describe this as an invariable clitic that may combine with any person and inflection in the verbal paradigm. What we can observe here is that this agreement marker is above the subject clitic in [Spec,TP] but below any topic marker (*Te* ‘you’ in (49a) and *La Maria* ‘Maria’ in (49b)).

(50) *La María, e la parla troppo*



<sup>37</sup> This may very well be true for the Carrara variety in (51), as well, due to the first- and third-person pronouns being morphologically identical and, therefore, not showing a singular/plural distinction.

Although Brandi & Cordin address the pragmatic function of this agreement morpheme in greater detail than the mechanism that drives it to arise, its syntax seems fairly straight forward. The licensing of this topic marker simply needs a nominal DP of some kind in its c-command domain to value it. This is what we find more generally with agreement markers such as those investigated in complementizer agreement constructions, and Galician fares no differently. It must not be the case that the Galician clitic is an argument of the verb, unlike the subject clitics mentioned for French (43), but instead must be a marker of agreement of some sort and, therefore, may not qualify as a subject clitic in the strict sense. The main problem with linking Galician C-AGR to the Fiorentino topic marker is that the former requires a host to its left that occupies C, whereas the latter does not. I address the distinction between clitics and agreement morphemes in more detail in Chapter 4.

### 2.3.2 Álvarez (2015)

Álvarez centers her account around a more sociolinguistic overview of this construction, citing a few small regional distinctions in the C-AGR construction.<sup>38</sup> These distinctions largely deal with the spell-out of the agreement morpheme. She claims that there are dialectal variants that only permit the singular, masculine form regardless of whether the subject is masculine or feminine, singular or plural as in (51).

(51) Velai      o      vai              ela /    van              elas  
        behold    CL<sub>M,SG</sub> go.PRS.3SG    she    go.PRS.3PL    them.FEM

‘There she goes/they go.’

Others, by contrast, only accept the singular forms (52). She also states that, for some of her informants, the morpheme must only be used with an animate referent (53).

---

<sup>38</sup> Although Galician is a bilingual community, there is no evidence that Spanish had any influence on this construction or its diachronic development. Therefore, I leave any discussion regarding bilingualism aside.

- (52) a. *Velaí o están teus pais*  
 behold CL<sub>M.SG</sub> be.PRS.3PL your parents  
 ‘There are your parents.’
- b. *Velaquí a veñen María e Xulia*  
 behold CL.F.SG come.PRS.3PL María and Xulia  
 ‘Here come María and Xulia.’
- (53) *Velaí (\*o) está meu coche*  
 behold CL<sub>M.SG</sub> be.PRS.3SG my car  
 ‘There is my car.’

Although I leave these dialectal variations aside, many of the theoretical approaches regarding how to account for these data will be discussed in Chapter 5.

Although she claims that her objective is not to solve the origin of the structure, she briefly describes the manner in which she believes it might have arisen. She outright rejects the hypothesis that *ai/aquí o vén* came from *velaí/velaquí o vén* despite noting the slight variation *veloaí/veloaquí* as mentioned in §2.2.1 of several dialectal areas. In contrast to what I propose in this thesis, she claims that the presentative use of *ai/aquí* analogically attributed to the “triumph and diffusion” of *velaí/velaquí*. Her account is based on two main factors: i) the presentative Galician-Portuguese *eis* (‘behold’), and ii) the analogical extension of the regional locative adverb *eí* (a dialectal variant of *ai*) to said presentative *eis*. Álvarez’s claim is that *eí*, being phonologically similar to *eis*, adopted its presentative property which, in turn, was extended to the standard variant *ai*. Below, I review both the data on *eis* and the dialectal variation of *ei/ai*, showing that neither of the aspects Álvarez presents holds up in accounting for the engenderment of this construction.

### 2.3.2.1 The Galician-Portuguese *eis* and other presentatives in Romance

The presentative item *eis* presents a similar distribution to other presentative words in Romance, notably the Italian *ecco* and Catalan *heus*.<sup>39,40,41</sup> Due to the formal work done on *ecco* mentioned below, I include all three in the data and theoretical analysis below in order to distinguish them from *velai/velaquí*.

Much like *velai/velaquí*, these three lexical items must always be sentence initial, and they always have a present-tense interpretation. They share some of the same properties as *velai/velaquí*, such as selecting a DP (54) and the complementizer *que* (55).

- (54) a. *Eis o menino!*  
 behold the child
- b. *Ecco il bambino!*  
 behold the child
- c. *Heus el nen!*  
 behold the child  
 ‘There is the child!’

---

<sup>39</sup> I leave aside a comparative analysis between these lexical items and the French *voilà/voici*, as the latter have a more extensive distribution than the aforementioned presentatives in that they may be embedded, used in questions, and used in substitution for certain copulas and existentials (Morin 1985).

<sup>40</sup> *Heus* is always paired with an adverb, *aquí/ací* (‘here’), and they are almost always in a string-adjacent relationship. In a search of the *Corpus textual informatitzat de la llengua catalana* (Rafel, dir.), only two cases arise in which *heus* and one of these adverbs are not found together: when there is cliticization to *heus* (i) and when there is an intervening DP (ii). Although (i) is obligatory (i.e., a pronoun may not cliticize to *aquí/ací*), (ii) seems stylistic and infrequent.

- (i) *Heus-les aquí*  
 behold-CL<sub>F,PL</sub> here  
 ‘Here they are.’
- (ii) *Heus una vida ací*  
 behold a life here  
 ‘Here is a life.’

<sup>41</sup> *Heus* is somewhat archaic in spoken Catalan, having given way to the more common presentative phrase, *vet aquí*. A full in-depth comparison and incorporation of *vet aquí* is well beyond the scope of this thesis. Therefore, I leave it to further research.

- (55) a. Eis que o mar lles devolveu este pago  
 behold COMP the sea CL<sub>DAT.PL</sub> return.PST.3SG this payment  
 ‘Here is the sea returning this payment to you...’
- b. Ecco que Liliane scrive un altro articolo  
 behold COMP Liliane write.PRS.3SG a other article  
 ‘Here is Liliane writing another article...’
- c. Heus acá que el tenia davant  
 hold here COMP CL<sub>M.SG</sub> have.IMPV.3SG in.front  
 ‘Here is what he had before him...’

Unlike *velai/velaquí*, however, these presentative words have verbal properties that extend beyond those of their Galician counterparts. Most notably, they may license clitics.

- (56) a. Ei-lo (aquí!)  
 behold-CL<sub>M.SG</sub> here
- b. Ecco-lo (qui!)  
 behold-CL<sub>M.SG</sub> here
- c. Heu-lo aquí!  
 behold-CL<sub>M.SG</sub> here  
 ‘Here it is!’

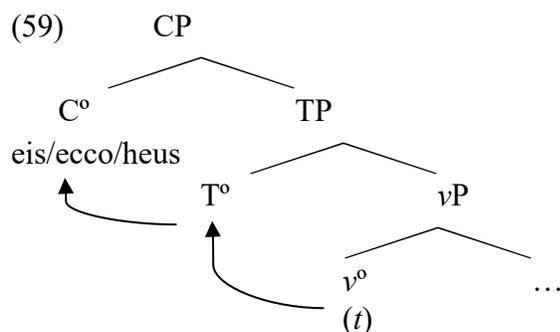
This is impermissible in Galician (57).

- (57) \*Velaquí as  
 behold CL<sub>FEM.PL</sub>  
 Intended: ‘Here they are!’

It is important to note that licensing clitics is distinct from hosting clitics, which does occur in Galician as in transitive constructions headed by *velai/velaqui* (16b) as well as the C-AGR construction. The clitics in (56) are clearly arguments of these verbal elements. In addition to this, they may not be in the same clause as verbs, unlike the intransitive and transitive examples we saw for *velai/velaqui* (16b).<sup>42</sup>

- (58) a. \*Ei-lo           tes!  
           behold-CL<sub>M.SG</sub> have.PRS.2SG
- b. \*Heu-lo           tens            aquí!  
           behold-CL<sub>M.SG</sub> have.PRS.2SG her
- c. \*Ecco-lo          hai!  
           behold-CL<sub>M.SG</sub> have.PRS.2SG
- Intended: “Here you have it!”

Following the work done on *ecco* by Zanuttini (2017), I claim that these elements are morphologically defective verbs based-generated in *v*, undergoing subsequent *v*-to-T-to-C movement.



<sup>42</sup> At one point in its diachrony, *eis* did accept verbal complements:

- i) Ei-lo           tes!  
       behold-CL<sub>M.SG</sub> have.PRS.2SG

This word is practically unrecognizable amongst Galician speakers that are not well-versed in the diachrony of the language, and it is no longer used in this way in present-day Portuguese either.

Following Zanuttini, I claim that these functional elements are indeed verbs that must pass through T in order to check a [+PRESENT] feature.

These presentative elements clearly differ from *velai/velaqui* both in their distributional properties and the derivational steps they take. Being presentatives, we would expect both sets of items above to end up in C and permit no preceding material, and this is borne out. Their differences essentially lie in their categorial nature. For *velai/velaqui*, their status as base-generated C-heads explains their ability to pair with verbs and their inability to license clitics. For the other group of presentatives, the fact that they may not select a verbal complement and license their own clitics makes them verbal candidates. Below I show how these data present us with a difficult bridge to reconstruct based on Álvarez's account.

### 2.3.2.2 Analogy between *eis* and *ei*

Álvarez's claims lead one to believe that when the Galician C-AGR structure was engendered, *eis* was still frequently used in the language, citing the Early Modern Period.<sup>43</sup> Her proposal relies on the form *eis* takes when it hosts a clitic pronoun, namely the dropping of the final -s in all forms (*eime* 'Here I am', *eite* 'Here you are', *eilo* 'Here he/it is', etc.). In turn, she claims that this bare form (*ei*) may be found without cliticization, citing an example from Lorenzo (1977):

- (60) ey    teu    irmão    vë                    a    ti  
       behold your    brother    come.PRS.3SG to    you  
       'There is your brother coming to you.'

---

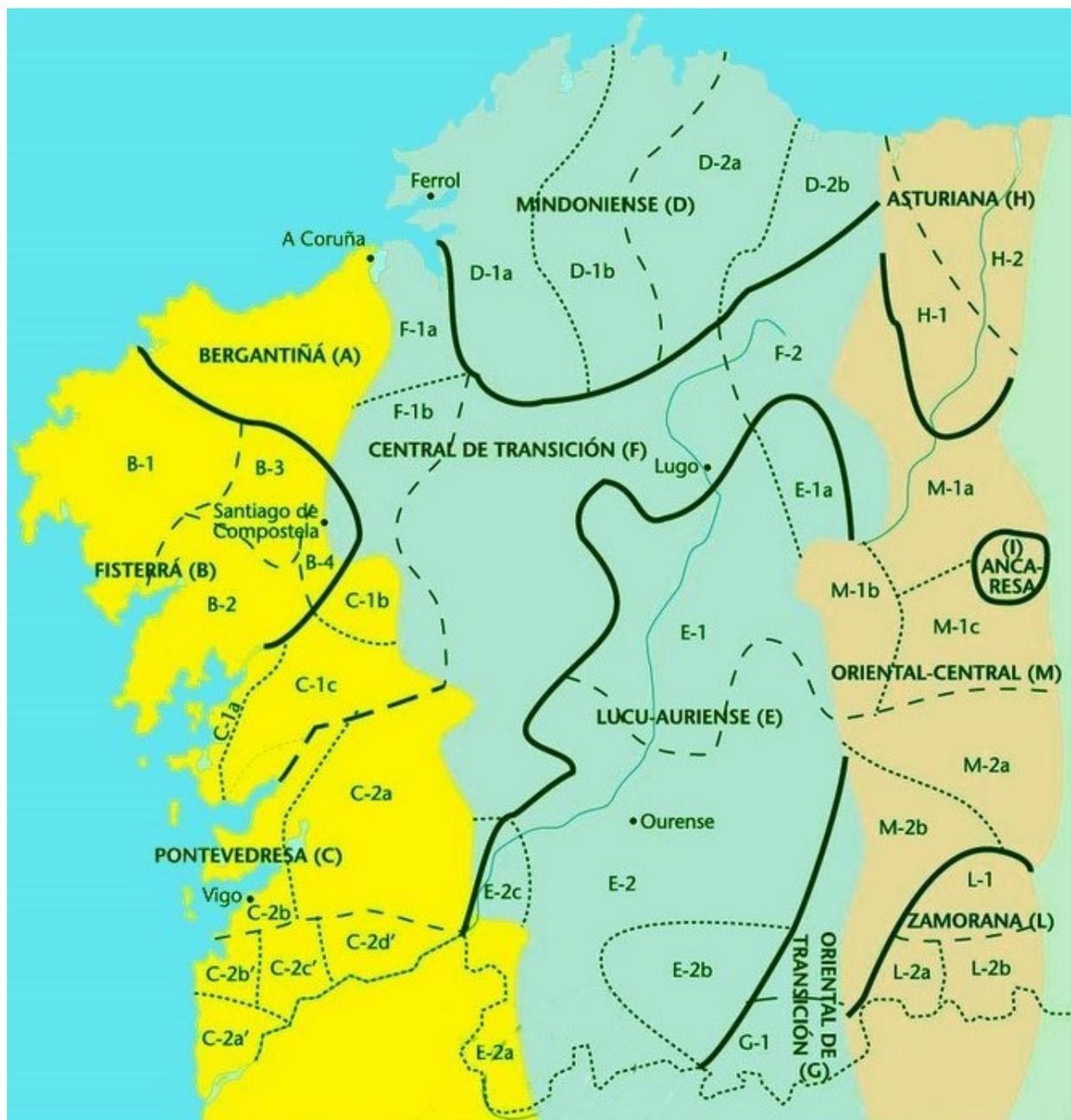
<sup>43</sup> The Early Modern Period (18<sup>th</sup> century) in Galician-Portuguese is thought of as the breach between these two languages, primarily due to orthographic differences that arose after *O Rexurdimento*, the Galician enlightenment period, of the mid-19<sup>th</sup> century. Available texts from before 1840 are almost non-existent, however, as this date marks the end of the *Séculos Escuros* ('Dark Ages') of Galician literature and culture.

Álvarez cites a key similarity to her analysis between the presentative *eis* and the deictic locative *eí* in that they both commonly select a DP complement in short, presentative-like constructions.

- (61) a. Ei-lo                    problema  
           behold-CL<sub>M.SG</sub> problem
- b. Eí        o problema  
           there    the problem
- ‘There is the problem.’

Once this analogy was established, *eí* essentially taking on the properties noted in § 2.3.2.1 and being used in place of *eis*, the speakers of the *eí* variant spread this construction to the those of the more common *aí* variant, which then spread throughout the areas where it is used today. Based on her proposal, we may assume that *eí* took on the syntactic behavior of *eis* in all aspects.

The first problem of this proposal is a geographic one. Galician dialectology is divided in different macro- and micro-areas as depicted in Figure 4.



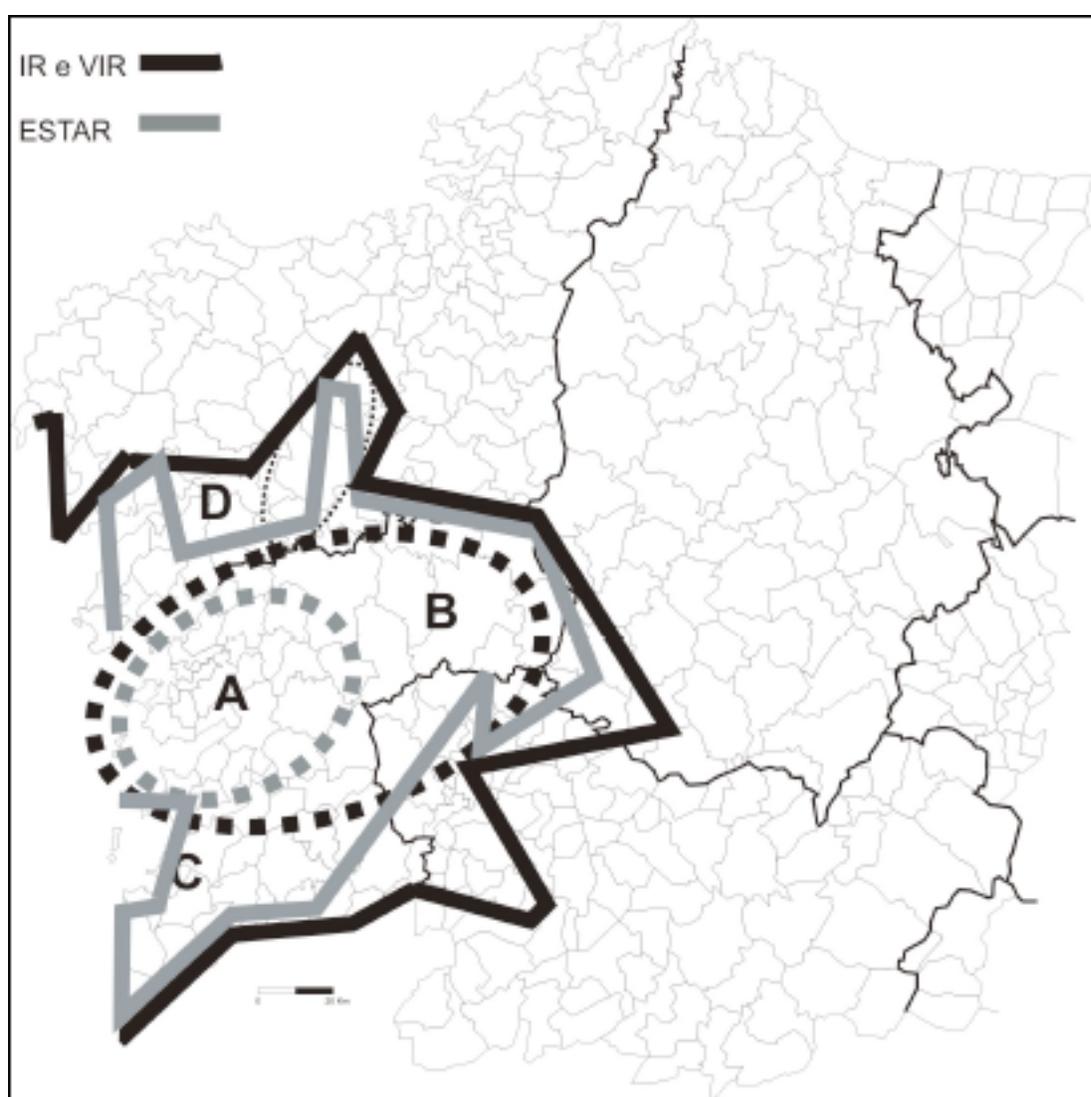
*Figure 4. Geographical distribution of Galician dialects*<sup>44</sup>

The divisions seen in Figure 4 are modeled after Fernández Rei (1990) and, therefore, I follow his analysis of the dialectal distribution of Galician. Galician dialectology is typically broken up

<sup>44</sup> <https://www.lingua.gal/o-galego/conhecendo/territorio-no-que-se-fala>

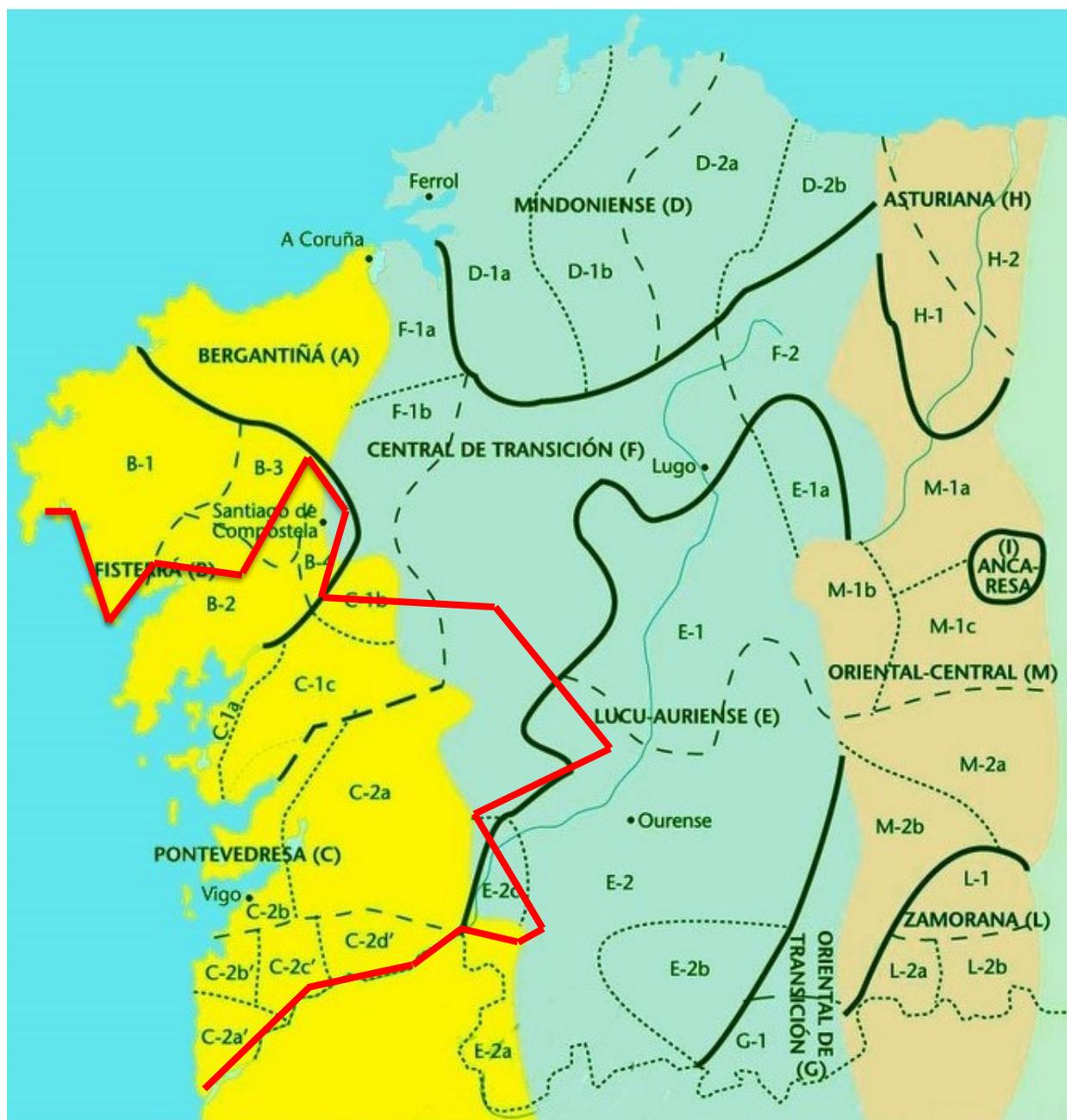
into three main blocks: the Western Block (yellow), the Central Block (blue), and the Eastern Block (brown). A different macro-division is that shown by the bolded lines which represents a more local subdivision of dialects, and a meso-division delineated by a bolded dotted line which further breaks up said dialects. Although there are further subdivisions noted, I leave them aside as they have no importance on my analysis here.

The region that Álvarez identifies as that of what I call Galician C-AGR forms part of the Western Block, as seen in Figure 5.



*Figure 5. Geographical distribution of Galician C-AGR (Álvarez 2015:211)*

Álvarez's geographical outline of Galician C-AGR is against the backdrop of the delineation of the four provinces of Galicia. Placing the outermost border (D) on the map in Figure 4, we may identify its dialectal borders.<sup>45</sup>

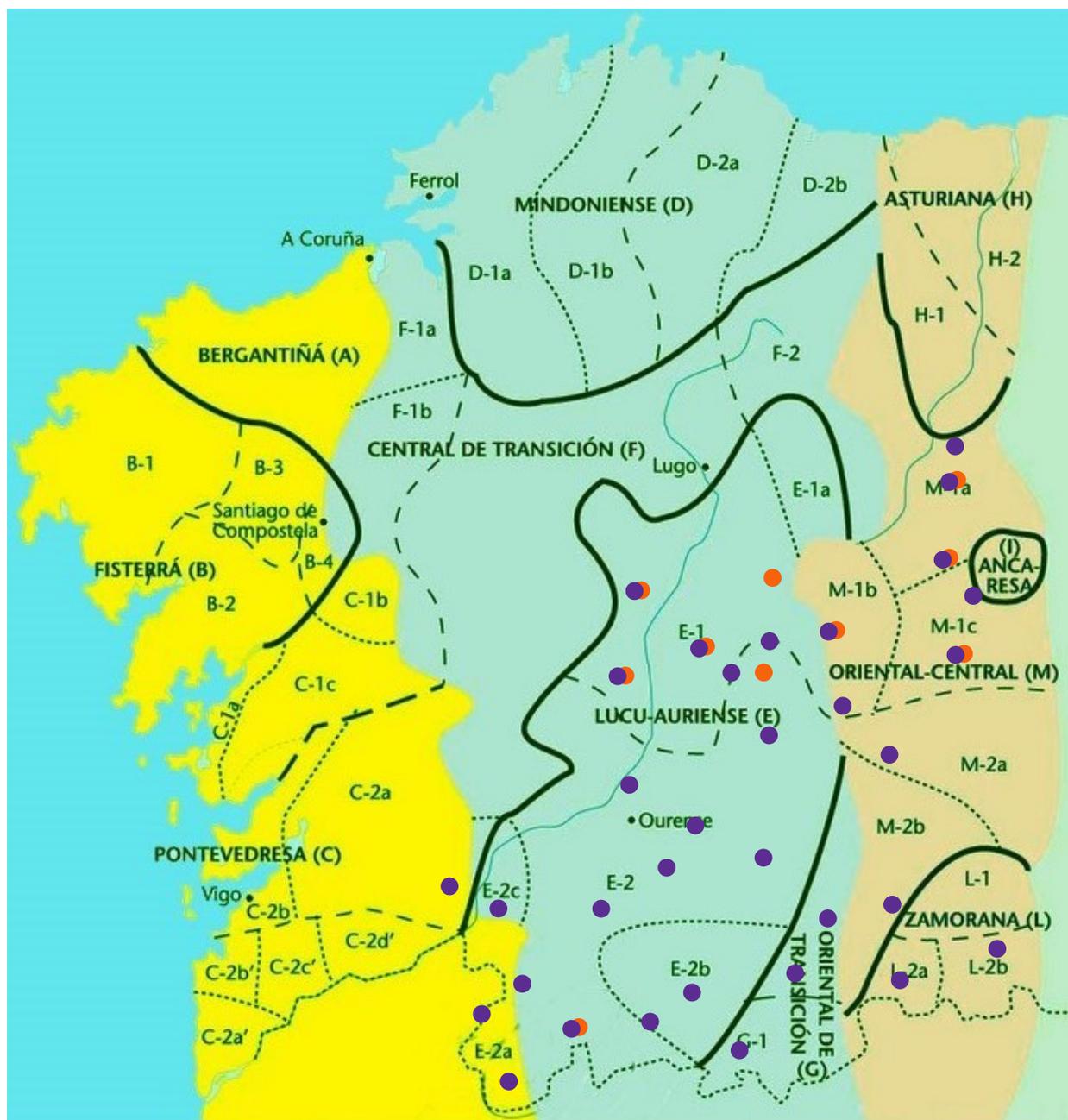


*Figure 6. Galician C-AGR within the purview of Galician dialects*

<sup>45</sup> I am not concerned with the other delineations Álvarez makes which are represented by the other lines that deal with verb usage. In this investigation, I am only concerned with the outermost geographical limits of the use of the construction.

In Figure 6, we see that Galician C-AGR mainly covers the dialects Fisterrá (B), Pontevedresa (C), the southern portion of the Central de Transición (F), and very limited areas of the Lucu-Auriense (E) region. This makes this phenomenon primarily one belonging to the Western Block with some presence in the Central Block, as well.

The dialectal variants *eí* and *eiquí*, on the other hand, are primarily a Central Block and Eastern Block phenomenon as seen in Figure 7.



*Figure 7. Distribution of dialectal variations eí and eiquí<sup>46</sup>*

The orange dots represent the geographical locations of *eí*, and the purple dots those of *eiquí*. The instances in which a purple dot overlaps an orange dot signify that these two variations coexist together in those particular locations. The first observation that should be noted is that

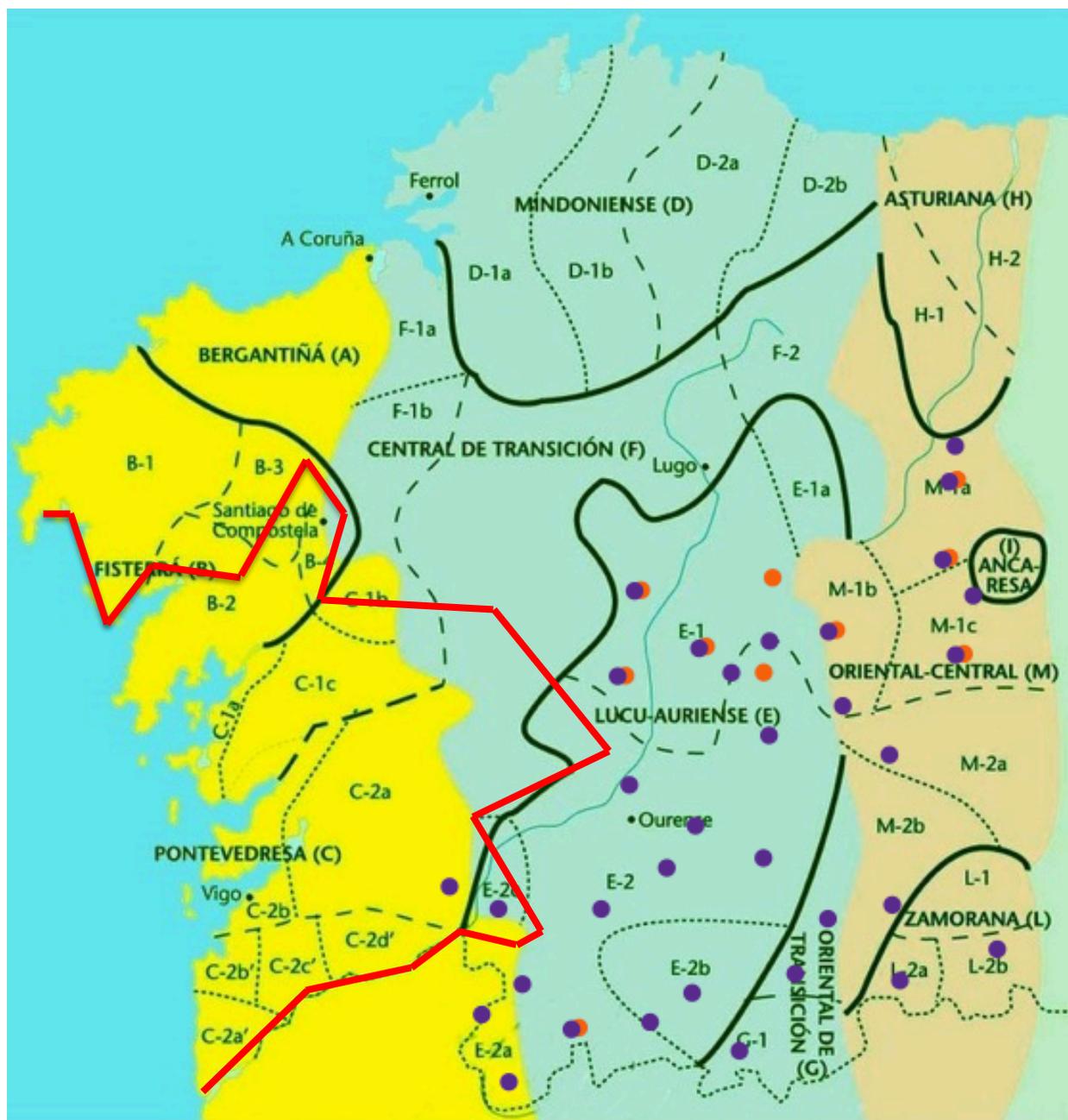
<sup>46</sup> These locations are confirmed via Maps 310-311 of the *Atlas Lingüístico Galego* (Fernández Rei, González González & Álvarez 1995).

*eiquí* seems a more widespread phenomenon than *eí*, which, to my knowledge, has not been stated explicitly in any previous work. Because the same phenomenon is at play in both variations, the raising of the central vowel [ä], it is often assumed that these variations are always found together. However, the map in Figure 7 shows that this must not be the case.<sup>47</sup>

The second observation of the map in Figure 7 is that the variations *eí* and *eiquí* clearly belong to the dialects of the southeastern portion of the Central Block and Eastern Block, being heavily identified as parts of the Lucu-Auriense (E), Orient-Central (M), and the Oriental de Transición (G) varieties and with minimal presence in the Zamorana (L) and Pontevedresa (C) varieties. This is in diametric opposition to what I showed in Figure 5 regarding the C-AGR phenomenon, which we saw spans over large portions of the Western Block dialects with minimal presence in the Central Block. I show the overlap of these two phenomena in Figure 8.

---

<sup>47</sup> As my focus is concerned with where *eí* and *eiquí* are present and not the potential mismatches of the two variations (i.e., where one may coexist with the standard variety *aí* or *aquí*), I have not included what the coexisting forms are in the locations in which *eí* and *eiquí* are not a part of the same geographical point.



*Figure 8. Distributions of Galician C-AGR and the dialect variations eí and eiquí*

In Figure 8, we see that the area of the C-AGR phenomenon and the territory where *eí* and *eiquí* coexist practically border one another. Aside from two locations signaled in southern area of the border between the Western and Central Blocks, and only one of them pertaining to the Western Block, there is no overlap of the two phenomena. Even if we were to extend the use of *eí* to the

entirety of the geographical area in which *eiquí* is found, something that the maps of the *Atlas Lingüístico Galego* do not support, the divide between these two phenomena is very clear and the idea that one of them directly influenced the other does not hold up.<sup>48</sup> It is also important to note that the idea that *eí* and *eiquí* have regressed over the centuries is highly unlikely. The variations *eí* and *eiquí* were identified in the very first descriptive grammars of Galician as pertaining mainly to the province of Ourense (Saco y Arce 1886, Lugrís Freire 1931).<sup>49</sup> This supports my position that speakers of the variant *eí* were not of the same speech community as those that belong to the C-AGR phenomenon. I find Álvarez's proposal that one of these phenomena influenced the other untenable due to the fact that they pertain to two geographically distinct areas with almost no evidence of cross-dialectal contact or influence.

A second reason that casts doubt upon Álvarez's idea above is one that relates to the diachronic documentation of Galician and the C-AGR phenomenon. Although historical corpus analyses may not be able to tell the entire story of a particular phenomenon's development and that the documented timeline along which a certain change might have taken place is most often delayed (Joseph 2011), there is a significant difference when it comes to documentation between the construction Álvarez believes surfaced first (*ai o vén*) and that which I claim surfaced first (*velai o vén*). According to the *TILG* corpus (Santamarina 2018) the first appearance of C-AGR with *ai* dates to 1979, as does that with *eí*. The first step that I propose in the diachronic development of this construction, (§1.6), first appeared in 1853. As I describe in greater detail in Chapter 3, the data of the *TILG* corpus point to very salient syntactic data that lead me attribute

---

<sup>48</sup> Another possible point of contention is that Álvarez (2015:212) claims that the C-AGR phenomenon is currently undergoing expansion, leading us to believe that the delineation of the area where Galician C-AGR is found might have originally been, for example, the size of areas (A) or (B) in Figure 5.

<sup>49</sup> Each of these references has a different reason for their mentioning this region specifically. Juan Antonio Saco y Arce, a native of Ourense, wrote a grammar that was heavily influenced by both the phonological variations of prepositions, determiners, etc., of the southern Central Block, as well as the vocabulary he defined. Manuel Lugrís Freire, being a native speaker of the Western Block, simply noted that “[T]he forms *eiquí*, *ehí*, *elí*, *atráis* are not found in the older documents and literary works: they are dialectal facts, principally of the south of Galicia” (73).

the creation of the C-AGR phenomenon to the string *ve-lo aí vén*, not *aí o vén*. Moreover, Saco y Arce (1886:194-195) notes the phonological variations of the complementizers *velai/velaqui* used in Ourense to this day, those of *velei* and *veleiquí*. He makes no reference to *eis*, but he does compare the use of these lexical items to the Spanish *he aquí*.<sup>50</sup> This fact, amongst those presented above, lead me to claim that the development of *velai/velaqui* was not only that which engendered the C-AGR construction but also that this development was not influenced by the presentative use of *eis*.

#### 2.4 Why a diachronic account is necessary

As I mentioned in Chapter 1, the idiosyncrasies of this construction may only be fully understood within the purview of diachrony. In addition to the data shown in §2.2, these functional C-heads participate in a construction that I believe bears a sharp resemblance to complementizer agreement (C-AGR) often found in Western Germanic, Irish, Nez Perce, and a few Bantu languages. The difference between Galician and these languages is that Galician C-AGR appears in matrix clauses only, akin to the C-AGR phenomenon described in Putnam & van Koppen (2009), which I address below. This is self-evident if we consider that *velai/velaqui* must be sentence initial (1).

Returning to the data briefly examined in Chapter 1 (1), Galician C-AGR is comprised of the C head *velai/velaqui*, a third-person accusative clitic (*o, a, os, as*), and an intransitive verb of state or motion (*ir* ‘go’, *andar* ‘walk’, *ficar* ‘remain’, *aparecer* ‘appear’, *quedar* ‘stay’, *estar* ‘be’).<sup>51</sup>

---

<sup>50</sup> This phrase is considered an archaism in both spoken and written Spanish, which is why I did not discuss it along with the other Romance presentatives in §2.3.2.1. Regardless, its interpretation may also be understood as ‘behold’, just like the others mentioned in that section.

<sup>51</sup> Álvarez (2015) claims that there are speakers that also accept *seguir* ‘follow,’ as well. However, none of my informants accept this verb in this construction, as *seguir* seems to take on a transitive feel in this construction.

(62) a. Velaquí a vén!  
 behold CL<sub>F.SG</sub> come.PRS.3SG

‘Here she comes!’

b. Velaí os van!  
 behold CL<sub>M.PL</sub> go.PRS.3PL

‘There they go!’

This construction also permits a postverbal DP that agrees in [NUMBER] and [GENDER] (63a-b).

This applies to the first- and second-persons, as well (64). This construction may also appear in the imperfect (65) in addition to the present.<sup>52</sup>

(63) a. Velaquí \*o/a está ela!  
 behold CL<sub>M.SG</sub>/CL<sub>F.SG</sub> be.PRS.3SG she

‘There she is!’

b. ‘Velaquí \*as/os andan os rapaces!  
 behold CL<sub>F.PL</sub>/CL<sub>M.PL</sub> walk.PRS.3PL the boys

‘There go the boys!’

(64) Velaquí os estamos nós!  
 behold CL<sub>M.PL</sub> be.PRS.1PL we

‘Here we are!’

(65) Velaquí os estaban cando chegou a tormenta.  
 behold CL<sub>M.PL</sub> be.IMPV.3PL when arrive.PST.3SG the storm

‘(And) they there were when the storm hit.’

---

<sup>52</sup> Álvarez (2015) also claims that this construction in the conditional sounds acceptable per her native intuitions. This tense may be largely idiolectal, as none of my informants accept this construction in this tense.

There are, however, a few oddities about the Galician C-AGR when compared to other known C-AGR varieties. First, this construction is not permissible with transitive predicates. When object-selecting predicates are used, the clitic is interpreted as the direct object.

(66) Velaquí a tes  
 behold CL<sub>F.SG</sub> have.PRS.2SG  
 ‘Here you have it’

(67) Velaquí os hai  
 behold CL<sub>M.SG</sub> be-EXT.PRS.3SG  
 ‘Here they are.’

Although the finer-grained details as to why this is will be discussed in Chapters 3 and 5, the ban on transitive verbs in Galician C-AGR also includes the invariable existential *hai* ‘there is/are’ (67), which pairs with transitive predicates in the sense that both select an object DP as their internal argument (Rodríguez-Mondoñedo 2007). Even when the object remains a lexical DP and the clitic agrees with the subject and not the object, the utterance is still ungrammatical and is never interpreted as signaling the subject via the agreeing clitic.

(68) Velaquí (\*o) ten as mazás (Xan)  
 behold CL<sub>M.SG</sub> have.PRS.3SG the apples Xan  
 ‘Here he (Xan) has the apples.’

The clitic in (68) does not agree with the object DP in [NUMBER] or [GENDER] and, therefore, cannot be an instance of clitic doubling (Preminger 2019). It does, however, agree with the suggested subject Xan, but the utterance is still ungrammatical. Not only is transitivity not an obstacle for other C-AGR varieties, but Nez Perce (Deal 2015) actually shows agreement with an object DP in addition to that of the subject.

(69) ke-pe-m-ex kaa *pro*<sub>subj</sub> cewcew-tée'nix *pro*<sub>obj</sub>  
 C-PL-2-1 then PRO.1PL telephone-TAM PRO.2SG

‘When we call you (sg.)’ (Deal 2015:8)

In (69), the  $\bar{A}$ -particle *ke* on *C* shows several inflections for [PERSON] as well as one for [NUMBER]. As mentioned above, the second-person inflection corresponds not to the subject, as is the case in most C-AGR varieties, but to the object. The idea that the head hosting the C-AGR inflection may host more than one agreement marker is a positive outlook on my approach here, as Galician C-AGR may often pair with a dative of interest, presumably on the same functional head.

(70) Velaquí **cho** vén!  
 behold **CL**<sub>DAT.2.SG</sub>-**CL**<sub>M.SG</sub> come.PRS.3SG

‘Here he comes (on you)!’

In (70), the C-AGR has cliticized to a second-person singular dative of interest *che*. Data such as those in (70) will aid in my analysis distinguishing clitics from true agreement markers in Chapter 5, although I address multiple probes on a single head, specifically with respect to cliticization, in both Chapters 3 and 5.

Another outlier amongst complementizer agreement phenomena found in Galician C-AGR is the lack of inflection for [PERSON] of the C-AGR morpheme. In every variety I have come across in the literature, there is a probe specified for person that minimally determines whether there is an agreement morpheme on *C*, which is usually a morphologically distinct agreement marker. Galician C-AGR, on the other hand, only inflects for [GENDER] and [NUMBER].

|                 | SG  | PL    |
|-----------------|-----|-------|
| 1 <sup>st</sup> | o/a | os/as |
| 2 <sup>nd</sup> | o/a | os/as |
| 3 <sup>rd</sup> | o/a | os/as |

Figure 9. Galician C-AGR paradigm

As shown in Figure 9, all of the singular inflections are morphologically the same, as are the plural ones. This contrasts with West Flemish, for example, which also shows inflection for all persons but whose agreement markers are all morphologically distinct based on person.

|                 | SG       | PL  |
|-----------------|----------|-----|
| 1 <sup>st</sup> | -k       | -me |
| 2 <sup>nd</sup> | -j       | -j  |
| 3 <sup>rd</sup> | -se/-tje | -ze |

Figure 10. West Flemish C-AGR paradigm (Haegeman 1992)

Each inflection in the West Flemish paradigm in Figure 10, however, makes a [PERSON] distinction. In Chapter 5, I address this contrast and the claim that the [PERSON] distinctions are expected in a language with poor verbal inflection such as West Flemish and, therefore, differs from Galician typologically, a language with rich verbal inflection.

The last point of contention in equating this phenomenon in Galician to complementizer agreement is the fact that it is a matrix-clause phenomenon. Having shown in §2.2 that *velai* and *velaqui* are matrix-clause complementizers based on the fact that they may never be embedded is at odds with the data we find in almost all other varieties of C-AGR due to the fact that complementizers are most commonly subordinating elements.

(71) Wisconsin Heritage German

a. Kanns                    du    mir            sagen    *afs*    du    morge    koms?  
 be-able.PRS.2SG    you    me.DAT    say    if-2SG    you    tomorrow    come.PRS.2SG

‘Can you tell me if you are coming tomorrow?’

b. Net    de Mann    wos                    du    ‘s Geld            geben    hast!  
 NEG    the man    whom-2SG    you    the money    given    have.PRS.2SG

‘Not the man to whom you gave the money!’                    (Bousquette 2014:565)

In (71a), the complementizer *ob* (‘if’) is inflected for the second-person singular. This also extends to relative clause indirect object pronoun *wo* (‘whom’) in (71b), which has presumably undergone movement to the left periphery. These data, much like the rest of the C-AGR literature, show that complementizer agreement is typically found on elements found in the C-domain of subordinate and relative clauses. The one exception may be from the data presented in Putnam & van Koppen (2009) regarding the *alls*-construction in Midwestern American English.

(72) Alls I know is he isn’t working hard enough.

In (72), the complementizer *all* is inflected with what Putnam & van Koppen claim is a truncated version of *as* that only appears on the first- and third-persons in English. Although the data here give my account a base on which to stand compared to other well-known complementizer agreement varieties that do not work this way, it remains to be understood how *velai* and *velaqui* came to adopt such an uncommon position in derivation. This is the focal point of Chapter 4.

## 2.5 Conclusion

In this chapter, I have analyzed in detail the syntactic distribution of the two complementizers *velai/velaqui* in addition to what I refer to as the Galician complementizer agreement phenomenon. Moreover, I have reviewed two accounts, one synchronic in nature and

one related to the possible origins of this construction. I have shown that the analysis by Longa, Lorenzo & Rigau (1996, 1998) of this construction's synchronic nature has numerous theoretical and empirical drawbacks that are unable to explain details such as why only a select few complementizers may head this construction and why intransitive verbs are the only ones that may participate in the C-AGR construction. The account given by Álvarez (2015) regarding how this phrase came to be presents several geohistorical issues that she fails to address, making her account equally porous. I have shown that her analogy-based account of the presentative *eis* and adverb *ei* is misguided based on the fact that these two phenomena occupy geographical regions of Galicia that do not overlap.

In Chapter 3, I discuss the beginning steps of this construction ((31) of §1.6). Upon explaining Step 1, I combine Westergaard's (2008, *et seq.*) *micro-cue* model with the data in Costa, Fiéis, & Lobo (2015) in explaining how children acquire accurate clitic positioning using the theory proposed by Raposo & Uriagereka (2005). In order to explain Step 2 to Step 3, I address the reanalysis from multi-word structures to single lexical items based on the findings in Hudson Kam & Newport (2005, 2009). In addition, I show how we may account for fixed clausal positioning from a language acquisition standpoint. I compare the data in Galician to other main-clause phenomena (the Spanish *dizque* 'It is said', Demonte & Fernández-Soriano 2013; the Romanian particle *hai*, Hill 2007, 2012) and show how these phenomena, too, may be explained by my approach to fixed positions within the clausal spine based on findings in Cassani, Grimm, Daelemans & Gillis (2018) and Cassani, Grimm, Gillis, & Daelemans (2019).

## CHAPTER 3

### REANALYSIS, CATEGORY ACQUISITION, AND LEARNABILITY

#### 3.1 Introduction

In this chapter, I address several phenomena related to learnability. With respect to language specific phenomena, I address issues related to learnability and acquisition of clitic placement in Western Iberian Romance. This topic has been addressed in the theoretical literature, as well as briefly in the language acquisition literature. However, there has been no attempt to combine the two approaches. Using child language acquisition data collected by Costa, Fiéis, & Lobo (2015), I provide an in-depth analysis of how we may account for the data formally. I also address several aspects of category acquisition and how it plays into both synchronic and diachronic reanalysis.<sup>53</sup> I examine the learnability of categories and discuss how difficult acquiring a particular category is based on its syntactic environment. Following observations from Cassani, Grimm, Daelemans & Gillis (2018) and Cassani, Grimm, Gillis, & Daelemans (2019), I show that functional elements in fixed sentential position (e.g. sentence-initial functional words) are harder to learn and, therefore, are primed contexts for change. The goal upon examining learnability in this manner proffers us with an explanatory understanding not only of how language acquisition proceeds in a way in which language transmission is complete (i.e., there are inherently no deviant structures), but also how deviance may be explained based on the input of the child. As the learning path of cliticization in Western Iberian is a long and puzzling one for learners, this will both set the basis for both tying in theoretical assumptions with actual

---

<sup>53</sup> Here I use the term “synchronic reanalysis” to refer to the repair a child realizes during her developmental linguistic journey.

psycholinguistic data regarding the topic as well set the foundation for much related discussion in Chapter 4.

### 3.2 The data

Before returning to the data specific to the steps of the C-AGR construction, a few comments about the proposed diachrony of Galician and about Galician more generally are in order. Galician culture and literature underwent a period of decline known as the *Séculos Escuros* ('Dark Ages'), spanning from the sixteenth to the eighteenth centuries. During this time, only a handful of Galician texts were written, all of them resembling well-established Portuguese orthography of the time. It was not until the *Rexurdimento* ('Renaissance') that Galician authors began writing in a language with an orthography they identified as representative of their own. The most recognized text marking the beginning of this period is *A gaita gallega* by Xoán Manuel Pintos in 1853, although there are known works that date to the second decade of nineteenth century.<sup>54</sup> This point in time is considered by many to be the birth of Modern Galician. Along with the resurgence of writing and a personal linguistic identity came the idea that the orthography no longer matched the phonological realization of the language. As observed in Freixeiro et al. (2005), there was much orthographic variation with respect to the realizations of particular syllables and sounds (e.g. the famed *gheada*, a phenomenon in which the glottal /h/ replaces the voiced plosive /g/). While this may be of interest in some investigations, the phenomena I have addressed and shall address in greater detail in this chapter remain unambiguous as to the pronunciation regardless of the exact orthography. That is, the morphophonological realization of certain cliticization patterns (those pronounced [lɔ], [lɐ],

---

<sup>54</sup> There is a group of writers known for preceding this date known as *os precursores* ('the forerunners') who were the first to write in Galician on the resurgence of interest regarding Galician culture and the language. The *TILG* (*Tesouro Informático da Lingua Galega*) from which I take a large portion of the diachronic data does indeed have relevant data for the construction under investigation in this thesis.

[los], and [lɛs]) may be orthographically realized as *-lo*, *l-o* or *lo* on the preceding host. However, there is no historical evidence that the forms proposed in the steps in Figure 11 were ever used interchangeably. Even in the face of an exploratory writing system as that of Galician after the *Séculos Escuros*, we should expect some sort of representational difference between morphophonological phenomena if this period is identified as Modern Galician.

The historical data used below for Steps 1 & 2 come from the *Tesouro Informatizado da Lingua Galega (TILG)* corpus (Santamarina 2018).<sup>55</sup> The earliest attestation of Step 1 is from 1853, that of Step 2 is from 1888, and what I identify as the clearest indication of the syntax of Step 3 is from 1877. Based on the proximity of these dates and the fact that they do not fall in line chronologically within the corpus data, I claim that, when writing returned at the beginning of the *Rexuridmento*, the reanalysis of this phenomenon was in full swing. Although the order of these steps and the processes they underwent are speculative due to the lack of historical data before this time period, I claim that there is enough evidence to recreate a viable pathway between the phenomena seen in Steps 1-4. References made below to ‘earlier stages’ of this construction refer to the time period in which Step 1 seemed to be the most prevalent, which corresponds to that of 1853-1930. Although there is significant overlap between the three stages during this time, Step 1 is hardly attested after this period.

Before moving forward, additional comments on the exactitude of the data with respect to their corresponding dates in the corpus are in order. First, the idea that the data in the proposed steps correlate to these dates exactly is unlikely. As mentioned in Joseph (2011), the historical recordings of diachronic phenomena rarely represent the precise timeline of when said phenomena take place. This is certainly the case of the C-AGR construction of Step 4, which

---

<sup>55</sup> The *TILG* corpus is comprised of 3,076 documents containing 30,924,082 words. Its earliest document is dated to 1612 and spans to 2018, the year of the latest version (4.1) of the corpus.

does not appear in the *TILG* corpus until 1991. Fernando Torreiro and Xosé Ramón Freixeiro (p.c.) both testify to this construction being present in their respective parishes growing up in the 1960s from the eldest members of the community, placing this phenomenon at the end of the 18<sup>th</sup>-century at the very latest. In Chapter 4, I claim that this is further evidence supporting the fact that the close diachronic proximity of Steps 1-3 and their coexistence together led to the engenderment of Step 4.

Second, the conjecture that the reanalysis of this phenomenon was well underway or possibly nearing completion by the time the *Rexurdimento* arose is supported by the simple occurrences of Step 1 and Step 2. Between 1853 and 1930, the total number of the observed instances of the strings proposed for Step 1 are seen in Table 2.

*Table 2. Occurrences of Step 1 based on clitic and adverb combinations*

| SURFACE STRING FORM                               | OCCURENCES |
|---------------------------------------------------|------------|
| Ve-lo ahí ('You see it <sub>M.SG</sub> there')    | 30         |
| Ve-la ahí ('You see it <sub>F.SG</sub> there')    | 10         |
| Ve-los ahí ('You see them <sub>M.PL</sub> there') | 7          |
| Ve-las ahí ('You see them <sub>F.PL</sub> here')  | 0          |
| Ve-lo aquí ('You see it <sub>M.SG</sub> here')    | 2          |
| Ve-la aquí ('You see it <sub>F.SG</sub> here')    | 1          |
| Ve-los aquí ('You see them <sub>M.PL</sub> here') | 1          |
| Ve-las aquí ('You see them <sub>F.PL</sub> here') | 1          |
| Ve-lo alí ('You see it <sub>M.SG</sub> there')    | 12         |
| Ve-la alí ('You see it <sub>F.SG</sub> there')    | 0          |
| Ve-las alí ('You see them <sub>F.PL</sub> there') | 1          |
| Ve-las alí ('You see them <sub>F.PL</sub> there') | 0          |

Limiting the data in Table 2 to the most common adverbs used in this construction (*ahí, aquí, alí*), we see a total of sixty-five occurrences across all possible combinations with clitics varying in gender and number. The reanalyzed form *velai*, however, shows two hundred nineteen results within the same time period. Based on these results alone, it seems that the transition from Step 1 to Step 3 had largely taken place.

### 3.2.1 Step 1

The data observed here is that which I propose corresponds to Step 1 in what I claim was a four-step diachronic process leading to the present-day C-AGR construction. I repeat the steps highlighted in Chapter 1 here for expository purposes.

Step 1: Biclausal, prosodically divided structure

Ve-lo                    aí        vai            Xan

see.PRS.2SG-CL<sub>ACC.M.SG</sub> there    go.PRS.3SG    Xan

‘You see him, there goes Xan.’

Step 2: Biclausal, prosodically unified structure

Ve-lo                    aí        vai            Xan

see.PRS.2SG-CL<sub>ACC.M.SG</sub> there    go.PRS.3SG    Xan

‘You see him there, Xan is going.’

Step 3: Monoclausal structure; reanalyzed verbal structure

Vel(o)aí    vai            Xan

behold    go.PRS.3SG    Xan

‘There goes Xan.’

Step 4: Monoclausal, main clause complementizer agreement structure

Vel(o)aí    o            vai            Xan

behold    CL<sub>ACC.M.SG</sub>    go.PRS.3SG    Xan

‘There goes Xan.’

#### *Figure 2. Proposed steps for Galician C-AGR*

The first point I wish to clarify is that of the declarative status of Step 1. Although the constructions under investigation adopted an emphatic, imperative-like nature, they began as

declarative sentences. This is clear based on the form of the accusative enclitic in declarative sentences compared to that of imperative constructions. Galician clitics have two primary clitic-determiner forms, known as “first forms” and “second forms” (Freixeiro 2006).<sup>56</sup>

*First forms*

|    | MASC. | FEM. |
|----|-------|------|
| SG | o     | a    |
| PL | os    | as   |

*Second forms*

|    | MASC. | FEM. |
|----|-------|------|
| SG | lo    | la   |
| PL | los   | las  |

Figure 11. First and second forms of Galician clitics & determiners

Unlike first forms, second forms only arise when the verbal inflection ends in *-r* or *-s*. Second forms are accompanied by the phonological loss of these consonants in the verb form. This morpho-phonosyntactic rule applies to infinitives (1a), the second-person singular (1b), the second-person plural (1c), and the first-person plural (1d). This is true of all tenses, including inflected infinitives (1e).

- (1) a. Podemos            leva-**lo**            alá.            (levar)  
           be-able.PRS.1PL    carry.INF-CL<sub>ACC.M.SG</sub>    there  
           ‘We are able to take it there.’

<sup>56</sup> I do not address “third forms,” which appear only when the verbal inflection is a diphthong and does not occur with the determiner clitic paradigm I discuss below.

- b. Viche-**la**                      cando      chegou?                      (viches)  
 see.PST.2SG-CL<sub>ACC.FEM.SG</sub>      when      arrive.PST.3SG  
 ‘Did you see her when she arrived?’
- c. Atoparade-**las**                      que      non      coñezo      (atoparades)  
 find.PST.PRFV.2PL-CL<sub>ACC.FEM.PL</sub> COMP NEG      know.PRS.1SG  
 ‘You had come across the ones I don’t know.’
- d. Viamo-**los**                      cada      venres                      (viamos)  
 see.IMPV.1PL-CL<sub>ACC.MASC.PL</sub>      each      Friday  
 ‘We used to see them every Friday.’
- e. De      facermo-**la**              ben,              aprobaremos              (facermos)  
 COMP do.INF.1PL-CL<sub>F.SG</sub>      well              pass.FUT.1PL  
 ‘If we do it well, we will pass.’

Unlike the tenses in (1), most ‘true imperative’ forms (i.e. second-person forms; cf. Platzack & Rosegren 1997) end in vowels and, thus, require the first forms.<sup>57</sup> Per the observations by these authors, ‘true imperatives’ are those that make reference to the second-person singular or plural, which may not be embedded or negated as in (2).

- (2) a. Toma-o  
 take.IMP.SG-CL<sub>ACC.M.SG</sub>  
 ‘Take it.’
- b. Comede-as  
 eat.IMP.PL-CL<sub>ACC.FEM.PL</sub>  
 ‘Eat them.’

---

<sup>57</sup> Exceptions are few, but irregular singular imperative forms that end in a diphthong such as *fai*, from *facer* (‘to do, to make’), would take neither first nor second forms. I leave these exceptions aside, as they are orthogonal to my point here.

As shown from the examples in (1), the verb of surface string *ve-lo* ('You see it') has an underlying form *ves* in contrast to the imperative form which, based on (2), would be *véo* ('Look at it'). Thus, based on this distinction, I propose that the data observed in Step 1 are declarative sentences.

There are several observations about Step 1 that are vital to my discussion here regarding the diachronic development toward the present-day construction. First, the construction in Step 1 permitted left-peripheral constituents, such as topicalized elements, that agreed with the accusative enclitic found on *ves* ('you see'). As I showed in §2.2.1, topicalization is not permitted in the present-day construction, unlike what we find in the data that represents Step 1 (3).

- (3) a. **O gaiteiro<sub>i</sub>,**            **vé-lo<sub>i</sub>**                    **ái**    **vai**                    **para**    **a**    **romaxe**  
           the bagpipe-player    see.PRS.2SG-CL<sub>M.SG</sub>    there    go.PRS.3SG    for    the pilgrimage  
           ‘The bagpipe player, you see him, there he goes.’
- b. **I**            **as cardas de lá<sub>i</sub>,**    **ve-las<sub>i</sub>**                    **elí’stán**  
           and    the cards    of wool    see.PRS.2SG-CL<sub>F.PL</sub>    there-be.PRS.3PL  
           ‘And the wool cards, you see them, there they are.’

As shown in §2.2.1, no material may precede the main-clause complementizers *velai/velaqui*. Moreover, the topicalized constituents of (3) clearly agree in [NUMBER] and [GENDER] with the accusative clitic. This will serve as the diachronic base for the C-AGR characteristic highlighted in §2.4 regarding the relation between the subject and the C-AGR clitic.

The extension of Step 1 is also shown based on the fact that this phrase was used in the second-person plural, as shown in (4).

(4) Os mociños de Galás<sub>i</sub>, véde-los<sub>i</sub> acolá van  
 the children.DIM of Galás see.PRS.2PL-CL<sub>M.PL</sub> over.there go.PRS.3PL

‘The little children of Galás, you see them, there they go’

As discussed above, the second form of the clitic indicates that this is the declarative *vedes* (‘you (pl.) see’), and not the imperative *vede* (‘see’), which would produce *vede-os*.

There are other data points worth mentioning that will reveal themselves to be pertinent in the diachronic development of this construction. First, there are clear patterns in the data with respect to what follows the verb *ver* (‘to see’). In the examples above (3-4), I have shown that the topicalized DPs in the left periphery were permitted during the time of Step 1. However, the most common variations of this construction were that as in (5) and (6).

(5) a. Vé-los aí van pola barca todos eles  
 see.PRS.2SG-CL<sub>M.PL</sub> there go.PRS.3PL by-the boat all they

‘You see them, there they go by boat, all of them.’

b. Vé-la alá vai, a raposa polo prado  
 see.PRS.2SG-CL<sub>F.SG</sub> there go.PRS.3SG the fox by-the field

‘You see it, there goes the fox through the field.’

c. E vé-lo acolá está Pexegueiro  
 and see.PRS.2SG-CL<sub>M.SG</sub> yonder be.PRS.3SG Pexegueiro

‘And you see him, over there is Pexegueiro.’

d. Vé-lo aí vai o mugarés  
 see.PRS.2SG-CL<sub>M.SG</sub> there go.PRS.3SG the mugarés

‘You see him, there goes the man from Mugaros.’

- (6) a. Vé-lo                      aí      vai                      de      porta      en      porta  
 see.PRS.2SG-CL<sub>M.SG</sub>      there      go.PRS.3SG      from      door      in      door  
 ‘You see him, there he goes from door to door.’
- b. Vé-la                      aí      vai                      tan      campante  
 see.PRS.2SG-CL<sub>M.SG</sub>      there      go.PRS.3SG      so      satisfied  
 ‘You see her, there she goes so satisfied (with herself).’

What all of the examples above have in common is the order of constituents. Following the *ve-lo aí* portion of the sentence, in (5) we see an intransitive verb in combination with a postverbal subject that agrees in [NUMBER] and [GENDER] with the accusative clitic argument of *ves* (‘you see’). Just as I described in §2.4 for the present-day construction, the postverbal subject is often elided just as in (6). In Chapter 4 I show how this strict order of constituents plays an important role in the development of the modern-day construction.

The observations above are worth noting when discussing what input children must have been presented with in the early stages of this construction. As I noted in §2.4, the complementizers *velai/velaqui* may also combine with a transitive verb outside of the Galician C-AGR construction. However, transitive examples with the construction highlighted above for Step 1 are very few.

- (7) a. Vé-lo<sub>i</sub>                      aí      cho<sub>i</sub>                      tran                      morto  
 see.PRS.2SG-CL<sub>M.SG</sub>      there      CL<sub>DAT.2SG-CL<sub>M.SG</sub></sub>      carry.PRS.3PL      dead.M.SG  
 ‘You see him, there they are carrying him to you, dead.’
- b. Vé-la<sub>i</sub>                      aí      tes                      pro<sub>i</sub>      na      fiestra  
 see.PRS.2SG-CL<sub>F.SG</sub>      there      have.PRS.2SG      pro      in-the      window  
 ‘You see her, there you have (her) in the window.’

Even in the scarcity of transitive examples, we may see clear patterns of co-indexation with the accusative clitic in the *ve-lo aí* portion of the sentence and the object nominal of the second transitive verb. In (7a), the accusative clitic of *ves* ('you see') is coindexed with the accusative clitic of the verb *tran* ('they bring'). The same is found in the example in (7b), a case of object drop in the second clause with *tes* ('you have'). Here the accusative clitic *la* is coindexed with *pro* that stands in for the object.<sup>58</sup>

### 3.2.2 Step 2

The proposal for Step 2 at first glance does not seem drastically different from that of Step 1. However, I will present data that support the assumptions regarding prosody and the implications for this proposal in the syntax. Until now, I have described the lone difference between Step 1 and Step 2 as the former having a prosodic break and the latter lacking one. I claim that the lack of prosodic break of Step 2 must be assumed from the data presented below.

- (8) a. Cala            Antón,    **que**    **vé-lo**                            aí    vén  
           hush.IMP.SG    Antón    COMP    see.PRS.2SG-CL<sub>M</sub>.SG    there    come.PRS.3SG  
           e        pode-te                            ouír!  
           and    be-able.PRS.3SG-CL<sub>2SG</sub>    hear.INF  
           'Shut up, Antón, because you see him, there he comes and he can hear you!'

---

<sup>58</sup> There is no formal account of object drop in Galician, and even the descriptive grammars are vague on the extent to which it is applied. A full analysis of object drop is orthogonal to my discussion here, but I can confirm that object drop is permissible in contexts in which the object dropped has just been mentioned in the discourse, much like in Portuguese (Raposo 1986, 1997), which is what we see in (7b). I leave further discussion of this topic for future research.

b. Veña                    **que**   **xa**                    **vé-los**                    alá    están  
 come.SUBJ.3SG    COMP   already    see.PRS.2SG-CL<sub>M</sub>.PL    there   be.PRS.3PL  
 os    carros!  
 the    haywains

‘Come on, because you already see them, there are the haywains!’

c. Calade,            diaños,            calade,            **que**   **vé-lo**                    alí  
 hush.IMP.PL    devils            hush.IMP.PL    COMP   see.PRS.2SG-CL<sub>M</sub>.SG    there  
 e    inda    vos    ha                    d’oir  
 and    still    CL<sub>2PL</sub>    have.PRS.3SG    COMP-hear.INF

‘Shut up, you devils, shut up, (because) you see him there and he can still hear you.’

In (8), we find instances of enclisis where we should find proclisis. Henceforth, I will refer to the structure in (8) as a ‘fixed-enclisis’ pattern.<sup>59</sup> As I describe below in §3.3.3 in Western Iberian (WI) complementizers such as *que* (‘that’) and adverbs such as *xa* (‘already’) in preverbal position serve as the prosodic host necessary for a clitic to be a in a “well-formed prosodic word at Spell-Out” (R&U 2005:650). As I will show, the learning task of dividing those preverbal constituents that trigger proclisis from those that permit enclisis is a daunting one. However, the data in (8) are specific to this construction during this particular stage in the diachronic development of the complementizers *velái/velaquí*.

Before addressing analyzing the data with respect to the learning path of the enclisis/proclisis split in WI, I should point out that there are two possibilities to what we see in the data in (8), both of which are considered and discussed at length at the end of this chapter upon discussing the reanalysis of the main-clause complementizers. First, we may analyze the *vé-lo aí* string as one lexical item, already having undergone reanalysis. This seems to be the case in (8c), where

<sup>59</sup> I make a finer distinction between these examples in §3.7 below.

the imperative form of *calar* ('to hush') is in the plural, yet we find the verb *ver* ('to see') in the singular. As I showed above in (4), Step 1 also boasted a plural variation in *vede-lo aí*, which we would expect in (8c) were this construction at the same point in its diachronic history as in Step 1. Were this the case, it would remain to be seen how to account for *vé-lo aí* appearing below the complementizer *que*, as this is not permitted in the modern grammar (cf. §2.2.1). In contrast, reanalysis must not have taken place in (8b), as we find a plural clitic agreeing in [NUMBER] and [GENDER] with the subject of the relative clause *os carros* ('the haywains'). Had reanalysis already occurred, we would expect lexically fixed forms as we find in present-day Galician. The example in (8c), therefore, is indicative of cliticization to a verbal structure rather than a fixed form as tentatively speculated in (8b). In §3.7, I conclude that these two structures are distinct possible realizations that must be accounted for by different syntactic derivations.

In the next sections that follow, I discuss the learning path of cliticization in Western Iberian, focusing on the findings of Costa, Fiéis & Lobo (2015) and how they apply to Galician in particular. Subsequently, I take an excursus on category learning and discuss the implications for reanalysis more generally. These steps are crucial in justifying and setting the stage for the "diachronic leap" that I propose that led to reanalysis. In Chapter 4, I discuss the full reanalysis of Step 3 and the implications that the overlap of Steps 1-3 had on the engenderment of Step 4.

### 3.3. Enclisis and proclisis in Western Iberian: the learning path

In this section, I describe the finite enclisis/proclisis split in Western Iberian. I present the data collected in Costa, Fiéis & Lobo (2015) (henceforth CFL) on the child L1 acquisition of cliticization patterns in European Portuguese. As European Portuguese and Galician are two of the three languages that make up Western Iberian, a variety of Romance that shows a finite enclisis/proclisis split, I assume that the learning path is the same for Galician children as it is for

those who grow up exposed to European Portuguese.<sup>60</sup> Subsequently, I tie in these data with the formal proposal by Raposo & Uriagereka (2005) and show how this approach is indeed viable from a language acquisition standpoint.

### 3.3.1 Enclisis or proclisis: What's in your left periphery?

Whether enclisis or proclisis is derived in finite clauses in Western Iberian is a question of what preverbal material is present. As I will show in §3.3.3, this amounts to what material is found in the left periphery of the clause. Following the evidence shown in §3.3.2, I claim that enclisis is the unmarked pattern in WI, so first I will present data on what derives proclisis.

In addition to *velai/velaqui*, there are a vast number of triggers for proclisis in finite clauses, all of which must be found in preverbal position: negative constituents including sentential negators (9a), negative adverbs (9b) and negative subjects (9c); quantified subjects such as those that begin with *todos* ('all') (10a) and *cada* ('each') (10b); some sentential adverbs such as *xa* ('already') (11a), *sempre* ('always') (11b), and *aínda* ('still') (11c); declarative focused constituents (12); constituents marked with a *wh*- word, including both question words (13a) and exclamation words (13b); and subordinate clauses with an overt complementizer such as *que* ('that') (14a), *porque* ('because') (14b), and the conditional *se* ('if') (14c).

- (9) a. **Non** o queremos /\*queremo-lo  
 NEG CL<sub>M,SG</sub> want.PRS.1PL want.PRS.1PL-CL<sub>M,SG</sub>  
 'We don't want it.'

---

<sup>60</sup> From a historical standpoint, Galician and European Portuguese were once one language, and many scholars consider this to be the case still. Based on the fact that these two languages share almost identical syntactic properties across the board, I claim that the use of European Portuguese child data to describe what occurs in Galician is appropriate.

- b. **Nunca** as vira /\*víraas  
 never CL<sub>F.PL</sub> see.PSTPERF.3SG see.PSTPERF.3SG-CL<sub>F.PL</sub>  
 ‘She had never seen them.’
- c. **Ninguén** mo dixera /\*dixéramo  
 no-one CL<sub>DAT.1SG-CL<sub>M.SG</sub></sub> say.PSTPERF.3SG say.PSTPERF.3SG-CL<sub>DAT.1SG-CL<sub>M.SG</sub></sub>  
 ‘No one had ever told me.’
- (10) a. **Todo o mundo** o fai /\*faino distinto  
 all the world CL<sub>M.SG</sub> do.PRS.3SG do.PRS.3SG-CL<sub>M.SG</sub> distinct  
 ‘Everyone does it differently.’
- b. **Cada can** che atouza /\*atózache  
 each dog CL<sub>DAT.2SG</sub> bark.PRS.3SG bark.PRS.3SG-CL<sub>DAT.2SG</sub>  
 ‘Every dog barks, ya know.’
- (11) a. **Xa** llo deu /\*duello?  
 already CL<sub>DAT.3SG-CL<sub>M.SG</sub></sub> give.PST.3SG give.PST.3SG-CL<sub>DAT.3SG-CL<sub>M.SG</sub></sub>  
 ‘Did she already give it to him?’
- b. **Sempre** che era /\*érase bon rapaz  
 always CL<sub>DAT.2SG</sub> be.IMPFV.3SG be.IMPFV.3SG-CL<sub>DAT.2SG</sub> good boy  
 ‘He was always a kid.’
- c. **Aínda** me chama /\*chámame todo-los días  
 still CL<sub>1SG</sub> call.PRS.3SG call.PRS.3SG-CL<sub>1SG</sub> all-the days  
 ‘She still calls me every day.’

- (12) **Moitas malleiras** me deu /\*deume miña nai  
 many beatings CL<sub>1SG</sub> give.PST.3SG give.PST.3SG-CL<sub>1SG</sub> my mother  
 ‘Many beatings my mother gave me.’
- (13) a. **Quen** me viu /\*viume?  
 who CL<sub>1SG</sub> see.PST.3SG see.PST.3SG-CL<sub>1SG</sub>  
 ‘Who saw me?’
- b. **Que ben** o fixeron /\*fixérono!  
 what well CL<sub>M.SG</sub> do.PST.3PL do.PST.3PL-CL<sub>M.SG</sub>  
 ‘How great did they do it!’
- (14) a. **Dixeran que** vos chegaban /\*chegábanvos  
 say.PSTPERF.3PL COMP CL<sub>DAT.2PL</sub> arrive.IMPFV.3PL arrive.IMPFV.3PL-CL<sub>DAT.2PL</sub>  
 dúas caixas  
 two boxes  
 ‘They had said that two boxes were arriving for you.’
- b. **Puidemos ir porque** nos chamaron /\*chamáronnos  
 be-able.PST.1PL go.INF because CL<sub>1PL</sub> call.PST.3PL call.PST.3PL-CL<sub>1PL</sub>  
 a tempo  
 at time  
 ‘We were able to go because they called us in a timely manner.’
- c. **Se** o ves /\*ve-lo, avísame  
 if CL<sub>M.SG</sub> see.PRS.2SG see.PRS.2SG-CL<sub>M.SG</sub> notify.IMP.SG-CL<sub>1SG</sub>  
 ‘If you see him, let me know.’

Enclisis occurs in every other context: non-quantified/focused subject constituents such as pronouns (15a) and standard DPs (15b), verb-first strings such as commands (16a) and normal declarative strings (16b), clitic left-dislocated (CLLD) constructions (17), and topicalized DPs (18).<sup>61</sup>

- (15) a. **Ela** chamouno /*\*o* chamou onte  
 she call.PST.3SG-CL<sub>M.SG</sub> CL<sub>M.SG</sub> call.PST.3SG yesterday  
 ‘She called him yesterday.’
- b. **Aqueles pícaros** atopáronse /*\* se* atoparon  
 those kids find.PST.3PL-CL<sub>REFL.3SG</sub> CL<sub>REFL.3SG</sub> find.PST.3PL  
 moi cansos despois de xogaren  
 very tired aftern COMP play.INF.3PL  
 ‘Those children found themselves very tired after playing.’
- (16) a. *pro* Falemo-lo /*\*O* falemos canto antes  
*pro* talk.IMP.1PL-CL<sub>M.SG</sub> CL<sub>M.SG</sub> talk.IMP.1PL when before  
 ‘Let’s talk about it as soon as possible.’
- b. *pro* Chegoulles /*\*Lles* chegou cedo á festa  
*pro* arrive.PRS.3SG-CL<sub>DAT.3PL</sub> CL<sub>DAT.3PL</sub> arrive.PRS.3SG early to-the party  
 ‘He arrived early to the party on them.’
- (17) **Meus pais,** recollinos /*\*os* recollín do centro  
 my parents collect.PST.1SG-CL<sub>M.PL</sub> CL<sub>M.PL</sub> collect.PST.1SG from-the center  
 ‘My parents, I picked them up from downtown.’

<sup>61</sup> It has been suggested by many authors (Barbosa 1995, 2000; Raposo 2000; Fernández-Rubiera 2013, *i.a.*) that all preverbal subjects are topics in main clauses. I follow this line of thinking and discuss this theoretical stance in detail in §3.3.3.

- (18) **Os alunos da semana pasada,** mandáronnos  
 the students of-the week last send.PST.3PL-CL<sub>DAT.1PL</sub> CL<sub>DAT.1PL</sub>  
 /\*nos mandaron os apontamentos  
 CL<sub>DAT.1PL</sub> send.PST.3PL the notes  
 ‘The students from last week, they sent us the notes.’

There are certain constituents that may permit both enclisis and proclisis and do not behave in the same uniform manner as outlined above.<sup>62</sup> For example, a handful of adverbs such as *agora* (‘now’) may trigger either enclisis or proclisis.

- (19) a. **Agora cho** levo  
 now CL<sub>DAT.2SG</sub>-CL<sub>M.SG</sub> carry.PRS.1SG  
 ‘I’ll take it to you in a bit.’
- b. **Agora lévocho**  
 now carry.PRS.1SG-CL<sub>DAT.2SG</sub>-CL<sub>M.SG</sub>  
 ‘I’ll take it to you right now.’

We should note that the translations in (19) are different depending on whether the clitic is proclitic or enclitic on the verb. This is not true in all cases, as there is no semantic change with other adverbs and adverbial phrases that permit both enclisis and proclisis such as *por iso* (‘because of that’) and *moitas veces* (‘a lot of times’). Below, I equate this to a difference between a topicalization and focalized positioning of the phrase, leading to a distinct prosody between the two options.

---

<sup>62</sup> Amongst this group of constituents that permits both enclisis and proclisis are inflected infinitivals. I leave them aside due to the fact that their variation is seen as purely stylistic and unrelated to the prosody and the syntactic positioning of a particular constituent as a host for the clitic, two aspects that are crucial to my claims below. Therefore, I leave the discussion of inflected infinitivals for future research.





six-year-olds, and 20 seven-year-olds. They tested 36 different structures that covered both enclitic and proclitic contexts, divided as in (23).<sup>63</sup>

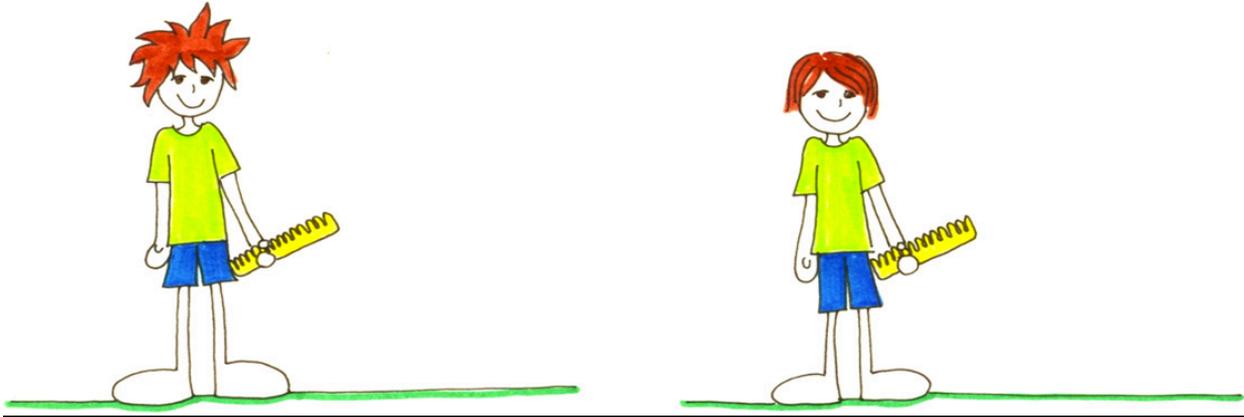
(23) Enclisis/proclisis conditions tested

- a) simple clauses with no proclisis trigger (expected: ENCLISIS)
- b) simple clauses with negation (expected: PROCLISIS)
- c) simple clauses with negative subjects (expected: PROCLISIS)
- d) simple clauses with quantified DPs (expected: PROCLISIS)
- e) simple clauses with preverbal adverb *já* ‘already’ (expected: PROCLISIS)
- f) finite subordinate complement clauses with *querer* ‘want’ (expected: PROCLISIS)
- g) subordinate adverbial clauses with *porque* ‘because’ (expected: PROCLISIS)

Questions were administered with the aid of pictures and without the presence of a clitic in order not to trigger a particular clitic placement in any given answer. Figures 12 and 13 describe the methodology.

---

<sup>63</sup> CFL also test coordinated clauses (a construction where enclisis is expected). However, they give no conclusions as to where it falls into the acquisition path regarding cliticization due to the lack of methodological cohesion with the rest of the tests. Therefore, I leave it out of my analysis moving forward.



[picture 1] This boy is untidy. Look at his hair. But he has a comb.

[picture 2] What did the boy do?

Expected answer: *Penteou-se*

comb.PST.3SG-CL<sub>REFL</sub>.3SG

'He combed himself.'

*Figure 12. Example of elicited production task for simple clauses (CFL 2015:18)*



[picture 2] These two girls have used their comb and now they are tidy. But this girl is still untidy. What didn't she do?

Expected answer: *Não se penteou*

NEG CL<sub>REFL.3SG</sub> comb.PST.3SG

'She didn't comb herself.'

*Figure 13. Example of elicited production task for negative clauses* (CFL 2015:18)

Both Figures 12 and 13 guarantee the presence of either proclisis or enclisis, based on the presence of or lack of a proclisis trigger based on the structures mentioned in (23).

### 3.3.2.2 Results and learning path hypotheses

The results of the elicited production tasks were codified according to several categories, only two of which are relevant here: i) proclisis, ii) enclisis, iii) doubling, iv) omission, and v) other.<sup>64</sup> I focus on proclisis and enclisis only, as the other options are orthogonal to my investigation in this dissertation. Therefore, Table 3 and Table 4 below show the global results found upon collapsing enclisis and proclisis contexts in which only answers with clitics are considered.

<sup>64</sup> Here, 'other' was classified under not just answers that presented no other potential options (enclisis, proclisis, doubling, omission) but also the unsuccessful methodological questions regarding coordinated structures. As these do not form part of my focus here, I leave this detail aside.

*Table 3. Percentage of answers obtained in enclisis contexts*

|             | ENCLISIS | PROCLISIS |
|-------------|----------|-----------|
| 5 year olds | 99.5%    | .5%       |
| 6 year olds | 100%     | 0%        |
| 7 year olds | 100%     | 0%        |

*Table 4. Percentage of answers obtained in proclisis contexts*

|             | ENCLISIS | PROCLISIS |
|-------------|----------|-----------|
| 5 year olds | 64%      | 36%       |
| 6 year olds | 46.2%    | 53.8%     |
| 7 year olds | 42.%     | 57.2%     |

There are multiple conclusions to be drawn, from which I will base a considerable portion of my hypotheses regarding the learning path children take in acquiring the correct clitic placement in §3.3.3. First, overgeneralization of proclisis in enclisis contexts was practically non-existent, with only one case of two-hundred sixty-four answers given in the five-year-old group. However, there was an elevated percentage of enclitic production (+40%) in contexts of proclisis from children of all groups, even the seven-year-olds. The biggest jump in proclitic accuracy came between five-year-old group and the six-year-old group, almost a 50% increase (36% to 53.8%). Due to the problematic nature of the proclitic contexts, the rates of proclisis with each syntactic condition (cf. 23) are of particular interest here and to my investigation moving forward. These are shown in Table 5.

*Table 5. Rates of proclisis based on syntactic condition*

|                              | 5 YEAR OLDS | 6 YEAR OLDS | 7 YEAR OLDS |
|------------------------------|-------------|-------------|-------------|
| NEGATION                     | 61%         | 87.5%       | 84.4%       |
| NEGATIVE SUBJECTS            | 48.3%       | 69.2%       | 70.9%       |
| COMPLEMENT CLAUSES           | 46.7%       | 74.6%       | 65.8%       |
| ADVERB <i>já</i> ('already') | 36%         | 58.2%       | 65.8%       |
| ADVERBIAL CLAUSES            | 14%         | 20%         | 32.9%       |
| QUANTIFIED SUBJECTS          | 9.9%        | 12.2%       | 21.3%       |

An important distinction made by CFL is that between *complement clauses* (those that are marked with *que* 'that') and *adverbial clauses* (those that are marked with *porque* 'because').<sup>65,66</sup> As I noted above (cf. 20), there is some variability with adverbial clauses that is not present with complement clauses, so this distinction is warranted. The results show that neither is every syntactic context the same with respect to accuracy of expected proclisis nor does each age group treat each syntactic condition the same way. There was a consistent increase in accuracy from the five-year-old group to the six-year-old group in all six contexts in Table 5. From the six-year-old group to the seven-year-old group, there was an increase in all but two contexts: negation and complement clauses. As both of these contexts induce proclisis categorically, these results are surprising, but most likely represent an anomalous result in the data. The consistently-low but ever-increasing percentages of the accuracy of proclitic placement across contexts of the adverb *já* ('already'), adverbial clauses, and quantified subjects present us

<sup>65</sup> We may include *aínda que* ('even though') in the categorization of *adverbial clauses* (cf. 20).

<sup>66</sup> Although there is variability as I showed above in (20), CFL do not specify whether their questions that triggered adverbial clauses contained a null, postverbal, or preverbal subject.

with the most intriguing details when discussing the learnability and acquisition of constituent type with respect to cliticization patterns, which I discuss below in further detail.

From the data above, the scale of development that CFL propose is the following:

- (24) NEGATION > NEGATIVE SUBJECTS > FINITE COMPLEMENT CLAUSES > ADVERB *já* >  
FINITE ADVERBIAL CLAUSES > QUANTIFIED SUBJECTS

According to the authors, there are both learnability and diachronic motivations behind the order in (24). However, I will focus solely on the learnability side, as I believe there are sufficient motivations that explain the results beyond the hypotheses made by CFL. Moreover, there are motivations specific to Galician that we may assume make this learning path even firmer for a Galician child.

Beginning with negation, this seems to be the most salient and easiest clitic directionality trigger that children are able to compute, namely due to the [+NEG] feature. As the data from CFL show, not all negative items are acquired at the same speed, however. What, then, distinguishes sentential negation with *não/non* ('no') from other lexical items specified for the feature [+NEG] that I mention in (9) such as negative subjects *ninguén* ('no one') or negative adverbials like *nunca* ('never')? My assumption deals with both structural and categorically relevant aspects. As claimed by Gupton (2014:205), *non* in Galician is clitic-like in the sense that it interacts at a syntactic as well as morphophonological level with the clitics it hosts. Based on the observations in Gupton (2010:221-222) regarding *non* having an X<sup>0</sup> status, Gravely & Gupton (2021), provide further evidence for this claim based on the phonological realization of a string such as in (25).

- (25) Non o vexo  
NEG CL<sub>M,SG</sub> see.PRS.1SG  
'I don't see it.'

The consonant-final *-n* in Galician is always velar, barring the interaction with a vowel of the subsequent lexical item when the latter is a clitic.<sup>67</sup> Therefore, the phonological realization of (25) is [no.nu.βe.ʃu] rather than [no.ŋo.βe.ʃu]. As the child receives more and more morphophonological evidence of her language, she must quickly come to the conclusion that there is something different about functional items such as *non* ('no'), *alguén* ('someone'), and verbal inflections such as *fan* ('they make'), as these all trigger the aforementioned phonological realization of the non-velar *-n*, whilst phrases such as that in (26) do not although they, too, end in *-n*.

- (26) [No chan] a atopara  
 in-the floor CL<sub>F.SG</sub> find.PST.PRFV.3SG  
 'On the floor I found her.'

In contrast to what we have seen for (25), the phonological realization of (26), [nu.tʃa.ŋa.ːto.pa.ra], is what we find when XP constituents precede clitics.

Interestingly enough, *ninguén* ('no one') is also realized without a velar *-n* when preceding accusative clitics, which would seem to suggest that it, too, is an X<sup>0</sup> element. What difference should there be between *non* ('no') and *ninguén* ('no one') to the learner? As hypothesized by CFL, the fact that *ninguén* falls into a category of possible grammatical subjects is surely a reason for which a child may hypothesize that *ninguén*, barring the phonological realization akin to that of *non* when preceding clitics, pertains to the larger category of subject DPs which, as we have seen, do not trigger proclisis. Even in the face of blatant evidence for the contrary, we see that even children at the age of seven years old in the CFL study still produce enclisis roughly 30% of the time with preverbal negative subjects.

---

<sup>67</sup> As originally noted by Álvarez Caccamo (1989) and discussed in further detail by Gravely (2020), this may apply to accusative clitics and determiner clitics alike in certain Galician varieties.

The third stage along the hypothesized learning path in (24), finite complement clauses, presents the most difficult one for which to draw a learning hypothesis. One view we may take deals with the erroneous grouping of complement clauses marked with *que* ('that') alongside those which CFL call "finite adverbial clauses" marked by *porque* ('because') or *aínda que* ('although'). Due to the fact that there are instances of finite enclisis with the latter two complementizers, a child may assume that there simply is no difference between those clauses marked with *que* ('that') and those marked with *porque* ('because'), regardless of the fact that she has only received input with finite enclisis with the latter complementizer. A second hypothesis not excluded by that described above involves the application of multiple derivational workspaces. Although not addressed as a possible hypothesis in CFL, this is quite plausible even in certain instances in adult grammar, as well, owing to the fact that multiple derivations may be created and, in turn, pieced together to form what may be perceived as a single surface string. We see this in examples from contemporary spoken Galician, as in (27), in which enclisis is common and sometimes even preferred with the phrase *é que* ('It is (the case) that').

(27) É                    que    íanse                    de chulos    e    agora están  
       be.PRS.3SG    COMP    go.IMPFV.3PL-CL<sub>REFL.3</sub>    of cool.PL    and    now    be.PRS.3PL  
       todos    cagados  
       all    terrified

'It's just that they were playing it cool and now they're all terrified.'

I abstained from including examples such as (27) in my analysis above, as I believe this type of surface-level complement clause is derived differently than that in (14a). Moreover, it is solely this complement-clause phrase that permits enclisis. Following much literature on multiple workspaces (Uriagereka 1999, Nunes 2004, Guimarães 2004, Lohndal 2014, *i.a.*), I claim that

(14a) and (27) differ in that the latter is formed by way of combining *é que* ('It is that') with the content of the lower clause. From a language acquisition standpoint, the concept of multiple workspaces may very well be an innate aspect of the human language faculty, as children must be able to process and produce constituents individually, only later merging them into a larger derivational structure.<sup>68</sup> Further support that this is indeed a vital part of the human language faculty comes from evidence of the need for this type of machinery in American Sign Language (ASL), as well (Emmorey, McCullough, Mehta & Grabowski 2014; Emmorey, Mehta, McCullough & Grabowski 2016). Specifically, for children, they may instantiate multiple workspaces in processing or production simply due to memory limitations or for reasons of computational efficiency, which is an additional possibility as to why children may produce finite enclisis in complement clauses.

The next category along the learning path hypothesized by CFL, *já* ('already'), presents a difficult acquisition process for different reasons than those hypothesized above. Although CFL only tested the lexical item *já*, there is a large group of adverbs that trigger proclisis in preverbal position, as I mentioned above in referencing the examples in (11). Amongst these are adverbs of manner (e.g. *simplemente* 'simply', *posiblemente* 'possibly'), sentential adverbs (*aínda* 'still', *ben* 'well', *sempre* 'always'), adverbs of location (*aquí* 'here', *alá* 'there'), and adverbial phrases (*por iso* 'because of that', *moitas veces* 'a lot of times.'). What makes this group of lexical items difficult to piece together in regard to finite enclisis-proclisis is the fact that the child must acquire them one by one, adding to a non-morphologically uniform category based on limited evidence. Because these adverbs appear both preverbally and in multiple postverbal positions, her exposure to these items in preverbal position may not be as frequent as necessary in order to

---

<sup>68</sup> I remain agnostic as to the intrinsic details of this process, as they vary slightly from theory to theory. What I wish to highlight here is the natural ability to create, hold, and combine multiple constituents in working memory, even those as large as entire clauses or extended adverbial phrases.

categorize them as proclisis triggers. This task, therefore, is akin to what Yang (2016) describes for irregular past tenses in English, but in the context of morphosyntax: the child needs a certain level of input in order to break away from her regularization biases. In the case of irregular past tenses in English, this is a question of verbal morphology. With respect to the present case, this is a matter of retreating from the unmarked pattern of enclisis. Possibly a more daunting task with this grouping of words and phrases is that of learning under what circumstances these adverbs may or may not trigger proclisis, as I have discussed above for (19). As discussed at length by Freixeiro (2006:694-708), the same word or phrase may induce proclisis or permit enclisis based on whether that phrase is topicalized (28a) or focalized (28b), essentially serving as a host for the clitic when focalized but not when topicalized and, hence, requiring enclisis on the verb.

- (28) a. Por iso, fancho ben  
 for that do.PRS.3PL-CL<sub>DAT.2SG</sub>-CL<sub>M.SG</sub> well  
 ‘Because of that, they do it for you well.’
- b. POR ISO cho fan ben  
 for that CL<sub>DAT.2SG</sub>-CL<sub>M.SG</sub> do.PRS.3PL well  
 ‘That’s why they do it for you well.’

As these two phrases are both prosodically and syntactically distinct, this may be yet another task the child must not only come to recognize but, in addition, receive sufficient input for in order to distinguish the two. In addition, because not all adverbs may be used in constructions with both enclisis and proclisis (e.g. *sempre* ‘always’ only induces proclisis in preverbal position), this is yet a further subcategorization which proves necessary for the child to make.

The last step of the learning path, *quantified subjects*, presents a challenge similar to that of negative subjects but in a less discrete manner. Although investigation concerning children’s

abilities to handle quantification on the same level as that of adults has shown to be shaky (Musolino et al. 2000), it seems to be the pragmatic use of quantification that holds them back (Musolino & Lidz 2006). Thus, it isn't clear that there is direct connection between the way in which children use or understand quantifiers and their understanding of them as triggers for proclisis. Considering the low rates of proclisis with quantified subjects among children in Table 5, it may be that because quantifiers are a small closed-class category, the phrases they head may simply fly under the radar as proclisis triggers and be grouped in with more common subjects in which enclisis is found. In this case, it may very well be the scarcity with which quantifier phrases are used in preverbal subject position.

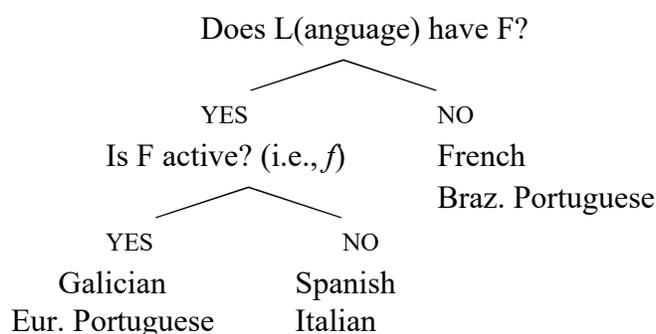
From the discussion of the five groups of proclisis triggers highlighted by CFL, we may identify several causes as to why children do not immediately reset their parameter for the finite enclisis-proclisis split in Galician. While some words may be harder to track due to their lack of uniformity (e.g. adverbs and adverbial phrases), children may resist retreating from the unmarked enclisis pattern based on the frequency with which these words and phrases are used in preverbal position. As I discuss in §3.4, frequency is the most important factor with respect to basic category learning (Cassani, Grimm, Daelemans & Gillis 2018) and this is likely the case for morphosyntactic phenomena, as well. Before discussing category learning, however, I link the data discussed in this section with the formal account of cliticization in Western Iberian as proposed by Raposo & Uriagereka (2005).

### 3.3.3 Cliticization in Western Iberian: Raposo & Uriagereka (2005)

The syntactic proposal put forth Raposo & Uriagereka (2005) (henceforth R&U) for cliticization in WI has several fundamental advantages over PF-based accounts to cliticization order (Bonet 1991, *et seq.*), particularly with respect to learnability and the identification of what

groups of constituents trigger what orders, as we examined above (cf. §3.3.2). Building off of work by Uriagereka (1988, 1995a, 1995b), these authors propose an ‘active’ left-peripheral projection *f* that attracts clitics in WI. As *f* is a syncretic projection, it is also responsible for hosting topicalized and focused constituents, as well. Unifying these two phenomena under one projection, they propose a kind of ‘language family’ parameter amongst Romance varieties with respect to whether *f* is licensed (or ‘active’ in their words).

(29) WI parameter (R&U 2005:644)



In (29), the question determining the presence (or lack thereof) of *F* immediately separates Eastern and Western Iberian from French and Brazilian Portuguese.<sup>69</sup> The lack of *F/f* in these languages means not only that clitics must be probed from another functional head (presumably *T*), but that these languages also lack the ability to topicalize or front focused constituents.<sup>70</sup> R&U then make a distinction between an ‘active’ (*f*) and ‘inactive’ (*F*) functional projection. They propose that for Spanish and Italian, *F* is present, which permits both a site for cliticization as well as a projection that permits dislocated constituents to move to its specifier.<sup>71</sup> This

<sup>69</sup> These languages seem to have English-type topicalization, a topic that does not concern my investigation here. I refer the interested reader to Kato & Raposo (2007) for a full analysis of this phenomenon.

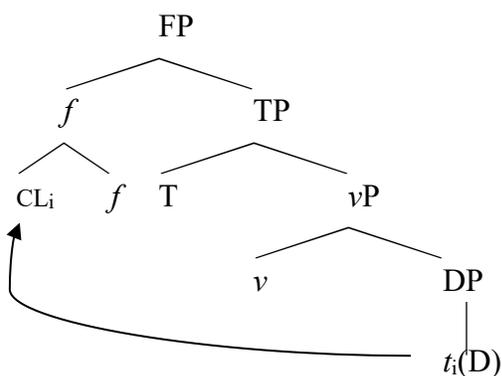
<sup>70</sup> This must not affect the ability to move *wh*-elements, however, which would assume a [+Q] on *C* if no other projection above *T* is present. This pure speculation based on where these authors left the discussion. As this bears no importance on my overall picture for their proposal of *f*, I leave this point aside.

<sup>71</sup> This is an unpopular view in more recent literature (Gallego 2016, Roberts 2010a, *i.a.*). Instead, the claim is that clitics adjoin to the verb in *v* before moving to *T*. I follow R&U, however, due to the rise in interpolated constructions in Spanish (at least the Peninsular variety) such as in (i).

dislocation, however, does not affect the position of the clitic with respect to the verb (i.e., these languages always have finite proclisis), hence the ‘inactive’ F. As we have seen in §3.3.2, Galician and European Portuguese are significantly different in that finite clauses alone do not determine whether enclisis or proclisis is present; it is the material (or lack thereof) that is found in the left-periphery of the clause. Due to this fact, these authors propose that F in these languages is ‘active’ (represented by *f*). This is the focus of my investigation moving forward.

Consider (30), with clitic movement independent of the verb on T.<sup>72</sup>

(30) *Clitic movement above the verb in WI*



Determining whether the clitic ends up enclitic or proclitic on the verb is a question of what material is able to appear to the clitic’s left. R&U make a notable observation (2005:656) that describes the nature of both enclisis and proclisis seen above: clitics in WI always lean left. That is, they may never be sentence initial and must always have some suitable material to their left as a host.<sup>73</sup> How this happens, however, is dependent on a few factors. Looking back at the

- 
- (i) Me    lo    **puto**    quiero    ya  
 CL<sub>REFL.1SG</sub> CL<sub>M.SG</sub>    fucking    want.PRS.1SG    already  
 ‘I fucking want it now.’

Although this issue for Spanish is orthogonal to my overall investigation in this thesis, I claim that the example in (i) is sufficient evidence for following the account proposed by R&U regarding the projection *f*.

<sup>72</sup> Below I show actual examples that lead to enclisis and proclisis on the verb (cf. 31). I leave this model as is for subsequent discussion on clitic movement in Chapter 5.

<sup>73</sup> There are exceptions to this statement, such as the example in (i).

situations that trigger proclisis, we have seen that focalized constituents trigger proclisis. Therefore, any phrase bearing a [+FOC] feature will bring said phrase to the clitic's left, serving as the host for the clitic. This is the case for the [+Q] of *wh*-elements, as well. Other constituents such as quantified subjects are said to be “focalized” (Freixeiro 2006:698). However, this terminology seems to refer to something beyond syntactic focus, at least with respect to the prosody typically associated with these constituents. In R&U, they referred to these constituents that were not focalized but still provoke proclisis on the verb as ‘affective phrases.’ Although these authors never specified what brought these constituents to the left periphery, I speculate that this may simply be an [ $\bar{A}$ ] feature present on *f*.<sup>74</sup> Last, base-generated heads in the left periphery (*velai/velaqui, que, porque*) may also serve as the host for the clitic. Their presence in the numeration acts as a signal to the derivation that there is indeed a host and no further movement is needed.

What occurs with topicalized or verb-first constructions to trigger the unmarked enclitic pattern?<sup>75</sup> For topics, it has been proposed by Barbosa (1995, 2000) and Raposo (2000) for European Portuguese and Fernández-Rubiera (2009, 2013) for Asturian that these are base-generated constituents and, thus, are not present during the construction of the derivation. In this way, topicalized constituents are not able to serve as a host because they are not available during the building of the clause. Thus, the derivational steps in topicalized constructions must mirror

---

|     |                                  |                     |                             |                           |                     |                    |                             |                        |
|-----|----------------------------------|---------------------|-----------------------------|---------------------------|---------------------|--------------------|-----------------------------|------------------------|
| (i) | <b>O</b>                         | que                 | sei                         | é                         | que                 | non                | sei                         | ren                    |
|     | <small>CL<sub>M.SG</sub></small> | <small>COMP</small> | <small>know.PRS.1SG</small> | <small>be.PRS.3SG</small> | <small>COMP</small> | <small>NEG</small> | <small>know.PRS.1SG</small> | <small>nothing</small> |

‘What I know is that I don’t know anything.’

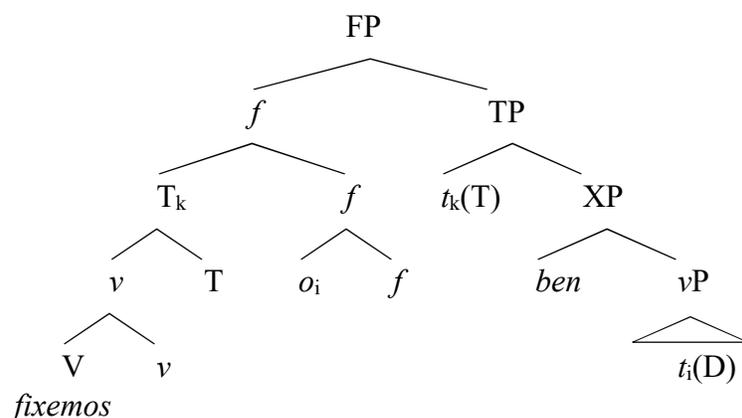
Assuming that the phrase *o que* (‘it that’) is externally merged as a complex head, this question goes away if we equate this structure to that of a DP, for which R&U argue (2005:666) determiners lean right for morphological reasons. As this is beyond the scope of my focus here, I leave this for further research.

<sup>74</sup> R&U present *f* as a syncretic projection that may host any number of features ([+Q], [+FOC], etc.), bringing the XP constituent bearing any of these features to [Spec, FP]. Nothing takes away from positing separate projections in a Rizzi-style (1997) left periphery, however, as Gupton (2012) showed. Here, an XP in the specifier of FocusP, for example, is still local enough to host the clitic.

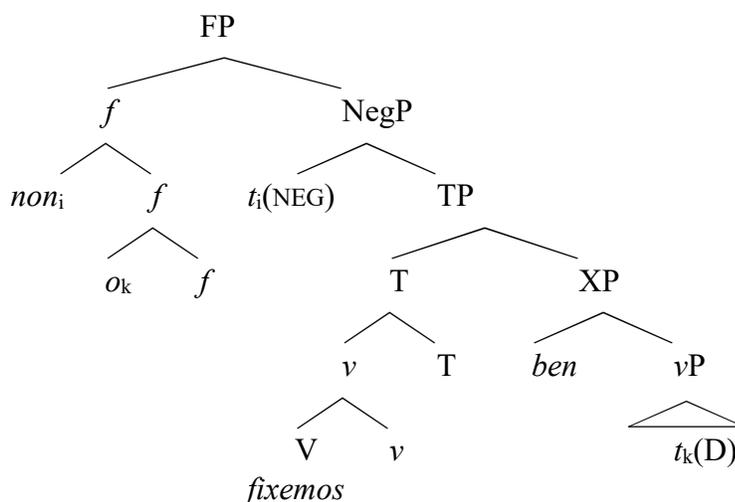
<sup>75</sup> Recall from fn. 60 that it is assumed in WI that preverbal subjects and other topicalized structures are base-generated topics (at least in matrix clauses).

verb-first constructions and how they are constructed. For these constructions, R&U propose a Last Resort movement by which the verb complex in T is “swallowed” to serve as a host for the clitic. More generally, we may say that if no constituent undergoes movement from below  $f$ , the derivation automatically selects the closest head (i.e. the verb) to serve as the host. This would predict that when the negator *non* (‘no’) is present, this functional item is swallowed to host the clitic instead of the verb. This is indeed borne out.

(31) a. *Fixemo-lo ben* (‘We did it well.’)



b. *Non o fixemos ben* (‘We didn’t do it well.’)



Per R&U’s proposal, once the clitic is probed by and moves to  $f$ , the derivation detects a problem. As there is no further material to be probed above  $f$  (e.g. a *wh*-element), there must be a

final step to save the derivation in order to avoid a clitic-first string. In (31a), the derivation looks down the clausal spine, identifies that T is the closest head with phonological material, and adjoins the verb complex to the clitic in *f*. In (31b), the first phonological head found is the sentential negator *non* ('no'), which is 'swallowed' to rescue the derivation in favor of T. We may write this rule as in (32).

(32) *Last Resort movement for clitics*

If there is no suitable material moved to the left of the clitic, take the first head with phonological material c-commanded by *f* and adjoin it to *f*.

The rule in (32) correctly predicts that a Last Resort movement will only occur if (a) there is no material suitable to host the clitic to be inserted into the derivation (e.g. a complementizer) or (b) there is no  $[\bar{A}]$  feature (e.g. [FOC]) attached to a constituent that will be moved to the left periphery as serve as the host.<sup>76</sup> Moreover, the specificity of the rule written for 'the first head with phonological material' correctly selects sentential negation when present and the verbal complex in T when absent. Below I discuss how the child may determine that this type of movement is necessary.

### 3.3.4 Tying in formality and acquisition

It is my claim that R&U's formal account of the derivational steps in WI cliticization patterns is advantageous to the learner. However, as we have seen based on the data in §3.2.2, coming to the conclusion that the clitic must move first is a slow and tedious process for children of WI

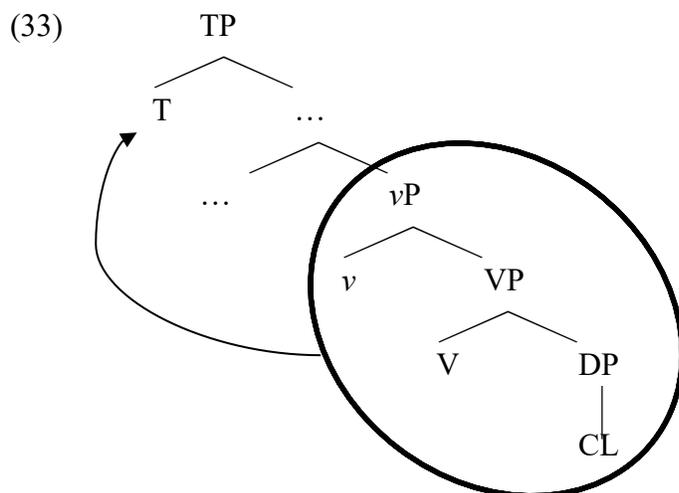
---

<sup>76</sup> The combination of both a complementizer and sentential negation is, of course, possible in both main and subordinate clauses:

- i) Dixo            **que**    **non**    o            vira  
 say.PST.3SG    COMP    NEG    CL<sub>M,SG</sub>    see.PST.PRFV.3SG  
 'He said that he hadn't seen it.'

In strings such as (i), it must be something else that moves *non* ('no') to a position in between the complementizer and the clitic. I leave this aside, as it bears little importance to my argument here.

languages. Initially, it seems that the child prefers a type of pied-piping of the entire  $vP$  to  $T$ , a movement that would render proclisis impossible.



The retreat from the movement in (33) could be due to more than one factor, as noted above (§3.3.2.2). As the child begins to assemble different groups associated with different cliticization patterns (e.g. *non* ‘no’), she may internalize a rule as in (34).

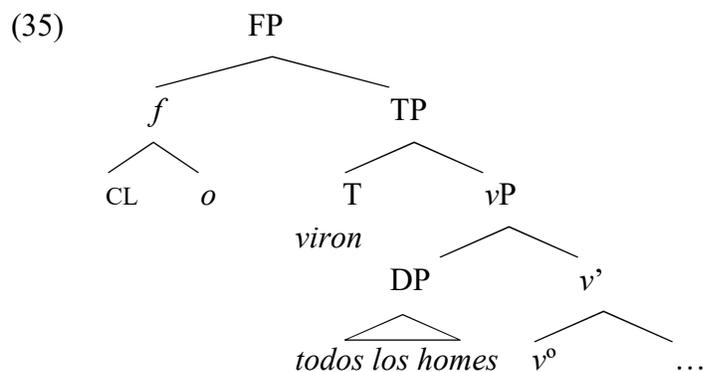
(34) *Move CL: non<sub>L</sub>*

When *non<sub>L</sub>* appears, move CL independently of the verb.

A few comments are in order. Following the line of acquisition regarding word order as proposed by Westergaard (2009, *et seq.*), these rules may go through multiple revisions in the initial stages of acquisition until the child learns to group different lexical items in particular categories (e.g. lexical items with a [NEG] feature). That is, in a more specific manner than hypothesized above (cf. 23), the child may initially create a separate rule for each individual word in her lexicon (i.e., *non<sub>L</sub>*, *ninguén<sub>L</sub>*, *xa<sub>L</sub>*, etc.) before retreating to a more generalized notion of the grouping of these words and phrases. This would be an internal repair strategy based on both category awareness

(inherent  $\phi$ -features) and syntactic positioning ( $\phi$ -features related to information structure).<sup>77</sup> The second thing worth mentioning is the rule in (34) does not guarantee that the clitic will find  $f$  or that this projection has even been postulated yet. Depending on the internal status of clitics as X<sup>0</sup>s or XPs, clitics may initially move to a position akin to [Spec, TP] following the postulation of the rule in (34). It seems plausible that a distinct position for clitics must be postulated only after the child has enough internal evidence to commit to such a projection, which may happen earlier than the precise division of all proclisis and/or enclisis triggers is made.

Once a child commits to  $f$  as the landing site for clitics, we may imagine that she is aware that the clitic must move independently of the verb. What may require further evidence and why we saw that children of WI languages struggle with fully resetting the enclisis-proclisis parameter may be due to two reasons that are not mutually exclusive. First, these assumptions are plausible if she has yet to commit fully to all of the lexical items that may trigger proclisis, particularly if they are on the tail end of the learning path in (23). Therefore, in the case of quantified DPs, after the clitic moves to  $f$  she may move the verb complex on T and, in turn, move the subject DP as if to attempt to recreate the derivational order in (34). Another factor may be an underlying preference for SV(O) order, as this is the most natural, default word order in Galician (Gupton 2014).



<sup>77</sup> It may be that these are one in the same, as posited by Baier (2017). I leave this technicality aside.

If the child has yet to internalize QPs as proclisis triggers, she may move both T and the DP in [Spec,  $\nu$ P] to the left of the clitic in *f*.<sup>78</sup>

The second reason is due to the type of movement (i.e., the motivation for the movement). Following R&U, ‘verb swallowing’ of the complex on T is a Last Resort movement. In the deviant child grammar, this seems to be movement similar to that of an [EPP] feature.<sup>79</sup> Although the reanalysis of movement in a featural sense is typically thought of as a probe related to  $\phi$ -features that eventually becomes an [EPP] feature (Simpson 2004), it may be the case that initially postulated [EPP] features are difficult to rid the child grammar of due to the extensive intake required to override misconceptions regarding the morphosyntax of word order, as with the case of cliticization in WI presented here.<sup>80</sup>

Ultimately, the child arrives at the conclusion that there are several steps to her parameter setting with respect to cliticization and the movements and/or filled syntactic positions that she must abide by upon building the derivation. This is seen below in (36).

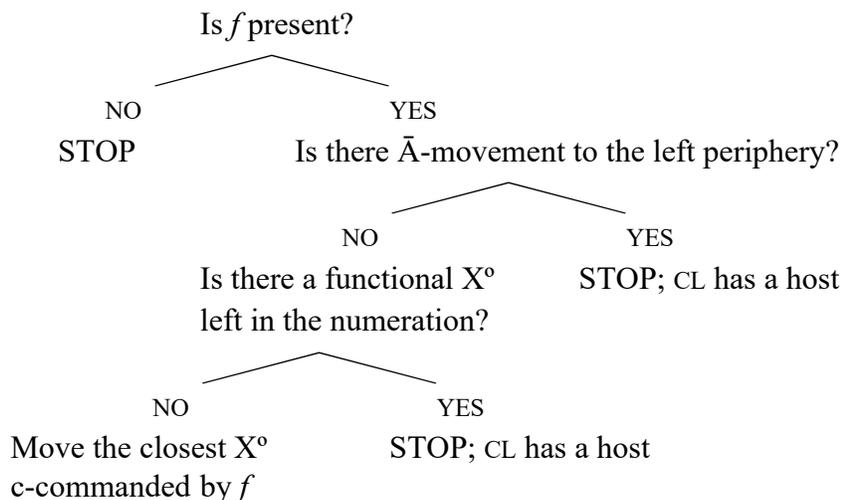
---

<sup>78</sup> This could be for a multitude of reasons. One may be that she has yet to associate hosts as XPs, a reasonable postulation considering the first three groups (and some words of the fourth group) along the learning path proposed by CFL (cf. 23) are all heads. A more speculative reason may be that there is a maturation process surrounding left-peripheral base-generated subjects and subjects that must base generate in the  $\nu$ P for reasons of quantification or information structure (e.g. VSO and VOS orders). I leave this point aside for now.

<sup>79</sup> As noted by Gallego (2016), calling the [EPP] a ‘feature’ is incongruous, as there is no checking procedure or subsequent deletion in Chomsky’s (2000, 2001) model. Here, I refer to an [EPP] feature as an internal mechanism that requires movement of some kind, such as that related to object shift out of the  $\nu$ P phase.

<sup>80</sup> Another possibility is that the child has postulated a categorial [+V] or [+T] feature. Regardless of its exact status, there does seem to be some feature that continues to move the verb to an enclitic host position with respect to the clitic.

## (36) Derivational cliticization parameter



The derivational decisions made by the child based on (36) explain the simplest and most accurate path in determining why constituents move and the derived order when more than one of these options is present. Thus, when both  $\bar{A}$ -movement and a complementizer appear to the left of the clitic as potential hosts, we know that it is the closest constituent, that which sits in [Spec, FP], that actually serves as the host.

- (37) Que MEUS PAIS che teño na casa  
 COMP my parents CL<sub>DAT.2SG</sub> have.PRS.1SG in-the house  
 ‘I’ve got my parents at home.’

As we have seen, child may require quite some time in order to distinguish constructions such as the main-clause sentence in (37) as permitting only proclisis and phrases such as *é que* (‘it’s that’) or *aínda que* (‘even though’), which may lead to proclisis or enclisis.

In the next section, I discuss how categories are determined based on findings in Cassani, Grimm, Daelemans & Gillis (2018) and Cassani, Grimm, Gillis, & Daelemans (2019). I return to the data and theoretical notions presented thus far in Chapter 4 at which point I discuss how the

factors surrounding category acquisition may affect children of WI along their learning path regarding cliticization.

### 3.4 An excursus on category learning and lexical categorization

In this section, I take an excursus on category learning and lexical categorization in child language acquisition to set the stage for syntactic reanalysis. I follow the observations made by Cassani, Grimm, Daelemans & Gillis (2018) and Cassani, Grimm, Gillis, & Daelemans (2019) regarding the importance of the three main factors that play a part in a child determining to what category a particular lexical item belongs.<sup>81</sup> Much of the work done here couples well with investigation done on what is often called ‘framing’ (Mintz 2003, *et seq.*), a term that references in what linguistic context a word does or does not find itself and, in turn, how this may aid or deter a child from making accurate determinations regarding what words pertain to what categories. Alongside the work by Cassani, I discuss the importance for distributional cues presented in the ‘flexible frames’ model (St. Claire, Monaghan, & Christiansen 2014). Because local distribution contexts provide the most information regarding the category of a lexical item, most of the work in this field is done with some sort of framing, typically that of a bi- or tri-gram template. The notions presented in both of these frameworks will provide a clear picture for the phenomenon under investigation here, particularly with respect to fixed-position lexical items such as *velai/velaqui*. Before making conclusions at the end of the chapter, I show how the contextual distribution of a word may be misleading with respect to its being assigned a category and, in turn, how it may obscure this assignment of a category label.

---

<sup>81</sup> These two references are extensions of much of the work done in Cassani (2019). As they are computational in nature, many of the finer details are fall outside of the line of investigation I present in this dissertation. However, as the results of these studies are directly applicable to the experience a child must undergo in her language acquisition journey, the conclusions drawn by these authors are insightful with respect to both normal category learning as well as lexical items that are reanalyzed and take on a new category label.

### 3.4.1 What factors are most important in category learning?

In the introduction to cue-based learning in language acquisition (§1.4.1), I briefly outlined the main assumptions regarding the first lexical items children learn in order to begin parsing and, in turn, grouping together the words that form well known categories.<sup>82</sup> Example (18), repeated in (38) below, is an instance in which only certain words have been acquired, undoubtedly based on their categorical status.

- (38) a. [DP o can] [VP vai] [PP por aí]  
           the dog go.PRS.3SG by there
- b. [DP o xxx] [VP vai] [PP por xxx]  
           the NOUN go.PRS.3SG by PLACE

Cassani et al. (2018) and Cassani et al. (2019) argue that there are three concrete aspects that go into determining the child’s categorization of lexical items: frequency, diversity, and predictability.<sup>83</sup> *Frequency* is typically associated with lexical learning in nouns (i.e., more frequent nouns are learned quicker), but its effect on functional items seems non-existent, at least in production-based investigations (Pine, Freudenthal, Krajewski, & Gobet 2013). *Diversity*, better thought of as *contextual diversity*, is “the degree of variability of the contexts in which a linguistic item occurs” (Cassani et al. 2018:2).<sup>84</sup> This is particularly important in parsing, as I discuss below, due to the information received being given in a linear and not a structured

---

<sup>82</sup> It is not an assumed fact within the psycholinguistic literature that categories even exist. Ambridge (2017, 2020) claims that categorization as presented here (i.e., fixed categorial labels) does not exist, but, instead, all that is needed are analogical generalizations. Cassani et al. (2018) and Cassani et al. (2019) do not assume strict categories as is generally assumed in generative theory. This, however, does not make their findings and my application of said findings any less significant.

<sup>83</sup> Historically, none of these three aspects related to learnability have been of primary focus in the generative literature, as the questions surrounding these three notions are typically answered by other frameworks. However, frequency has recently been cited as increasingly important in the diachronic generative work regarding certain phenomena (Biberauer & Roberts 2017, Roberts 2007, *i.a.*) thanks to interdisciplinary work that incorporates findings from computational linguistics (Yang 2002, 2016).

<sup>84</sup> As the work cited above is computational in nature, the understanding of contextual diversity in Cassani et al. (2018:3) is more specifically based on *the number* of variable contexts in addition to the types of contexts detected.

manner. Once words are processed and reanalyzed in a tree-type structure, it may be that the interaction of entire constituents and not simply the linear string of words initially presented becomes crucial for category learning, as well. Finally, *predictability* is “the average conditional probability of a context given all the words it co-occurs with” (Cassani et al. 2019:5). As indicated in §1.3.3, predictability is a key component of not only category learning, but the overall process a child goes through during lexical learning.<sup>85</sup>

Although it is well known that these three aspects of language acquisition are vital to optimal learning, the way they interact is contradictory on numerous levels. Beginning with predictability, the general understanding is that a word easy to predict is difficult to categorize. In turn, a word that combines with many other words is contextually more diverse and, thus, more useful to the learner.<sup>86</sup> As we will see below, lexical items that have fixed syntactic positions such as *velai/velaqui* or those that consistently follow a particular functional item are harder to categorize due to a limited syntactic context. Predictability is seen as the least important aspect in the eyes of Cassani et al. (2019), which they attribute to a child’s innate modeling of usefulness based on unexceptional attention to detail. As mentioned in §1.3.2, that children are subjected to a *Goldilocks effect* in several aspects of biological development (Kidd, Piantadosi & Aslin 2012, 2014) leads Cassani et al. (2019) to claim that children may be sensitive to a modest degree of predictability. Therefore, they may quickly deem a particular

---

<sup>85</sup> Predictability in adult speech forms a significant part of computational (Hale 2011) and psycholinguistic work and, therefore, must serve as a developmental staple in a child’s mind and, thus, linguistic experience.

<sup>86</sup> That a lexical item may be ‘useful’ is another aspect of language acquisition that is a non-factor in the generative literature. Here, we may understand a lexical item to be ‘useful’ if it aids in learning of some kind (e.g. its contextual diversity or frequency being high). Although it may seem trivial, not all input is useful (i.e., not all input is pushed through to the perceptual intake engine of the language acquisition device).

piece of information useless or trivial if they are confident in their commitment to it, just as they may ignore information that they are unable to compute or track well.<sup>87</sup>

Turning to contextual diversity, this is what Cassani et al. (2018) deem the most important aspect of categorization. A word that may appear in multiple contexts (i.e., surrounded by a multitude of different words/categories) is easier to categorize because there is more evidence for where it is able to appear and, thus, what type of word it is. As acquisition is predicated on learning from a linearized string and not an already structured representation, this combination does not inherently reference internal-constituent combinations (e.g. determiners that select an adjective or noun first). Clearly, these facts are at odds with predictability: a word that may appear in a multitude of syntactic positions will be harder to predict due to its flexibility in the clause. For example, the fact that determiners in a head-initial language such as Galician may be sentence initial (39a), preverbal (but follow some material) (39b), postverbal (39c), postverbal (but not adjacent to the verb) (39d), etc., give them a high contextual diversity.

- (39) a. **O** can veu lentamente  
           the dog come.PST.3SG slowly
- b. Lentamente, **o** can veu  
           slowly the dog come.PST.3SG
- c. Veu **o** can lentamente  
           come.PST.3SG the dog slowly
- d. Veu lentamente **o** can  
           come.PST.3SG slowly the dog

‘The dog approached slowly.’

---

<sup>87</sup> I leave an explanation for how predictability might have played a part in the phenomena discussed in this thesis in Chapter 4.

The variability in (38) shows that it should be difficult to predict where the determiner *o* ('the') may arise and, therefore, its contextual diversity is high. This observation bodes well for how predictability and contextual diversity complement one another despite the negative effect the former seems to have on the categorization process. However, as Cassani et al. (2018:26) note, this relationship is intuitive and, in the grand scheme of things, productive for the child.

The last aspect of the acquisition of categorization, frequency, is seen as middle of the road with respect to importance in category learning. Frequency may be more important for lexical learning and syntactic positioning (i.e., what syntactic position gives a particular interpretation), although it does play an important role up until a certain point with respect to categorization. That is, for a category to be learned, it must appear a certain amount of times. As Cassani et al. (2019:18) note, what we may derive from this is that contexts that occur infrequently cannot co-occur with many different words, which makes them homogenous in terms of their diversity and, thus, more difficult to learn. This will be an important aspect of how a particular set of input determines certain diachronic changes and the syntax of various functional items. Above, we see that frequency aids in contextual diversity, as a lexical item that appears more often will provide more exposure to the child as to what its underlying category is. Frequency does little for predictability, however, as the most frequent items are often the hardest to predict. Cassani et al. (2018, 2019) determine that the importance for categorization is as in (40).

(40) Contextual Diversity > Frequency > Predictability

Moving forward, we shall see that both contextual diversity and, to a certain extent, frequency are key factors in accounting for reanalysis and diachronic phenomena.

### 3.4.2 How input influences categorization

Categories are formed based on the understanding of the child that any lexical item  $LI_1$  pertaining to category  $C$  may be exchanged for any other lexical item (e.g.  $LI_2$ ) of said category with no syntactic penalty. As we have seen, when the distributional context of a particular lexical item is highly predictable (i.e., limited syntactically), it makes categorization more difficult due to the limited evidence the child has in order to determine what kind of element it is. Although the use of a particular lexical item may be adult-like in a child's speech, the internal syntax and categorization of said lexical item may be far from that of an adult. In Chapter 4, I explore how strings that systematically appear in sentence-initial position may undergo reanalysis as a single functional item. In some cases, such as that of *velai/velaqui*, this functional item may be generated on the highest head previously occupied by the first lexical item in the aforementioned string. From there, it may undergo further reanalysis to a higher functional head, but this is not obligatory. The fact that this further reanalysis is not mandatory should tell us something both about why it happens as well as why it does not. Based on the work from Demonte & Fernández Soriano (2014) on the Spanish *dizque* ('It is said'), I show in §3.5 that multiple-word strings that undergo reanalysis and become functional items based on sentence- or clause-initial use do not always become sentence-initial functional items. The goal upon examining the distribution of this functional item is to argue that aspects related to contextual diversity are important and even shape the distribution of words that are in the process of undergoing change.<sup>88</sup> These developments are particularly telling when compared to Step 2 as presented above for which the analysis of *dizque* will be relevant in accounting for the two scenarios briefly discussed in §3.2.2. In contrast to what we shall see in Chapter 4, this is not always enough to provoke the 'last step'

---

<sup>88</sup> This is expected, of course, if we take diachronic reanalysis to be a product of individual synchronic learning, albeit of a slightly different variation of the target grammar. It may also be applicable to a synchronic grammar in which there is a true syntactic variation.

of diachronic change. Thus, we shall see that the combination of the pragmatic-syntactic judgement of the child in combination with abnormally high syntactic variation in the input is a viable path in order to describe how this change may occur. However, identifying both patterns may allow us to make further predictions about the intersection between language use, parsing, and diachronic phenomena.

### 3.5 The syntax of *dizque*

In this subsection, I outline the work in Demonte & Fernández-Soriano (henceforth D&FS). As much of their investigation deals with semantic notions that are orthogonal to my investigation here, I focus on their comparison of *dizque* to the ‘evidential *que*’ or ‘reportative *que*’ (cf. §2.2.2) and, in turn, discuss the syntax of each.<sup>89</sup> I argue against the authors’ syntactic analysis of *dizque*, suggesting that their semantics-based argument for the syntactic position of this functional item overgenerates. Instead, I claim that *dizque* must be within the syntax proper due to the fact that it may be embedded by a complementizer located in Force.

The functional form *dizque* (‘It is said’) stems from the combination of the shortened form of the present indicative *dicen* (‘they say’), *diz*, and the complementizer *que*.<sup>90</sup> D&FS claim that *dizque* was first a reportative evidential.

---

<sup>89</sup> The stance taken by these authors here is different from that of Demonte & Soriano Fernández (2009) which was cited earlier in §2.2.2. In the investigation under focus here, it seems that what I have referred to as the ‘reportative *que*’ may also be blanketed under the term ‘evidential *que*’ as a subcategorization of ‘reportative evidentials’ more largely. Although D&FS (2014) take these two to be one in the same and, thus, should be applicable to the analysis given in §2.2.2, I review the specifics of this investigation in particular. I use these terms interchangeably below.

<sup>90</sup> D&FS, *apud* Kany (1944), note that *diz* was an informal variation of *dicen* that had the modern-day interpretation of the passive voice construction *se dice* (‘It is said’).

(41) a. ... vase a la comedia, que **diz que** estaba  
 go.PRS.3SG-CL<sub>REFL</sub>.3 to the comedy COMP DIZQUE be.IMPFV.3SG  
 cuajá...  
 crowded

‘... (he) went to the comedy that it is said was crowded...’

b. ... escuderos & otras personas no usadas de tener juzgado, los  
 squires and other people NEG used of have.INF judged the  
 cuales **dizque** fazen injusticias & estorsiones a los  
 which DIZQUE do.PRS.3PL injustices and extortion to the  
 vezinos...  
 neighbors

‘... squires and other people not used to having judged, those which it is said that do their neighbors injustice and extort them.’

c. ¡Con indutzria **dizque** le engolda!  
 with industry DIZQUE CL<sub>DAT</sub>.3SG feed.PRS.3SG

‘With industry, it is said that they feed him!’

(D&FS 2014:215-216)

An important observation regarding the syntax of *dizque* can be observed in the constituents that precede it in each example above. In (41a), *dizque* is selected by the aforementioned ‘evidential’ or ‘reportative *que*’, whereas in (41b) it follows a relative DP constituent, and in (41c) it is preceded by a focused adverbial constituent. It is worth noting that in all three cases *dizque* selects a finite verb. This will prove important for my analysis below.

As a pure evidential, these authors claim that over time this came to adopt a meaning of doubt and, thus, a more restricted syntactic use. Although they provide no evidence in the data presented, they claim that reported evidentials may be used with first-person subjects, whereas this is ungrammatical with *dizque*.

D&FS agree with previous investigative conclusions surrounding the semantics of *dizque* regarding its use from “encoding reported speech and hearsay with a notion of doubt implied in some contexts, to a marker of epistemic modality, encoding extensions of the notion of doubt implied in its evidential use and nothing about source of information” (Travis 2006:1269). While much of the focus D&FS give to this topic deals with its influence from other evidentiality systems in Quechua and Aymara, these authors also contrast *dizque* with the ‘reportative *que*’ outlined in §2.2.2, citing four crucial properties of reportative evidentials.<sup>91</sup>

First, reportative evidentials never refer to something said by the speaker or hearer. Second, reportative evidentials most naturally occur in declarative sentences, although they may be found in questions in a limited capacity in languages with an evidential system (Aikhenvald 2004 *apud* D&FS 2014:221). Third, the information relayed via reportative evidentials may not be affected by negation or dissention, as these linguistic operations “can only access the proposition introduced by the evidential, not the ‘source of the information’ it refers to” (D&FS 2014:221). Fourth, these evidential constructions hold additional meanings when involved with a first-person subject. D&FS refer to this as an aspect of ‘unprepared mind’ or surprise (i.e., new information). Based on observations by Lim (2010:60-63), these authors claim that this is also the case for the ‘evidential *que*’ in Spanish.

---

<sup>91</sup> D&FS cite a fifth, highlighting the use of reportative evidentials in folklore tales. Given that it is unrelated to the discussion regarding syntax, I leave it aside.

(42) *Upon realizing that the speaker has won the lottery, she proclaims:*

(Oye),            **que**    he                    ganado    la    loteria!

hear.PRS.3SG    COMP    have.PRS.1SG    win.PRTCP    the lottery

‘Hey, I’ve won the lottery!’

Although the use of *que* in evidential situations of this sort are not available in all Romance languages, this does in fact seem to be a property of Western Romance (Spanish, Galician, Portuguese, Asturian), as this construction with the same contextual meaning is also frequent in Galician.<sup>92</sup>

(43) *Upon finding out the speaker has been accepted to an academic conference:*

Ouviches,        **que**    me        aceptaron        a    proposta!

hear.PST.2SG    COMP    CL<sub>1SG</sub>    accept.PST.3PL    the proposal

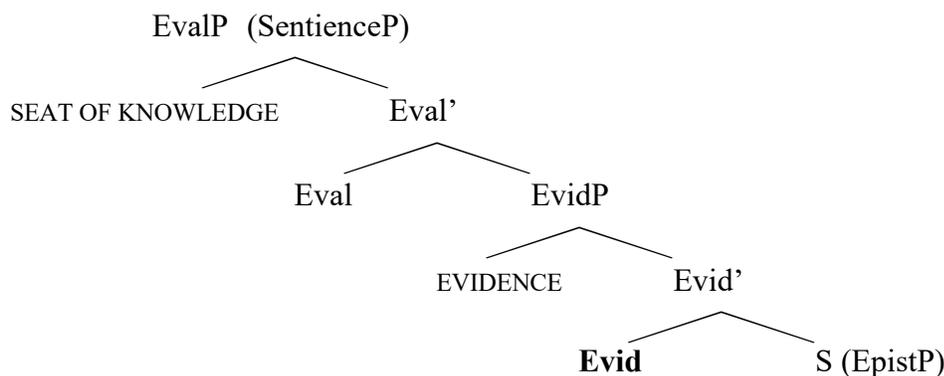
‘Hey, they accepted my proposal!’

Based on these facts, D&FS propose the ‘evidential *que*’ surfaces as the head of an Evidentiality Phrase (EvidP), located within a Speech Act Phrase (Speas & Tenny 2003) above CP. These projections above CP are “the inventory of grammatically relevant pragmatic roles” (Speas & Tenny 2003:315).

---

<sup>92</sup> Mattia Pistone (p.c.) informs me that the complementizer *che* is not used this way in standard Italian, hence my claim that this evidential use is not spread throughout all Romance. There are complementizers, however, used in jussive structures in extreme and upper southern Italian dialects (ESIDs and USIDs) (Colasanti & Silvestri 2018), where it takes on a different meaning than that discussed here. As a further cross-linguistic comparison would take me too far afield for the purposes of this thesis, I leave this for further research.

(44) *Speech Act Phrase* (D&FS 2014:223)



Although the semantic descriptions given above may lead one to assume this particular use of the complementizer *que* is beyond the scope of the syntax, there are reasons to believe this is not the case. Let us return to the case of the Galician example above in (43). Were this solely a pragmatic use, as one would expect from a discourse-related speech act particle, there would be no interaction with the syntax in any way. This is, however, not what we find. Notice that the first-person object clitic *me* is in a proclitic position. As discussed at length in §3.3.1, only syntactic elements that are in a relationship local enough to the clitic may host it. Regardless of the status of the constituent (i.e., be it an  $X^0$  or XP element), we have seen that only elements within the CP may host clitics. Were this ‘evidential *que*’ part of a projection based solely on a pragmatics-oriented speech act such as that described by Hill (2007, 2012), we should not expect it to interact with the syntax in such a direct manner. Therefore, what D&FS deem the ‘evidential *que*’ must be in the same structural adjacency as the ‘reportative *que*’, which I showed in §2.2.2 to be on C.

As we saw in §2.2.2, the complementizer *que* may appear on multiple C-oriented heads. Based on the data shown in D&FS, this appears to be the case for *dizque*, as well. Whilst *dizque* may appear in a sentence- or clause-initial position (45), the examples in (41) show that *dizque* may be selected by several constituents in the left periphery.

(45) Y     **dizque**    dende en tres    días    el dicho Cartagena  
 and    DIZQUE    since in three    days    the said    Cartagena  
 no    lo    tomó            a    saludar  
 NEG   CL<sub>M,SG</sub> take.PST.3SG to    greet.INF

‘And it is said that since then the aforementioned Cartagena did not greet him for three days.’

(D&FS 2014:216)

Above I highlighted the fact that the data in (41) showed that *dizque* may be embedded by multiple left-peripheral constituents, and in every scenario, it selects a finite verb. These data lead me to postulate that *dizque* is indeed a complementizer that occupies a low CP head such as Fin/*f*.<sup>93</sup> Using the example in (41a), we may imagine this structurally as in (46).

(46) [CP que [FP *dizque* [TP estaba ...]]]

What does this theoretical stance on *dizque* buy us? If we assume that *dizque* is a C-oriented element based on the aforementioned observations, this creates several parallels between what we saw with Steps 1 & 2 in §3.2.2. Pairing these observations with the fact that the contextual diversity of a word is the primary factor in a child determining its categorial status, there should be clear signs as to where *dizque* is located along the clausal spine. Detecting that constituents such as relative pronouns, topics, and focus-fronted constituents may appear before it while a finite verb follows it should be clear indicators that i) the string processed in the input is all a part of one clause, ii) *dizque* is not a verb, and iii) it must appear in a hierarchical position higher than the verb in T.

---

<sup>93</sup> The illocutionary force that D&FS describe for this functional item coincides with that described for *f* as noted in Raposo & Uriagereka (2005)

### 3.6 The two faces of Step 2

In §3.2.2, I identified what appear to be two different constructions that contradict the cliticization patterns reviewed in §3.3.1. I will refer to one of them as the ‘fixed enclisis’ construction (47a) and the other as ‘low *velai/velaqui*’ (47b).

- (47) a. Veña                    **que**    **xa**            **vé-los**                    alá    están  
           come.SUBJ.3SG    COMP    already    see.PRS.2SG-CL<sub>M,PL</sub>    there    be.PRS.3PL  
           os        carros!  
           the        haywains  
           ‘Come on, because you already see them, there are the haywains!’
- b. E        cadra            ben    **que**    **velaí**            vén                    Alberte    e  
           and    fit.PRS.3SG    well    COMP    behold    come.PRS.3SG    Alberte    and  
           vou-lle                    dar        as        boas    tardes  
           go.PRS.1SG-CL<sub>DAT,3SG</sub>    give.INF    the        good    afternoon  
           ‘And it works out well that here comes Alberte and I’m going to tell him good  
           afternoon.’

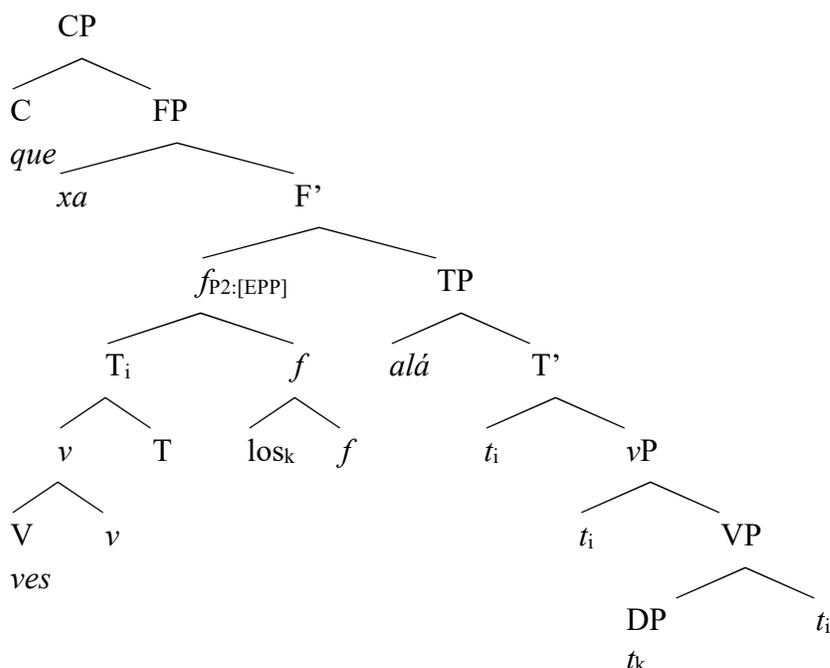
In both instances, *ve-lo* shows an enclitic structure preceded by a complementizer *que* which, as we have seen, should always result in proclisis.<sup>94</sup> The difference that I wish to distinguish is the following. With the ‘fixed enclisis’ construction, reanalysis could not have taken place due to the plural number and masculine gender features of the clitic *los* compared to those of the coindexed DP *os carros* (‘the haywains’). Therefore, we must assume that the same derivational procedures are in play here as with regular enclisis. The difference between this construction and what I propose for the ‘low *velai/velaqui*’ construction is comparable to the situation we saw with

<sup>94</sup> Having addressed this issue in more than one instance, it should be noted that this *que* is the ‘evidential’ or ‘reportative *que*’ I have presented above and in §2.2.2. This is not an instance of *e que* (‘It is just that’) which we saw permits enclisis regularly. Enclisis is never found with the ‘evidential’ or ‘reportative *que*’ however.



Due to the fact that enclisis, as I have argued for it, comes about as a Last Resort strategy, essentially a derivation-saving mechanism, we must account for the data in (47a) in a different manner. There are two ways I propose this might have happened. The first possibility is what Uriagereka (1988, 1995b) originally proposed for verb movement in order to account for enclisis: checking of morphological  $\phi$ -features. This is similar to the motivation of clitic movement to  $f$  in Raposo & Uriagereka (2005). Using the data from (47a), I show this morphological dependency below in (50).

(50) *Verb movement to F: a morphological dependency* (modified from Uriagereka 1995b:101)



It is worth noting that the idea that verb movement to  $f$  was predicated on the checking of morphological  $\phi$ -features was not lost in R&U's (2005) proposal. In contrast to Uriagereka's original ideas, however, this mechanism was claimed to account for the checking of morphological agreement on inflected infinitivals in Western Iberian. As T has no [PERSON] or [NUMBER] features when the verbal predicate it selects is an infinitive, R&U claim that the overt

agreement comes from checking  $\varphi$ -features on the structurally higher  $f$ .<sup>95</sup> While fundamentally different, it is worth mentioning that R&U's idea for inflected infinitivals stemmed from the earlier work by Uriagereka (1998, 1995b). For the example in (47a), this would equate to the derivation retaining this  $\varphi$ -probe that normally would be absent when a complementizer is present in C or an XP is in the specifier of  $f$ .

The second possibility deals with an approach that could also explain enclisis as an alternative to the Last Resort approach and is the one I argue for here. In accounting for how the four main Person Case Constraint (PCC) types may surface, Deal (2020) proposes a model she calls *dynamic interaction*. Under this model, the interaction features by a probe need not be fixed but, as she claims, may change during the course of the derivation.<sup>96,97</sup> Applying this idea to the derivational cliticization parameter in (36), we must assume that  $f$  has more than one probe: one for clitics and one in order to satisfy movement of a constituent to host the clitic (be it an  $X^0$  or XP).<sup>98</sup> For simplicity purposes, I label these probes P1 and P2, respectively. The idea that a head  $H^0$  may probe for two separate goals of two separate featural specifications has most recently been explored in Roberts (2010a), where he accounts for two types of head movement within the  $vP$  (i.e., cliticization and V-to- $v$  movement).<sup>99</sup>

---

<sup>95</sup> R&U do not distinguish the morphological  $\varphi$ -features that attract clitics from those that require inflected infinitivals to raise above T, so I shall assume they are one and the same. In Chapter 5, I take a different approach to what requires clitics to raise, arguing for a more fine-grained syntactic approach.

<sup>96</sup> I save her examples with clitics for Chapter 5 upon adopting her account for cliticization in Galician. For now, I focus specifically on the implementation of *dynamic interaction* with respect to a mechanism that may replace the Last Resort movement of the verb on T.

<sup>97</sup> A topic of much interest in early minimalism (e.g. Pesetsky & Torrego 2007) was how the valuation of (uninterpretable) features took place. Deal's mention of 'interaction of features' in this sense dates back to Deal (2015) in which she claims that Agree is broken into three stages: search, copy, and valuation. Features that interact, thus, are those that are copied from a goal and undergo valuation on a probe.

<sup>98</sup> Theoretically, this could be reduced to one probe that first seeks to agree with the clitics and then begins to probe for a constituent to serve as the host to the left of the clitic. As the outcome should still remain the same, I utilize the idea of two probes for expository purposes.

<sup>99</sup> This must also be the case for accounts that claim cliticization happens to T instead of  $v$  in Romance, where T would agree both with the clitic and the subject via two separate probes.

Assuming that P1 is exhausted and attracts all relevant clitics to *f*, P2 searches in its c-command domain for a potential goal to bring to [Spec, FP]. This probe may search for a feature as simple as [ $\bar{A}$ ] (Baier 2017) or it may cycle through a series of features that would serve to bring a constituent to the left of the clitic.<sup>100,101</sup> Another alternative for the nature of this feature is that *f* has a probe for clitics and, in addition, an [EPP] feature. As [EPP] features may be satisfied by any number of constituent types, this also seems feasible for what I have called P2. The importance for what P2 represents with respect to the construction represented in (47a) is that we may imagine when the Numeration has a base-generated functional head such as the complementizer *que* left in its inventory, P2 will simply not be warranted. This type of overgeneration is what we would expect of a child who has determined that clitic movement must be independent of verb movement but, nevertheless, produces enclisis.<sup>102</sup> In the case of the phenomenon in question, however, this is unexpected from adult grammars. I claim that the cases of stage 2 as in (47a) must have proceeded with P2 on *f* regardless of the presence of the complementizer.

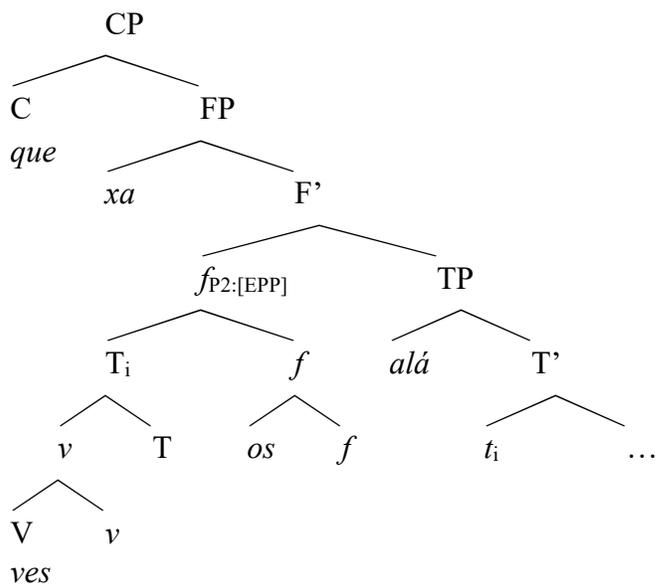
---

<sup>100</sup> The idea of a feature as unspecified as [ $\bar{A}$ ] may be necessary if one assumes that adverbs base generate on a series of heads below *f* (Cinque 1999) and must move to its specifier. As adverbs do not possess canonical DP-like  $\phi$ -features, this would be the only explanation for their movement.

<sup>101</sup> For the second option, recall that the failure of a probe to find a viable goal does not necessarily result in a derivational crash (§1.5).

<sup>102</sup> This is a simple alternative to what I showed above in (33), possibly representing a more advanced stage of a child's understanding of how cliticization patterns must proceed.

(51) ... *que xa vé-los alá...*



In (51), I use the [EPP] option to represent P2 for expository purposes. We may see here that the verb should not move due to the proximity of *xa* ('already'), but this is precisely what we find. The theoretical explanation for this stage of Step 2 will provide us with crucial information regarding the reanalysis of the complementizers in Chapter 4.

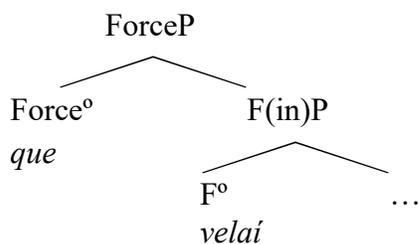
Finally, what I consider to be the last stage of Step 2 is that which we see in (47b), repeated here for convenience.

(52) b. E      *cadra*      *ben*      **que**      **velaí**      *vén*      *Alberte*      *e*  
 and    fit.PRS.3SG    well    COMP    behold    come.PRS.3SG    Alberte    and  
*vou-lle*      *dar*      *as*      *boas*      *tardes*  
 go.PRS.1SG-CL<sub>DAT</sub>.3SG    give.INF    the    good    afternoon  
 'And it works out well that here comes Alberte and I'm going to tell him good  
 afternoon.'

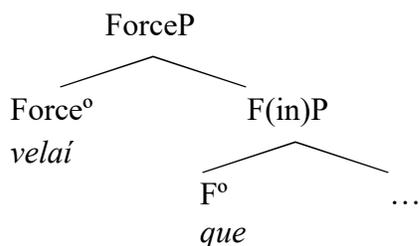
I consider this a reanalyzed form of *vé-lo aí* that is characteristic of the analysis provided above for *dizque*. That is, reanalysis has taken place, but *velaí/velaqui* are still in a position to be

selected by a higher head (e.g., the complementizer *que*). Although this may seem a better fit for Step 3, the complementizer I consider for Step 2 is different from that in Step 3 in the sense that it has both structural and syntactic differences although their pragmatic function seems to be the same.

(53) a. *Velai/Velaquí* in Step 2



b. *Velai/Velaquí* in Step 3



First, what I have called ‘low *velai/velaquí*’ may be selected by *que*, whereas the present-day complementizers may not (§2.2.1). Moreover, the ‘low *velai/velaquí*’ are never found in the C-AGR construction. This last point may be concomitant of the fact that the C-AGR construction is the result of not only very mixed data, as I show in Chapter 4, but also only arises once *velai/velaquí* are reanalyzed to the syntactic position that I argued for in §2.2.1 as the head of ForceP.

### 3.7 Conclusion

In this chapter I have provided a detailed analysis of cliticization in Western Iberian and coupled these data with theoretical and empirical facts from both adult and child grammars. Most

importantly, I have made important connections with empirical child data that show children's preference for enclisis and their retreat from it into a theoretical model that explains the assumptions children must come to in order to match the cliticization patterns of the adult grammar. On a separate note, I have discussed the basics of category learning and the three primary aspects that go into placing a categorial status on lexical items. I then discussed and provided a new analysis for the complementizer *dizque* in order to provide a backdrop for my analysis of Step 2 in which I claim there are three separate stages that can be individually identified with respect to their syntax. Here, I claimed that *dizque* was reanalyzed from a multi-word verbal phrase to a functional item in C rather than functioning as discourse-related material in an extended Speech-Act Phrase above CP.

In Chapter 4, I discuss the reanalysis of *velai/velaqui* to that of the present-day grammar. I then discuss the projected scenario for a Galician child during the time in which Steps 1-3 coincided, leading to syntactic reanalysis. I shall argue the mixed syntax of these constructions being present in the input epitomizes the situation argued for by Hudson Kam & Newport (2005, 2009) with regard to requirements for diachronic change from a child innovator approach. Moreover, I explain that Step 4 was engendered from this very situation, as well, providing further evidence for the idea of conceptual innovation via the aspects argued for by these authors.

## CHAPTER 4

### NEW CATEGORIES, NEW GRAMMAR

#### 4.1 Introduction

This chapter addresses three types of change: multi-word category reanalysis, structural reanalysis, and what I will refer to as endogenous reanalysis. Unlike the previous two changes, I argue that the latter seldom arises due to the child undergoing repair in order to match the target grammar. There are clear cases, however, in which this is the only rational explanation for how certain syntactic phenomena arise. In the grammar change literature, the discussion of what changes may be considered exogenous and which are purely endogenous have been largely inconclusive. From a *Child Innovator* approach, I argue that all change is endogenous in the strict sense. However, some changes are more directly affected by the input a child receives than others, supporting the longstanding idea that there are changes that come about via ‘transmission failure’. In turn, I claim that there are also changes that are endogenous based on the input and, therefore, internal biases, as well as changes that are endogenous based on misrepresented inductions of the how the input is processed and fed through to the perceptual intake.

Specific to the phenomenon under investigation in this this dissertation, this chapter also addresses two main topics related to the claims above. First, I build off of the aspects in Chapter 3 with respect to categories and reanalysis. In addition to discussing how new categories are formed, be these novel categorial labels or the reanalyzing of a surface string to a well-attested category in the grammar, I give evidence for the fact that the syntactic-pragmatic experience of a given lexical is enough to account for what have traditionally been labeled different ‘types’ of

change based on their semantic outcomes. My analysis here supports the idea that learning, and in turn change, is dependent on the syntactic experience and induction of the structural aspect of the input. This is in contrast to the common assumption that semantic change is what provokes eventual structural change (Bybee 2003a, 2003b; Heine & Kuteva 2002, 2012; Traugott & Dasher 2001; Traugott 2012; *i.a.*). The observations I make aid in the ongoing discussion how and why reanalyzed words are more restricted in their syntax than their semantics. Specific to the main phenomena under investigation in this thesis, this provides a solid explanation as to why *velai/velaqui* underwent further movement up the clausal spine after reanalysis from a multi-word string to a single function item had already taken place.

Second, I discuss the possible avenues by which the Galician C-AGR construction was engendered. Ultimately, I argue that Step 4 must have arisen due to an extremely varied syntactic experience composed of Steps 1-3. Therefore, Step 4 should be analyzed as an unexpected amalgamation of several of the properties found in the previous three constructions. My approach here offers an understanding of how novel constructions may come about and the individuality that drives such change. Moreover, it also permits several hypotheses regarding how children may be able to repair a previously ill-formed structure based on adult grammar and where these processes fit into the general discussion of diachronic change.

#### 4.2 Different processes, the same change

In Chapter 1, I explained the model put forth by Pearl (2020a) based on the original ideas of Gagliardi & Lidz (2015) regarding the internal workings of the language acquisition device (LAD) (Figure 3). The purpose there was not only to explain the internal mechanisms that drive each part of learning (structure building, internal structural biases, creation of overhypotheses, constrained hypothesis spaces), but also to support the varied outcomes of the basic scenarios

that children encounter during their linguistic experience. Let us revisit the three outcomes posited by Phillips & Ehrenhofer (2015:416-417) discussed in §1.3.3, repeated below, regarding how the degree of the child's parsing success may influence her grammar.

(1) *Three parsing outcomes*

- a. A learner may fail to assign a parse to an arriving string, either because it is too complex, or it arrives too quickly;
- b. A learner may systematically assign an incorrect parse (per the established adult grammar) to a particular sentence type due to parsing biases or reanalysis failure;
- c. A learner that parses an input string successfully so that it may reach the perceptual intake mechanism of the LAD may extract more or less information from said input string depending on her ability to predict the contents of the sentence as it unfolds.

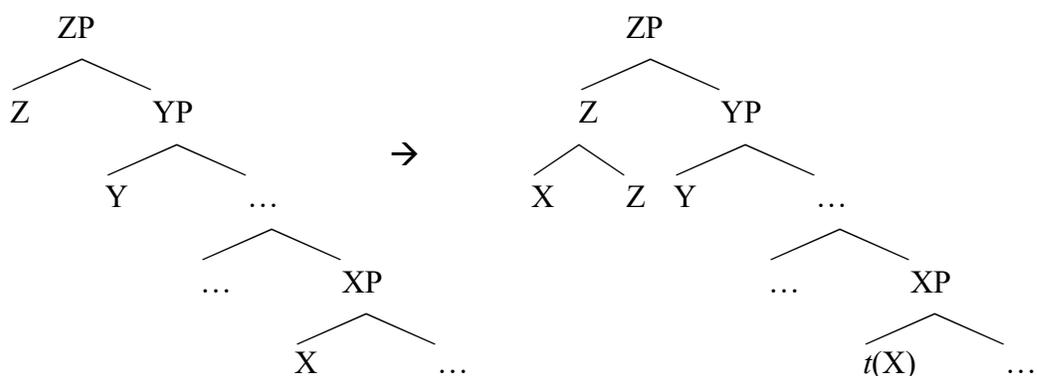
As discussed in Chapter 1, outcome (1b) could be highly disruptive during development if a particular string is continuously misparsed. In Chapter 1, I claimed that this is the primary avenue by which change arises, what is referred to diachronically as reanalysis. Change may arise via multiple pathways, however, and the identification of these pathways has led investigators of diachronic phenomena to suggest that change may occur in one of two ways. The first type of change is *exogenous*, which references a change that may be directly influenced by the input a child receives. The change may be provoked based on the lack of robust evidence in the input so that there is no firm postulation for a particular parameter setting or due to a multilingual situation in which one grammar seemingly takes precedence over the other during the acquisition period.<sup>103</sup> In the first instance in particular, grammar change is often thought of not only as feature reanalysis (van Gelderen 2009), but the loss of features altogether.

---

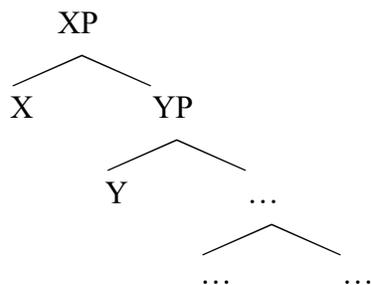
<sup>103</sup> I do not address language-induced change in this dissertation, as this type of change seems to be far more unconstrained compared to monolingual change (Fuß 2017).

Concomitantly, the picture that emerges is that reanalysis is upward in the tree, owing to a ‘Merge-over-Move’ (or External Merge over Internal Merge) situation (Roberts & Roussou 2003, Roberts 2007, *i.a.*).

(2) a. *Step 1*

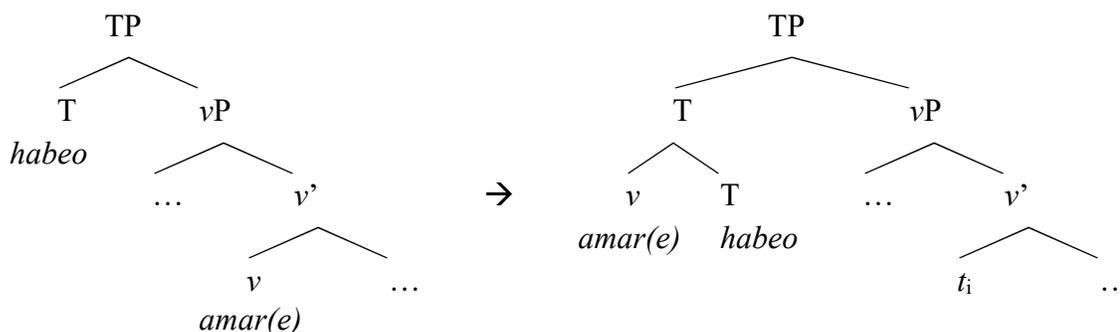


b. *Step 2*

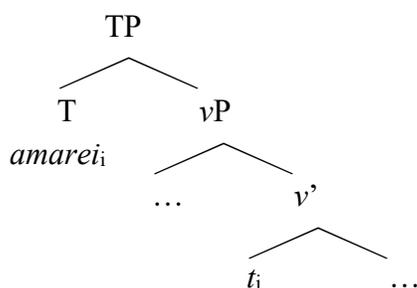


Roberts (2007) describes the change from (2a) to (2b) as ‘X replacing Z’. X need not replace Z in the strict sense, but instead, the amalgamation of X and Z may create a new lexical item. Moreover, as we have seen with the reanalysis of *velai/velaqui*, more than two lexical items may be reanalyzed as a single lexical item, creating a broader picture than that painted in (1). A common change of this type has been identified for the Romance futures (Roberts 1993, 2010b; Roberts & Roussou 2003; Graham 2015).

(3) a. *Amar(e) habeo* ('I will love'; Latin)



b. *Amarei* ('I will love'; Galician)



The transition from (3a) to (3b) entailed the reanalysis of the complex head on T in (3a) as a single lexical item, creating a new pattern of synthetic future conjugations.<sup>104</sup>

The second type of change is *endogenous*, which suggests a scenario in which children create novel structures with no direct evidence in their input for such structures. This type of change is the model scenario when one considers the outcome of when children see structural generalizations that are not actually present in their input. In some cases, such as that discussed by Belletti (2017) as well as my proposal for the engenderment of Step 4, there are pathways by which one may be able to account for why these changes might have arisen although they are presented in a less cohesive and more arbitrary manner diachronically.<sup>105</sup> Moreover, these

<sup>104</sup> Roberts & Roussou (2003) & Roberts (2010b) discuss this change having undergone three different micro-reanalyses where these authors focus on the eventual inflection as a clitic-like element before fusing with the lexical verb. The sketch I give in (3) is a succinct version of their proposal.

<sup>105</sup> As Joseph (2011) notes, the fact that the historical data do not clearly describe the change hypothesized is one of the principal issues with utilizing historical data in attempting to explain diachronic change.

changes are rarely cross-linguistically attested, an expected result if we consider these them to arise as language-specific changes.

(4) Il coniglio a i'pinguino lo tocca  
 the rabbit DOM the-penguin CL<sub>M.SG</sub> touch.PRS.3SG

‘The rabbit touches the penguin.’ (lit. ‘The rabbit to the penguin (he) touches him.’)

In the child data taken from Belletti (2017:2) in (4), the object *i'pinguino* (‘the penguin’) is differentially marked. As I discuss below in §4.3.1, this is not a property of Standard Italian. There is reason to believe that these deviant strings such as that in (4) represent a clear case of exogenous change due to internal pressures in the developing grammar of the child.

#### 4.2.1 Accounting for exogenous and endogenous changes in learning

How are these changes distinguished from the position of language acquisition? I argue that all change is endogenous in the strict sense, as every grammar is developed and created anew in the mind of the child regardless of our ability to identify the change as caused by the input. Although transmission is by and large highly successful, the grammar that the child acquires should be viewed as a new one regardless of how indistinguishable it is from the grammar of the input. The position taken by those in support of the *Child Innovator Approach* (Hale 1998, 2007; Courneau 2014, 2015, 2017; Pearl 2007; Lightfoot 1979, 2020; *i.a.*) directly mirrors that which we find in the language acquisition literature. Han, Musolino, & Lidz (2016) note that the knowledge of children and their parents is independent of one another. However, they also highlight the fact that “[t]he combination of unpredictable variation and parent-child independence suggests that the relevant structural feature is supplied by each learner independent

of experience with language” (942).<sup>106</sup> As I will show, this includes internal pressures and firm biases, some of which never undergo repair during the development of the grammar.

Were we to try differentiate the types of change from an acquisition standpoint, there are subtle differences between the experience of the child and the internal decisions that the child makes with the input received that may aid in understanding what linguistic information is exogenous or endogenous. Phillips (2012:282) highlights a basic list of prerequisites for transmission to take place accurately, with the child grammar acquiring as many of the aspects of the adult grammar as possible.

- (5) a. The learner must be exposed to relevant linguistic input (Input I)
- b. The learner must experience that input in suitably informative contexts (Input II)
- c. The learner must attend to these informative situations (Uptake I)
- d. The learner must appropriately encode these informative situations (Uptake II)
- e. The learner must store relevant examples and perform appropriate computations over them, in order to form grammatical hypotheses (Computation I)
- f. The learner’s computation may be guided by a constrained hypothesis space (Computation II)

The prerequisites (5a) and (5b) are the only true external aspects of a child’s linguistic experience, but they are not typically discussed as causes for change. This is primarily because either scenario here would entail no relevantly intelligible linguistic experience or one that was so impoverished or infrequent that the child would be unable to formulate hypotheses. The aspects addressed in (5c) and (5d) represent the initial stage of the LAD model in Chapter 1

---

<sup>106</sup> The authors’ use of ‘feature’ here should not be confused with the  $\phi$ -features in syntax, which I have claimed must be learned based on the evidence and the conclusions children make about their language based on their experience. Han et al. use the term ‘feature’ to describe the internal structural mechanisms of derivational building and, thus, a Third Factor in the Chomskyan literature.

(Figure 3). Specifically, (5c) deals with the uptake process of interpreting the incoming information that is fed into and deconstructed by the perceptual encoding device. As discussed in §1.3.1, based on the current linguistic development of the child, this encoding may be more or less accurate and it is at this stage that the incoming information is either processed and passed through to the perceptual intake device or discarded as unusable. It follows, thus, that the prerequisite (5d) is parasitic on (5c), as this is when the encoding is passed through and given a structural representation for further processes. It is at this point when the information is passed from the perceptual intake to the inference engine. The structure fed to the inference engine then interacts with the prerequisites in (5e) and (5f). The structure the child commits to in the perceptual intake is then put up against the already hypothesized aspects of the grammar dealing with the phenomena present in the string just analyzed. As Yang (2016) has shown, there may be severe consequences for whether the structure of string  $S$  matches what has previously been posited in the developing grammar or not. A newly hypothesized rule  $R$  regarding phenomenon  $P$  will permit few exceptions early on in the child's experience with it. As the productivity of  $R$  is dependent on whether the child deems the syntax in the incoming strings compatible with her previous experience, stepping beyond the margin for error may lead to two possibilities: either she will determine that the strings that are not compatible are indeed productive must be subsumed under a separate rule  $R_1$  or she will withdrawal from the hypotheses that established  $R$  in the first place.<sup>107</sup> All of these factors and possibilities are assumed under the interaction of the newly interpreted intake and the hypotheses and rules that currently govern the developing grammar, as highlighted in (5e). Finally, the effects seen by the implementation of (5f), as

---

<sup>107</sup> These are general claims that work for both productive morphology, the focus of Yang's work, as well as productive syntax constructions. There is no reason *a priori* to assume that the syntax, which typically shows more consistent flexibility amongst cross-linguistic phenomena, would be more restrictive in the sense of rule building than morphology.

expected, simply constrain the hypotheses the child is able to assume to be plausible about her language.<sup>108</sup> While certain movement has been shown to be robustly impossible cross-linguistically (e.g. the Final-over-Final Condition in Sheehan, Biberauer, Roberts & Holmberg 2017), these constraints should simply limit, not direct, a child's outlook on the grammatical possibilities in her language.

As Phillips (2012:282) suggests, the most common sources of variation and, in turn, change are (5a-c).<sup>109</sup> Were we to divide *exogenous* and *endogenous* change based on the linguistic experience of a child as detailed above, *exogenous* change would take place in external-driven portions of her experience. Phillips also states, however, that "... it is also conceivable that there might be variation in (5d-f), either as a consequence of experience or as a consequence of genetic variation..." Let us return to (5d) and (5e) specifically. Omaki & Lidz (2015) cite two linking problems for tracking accurate acquisition in children. The first deals with the fact that the full spectrum of linguistic knowledge of a child is impossible to readily deduce with behavioral research data. This impediment only grows larger in a diachronic context. The second, which is highly pertinent to the discussion of the prerequisites outlined in (5), is that there may be considerable and often consistent divergence between the input signal in a child's linguistic environment and the intake that she processes and feeds into her developing grammar. Is it plausible that some hypotheses posited by the child are simply not present based on the evidence in her input? Certain diachronic change as well as synchronic child data suggest that this is indeed possible and that children have expectations regarding how their input *should* be

---

<sup>108</sup> There are two lines of thought as to where these constraints originate. Chomsky (1965, *et seq.*) proposed that these constraints are innate properties of the LAD. Perfors, Tenenbaum, & Regier (2011) claim that these constraints may be formed in the development of a given grammar based on Bayesian learning model, nullifying the need for these restrictions to be part of the innate language faculty.

<sup>109</sup> It is worth noting that the failure of the prerequisite (5a) is highly implausible, although it has been considered in several discussions related to how this scenario would actually play out (Walkden 2011, 2012; Roberts 2007:233 on *P-ambiguity*).

structured (Lidz, Gleitman, & Gleitman 2003; Lidz & Gagliardi 2015). These aspects are at the heart of the endogenous changes I discuss in §4.3.

#### 4.3 Reanalysis and highly variable input

In this section, I discuss the aspects of grammar change in acquisition as addressed in Hudson Kam & Newport (2005). In particular, I address the effects that exceptionally high variable input may have on a developing grammar, particularly with respect to regularization. I apply these basic ideas to the scenario discussed in Chapter 3 in which the child may receive input for several strings that are minimally distinct from a structural standpoint and, thus, forces reanalysis in order to regularize the data in the intake.

##### 4.3.1 Hudson Kam & Newport (2005)

The investigation in Hudson Kam & Newport (2005) (henceforth HK&N) focuses on answering two primary questions: i) What do learners acquire when their input contains inconsistent grammatical morphemes?, and ii) do learners acquire variability veridically or do they regularize it as their grammars develop?<sup>110</sup> In a series of two artificial language experiments performed on adults and children alike, these authors concluded that adults acquire new, highly variable morphological patterns very differently than children. These authors found that children learn differently than adults in that they are more apt to regularize inconsistent input whereas adults simply learn and repeat the patterns they learn.<sup>111</sup> For adults, this does not entail that they learn all of the relevant paradigm, but simply that they accurately reproduce what they have learned. In contrast, children may acquire part of a relevant paradigm and use these forms to regularize their internal grammar, ultimately bypassing the data they receive that they deem too

---

<sup>110</sup> *Veridically* in this sense is not linked to a child's commitment to truth conditions. Instead, it should be interpreted here as whether children imply a 'what you see is what you get' approach with respect to the input being true to form.

<sup>111</sup> These claims have been supported by further computational analyses such as Yang (2016 §3.5.2).

inconsistent or variable. Much of the work prior to HK&N (2005) claimed that these regularizations were due to the input being probabilistic, suggesting that this type of regularization should only arise with input that is highly variable and inconsistent. This may be because children overmatch certain forms whilst abandoning others (Stevenson & Weir 1959), or because they use their own novel patterns (Goldowsky 1995). Although more recent investigation confirms that learning cannot be probabilistic in nature (*pace* Ferreira & Patson 2007), these two patterns seem to be able to account for all types of grammar change.

Using an artificial language, two experiments were done, one with adults and one with children. In Experiment 1, forty-three native English speakers (average of 22;7 years old) were presented with data from an (neg)-V-S-O language consisting of fifty-one words: thirty-six nouns, seven intransitive verbs, five transitive verbs, one negative, and two determiners (one for each of the two noun classes). All aspects of the input were consistent except for the determiner-noun combinations. For half of the participants, the two noun classes were divided based on the count/mass distinction, nine mass nouns and twenty-seven count nouns, which was based on determiner selection. For the other half, there was an arbitrary gender component based on the determiner use. The participants were only exposed to the auditory aspect of the language, mimicking English prosody and phonology, that narrated scenes from a video as in (6).

(6) /flm ronmawt po blɛrɛnfal po/  
hit bowling-ball DET2 bowling-pin DET2

‘The bowling ball hits the bowling pin.’

Participants were broken up into groups based on exposure of determiner use: a low input group (forty-five percent of the time), a mid-input group (sixty percent of the time), a high input group (seventy-five percent of the time), and a perfect input group (one-hundred percent of the time).

After six sessions of input exposure, the participants were given four tests in order to test their performance. The first was a vocabulary test in order to examine the participants' knowledge of the vocabulary which was crucial in order to take the other tests. The second was a sentence completion task centered around the consistent grammatical aspects of the language. These two tests were used in order to determine if the participants had acquired enough of the input in order to be examined based on their grammatical knowledge. The third was a grammaticality judgement task based on determiner usage in which the participants rated the sentence on a four-point scale based on how much they liked or disliked the sentence. The thirty-six sentences of the grammaticality judgement task consisted of three variations of twelve base sentences: one with the correct determiner placement (postnominal) (7a), one with incorrect determiner placement (prenominal) (7b), and one with no determiner (bare noun) (7c).

- (7) a. /gɛrn fɛrluka po/  
       fall girl DET2
- b. /gɛrn po fɛrluka/  
       fall the girl
- c. /gɛrn fɛrluka/  
       fall girl

Intended: 'The girl falls.'

Four of the twelve base sentences had varying determiner selection with a transitive subject, four had varying determiner selection with a transitive object, and four were intransitive. The last test was a forced choice grammar test in which participants listened to sixteen pairs of sentences and were asked to select the sentence that sounded most like a sentence from the language to which

they had been exposed. The pairs were two versions of the same sentence, one grammatical and one ungrammatical.

The results of the two grammar tests directly reflected the amount of exposure to determiner-noun combinations that each participant group received, as noted in Table 6 (Hudson Kam & Newport 2005:171).

*Table 6. Percentage of production systematicity by input group*<sup>112</sup>

| Input Group | Production Type |                     |               |
|-------------|-----------------|---------------------|---------------|
|             | Systematic User | Systematic Non-User | Variable User |
| 100%        | 100             | 0                   | 0             |
| 75%         | 11.1            | 11.1                | 77.8          |
| 60%         | 0               | 25                  | 75            |
| 45%         | 0               | 0                   | 100           |

As Table 6 shows, participants used determiners in their production with a high correlation to how often they heard them in the input (i.e., those that experienced more accurate determiner input produced them more accurately). This led HK&N to claim that “[adults] learn to reproduce the amount and type of variation that is present in their input, but they do not regularize the language as they learn it” (173).

In Experiment 2, nineteen children (average of 6;4 years old) and eight adults (average of 20;1 years old) were exposed to an artificial language *Sillyspeak* containing seventeen words: four verbs, twelve nouns, and one determiner.<sup>113</sup> As in Experiment 1, all input was consistent

<sup>112</sup> *Systematic users* were those who used determiners with all NPs, *systematic non users* were participants who used no determiners at all, and *variable users* were those who used determiners variably with no recognizable systematic pattern, akin to the inconsistent input they were provided.

<sup>113</sup> There were several differences between the two experiments, two that are worth noting with respect to the methodology and the final results. First, the children were taught the vocabulary words (although no explicit grammar was taught) in place of the vocabulary tests done with the adults in order to assure they had learned a sufficient amount of vocabulary to complete the last two tasks. Second, four of the nineteen children failed to learn

outside of determiner use. Participants were broken up into two groups with respect to exposure of determiner use: an *inconsistent* group (sixty percent of the time) and a *completely consistent* group (one-hundred percent of the time). As in Experiment 1, the participants of Experiment 2 were also exposed to the language over a period of six sessions. The first test, a sentence completion task, came in two parts. First, children were asked to name a series of toys whose name they had learned in order to ensure that the child recognized the lexical items in her test set. Much like the grammatical judgement test in Experiment 1, Experiment 2 presented the children with eighteen total sentences consisting of three variations of six base sentences. As in Experiment 1, only one of the variations was correct, whilst one contained a misplaced determiner and one had no determiner. Children were asked to rate these (one through four) based on the degree of smiley faces (sad, slightly sad, slightly happy, happy) depending on how much it sounded like a Sillyspeak sentence.<sup>114</sup> The next test, a general grammar test, used the same set of faces in order to judge how much the children liked certain grammatical patterns related to determiner use in both transitive and intransitive sentences. The third test was production based. It was here that HK&N noted the significant difference between adults and children of the inconsistent input groups (i.e., the sixty percent groups). Their conclusion for the results of this test was that “[c]hildren are very likely to produce consistent patterns, even when receiving inconsistent input; adults, however, are not. Adults are systematic when their input is consistent and variable when their input contains variation” (181). Returning to the fact that language acquisition is not probabilistic in the sense suggested by Ferreira & Patson (2007), this inherently implies that the morphosyntactic rules children adhere to are, in fact, not the result of

---

enough words to produce any sentences and did not complete the study. All eight adults that began Experiment 2 completed it.

<sup>114</sup> For the equivalent of judgements 2 and 3 from Experiment 1 are the slightly sad and slightly happy faces provided to the children in Experiment 2.

a “good enough” attitude but, instead, the product of the grammar to which they have committed. As we have seen, this may very well be deviant with respect to the adult grammar and may or may not undergo repair as children learn more about their language(s).<sup>115</sup> The final test, one of determiner judgement, mirrored the third test from Experiment 1 in that the children were presented with eighteen sentences with three variations of each of the base sentences: one in which the determiner placement was correct, one in which it was not, and one in which there was no determiner. The results highlighted the fact that children in the sixty-percent condition group did not prefer sentence types that they had never heard (i.e., those with a determiner in the incorrect place). HK&N claim that while adults prefer to internalize a grammar that differentiates structures that are more or less common, children often extract knowledge that leans toward the structure they perceive to be most common.

Although the experiments in HK&N (2005) highlight the aforementioned aspects of acquisition and potential change at the morphology-syntax interface, I argue that the same patterns and mechanisms are at work in the syntax proper.<sup>116</sup> For example, where HK&N argue that certain determiner-noun class combinations may be deemed too difficult to learn based on the input provided to the learner, I show that the same may be suggested for purely syntactic phenomena when a learner is faced with input that are unable to be subsumed under her established hypothesis regarding the syntactic behavior for a particular lexical item. Pearl (2007)

---

<sup>115</sup> This meaning of *probabilistic* should not be confused with that which is often mentioned in the computational linguistics literature (e.g. Yang 2016). There, probabilistic is understood purely mathematically with respect to the percentages of what has been presented and, in turn, committed to by the learner based on her interpretation of the input. In a model that supports a computational component such as Yang’s *Tolerance Principle* or Bayesian learning and overhypotheses as mentioned in §1.2.1, the probability of grammar *G* being adopted is based on the internal percentage of committed instances of a particular morphosyntactic feature that support the acquisition of grammar *G* instead of another. These computational aspects coupled with a child’s running commitment to her structural parses have proven adequate beyond the ideas of *transformational learning* (Gibson & Wexler 1994) which led to the idea that a single misparse could derail a child’s entire grammar with respect to a certain phenomenon. I refer the interested reader to Chapter 1 of Yang (2016) for a more in-depth analysis of this.

<sup>116</sup> Much like the natural language data from Tsez in Gagliardi & Lidz (2014) and Gagliardi, Feldman & Lidz (2017), the artificial patterns in HK&N (2005) deal with determiner class-noun class dependencies. Although there are strong morphological implications, many of the aspects discussed are syntactic in nature.

and Pearl & Weinberg (2007) have explored these aspects of acquisition as an avenue for the change in word order in English from SOV to SVO, a change in constituent order derived from the loss of features (Biberauer & Roberts 2016, 2017).

I suggest that these patterns are present in child language acquisition with phenomena that have yet to provoke diachronic change (i.e., those that are currently stable in the modern grammars of the respective languages). This further supports the idea that both diachronic phenomena and deviant structures in child language are inextricably linked and result from the same fundamental mechanisms (Cournane 2017, Hale 1998).<sup>117</sup>

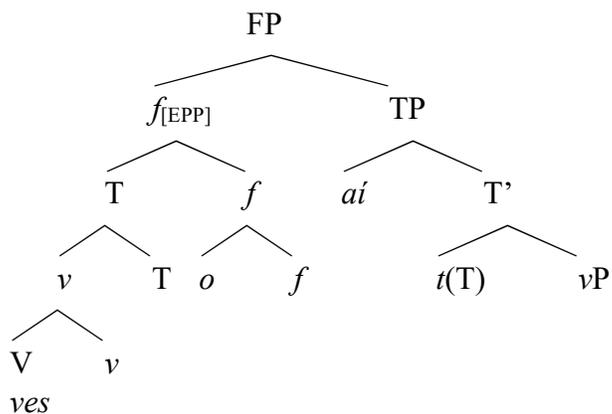
#### 4.3.2. The reanalysis of *velai/velaqui*

In §3.7, I argued for two variations of Step 2 based on data that showed a ‘fixed enclisis’ construction under embedding of the complementizer *que* and T-oriented adverb *xa* (‘already’). The mixed input scenario that I claim forced reanalysis encompasses both of these variations in addition to abundantly regular, sentence-initial use of this expression. As I showed in §3.2.1, Step 1 did permit CLLDed objects, but the frequency was not robust and this construction seemed to disappear as the variation in Step 2 grew. The main constructions I focus on, thus, are those structured below in (8-10).

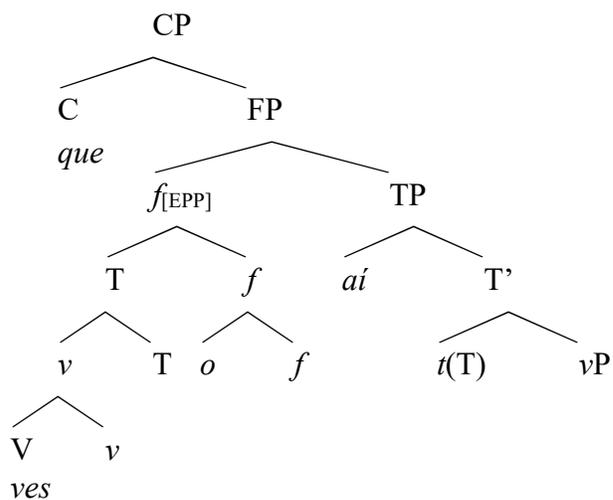
---

<sup>117</sup> As noted by Cournane (2017:13), this does not entail that the patterns themselves are mimicked in the same manner in child language acquisition (i.e., that the proposed steps in diachronic developments should pattern the same way and in the same order in child language), a view supported by investigators that do not advocate for the *Child Innovator Approach* (e.g. Diessel 2012).

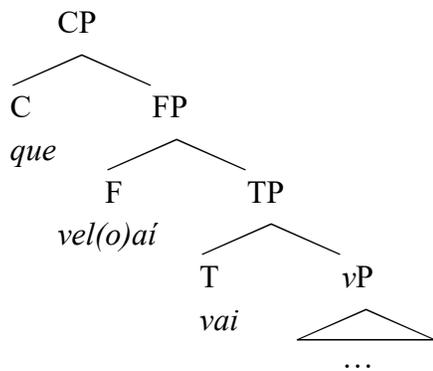
(8) *Ve-lo aí* (Clause-initial enclisis)



(9) ... *que ve-lo aí* (Enclisis under a filled C)



(10) ... *que vel(o)aí vai* (Reanalyzed complementizer in F)



HK&N (2005:184) argue that children “have a strong tendency to impose systematicity on inconsistent input and sometimes impose patterns on this input that are not produced by adult learners.” This overlaps with what I proposed in §4.2.1 regarding endogenous change and the examples I show below in §4.4. However, HK&N highlight the fact that there are also general learning mechanisms that fail to interact in an optimal manner with certain data, arguably because of immature grammatical development (Newport 1990). This seems to be particularly clear when there are minimally variable forms used across a speech community that may become regularized and consistent within a given grammar. This, I claim, is what happens in traditional instances of multi-word reanalysis in order to form a single functional item. This slightly contrasts with instances of reanalysis that remain morphosyntactically variable after having undergone change, as was the case of the Romance futures (Graham 2015, Roberts 1993, 2007, 2010, *i.a.*). Here, there was certainly reanalysis of morpho-phonosyntactic boundaries. However, as the respective morphology was productive enough for the learners of these speech communities to maintain relevant [PERSON] distinctions, the syntactic reanalysis of these forms maintained their variability.<sup>118</sup>

Let us imagine that (3-5) were frequent in the input. The loss of topicalized objects in CLLD constructions alone places (3) in an optimal position to be reanalyzed as a single lexical item. However, adding situations of enclisis under embedding as in (4) places the learner’s hypotheses regarding (3) as another instance of finite enclisis in a disadvantageous position, as this verb construction in particular is the only one she will experience in the input that does not show proclisis under the embedding of the complementizer *que* (§3.7). In a way, this places the structures (8) and (9) on the same level syntactically and, thus, makes them largely

---

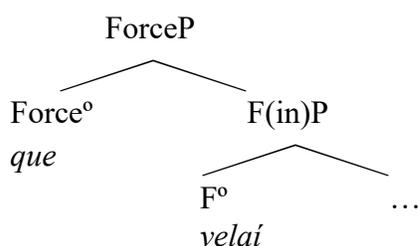
<sup>118</sup> [PERSON] distinctions seem to be more persistent historically than those of [GENDER] or even [NUMBER] (Haspelmath 2004).

indistinguishable from (10). It is clear here, as well, that what is predicted for formal reanalysis (Roberts & Roussou 2003) does not hold up for the reanalyzed *velai/velaqui*: these two functional items remain in F during what I described as Stage 2 of Step 2 (§3.7) as in (10), whereas it is predicted that reanalysis always entails movement up the tree.

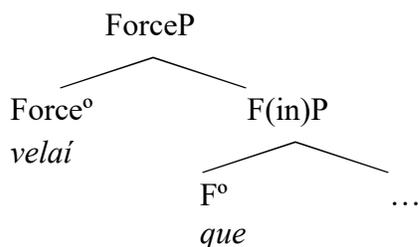
#### 4.3.3 Movement motivated without features

In §3.7 I described two syntactically distinct stages the reanalyzed complementizers *velai/velaqui*, repeated below.

(11) a. *Step 2*



b. *Step 3*



Aside from their base-generated position, there are several syntactic properties of (11b) that are absent in (11a). Most importantly for this investigation, there is no record of the structure in (11a) showing the C-AGR phenomenon. Moreover, based on the evidence shown in §2.2.2, this should be impossible. Let us recall that in the modern grammar, cliticization, and in turn C-AGR, is ruled out when *velai/velaqui* selects the complementizer *que* in F.

- (12) \*Velaquí que a vexo  
 behold COMP CL<sub>F.SG</sub> see.PRS.1SG

Intended: ‘Here she is, I see her.’

As I claimed in §2.2.2, that the clitic and the complementizer *que* are in complementary distribution is due to the fact that both occupy the same position in *F/f*.<sup>119</sup>

What might have motivated the switch of *velai/velaquí* and *que*? I claim that the principle reason lies in the pragmatic interpretation of the reanalyzed complementizers as presentatives. As Zanuttini (2017) and Wood et al. (2015) note, presentatives are always sentence initial, which is why they reject syntactic environments with negation or embedding, for example. Although I have rejected the notion that semantic change drives syntactic change (§1.4), there is evidence from the psycholinguistic literature that suggests children are sensitive not only to the structure of their language but in what contexts it is used (Harrigan et al. 2016, 2018, 2019; Hacquard & Lidz 2019). If presentatives have almost identical pragmatic and syntactic properties cross-linguistically, we may imagine there is something inherent about their nature that drives this behavior and permits children to learn them easily. As I highlighted above, the frequency of the sentence-initial nature of Step 1, the most syntactically variable of all the stages I propose for the diachrony of this construction, was very robust. Moreover, it is never found with negation or as an embedded clause. This high sentence-initial syntactic frequency and the lack of negation and embedding alone might have forced not only reanalysis but also a prompt alternative hypothesis regarding the left-peripheral nature of these main-clause complementizers that led to establishing

---

<sup>119</sup> It may also be that the probe on *velai/velaquí* is unable to probe past the phrase head *f* (cf. Fernández Rubiera 2013, Gravely 2020 for evidence of *f* as a phrase head in Asturian and Galician respectively) or that there is simply no probe introduced by *velai/velaquí*. I address the second possibility in Chapter 6 and leave the former for future research.

them as presentatives.<sup>120</sup> A transition in hypothesis regarding the loss or increased scarcity of contradictory evidence is a shift from what Roberts (2007:133) called *strong P-ambiguity* to *weak P-ambiguity*.

(13) *P-ambiguity*

- a. A substring of the input text S is strongly P-ambiguous with respect to a parameter  $p_i$  just in case a grammar can have  $p_i$  set to either value and assigns a well-formed representation to S.
- b. A strongly P-ambiguous string may express either value of  $p_i$  and therefore trigger either value of  $p_i$ .
- c. A weakly P-ambiguous string expresses neither value of  $p_i$  and therefore triggers neither value of  $p_i$ .

Applying the term parameter in the strict sense here to simply refer to the possible syntactic position(s) of a lexical item, it is clear that *strong P-ambiguity* is that which we would expect to force reanalysis. The notion of ‘strength’ here is not elaborated on in depth by Roberts. However, we may imagine that there must be a certain level of ambiguity with respect to the hypothesis the child draws from her input in order to consider multiple options statistically viable. In the case of *velai/velaqui*, this is would reference either the sentence-initial behavior seen in (11b) or the second-position structure in (11a). For most cases, one would assume that the child would select the most frequent statistical option, as alluded to by Yang (2016). However, as I show in the next section, this is not always the case. In turn, a *weak P-ambiguity* would show no mixed hypothesis at all. In the formulation in (13), Roberts only highlights the instances in which no parameter is set (i.e., no phenomenon is detected). However, this terminology is also valid for instances in

---

<sup>120</sup> I use the reference of the left periphery due to its position in the adult grammar. However, as hypothesized for by the *truncation hypothesis* (Rizzi 1994), this may be simply the highest projection established in the developing grammar (e.g. TP) for children.

which there is only one clear parameter setting in the input, thus, the child easily comes to a firm conclusion about the syntactic behavior of said parameter.

#### 4.4 How arbitrary is endogenous change?

In this section, I discuss data originally collected in Manetti & Belletti (2017) and discussed in Belletti (2017) with respect to deviant structures in child language from Italian. Belletti claims that internal pressures related to computational complexity may lead to liberal creation of grammar by the child. The conclusions from this section will provide a solid base for which I propose how the modern-day Galician C-AGR construction (Step 4) was engendered.

##### 4.4.1 When children create grammar

Belletti (2017) discusses deviant structures in child data collected from Manetti & Belletti (2017) regarding the non-standard use of differential object marking and scrambling positions in Italian. In standard Italian, full DP object nominals are not differentially marked regardless of their position in the clause.

- (14) a. Mattia      ama                  Francesca  
        Mattia      love.PRS.3SG Francesca  
        ‘Mattia loves Francesca.’
- b. Francesca<sub>i</sub>,    la<sub>i</sub>      ama                  Mattia  
        Francesca      CL<sub>F.SG</sub> love.PRS.3SG Mattia  
        ‘Francesca, Mattia loves her.’

Moreover, Italian seems to have restricted clausal positions for objects outside of SVO and CLLD structures (Belletti 2004). The data from Manetti & Belletti, however, show that children both differentially mark their objects and scramble them to a position not permitted in the adult

grammar.<sup>121</sup> The design of the questions for Experiment 1 posed to the children participants in Manetti & Belletti showed pictures of a rabbit touching a penguin and a giraffe licking a cow, followed by the question in (15).

- (15) Che cosa succede ai miei amici, il pinguino e la mucca?  
 what thing happen.PRS.3SG to-the my friends the penguin and the cow  
 ‘What is happening to my friends the penguin and the cow?’

They showed that the answers to (15) were often CLLD constructions, which is perfectly acceptable in the adult grammar. However, the way in which they formulated the CLLD sentences was unlike what we see above in (14b).

- (16) Il coniglio a i’pinguino lo tocca  
 the rabbit DOM the-penguin CL<sub>M,SG</sub> touch.PRS.3SG  
 ‘The rabbit touches the penguin.’ (lit. ‘The rabbit to the penguin (he) touches him.’)

Two things are noteworthy about the string in (16). First, the subject remains in subject-initial position rather than following the verb. There were also instances of CLLD-S-V order. Second, the object is differentially marked which, as discussed above, is not a characteristic of standard Italian. This phenomenon is what Belletti (2017, 2018) refers to as *a*-Topics, a phenomenon previously discussed in great detail in Belletti & Rizzi (1988). These authors claimed that these DOMed objects are only marginally acceptable in psych-predicates of the ‘worry class’ (17a), whereas they are ungrammatical for all speakers in traditional transitive constructions (17b).<sup>122</sup>

<sup>121</sup> Belletti (2018) notes that while there are variations of Italian that utilize the differential object marker *a* in this manner (e.g. southern varieties such as Calabrese; Ledgeway, Schifano, & Silverstri 2019), this is not the case in standard Italian (Tuscan). As the children in Manetti & Belletti (2017) were from this area, there is no reason to attribute these deviant structures to variety-induced change/influence.

<sup>122</sup> It is worth noting that the observations in Belletti & Rizzi (1988) precede the exploration of non-standard varieties of Italian. When Belletti (2018) returned to the investigation of the *a*-Topic phenomenon, she noted that this was a characteristic of southern Italian varieties, a topic of much research in the past decade (e.g. Benincà, Ledgeway, & Vincent 2014).

- (17) a. ?A Mattia questi argomenti non l'hanno convinto  
 DOM Mattia these arguments NEG CL<sub>M.SG</sub>-have.PRS.3PL convince.PRT

Intended: 'As for Mattia, these arguments haven't convinced him.'

- b. (\*A) Mattia la gente non lo conosce  
 DOM Mattia the people NEG CL<sub>M.SG</sub> know.PRS.3SG

'As for Mattia, people don't know him.'

Belletti (2017:3) shows that the position of the subject also had an influence on the percentage of *a*-Topics. When the subject preceded the verb, either before or after the *a*-Topic object, eighty-eight percent of CLLD objects were *a*-Topics. When the subject was postverbal or null, however, this number decreased to forty-eight percent, almost equal to the number of non-DOMed objects. The author relates the need to differentially mark objects in this position to the poor comprehension performance of children to accurately parse object relatives of the same number feature, as in (18).

- (18) Il cane che il gatto morde  
 the dog.SG COMP the cat.SG bite.PRS.3SG

'The dog that the cat bites.'

Based on the experiment findings of Friedmann, Belletti, & Rizzi (2009) and Belletti, Friedmann, Brunato, & Rizzi (2012), children until eight or nine years old parse this object relative construction in such a way that the object and the subject nominals are seen as competing for agreement with the verb. Belletti attributes this to a false position of Relativized Minimality (RM; Rizzi 1990, 2013).

- (19) [X ... [Z ... [Y ...]]]  
 ↑

Belletti explains that a constituent originating in Y that forms a dependency with the target position in X across an intervening position Z cannot hold if X and Z share relevant features. Thus, an object relative string as in (18) would be modeled as in (20).

|      |               |                   |      |               |       |              |       |
|------|---------------|-------------------|------|---------------|-------|--------------|-------|
| (20) | Il            | cane <sub>i</sub> | che  | il            | gatto | morde        | $t_i$ |
|      | the           | dog               | COMP | the           | cat   | bite.PRS.3SG | $t$   |
|      | $X_{\phi SG}$ |                   |      | $Z_{\phi SG}$ |       |              | Y     |

Belletti's observations lead the reader to believe that, because the relative head (X) and the restricted subject (Z) are in a relation of inclusion, both are within the feature set of the target and, in turn, movement. Moreover, she notes that RM alone is able to account for the reason why when the object relative and the subject of the relative are not identical in features, the ambiguity is removed.

|      |               |                   |      |               |         |               |       |
|------|---------------|-------------------|------|---------------|---------|---------------|-------|
| (21) | Il            | cane <sub>i</sub> | che  | i             | cavalli | rincorrono    | $t_i$ |
|      | the           | dog               | COMP | the           | horses  | chase.PRS.3PL | $t$   |
|      | $X_{\phi SG}$ |                   |      | $Z_{\phi PL}$ |         |               | Y     |

Comparing these facts to the data from Manetti & Belletti (2017), Belletti (2018:4-5) concludes that there are two possible reasons as to why speakers need to disambiguate the object from the subject in the *a*-Topic fashion seen in (16). First, the differential marking establishes an 'empathetic relation' with the object, mirroring the affected interpretation seen in (13a) above. Second, due to the decrease in the percentage of *a*-Topic use when the subject was not preverbal, there seems to be a struggle within the syntactic configuration posited by the child when both the topicalized subject and preposed object are preverbal.

Due to the fact that these children (4;1-5;11 years) were expected to have no influence from southern varieties, I disagree with the first argument from Belletti as to the internal reasoning of

the use of *a*-Topics in the children’s data.<sup>123</sup> The lack of influence from other varieties highlights how a child’s experience may not be able to explain how she arrives at a particular conclusion related to her input, an acknowledgement she makes at the end of her conclusions regarding the data (Belletti 2017:6, fn. 8). The second postulation by Belletti, however, has clear theoretical explanations based on cross-linguistic evidence. If we consider differential marking to be a syntactic phenomenon (Ormazabal & Romero 2007, *et seq.*, Kalin 2018; *pace* López 2012), an object shifted above the verb would be in the same phase as the preverbal subject. Under a theory such as the *Distinctness Condition* (Richards 2010), two nominals of the same label within the same phase will cause a derivational crash.<sup>124</sup> Specific to differential marking, this has been claimed for object nominals in Galician that bear an [ANIMATE] feature when in the same phase as a postverbal subject (Gravely 2021). Much like the Italian data in Belletti (2017) suggests, there is no need for differential marking in Galician when the subject is preverbal and the object is postverbal (i.e., when they are not in the same phase).

- (22) a. Os pais axudan \*ós / os nenos  
           the parents help.PRS.3PL DOM-the/ the kids
- b. Axudan ós / \*os nenos os pais  
           help.PRS.3PL DOM-the/ the kids the parents
- ‘Parents help their children.’

The exact details are a bit more complex than what is portrayed here. However, considering Galician, like Italian, does not generally mark full object DPs, the scarcity and position-based

---

<sup>123</sup> I call attention to this observation due to the fact that influence from southern varieties in which *a*-Topics are used, albeit sparingly, would be a case for change via contact and, thus, a plausible reason as to why these children utilized a grammatical phenomenon that is not characteristic of their variety.

<sup>124</sup> Richards refers to either the label (category) or sublabel ( $\phi$ -features) of the two constituents that can cause the crash, citing examples from both.

implementation shown in both adult Galician and child Italian grammars draws certain parallels to distinguishing ‘subject-like’ objects going back to observations in Aissen (2003).

Although this type of deviant structure in child Italian is not unconstrained, it should still be seen as purely endogenous due to the fact that there is no input the children might have received that would lead to their postulation of *a*-Topic marking. Instead, we may assume that a developing grammar is more sensitive to or more dependent on its ability to distinguish arguments internally, having experienced fewer data on which to build hypotheses and, in turn, come to conclusions that would lead the learner to tolerate adult-type constructions (e.g. those with multiple topics). Finally, Belletti (2017:6) highlights that “children’s grammatical creativity in newly identified contexts”, which we may expect to arise and “not necessarily require a rich input to be put into work.” I claim that this creativity is at the heart of endogenous change. Although changes of this sort may undergo repair and be lost before the child’s grammar is solidified, I claim that endogenous change happens when repair does not take place.

#### 4.4.2 Galician C-AGR: from verbal predicate to complementizer

Much like the deviant syntactic behavior of the data from Italian in §4.4.1, I propose that Galician C-AGR was engendered by a misanalysis of *velai/velaqui* as a verbal predicate. In §4.3.2, I discussed the effects of high variability with respect to reanalysis as a whole, as well as the result of multi-word strings reanalyzed as a single lexical item. In addition to this instance of reanalysis, I claim that a second phenomenon surfaced as a result of the mixed input described in §4.3.2. I argue that two main factors gave evidence for the present-day C-AGR construction to surface as it did. First, I claim that *velai/velaqui* was initially analyzed as a verb instead of a left-peripheral functional head. In §2.3.2.1, I compared *velai/velaqui* to other presentatives in Romance, showing their similarities and highlighting the only difference amongst them: unlike

*velai/velaqui*, they are all inflectionless functional verbs. Many of the diagnostics to determine their status as verbs are also possible for *velai/velaqui*, making the link between the proposal of these two functional items as verbs a plausible one. Second, postulating *velai/velaqui* as verbal predicates puts them in the same class as the rest of Romance presentatives in that they should permit cliticization. Based on the strict word order of this construction throughout its diachronic developments, the postulation of the string in (23) almost identically mirrors the robust data we have seen from Step 1 (24): [<sub>CLAUSE 1</sub> V CL<sub>i</sub> [<sub>CLAUSE 2</sub> V SUBJ<sub>i</sub>].

(23) Velai      o      vai              Xan  
       behold    CL<sub>M,SG</sub> go.PRS.3SG    Xan

(24) Ve-lo                      aí      vai              Xan  
       see.PRS.2SG-CL<sub>M,SG</sub>    there go.PRS.3SG    Xan

Taking this hypothesis as a starting point, the ideas presented above regarding child regularization due to highly variable input also works in favor of the idea that ‘verbal’ *velai/velaqui* would select an accusative clitic as its internal argument of *ves* (‘you see’) in the original string.

In attempting to explain the unexpected agreement between the clitic and the subject in the Galician C-AGR construction, Freixeiro (2006:132, fn. 98) suggests several possibilities:

“A possibility in order to explain the origin of this accusative[-like] construction may lie in the primitive function of the clitic as a direct object of the verb *ver* (*Velo aí está*), which, after fusing with the locative adverb *aí*, giving rise to the presentative adverb *velai* (the same for *velaqui*, vid. Ferreiro 1999:354), lost its phonic identity, provoking the recovery of the full form (*Velai está* → *Velai o está*)... It is also possible [to link] a certain similarity to the

construction *Aí o tes*, but in this case *o* is clearly the direct object; the pronoun is not coreferential with the subject here (*Aí o ten (ela)*).”

The first postulation by Freixeiro is a functional resultative summary of what I propose. His other hypothesis, based on analogy, is one supported by learning strategies that rely on stored exemplars as in Ambridge (2020). In a model such as this, children use and learn from “stored exemplars and the ability to analogize across them on the fly in comprehension or production” (Ambridge 2020:3). This entails the overlap of syntactic and semantic similarities across entire strings of words, with no regard for individualization of internal mechanisms and a sole focus on inferences based on surface form. As Adger (2020) points out for recursivity and Koring, Giblin, Thornton, & Crain (2020) point out for similar syntactic structures with vastly different semantic meanings, a learning system that is focused entirely on surface-level forms of entire utterance strings rather than abstract linguistic principles fails to account for numerous syntactic phenomena in addition to the discrepancies in meaning between two superficially similar strings.

Ruling out analogy as a source of the C-AGR construction allows us to return to the perceived similarities Galician children of the time might have regularized. From a parsing standpoint, it is reasonable to assume that the reanalyzed *vel(o)ai* might have been attributed with the same syntactic properties (e.g. a categorial [+V] feature) as the verbal variations of the *ve-lo ai* string. This becomes more plausible when we consider there are modern-day varieties of Galician whose reanalyzed complementizers are pronounced [be.lʊ.a.i] and [be.lʊ.a.ki] as I mentioned in §2.2.1. Further support lies in the fact that this construction never appeared under embedding and, thus, never showed a proclitic construction, as in the unattested example in (25).

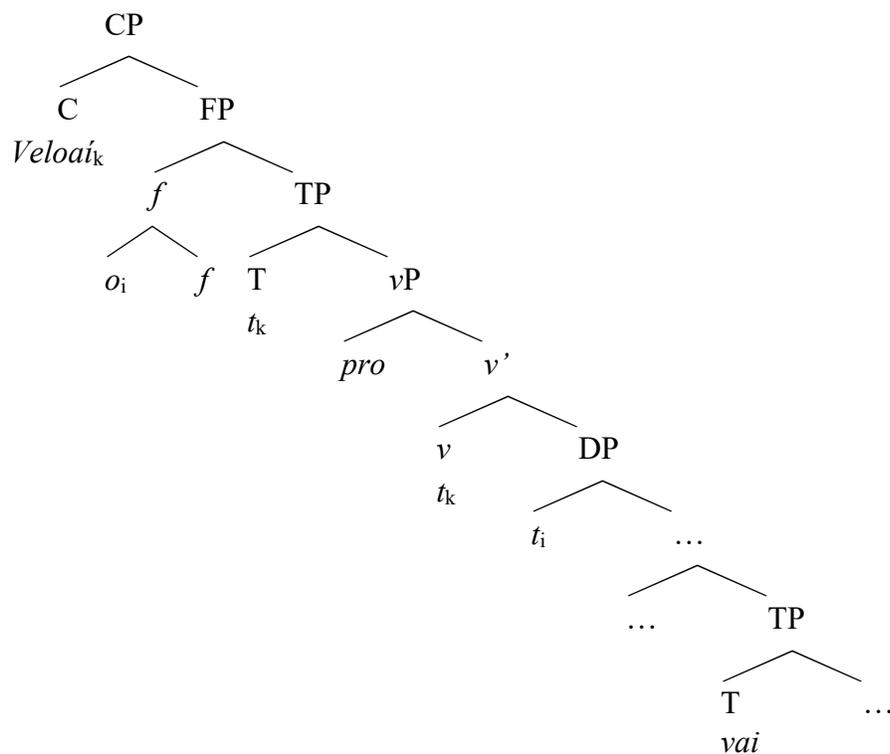
- (25) \*Que o ves aí vai  
 COMP CL<sub>M,SG</sub> see.PRS.2SG there go.PRS.3SG

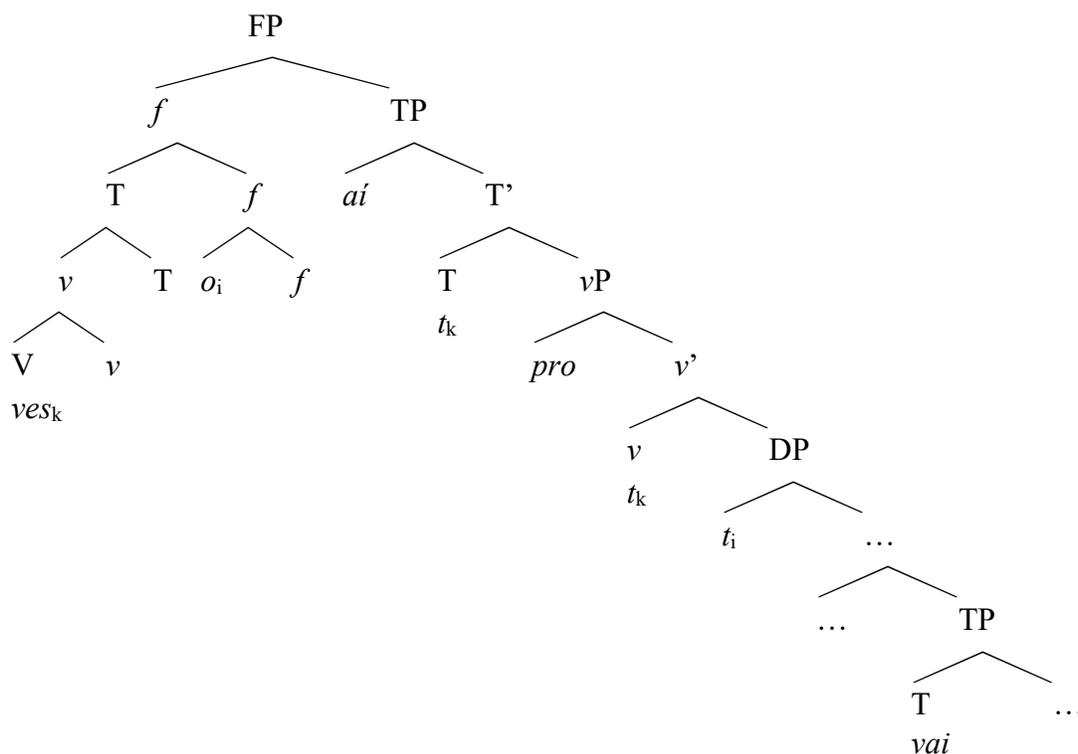
Intended: ‘That you see him, there he goes!’

I claim that the lack of evidence for (25) makes the maintenance of the clitic derivationally simpler after the repair of *velai/velaqui* from verbs to main-clause complementizers. As I show in Chapter 5, the way C-AGR is derived in Galician is much like another cliticization pattern found in the language, namely that of determiner cliticization (Uriagereka 1996).

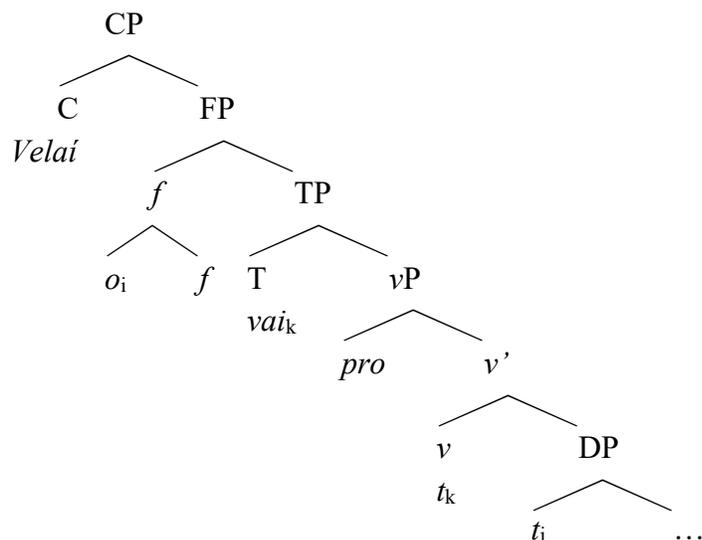
Building on what I showed for other Romance presentatives in §2.3.2.1, I claim that the hypothesized verbal presentatives here must have contrasted minimally with the string in Step 1 with respect to its overall structure.

- (26) *Vel(o)aí<sub>[+V]</sub> o vai*



(27) *Ve-lo ai vai*

The models in (26) and (27) leave us with a similar structural comparison that begs the question that will provide us with much of the focus for the remainder of this dissertation: Did children treat the derivational steps for cliticization in the same manner in (26) as in (27)? Based on the arguments above (cf. 25), one would imagine not. However, if they did, then an additional repair would be needed, which might have aided in the reanalysis of the structure in (26) to that of the present-day construction, with the analysis that I propose in (28).

(28) *Velai o vai*

The arguments for the reduction of clausal boundaries in the diachronic literature abound (Roberts 1993, 2007; Roberts & Roussou 2003; Kuteva 2000; Klausenberger 2000; Hopper & Traugott 2003, §7; *i.a.*). Although double- to single-clause reanalysis typically restructures the syntactic properties of the newly reanalyzed clause, this does not seem to be the case in the development of Galician C-AGR. Barring the CLLD constructions from Step 1, there is a firm order in which constituents are placed. In addition to the reanalysis of (26) as (28), it is clear that this word order remained a constant due to the fact that children only received input in which the subject was postverbal and coreferential with the accusative clitic, either with the verb *ves* (27) or the ‘verbal’ complementizer in (26). This fixed-position string of constituents throughout the development of the phrase explains the data that we examined in §2.2.1, where the modern grammar does not permit preverbal subjects in constructions in which *velai/velaqui* select a verb, regardless of whether it is a C-AGR construction or a transitive verb.<sup>125</sup>

<sup>125</sup> I claim this is also a concomitant of the fact that postverbal subjects, arguably in their base-generated position in [Spec,vP], are focused constituents (Belletti 2004, *i.a.*). This strengthens the position that the postverbal subject in Galician C-AGR constructions represent the real-world entity being focused on in the discourse, an inherent property of presentative elements cross-linguistically.

#### 4.5 Conclusion

In this chapter, I discussed reanalysis as a whole from the *Child Innovator Approach*, highlighting the two types of change that may occur in the learning experience: *exogenous* and *endogenous* change. I claimed that all change is technically endogenous, although certain parts of the learning process may be influenced more or less by external forces and, thus, exogenous change may be claimed for those in which specific input patterns lead children to reanalyze said input in order to form new grammatical hypotheses. Citing claims from Hudson Kam & Newport (2005), I claimed that a common source of exogenous change is due to highly variable input that differs slightly in such a way that children opt to regularize the input as they see fit. I showed that the variable forms for *velai/velaqui* were often negligible from a surface-string standpoint, for which regularization and, in turn reanalysis, must have been an optimal choice for young learners. Moreover, I discussed how *velai/velaqui* came to be sentence-initial functional items only, going from a stage in which *que* in Force selected them in Fin to the reverse scenario.

For endogenous change, I reviewed data from Belletti (2017) regarding creative grammar implementation of children in Italian that can only be considered endogenous, as there is no evidence for this grammar in the adult language. I claimed that this point of view is relevant for how the Galician C-AGR construction was engendered, as *velai/velaqui* were essentially miscategorized as ‘verbal’ presentatives, much like the rest of those found in Romance (§2.3.2.1).

In Chapter 5, I review ample work regarding clitics and agreement affixes, both from the formal diachronic and synchronic literature. In order to properly identify how Galician C-AGR is derived, I analyze several theoretical claims regarding both clitic doubling and pure agreement, showing that all of them fall short of accounting for the phenomenon in question. I conclude with

a discussion regarding the transition from accusative clitic argument to subject agreement marker from a learning perspective and how it should be analyzed diachronically.

## CHAPTER 5

## CLITICS, AGREEMENT, AND COMPLEMENTIZER AGREEMENT

5.1 Introduction

This chapter addresses several aspects of agreement with respect to clitics both in a diachronic and synchronic light. I review the long line of literature that distinguishes affixal morphemes from clitic morphemes, what many investigations have referenced as the difference between true agreement and cliticization. Specifically, I highlight instances in which clitic doubling appears to be agreement on the surface and how to account for this anomaly with several cross-linguistic examples within Romance. In turn, I discuss the theories related to affixes and clitics for the C-AGR phenomenon, many of which are centered around whether said phenomenon is indeed syntactic or post-syntactic. Coupling these two theoretical lines of research, I discuss where Galician C-AGR does and does not fit in with the theories proposed. Galician C-AGR is particularly appealing for both aspects of investigation due to the fact that it appears to be an instance of true agreement realized as a clitic morphologically identical to accusative clitics in a language that does not have subject clitics or free-standing agreement morphemes. From a diachronic perspective, I show the challenges for claiming that the Galician C-AGR phenomenon follows the cyclical patterns found in the literature on change. Specifically, whereas there is a tendency for nominal constituents to follow the pattern  $DP > D^{\circ} > \Phi$ -AGREEMENT, there are arguments for and against claiming that the Galician C-AGR morpheme has undergone the change from a  $D^{\circ}$  argument to a true agreement morpheme.

I break this chapter up into two main parts: the review of clitic doubling vs. true agreement (or syntactic cliticization and affixation more generally) in §5.2-§5.7 and how both the proposals and the data for the Galician C-AGR phenomenon hold up based on the C-AGR literature more generally in §5.8-§5.9.

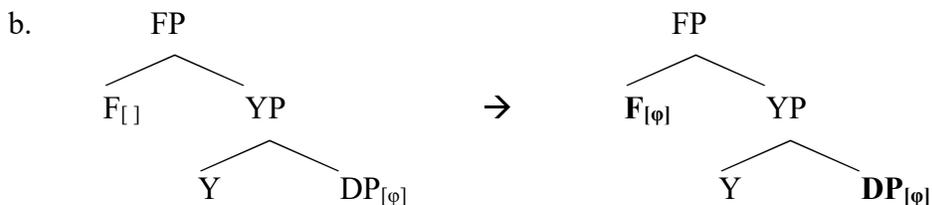
## 5.2 Clitic or affix? The longstanding debate

Since the seminal work of Zwicky & Pullum (1983) the literature addressing how clitics may be distinguished from affixes abounds in the generative literature (Anagnostopoulou 2016; Baker & Kramer 2016; Compton 2016; Harizanov 2014; Kramer 2014; Nevins 2011; Preminger 2009, 2011, 2014, 2019; Yuan 2020, 2021; *i.a.*). These investigations have taken two primary directions: i) to show how the natural language under investigation does or does not correspond to the specifications outline in Zwicky & Pullum, and/or ii) to provide novel theoretical views on how clitics and affixes are distinguished. Furthermore, the discussion of clitics has an important subcomponent, particularly for my investigation here, namely whether the clitic is a doubled clitic or not. Although there are numerous individual components to unpack, one thing is crucial for deriving both clitics and affixes: both are triggered by Agree. Therefore, one must look for other diagnostics in order to determine how different formatives surface and, in turn, what internal mechanisms the syntax utilizes in order to derive them. Regardless of the phenomenon determined to take place, a vast number of the approaches followed in this chapter agree that there must be agreement between a functional head and (minimally) a D<sup>o</sup> (or a DP).<sup>126</sup> Thus, the tradition of distinguishing affixal (agreement) from clitic doubling, for example, may be summarized as in Kramer (2014:598).

---

<sup>126</sup> I continue to use (*true*) *agreement* to refer to the valuation of a probe by a goal in its c-command domain, *sans* movement, although clitic doubling (and clitic movement more generally) is dependent on agreement, as well. I elaborate on this in greater detail in §5.4.

- (1) a. *Agreement*: an obligatory affixal morpheme that is the realization of valued  $\phi$ -features on a functional head F (e.g. *v*, Asp, T, etc.)

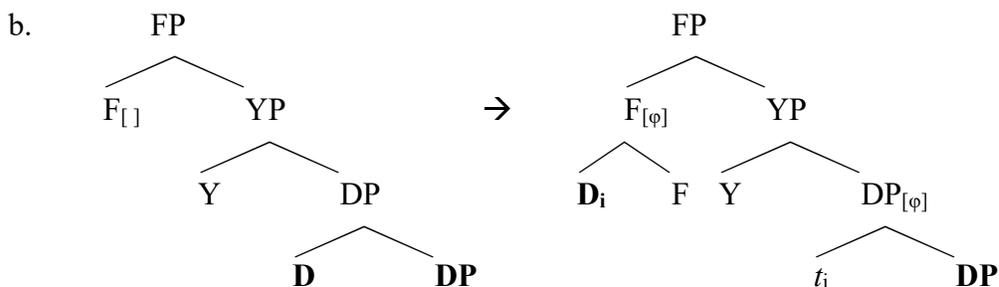


- c. ke-**m** kaa *pro*<sub>SUBJ</sub> cewcew-téetum *pro*<sub>OBJ</sub>  
 C-2 then *pro*<sub>2SG</sub> telephone-TAM *pro*<sub>1SG</sub>

‘When you call me.’

(Nez Perce; Deal 2015:6)

- (2) a. *Clitic doubling*: an optional morphophonological  $D^{\circ}$  clitic that has moved to a functional head F (e.g. *v*, Asp, T, etc.)



- c. **Lo**<sub>i</sub> vimos **al** **niño**<sub>i</sub> (Argentine Spanish)  
 CL<sub>M,SG</sub> see.PST.1PL DOM-the boy

‘We saw the boy.’

Along with the derivations in (1) and (2), other effects concomitantly arise based on how the derivation must proceed in each case.

Before I revisit the importance of this distinction for the Galician C-AGR phenomenon, let us review some of the basic notions that have shed much theoretical light on distinguishing true agreement and cliticization.

### 5.2.1 Clitics and affixes under the guise of traditional description

The distinctions held between clitics and affixes have been analyzed in many traditional descriptions, but none of them have influenced the work on separating these phenomena like that of Zwicky & Pullum (1983) (henceforth Z&P). These authors noted six criteria that they claimed distinguished clitics from affixes (3).

- (3) a. Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.
- b. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.
- c. Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.
- d. Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.
- e. Syntactic rules can affect affixed words but cannot affect clitic groups.
- f. Clitics can attach to material already containing clitics, but affixes cannot.

Let us examine each of these diagnostics individually. (3a) seems correct, particularly for a language such as Galician in which the host may be any number of constituents in the left periphery (§3.3.1). (3b) highlights the fact that cliticization is not affected by aspects such as verb tense (i.e., affixes are not always tense variant in the same way clitics are (Nevins 2011, Kramer 2014)). Affixes, on the other hand, may or may not arise based on the combinations with their host.

- (4) a. Hy      brûk-t  
           he      use-3SG  
           ‘He uses.’

- b. Hy brûk-te-Ø  
 he use-PST-Ø

‘He used.’

(Frisian; Dyk 2020 *apud* van Alem 2020:13)

In (4), the 3<sup>rd</sup>-person singular affix *-t* is necessary in the construction in the present in (4a), but it disappears when the past tense affix *-te* is inserted on the verbal stem.

(3c) seems to suggest that affixes, but not clitics, may vary morphophonologically based on their host. As noted by Nevins (2011:959), it would be extremely difficult to find a clitic that does not undergo some degree of allomorphy. This is exceptionally true for Galician clitics as highlighted in (5).

- (5) a. Tócoa

play.PRS.1SG-CL<sub>F,SG</sub>

‘I play it.’

- b. Toca-la

play.INF-CL<sub>F,SG</sub>

‘To play it.’

- c. Vaina tocar

go.PRS.3SG-CL<sub>F,SG</sub> play.INF

‘She is going to play it.’

- d. Non a toco

NEG CL<sub>F,SG</sub> play.PRS.1SG

‘I don’t play it.’

The examples in (5) highlight the allomorphy referenced in §3.3.1 & §3.3.2.2. In (5a), a host ending in a vowel permits the clitic to maintain its ‘first form’, which is morphologically

identical to determiner ‘first forms’. In (5b), the same is true of verbal inflections ending in /r/ or /s/, which give rise to ‘second forms’, which are also applicable in cases of determiner cliticization. In (5c), clitics, but not determiners, may take a ‘third form’ when the verbal inflection ends in a diphthong. Finally, clitics that combine with certain constituents that end in a velar consonant /ŋ/ neutralize this sound as in (5d), forcing syllable reconstruction from \*/noŋ.u/ to /no.no/. Thus, the diagnostic in (3c) seems far from valid.

(3d) highlights the fact that agreement may be able to account for semantic idiosyncrasies, as in the Galician example in (6), whereas clitics are more consistent in this sense.

- (6) As      mulleres      somos      da      parroquia.  
       the      women      be.PRS.1PL      of-the      parish  
       ‘Us women are from the parish.’

As I shall show, the Galician C-AGR morpheme behaves in a similar fashion, as does other clitic phenomena from Zdrojewski & Sánchez (2014), which I review in §5.7. (3e) is expected based on the derivational composition represented in (1) and (2) due to the fact that affixes are valued functional heads that combine with a verbal stem along the clausal spine. Moreover, the fact that instances of true  $\phi$ -feature agreement are affected by notions such as animacy, definiteness, or specificity, whereas cliticization phenomena (e.g. clitic doubling) is not further promotes the divide between the two (Preminger 2019:13). Finally, (3f) prohibits the valuation of a probe on a functional head that has already served as a host for cliticization. If true, this must say something about the derivational steps taken when accounting for the combination of affixes and clitics on the same host.<sup>127</sup> Let us observe a potential case of this in (7).

---

<sup>127</sup> This criterion has not been of much theoretical interest. However, Yuan (2021:157) remarks that these distinctions “are not universal and do not obviously follow from any theoretically grounded differences between the two.” This is a topic that I explore more in depth in §5.7-5.9.

(7) Se lo<sub>i</sub> llevó una caja<sub>i</sub>  
 CL<sub>DAT.3</sub> CL<sub>M.SG</sub> carry.PST.3SG a<sub>F.SG</sub> box<sub>F.SG</sub>

‘She took a box to him.’ (Luján 1987:115 *apud* Mayer 2017:149)

In §5.7, I argue that the morpheme *lo* in (7) may surface an instance of true agreement rather than cliticization based on the fact that it fails to meet several morphosyntactic criteria with respect to its doubled DP (e.g. it does not match the  $\phi$ -features of *una caja* ‘a box’). In this respect, it surfaces closer to its verbal host than other clitic material (the indirect object *se*). The notion behind (3f) will be of particular interest in this chapter due to the unusual behavior of the Galician C-AGR morpheme in §5.7.

Before moving on, it should be noted that recent theoretical work on the distinction between pure agreement and clitic doubling has heavily downplayed the aforementioned criteria. In addition to the claims mentioned above by Yuan (2021), Nevins (2011), for example, asserts that all of these diagnostics should be discarded as viable traits distinguishing affixes and clitics due to their heavy focus on morphophonological characteristics as opposed to their syntactic properties. I review the latter in §5.4-5.6 and apply them to the data presented on Galician C-AGR in §5.3 in §5.7.

### 5.3 A review of Galician C-AGR

In order to apply the theoretical underpinnings of the morphosyntactic properties analyzed in the subsequent sections, let us return to the Galician C-AGR data from Chapter 2.

First, the C-AGR agreement morpheme has the same morphological makeup as first-form accusative clitics in Galician. As they mark agreement with the subject in intransitive-verb constructions, they only agree with the subject in [NUMBER] and [GENDER] (8).

- (8) a. *Velaí o<sub>i</sub> vai el<sub>i</sub>*  
 behold CL<sub>M,SG</sub> go.PRS.3SG he<sub>M,SG</sub>  
 ‘There he goes!’
- b. *Velaquí as<sub>i</sub> veñen as nenas<sub>i</sub>*  
 behold CL<sub>F,PL</sub> come.PRS.3PL the girls<sub>F,PL</sub>  
 ‘Here come the girls!’

However, in addition to third-person subjects, they may agree with subjects that are 1<sup>st</sup>- and 2<sup>nd</sup>-persons, which are not morphologically marked for [GENDER] (9). As expected in a *pro*-drop language such as Galician, the subject is not obligatory (10).

- (9) a. *Velaquí os<sub>i</sub> estamos nós<sub>i</sub>*  
 behold CL<sub>MPL</sub> be.PRS.1PL we  
 ‘Here we are!’
- b. *Velaí as<sub>i</sub> ides vós<sub>i</sub>*  
 behold CL<sub>F,PL</sub> go.PRS.2SG you<sub>PL</sub>  
 ‘There you (PL) go!’
- (10) a. *Velaquí o vai*  
 behold CL<sub>M,SG</sub> go.PRS.3SG  
 ‘There he goes!’
- b. *Velaquí as veñen*  
 behold CL<sub>F,PL</sub> come.PRS.3PL  
 ‘Here they come!’

The subjects co-indexed with this agreement morpheme may also be inanimate (11a) and indefinite (11b).

- (11) a. *Velaí o vai o teu coche*  
 behold  $CL_{M,SG}$  go.PRS.3SG the your car  
 ‘There goes your car!’
- b. *Velaquí as veñen unhas nenas da escola*  
 behold  $CL_{F,PL}$  come.PRS.3PL some girls from-the school  
 ‘Here come some of the girls from school!’

In addition to the present tense, this construction may also be found in the imperfect.<sup>128</sup>

- (12) *Velaí os estabamos cando chegou, cansos e con fame*  
 behold  $CL_{M,PL}$  be.IMPV.1PL when arrive.PRS.3SG tired $_{M,PL}$  and with hunger  
 ‘(And) there we were when he arrived, tired and hungry.’

This construction also boasts the ability to incorporate a dative of interest in combination with the agreement morpheme.

- (13) a. *Velaí cho vai teu neno*  
 behold  $CL_{DAT,2SG}-CL_{M,SG}$  go.PRS.3SG your kid $_{M,SG}$   
 ‘There goes your boy (on you)!’
- b. *Velaquí no-la está*  
 behold  $CL_{DAT,1PL}-CL_{F,SG}$  be.PRS.3SG  
 ‘Here she is (on us)!’

This dative may not be doubled by a lexical DP (14a), as is normally the case with both argumental datives (14b) and non-argumental datives (14c).<sup>129</sup>

<sup>128</sup> As mentioned in §2.4 (fn. 51), Álvarez (2015) claims that this construction may be used in the conditional, as well, which I have claimed should be seen as largely idiolectal.

<sup>129</sup> I use the term *non-argumental dative* loosely here, referring to any of the applicative positions argued for by Cuervo (2003, 2010, 2020) above the LowAppl position within the vP.

- (14) a. Velaquí (\*a min) mo vén (\*a min)  
 behold DAT me CL<sub>DAT.1SG-CL<sub>M.SG</sub></sub> come.PRS.3SG DAT me  
 Intended: ‘Here he comes (on me).’
- b. Déronlle<sub>i</sub> o regalo a Uxía<sub>i</sub>  
 give.PST.3PL-CL<sub>DAT.3.SG</sub> the gift DAT Uxía  
 ‘They gave Uxía the gift.’
- c. A min armouseme<sub>i</sub> a gorda  
 to me fall.PST.3SG-CL<sub>REFL.3-CL<sub>DAT.1SG</sub></sub> DAT fat  
 ‘The shit hit the fan (on me).’ (Lit. ‘The fat one was put together on me.’)

I discuss the reasons that doubling is impossible in §5.7.

The final point to note is that, much like what has been shown for the C-AGR phenomenon in Germanic (e.g. Fuß 2014), this morpheme is illicit when either the complementizer or the verb in T is elided.

- (15) a. Velaí o vai  
 behold CL<sub>M.SG</sub> go.PRS.3SG
- b. \*Velaí o  
 behold CL<sub>M.SG</sub>
- c. \*Vaino  
 go.PRS.3SG-CL<sub>M.SG</sub>  
 Intended: ‘There he goes.’

I discuss the theoretical implications for this in §5.8.

The final aspect worth mentioning is the fact that, much like the subject, this morpheme is optional.

- (16) a. Velaí (o<sub>i</sub>) vai (e<sub>i</sub>)  
 behold CL<sub>M,SG</sub> go.PRS.3SG he<sub>M,SG</sub>  
 ‘There he goes!’
- b. Velaquí (as<sub>i</sub>) veñen (as nenas<sub>i</sub>)  
 behold CL<sub>F,PL</sub> come.PRS.3PL **the girls**<sub>F,PL</sub>  
 ‘Here come the girls!’

This idiosyncrasy will prove particularly important when I analyze the derivational requirements for clitic doubling and agreement in the next section.

It is worth noting that Galician, unlike several of the Northern Italian Dialects (NIDs) or French, does not have subject clitics. Moreover, it does not have clitic doubling in the sense of Anagnostopoulou (2003), *i.a.* Below, I highlight several theoretical observations brought forth and tested by numerous investigators and examine how they match up with the Galician data presented above. I show that the Galician C-AGR phenomenon shows a mix of affix (or pure agreement-based) and clitic (or doubled-argument) characteristics that makes it difficult to classify as one or the other. Ultimately, I will present two possible approaches for the generation of this agreement morpheme, one as a clitic and the other as an instance of pure agreement. In order to do so, I must first lay out theories that have recently paved the way for a purely morphosyntactic analysis of both clitic doubling and true agreement.

#### 5.4 Preminger (2009, 2014, 2019)

As I show in this chapter, the notion that one of these operations is obligatory (agreement) and one is optional (clitic doubling) has been debunked on numerous fronts. This, in turn, has led investigators to question longstanding descriptive generalizations in order to paint a more accurate picture theoretically. This precision has led to the conclusion that many instances of

what was previously thought of as affixal agreement are, in fact, clitic doubling. The retreat from this affix-centered thinking with respect to agreement patterns and their featural make-up has allowed investigators to come to the conclusion that agreement on the verbal stem in many agglutinating languages, for example, is actually an instance of clitic doubling.

- (17) Guraso-e-k<sub>j</sub>            niri<sub>i</sub>    belarritako    ederr-ak            erosi  
 parent(s)-ART<sub>PL</sub>-ERG    me<sub>DAT</sub>    earring(s)    beautiful-ART<sub>PL</sub>    bought  
 d-    i-            zki-            **da**<sub>i</sub>-            **te**<sub>j</sub>  
 3.ABS-  $\sqrt{AUX}$ -    PL.ABS-    1.SG.DAT- 3PL.ERG

‘(My) parents have bought me beautiful earrings.’

As shown in (17), finite auxiliaries in Basque are able to carry agreement morphology that corresponds to the  $\varphi$ -features of each core argument (absolute, ergative, dative). Upon dissecting earlier work by Laka (1996), Preminger (2009, 2014) notes that both the dative *niri* (‘me’) and ergative *guraso-e-k* (‘parents’) arguments are in fact clitic doubled by the morphemes *da* and *te*, respectively. In contrast, he argues that the absolute argument *belarritako ederr-ak* (‘my beautiful earrings’) is an instance of true agreement, with the absolute person marker *d-* prefixed on the auxiliary root and its plural number affix following it. Preminger claims that an inherent property of clitic doubling is that these morphemes are coarse (i.e., they have the same  $\varphi$ -set as their doubled constituents). He refers to this as the *coarseness property of clitic doubling*, outlined in (18).

(18) *The coarseness property of clitic doubling*

If CL<sup>o</sup> is the result of clitic doubling of some noun phrase  $\alpha$ , then CL<sup>o</sup> will reflect the full set of  $\varphi$ -features on  $\alpha$ . (Preminger 2014:51)

The property in (18) starkly contrasts with agreement in the sense that the result of a probe triggering pure agreement is unary, even when multiple probes are hosted on the same functional head (e.g. D'Alessandro 2008).

In more recent work, Preminger (2019) notes several factors regarding clitic doubling (and cliticization more generally) that bring this phenomenon in line with agreement. First and most importantly, regardless of the mechanism by which a DP is doubled, it must, at its core, be an instantiation of Agree. Preminger contends that a functional head establishes a featural relation with a DP bearing suitable  $\phi$ -features, which has two fundamental outputs. First, the derivation must decide if what should move is the DP or simply the  $D^{\circ}$ . This, he claims, may be reduced to a simple principle in (19).

(19) *Minimal remerge*

If  $X^{\circ}$  is moveable, move only  $X^{\circ}$ . (Preminger 2019:28)

Second, if only  $D^{\circ}$  may be moved, this creates a scenario in which the doubling of a DP argument has additional syntactic results, such as creating new antecedents for binding (Anagnostopoulou 2003, *i.a.*).

Preminger offers an additional hypothesis: clitic doubling is an instance of true doubling of the  $\phi$ -features of the DP. Building off of work by Nunes (2004) and Bošković & Nunes (2007) regarding phonological chain reduction, Preminger claims that clitic doubling can be derived from the conditions in (20).

(20) *Conditions on phonological chain reduction of head movement*

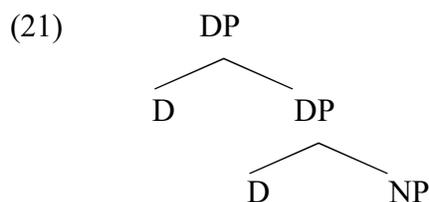
Let  $X^{\circ}$  be a head that undergoes movement to  $Y^{\circ}$ , and let  $\alpha$  be the lower copy of  $X^{\circ}$ .  $\alpha$  will be phonologically deleted iff either of the following conditions is met:

- i)  $\alpha$  and  $Y^{\circ}$  are not separated by a phasal maximal projection (incl. XP)

- ii)  $X^0$  and  $Y^0$  are part of the same extended projection (Grimshaw 2000), and  $Y^0 \alpha$  in the surface structure (i.e., no constituent containing  $\alpha$  but not  $Y^0$  has undergone subsequent movement to a position above  $Y^0$ )

(Preminger 2019:20)

This scenario is an attempt to capture both maximally local head movement (e.g.  $v$  to T or T to C) as well as long distance head movement (e.g.  $v$  to C in Breton; Roberts 2004). This approach clearly differs from the ‘Big DP’ analysis (Torrego 1992, Uriagereka 1995a, Arregi & Nevins 2012, *i.a.*) in which a  $D^0$  moves out of a larger DP with the doubled DP as its sister.



Although Preminger does not address these data directly, this approach may also account for instances of clitic climbing in which both the upper and the lower copies are pronounced as in Chilean Spanish (Belloro 2007).

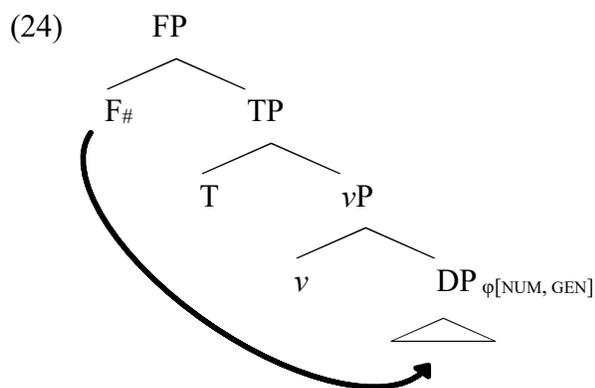
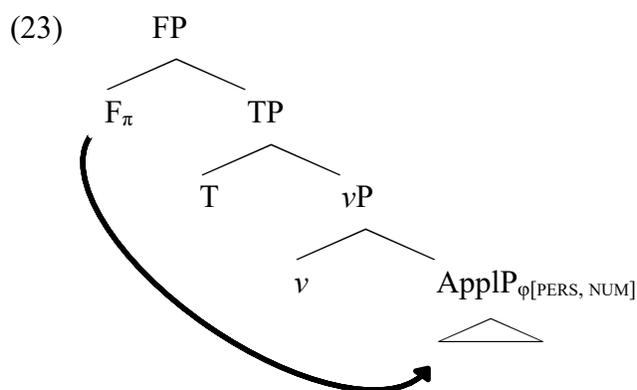
- (22) Lo      quiero      hacerlo  
CL<sub>M,SG</sub> want.PRS.1SG do.INF-CL<sub>M,SG</sub>

As in the examples of clitic doubling in which the clitic is identical to the determiner, the clitic preceding *quiero* (‘I want’) is an identical copy of that in the enclitic position on *hacer* (‘to do’) in (22).

Once we reduce true agreement and cliticization to byproducts of Agree, Preminger notes that the “distinction [between the two] is immaterial if probes always probe for, and copy,  $\phi$ -sets

in their entirety” (2014:51).<sup>130</sup> Accounting for this difference, thus, equates to different probes being specified for specific features (or in a feature geometry, as adopted in Chapter 1 §1.5.1, ex. 31, the point to which the probe is relativized). Several authors note that, due to this coarseness, clitic doubling must be an instance of pronominalization (Harizanov 2014; Kramer 2014; Preminger 2014, 2019; *i.a.*).

Inspired by seminal work by Béjar & Rezac (2003, 2009), Preminger (2014:51) notes, however, that we should expect mismatches in the actual probe and the realization of the agreement as a clitic. This is indeed what we find with both dative (23) and accusative (24) cliticization.



For the dative in (23), the probe may be first or second person, corresponding to the feature for which probe is specified. However, this will be realized along with the [NUMBER] specification of

<sup>130</sup> As I discuss in §5.7, this is not true for all probes. The idea that all probes are atomic, particularly those that deal with agreement on the verb, has been shown to be false, even in Romance.

the applicative argument, be that singular or plural. Therefore, the first-person clitic is able to surface as either *me* or *nos* in Galician depending on the inherent [NUMBER] feature although the probe is solely specified for [PERSON]. The same is seen in (24) for third-person accusative arguments where the probe is specified for [NUMBER], and, as in Romance, a [GENDER] feature may be copied onto the probe, as well.<sup>131</sup>

(25) *Galician clitic probes and their morphological realizations*

- a. Probe: [PERSON] →  $\phi$ -features copied: [PERSON], [NUMBER]
- b. Probe: [NUMBER] →  $\phi$ -features copied: [NUMBER], [GENDER]

Based on these claims, this is what occurs in instances of clitic doubling in Argentine Spanish.

- (26)  $La_i$  oían a Paca<sub>i</sub> / a la niña<sub>i</sub> / a la gata<sub>i</sub>  
CL<sub>F.SG</sub> hear.IMPV.3PL DOM Paca DOM the girl DOM the cat  
 ‘They heard Paca/the girl/the cat.’ (Suñer 1988:396)

In (26), the accusative D° *la* serves as a suitable goal for a probe specified for [NUMBER], resulting in a similar scenario to what we see above for Galician in (25b), here being singular and feminine.

### 5.5 Nevins (2011)

The most important part about Preminger’s approach is that clitic doubling should be considered an instance of true doubling of the  $\phi$ -features of the DP argument. Thus, it would seem that any semblance of doubling without this coarseness requirement is an instance of agreement. This is an important distinction to make and one that has not been addressed specifically in the literature regarding the internal mechanics of Agree-based procedures as we

<sup>131</sup> As we will see in (26), this should not be considered *partial agreement*. This is simply the type of copying mechanism that certain probes have that allows them to probe for one feature but copy the entire  $\phi$ -set.

have seen between true agreement and clitic doubling.<sup>132</sup> With respect to clitic doubling, there is a seemingly parallel situation found in several South American dialects of Spanish (e.g. Andean Spanish in (27)), as well as Peninsular Spanish with dative clitic doubling (28). In these varieties of Spanish, what is often referred to as “clitic doubling” surfaces as a type of invariable object agreement.

(27) Eso también lo<sub>i</sub> mata las plantas<sub>i</sub>  
 that also CL<sub>M.SG</sub> kill.PRS.3SG the<sub>F.PL</sub> plants<sub>F.PL</sub>  
 ‘That kills the plants, too.’ (Zdrojewski & Sánchez 2014:165)

(28) No le<sub>i</sub> pegues (golpes) a los niños<sub>i</sub>  
 NEG CL<sub>DAT.3SG</sub> hit.SUBJ.2SG blow to=DAT the<sub>M.PL</sub> kids<sub>M.PL</sub>  
 ‘Don’t hit the kids.’ (Ausín & Fernández-Rubiera 2021:140)

As these examples are not true instances of clitic doubling based on the diagnostics taken above, one would expect them to behave as agreement markers rather than clitics. That is, where we would expect *las* to be the morpheme in a clitic doubling situation as both *las* and *las plantas* (‘the plants’) bear the same  $\phi$ -features ([PLURAL] and [FEMININE]), the accusative clitic *lo* and its coreferential DP do not. Similarly, we would expect the plural dative *les* to double the lexical DP (*a los niños* (‘to the kids’), as they both share the  $\phi$ -set of 3<sup>rd</sup>-person and [PLURAL]. Instead, there is a mismatch between the  $\phi$ -features based on the fact that *le* is singular but is still co-indexed with the dative-marked DP. I address these phenomena further in §5.7.

---

<sup>132</sup> The literature on distinguishing clitic doubling from true agreement almost always addresses phenomena that conclude the opposite (i.e., what was once considered true agreement is actually clitic doubling). This has been shown not only for Basque (Preminger 2009, Arregi & Nevins 2012), but also for Amharic (Kramer 2014, Baker & Kramer 2018), Inuit (Yuan 2021), Choctaw (Tyler 2019), and Limburgian (van Alem 2020). I take this to be a matter of theoretical revelation rather than an instance of change, particularly as change between these two phenomena has shown to go the opposite way (cliticization > agreement; van Gelderen 2011).

There are, however, a few drawbacks to this analysis based on a further theorem akin to those discussed above in §5.4. For example, Nevins (2011:960) claims that Romance clitics cannot be affixes due to the fact that they may undergo clitic climbing within a variety of constructions such as modals (29) and imperatives (30) in Spanish. That is, features related to affixes are not thought to “percolate” throughout the clause but, instead, have a fixed place along the clausal spine.

- (29) a. **Nos** podemos ver esta tarde  
 CL<sub>1SG</sub> be-able.PRS.1PL see.INF this afternoon
- b. Podemos **vern**os esta tarde  
 be-able.PRS.1PL see.INF-CL<sub>1SG</sub> this afternoon
- ‘We can see one another this evening.’

- (30) a. Empezad a **vestiros**  
 begin.IMP.2PL to dress.INF-CL<sub>REFL.2PL</sub>
- b. Empezaos a **vestir**  
 begin.IMP.2PL-CL<sub>REFL.2PL</sub> to dress.INF
- ‘Begin to dress yourselves!’

In (29a) and (30a), these clitics attach to their host high in the clause, whereas they attach low in (29b) and (30b). In Cardinaletti & Shlonsky (2004), this is possible based on two positions, one above T and one below it, to which a clitic may attach to its host. In contrast, affixes do not have multiple projections on which they may attach to their hosts.

In addition to this observation, Nevins claims that there are three additional characteristics of clitics that are not shared with agreement markers. First, doubled clitics, as in the case of regular syntactic clitics, must be tense-invariant due to the mere fact that they are pronominal

elements.<sup>133</sup> There have been claims (e.g. Kramer 2014 for Amharic, Woolford 2010 for Lakota, Arregi & Nevins 2012 for Basque) that there must only be one instance of true agreement per clause, which would leave us with any object-marking morpheme as a clitic rather than an instance of agreement if we assume that agreement targets subjects cross-linguistically (Kalin 2018).<sup>134</sup> Second, Nevins claims that person complementarity effects are found only with clitics (e.g. the Person-Case Constraint) and never with instances of true agreement.

(31) a. Me / Te lo presentaron  
 CL<sub>DAT.1.SG</sub> CL<sub>DAT.2.SG</sub> CL<sub>M.SG</sub> introduce.PST.3PL

‘They introduced him to me/you.’

b. \*Te me presentaron  
 CL<sub>2.SG</sub> CL<sub>1.SG</sub> introduce.PST.3PL

Intended: ‘They introduced you to me/They introduced me to you.’

The PCC has been identified to have at least four concrete variations, all of them based on the constraints placed on [PERSON] combinations in clitics. The example from Spanish (31a, b), a *Strong PCC* language, shows that in combinations of dative and accusative clitics, the direct object must be 3<sup>rd</sup>-person.

Third, he claims that instances of omnivorous number agreement are only found with 1<sup>st</sup>- and 2<sup>nd</sup>-person arguments.<sup>135</sup>

<sup>133</sup> Preminger (2019:14) notes that syntactic clitics, too, should be considered instances of clitic doubling, as they are thought to always double *pro* when there is no overt doubled DP.

<sup>134</sup> See Oxford (2014) for a rebuttal to this proposal based on evidence from Algonquian.

<sup>135</sup> Possibly the most important implication of this observation is the fact that 3<sup>rd</sup>-person arguments do not generate a ‘Big DP’ structure with a clitic. Although intriguing, testing this is beyond the scope of this chapter. I leave it for further investigation.

(32) a. g- xedav

2.OBJ- saw

‘I saw you/He saw you.’

b. g- xedav- t

2.OBJ- saw- PL

‘I saw you(PL)/We saw you(PL)/He saw you(PL)/We saw you(PL).’

Nevins’ argument is that probes that search for [PLURAL] are omnivorous in that they may agree with either the object or the subject, as in Georgian (32a, b). In §5.8, I show that this is true of complementizer agreement in Nez Perce, as well.

These diagnostics are quite different from those presented in §5.4. However, as they focus on the morphosyntactic properties of the clitic vs. affix distinction rather than the morphophonological properties reviewed in §5.2.1, these should be considered viable theories against which to test the mismatched data from Andean Spanish in (27) and (28), as well as the C-AGR morpheme in Galician.

### 5.6 Kramer (2014)

As mentioned above, Kramer (2014) identifies the phenomenon of object agreement in Amharic which, she claims, is actually clitic doubling despite its several characteristics of true agreement. Perhaps the most convincing of all the arguments for this being a case of clitic doubling is the fact that it is optional.<sup>136</sup>

(33) Almaz tamari-w-in ayy-ätʃʃ(-iw)

Almaz.F student-DEF.M-ACC see-3F.SG.SUBJ-(3M.SG.OBJ)

‘Almaz saw the male student.’

<sup>136</sup> It is worth noting that Preminger (2019:13) contends that clitic doubling is normally not optional. However, the optionality of clitic-like morphemes in constructions such as (30) has always pointed toward instances of clitic doubling rather than cases of agreement.

Moreover, similarly to much of what has been proposed for clitic doubling in Argentine Spanish (34), the Amharic object marker is subject to the features of the DP. In the case of Amharic, the object marker may only cross-reference specific DPs (35).

- (34) a. Lo vi a Eduardo  
 CL<sub>M.SG</sub> see.PST.1SG DOM Eduardo  
 ‘I saw Eduardo.’
- b. (\*Lo) vi un lápiz  
 CL<sub>M.SG</sub> see.PST.1SG a pencil  
 ‘I saw a pencil.’
- (35) a. Almaz doro wät’-u-n bäll-atʃtʃ-iw  
 Imaz.F chicken stew-DEF.M-ACC eat-3F.SG-**3M.SG.OBJ**  
 ‘Almaz ate the chicken stew.’
- b. \*Almaz doro wät’ bäll-atʃtʃ(\*-iw)  
 Almaz.F chicken stew eat-3F.SG-**3M.SG.OBJ**  
 ‘Almaz ate chicken stew.’

Preminger (2019:13) claims that clitic doubling is in fact not predicated on the DP bearing particular features such as [+DEFINITE], [+ANIMATE], or [+SPECIFIC] in order to control clitic doubling but, instead, that these properties are simply the result of movement of the DP to a particular position from which then clitic doubling is possible (or obligatory).<sup>137</sup> Although this topic in particular is far from clear, Kramer claims that the fact that the data in (35) are subject to  $\varnothing$ -feature restrictions seems to point to the fact that this phenomenon, like agreement, is sensitive to the featural makeup of the argument.

<sup>137</sup> This entails a connection between clitic doubling and projections related to object shift, as proposed by many investigators (in particular, López 2012). As this has no direct bearing on my accounting for Galician C-AGR, I leave it aside for now.

The last situation in which the Amharic object marker follows a closer path toward agreement rather than clitic doubling is the fact that it is limited to locality restrictions. That is, it cross-references the highest internal argument (the Goal in ditransitive clauses and the Theme in transitive clauses), which seems to follow the line of thinking that agreement with *v* is subject to locality constraints.

- (36) a. Girma lä-Almaz<sub>i</sub> mäs'hauf-u-n sät't'-at<sub>i</sub>  
 Girma.M DAT-Almaz.F book-DEF.M-ACC give-(3M.SG.SUBJ)-**3F.SG.OBJ**
- b. \*Girma lä-Almaz<sub>i</sub> mäs'hauf-u-n<sub>i</sub> sät't'-ä-w<sub>i</sub>  
 Girma.M DAT-Almaz.F book-DEF.M-ACC give-(3M.SG.SUBJ)-**3M.SG.OBJ**
- ‘Girma gave the book to Almaz.’

This type of locality restriction is not seen with true clitic doubling, as both Theme and Goal arguments are able to (and sometimes must) undergo doubling as in Romance, as in the Romanian examples from Cornilescu (2020:119) in (37).

- (37) a. \*Comisia **le-a** repartizat pe mai mulți  
 board.the CL<sub>DAT.3PL</sub>-have.PRS.3SG assign.PRTP DOM more many  
 medici rezidenți unor foști profesori de-ai lor  
 medical residents some<sub>DAT.PL</sub> former professors of-GEN theirs  
 ‘The committee assigned several residents to some former professors of theirs.’
- b. Comisia **i** **l-a** repartizat pe fiecare  
 board.the CL<sub>DAT.3SG</sub> CL<sub>3SG.M</sub>-have.PRS.3SG assign.PRTP DOM each  
 resident unei foste profesoare a lui  
 resident some<sub>DAT.3SG.F</sub> former professor GEN his  
 ‘The committee assigned every medical resident to some former professor of his.’

In (37a), the dative argument *unor foști profesori* ('some former professors') is the only argument doubled (by the clitic *le*), resulting in ungrammaticality. A simple repair strategy is doubling the accusative DP *fiecare resident* ('each resident') in addition to the doubled dative as in (37b). Clitic doubling, thus, is not subject to locality constraints as one would expect in instances of true agreement. Instead, both internal arguments may undergo doubling in some varieties, showing the lack of competition for an agreement slot as in the Amharic example in (36).

Having reviewed the three accounts above, we can now apply the proposed morphosyntactic diagnostics to the Galician C-AGR data in §5.3: the coarseness property of clitic doubling (Preminger 2014), tense invariance (Nevins 2011), person complementarity (Nevins (2011), optionality (Kramer 2014), and property restrictions of doubled DPs (Kramer 2014).

### 5.7 Theoretical underpinnings of Galician C-AGR

At first glance, it appears that the C-AGR morpheme *a/os* in (38a, b) is a viable case of clitic doubling of the subject due to the fact that it agrees in [NUMBER] and [GENDER] with the subject.

- (38) a. Velaquí a vén (ela)  
 behold CL<sub>F.SG</sub> come.PRS.3SG she  
 'Here she comes!'
- b. Velaí os están (os nenos)  
 behold CL<sub>M.PL</sub> be.PRS.3PL the boys  
 'There are the kids!'

In particular for the condition in (20), repeated below in (39), this seems to be an optimal strategy for deriving the Galician C-AGR morpheme.

(39) *Conditions on phonological chain reduction of head movement*

Let  $X^{\circ}$  be a head that undergoes movement to  $Y^{\circ}$ , and let  $\alpha$  be the lower copy of  $X^{\circ}$ .  $\alpha$  will be phonologically deleted iff either of the following conditions is met:

- i)  $\alpha$  and  $Y^{\circ}$  are not separated by a phasal maximal projection (incl. XP)
- ii)  $X^{\circ}$  and  $Y^{\circ}$  are part of the same extended projection (Grimshaw 2000), and  $Y^{\circ}$   $\alpha$  in the surface structure (i.e., no constituent containing  $\alpha$  but not  $Y^{\circ}$  has undergone subsequent movement to a position above  $Y^{\circ}$ )

(Preminger 2019:20)

The problem arises when the subject is not 3<sup>rd</sup>-person. In these instances, the condition in (39) overgenerates, as we would expect the clitic to also bear a 1<sup>st</sup>- or 2<sup>nd</sup>-person feature, contrary to fact (40a, b).

- (40) a. \*Velaquí nos estamos  
 behold CL<sub>1PL</sub> be.PRS.1PL  
 Intended: ‘Here we are!’
- b. \*Velaquí vos estades  
 behold CL<sub>2PL</sub> be.PRS.1PL  
 Intended: ‘Here you are!’

Therefore, we may quickly conclude that the C-AGR morpheme cannot be a true doubled morpheme of the subject per the *coarseness property of clitic doubling* (18).

Turning to the diagnostics from Nevins (2011), I have shown that the subject agreement morpheme is attested in both the present and the imperfect (41), meaning that this is a tense-invariant phenomenon.

- (41) Velaí os estamos/estabamos na beira do río  
 behold CL<sub>M,PL</sub> be.PRS.1PL/be.IMPV.1PL on-the bank of-the river  
 ‘Here we are/were on the riverbank.’

Second, this morpheme does not seem to be limited in any way by any person or person-combination restrictions, as shown by its ability to combine with a dative specified for any person (42).

- (42) Velaí mo / cho / llo / no-lo /  
 behold CL<sub>DAT.1SG-CL<sub>M,SG</sub></sub> CL<sub>DAT.2SG-CL<sub>M,SG</sub></sub> CL<sub>DAT.3SG-CL<sub>M,SG</sub></sub> CL<sub>DAT.1PL-CL<sub>M,SG</sub></sub>  
 vo-lo / lle-lo vai  
 CL<sub>DAT.2PL-CL<sub>M,SG</sub></sub> CL<sub>DAT.3PL-CL<sub>M,SG</sub></sub> go.PRS.3SG  
 ‘There he goes (on me/you(SG)/her/us/you(PL)/them)!’

Third, as this morpheme is restricted to the 3<sup>rd</sup>-person, it predictably does not show specifications/restrictions for omnivorous number agreement.

As for the diagnostics in Kramer (2014), much like the Amharic object clitic, the Galician C-AGR morpheme is optional. Grammaticality is not predicated on its surfacing in the clause.

- (43) Velaquí (os) van  
 behold CL<sub>M,PL</sub> go.PRS.3PL  
 ‘Here they go!’

Unlike the object marker in Amharic, the Galician C-AGR morpheme does not seem restricted to properties of the DP such as animacy or specificity, as subjects lacking both of these characteristics

- (44) Velái o vai un bote po-la ría  
 behold CL<sub>M.SG</sub> go.PRS.3SG a dinghy by-the estuary  
 ‘There goes a dinghy along the estuary!’

The last observation by Kramer, that the object marker is subject to locality conditions, is not a factor in the Galician C-AGR constructions due to the fact that i) it is monoclausal and ii) it never participates in transitive constructions.

Concluding the observations of the morphosyntactic diagnostics proposed by these three authors, we may see that the Galician C-AGR morpheme does indeed have properties related to cliticization as well as those related to pure agreement.

*Table 7. Morphosyntactic diagnostics results of Galician C-AGR*<sup>138</sup>

|                                                | Cliticization | Agreement |
|------------------------------------------------|---------------|-----------|
| True clitic double (Preminger 2014)            |               | ✓         |
| Tense-invariant (Nevins 2011)                  | ✓             |           |
| Person complementarity (Nevins 2011)           |               | ✓         |
| Optionality (Kramer 2014)                      |               | ✓         |
| Doubled DP property restrictions (Kramer 2014) |               | ✓         |

Based on the observations in Table 7, we see that all but one of the relevant morphosyntactic diagnostics point to the Galician C-AGR morpheme being an instance of agreement. There are, however, a few additional points worth noting. First, although the condition on phonological reduction (39) overgenerates for this phenomenon, a ‘Big DP’ analysis is not out of reach from being a viable source in deriving the referentiality between the C-AGR morpheme and the

<sup>138</sup> Due to the properties of this construction, I do not include the phenomena of omnivorous number (Nevins 2011) and locality restrictions (Kramer 2014) in Table 7.

subject DP. In fact, I shall argue for this analysis, or some derivative of it, as a potential option if one were to conclude that Galician C-AGR morpheme is indeed a clitic. Second, the person complementarity in Nevins (2011) stems from observations regarding the Person-Case Constraint (PCC) in languages that exhibit *Strong PCC* effects like Spanish (45).

- (45) \*Te me mandaron  
 CL<sub>2.SG</sub> CL<sub>1SG</sub> send.PST.3PL

Intended: ‘They sent you to me/me to you.’

I claim that Galician is a true *Weak PCC* language based on the data below (46), which has consequences for the head that probes these internal arguments (what has been proposed thus far is *f*; cf. §3.3.3).

- (46) a. Presentaron-**che-me**  
 introduce.PST.3PL-CL<sub>DAT.2SG</sub>-CL<sub>1SG</sub>  
 ‘They introduced me to you.’
- b. Presentaron-**te-me**  
 introduce.PST.3PL-CL<sub>2SG</sub>-CL<sub>DAT.1SG</sub>  
 ‘They introduced you to me.’

Following the specification(s) of probing heads as in Deal (2015, 2020), Agree takes place in three stages: *search*, *copying*, and *valuation*. Between these stages, however, lies an inherent property of Agree, what Deal calls *interaction* and *satisfaction*. A probe *interacts* with its syntactic environment (i.e., its c-command domain), but it need not be *satisfied* by the first goal it encounters. This is formalized in (47).

- (47) A probe may interact with feature set F even if it may only be satisfied by feature set G, where  $F, G \subseteq \varphi$  (the set of  $\varphi$ -features) and  $F \neq G$ . (Deal 2015:2)

For the approach to the four main PCC types in Deal (2020), it is the *satisfaction* of the probe that refuses further cliticization of internal arguments in the *Strong PCC*.<sup>139</sup> Her proposal for the *Weak PCC*, however, identifies the probing head as *insatiable* with respect to its *satisfaction* condition (i.e., a 1<sup>st</sup>-person direct object clitic does not rule out the incorporation of a 2<sup>nd</sup>-person indirect object clitic).

This may seem irrelevant for the phenomenon at hand considering the fact that the C-AGR morpheme would double the external argument rather than an internal argument. However, considering clitics in Western Iberian are probed by *f*, any and all clitics must adhere to some sort of featural constraint set by the probing head. If we consider the syntax blind to notions such as ‘internal’ or ‘external’ argument (Preminger 2014, Kalin 2018, *i.a.*), then functional heads should blindly agree with DPs bearing features for which they are specified. Therefore, if the Galician C-AGR morpheme is a clitic, then it must be probed by *f*. Due to the fact that *f* in Galician has no *satisfaction* constraint (its probe being *insatiable*), regardless of the dative+C-AGR combination, the result will be grammatical. Although a complex way to approach this phenomenon, it is important to address the internal mechanisms (particularly for this phenomenon as a clitic, its combination with other clitics) by which the derivation may arise. Although we see in (42) that the Galician C-AGR phenomenon is not subject to person complementarity restrictions, I claim that this is largely due to the fact that the probe on *f* is liberal with respect to the clitics with which it may agree, rendering Nevins’ person complementarity argument a moot point here.

Before moving on to how the construction under investigation fits in the with C-AGR phenomenon more generally, I wish to address one of the criteria proposed by Zwicky & Pullum

---

<sup>139</sup> The *Strong PCC*, in contrast to the *Weak PCC*, essentially states that the direct object must be 3<sup>rd</sup>-person, ruling out strings as in (44).



(49) Eso también lo<sub>i</sub> mata las plantas<sub>i</sub>  
 that also CL<sub>M,SG</sub> kill.PRS.3SG the<sub>F,PL</sub> plants<sub>F,PL</sub>

‘That kills the plants, too.’ (Zdrojewski & Sánchez 2014:165)

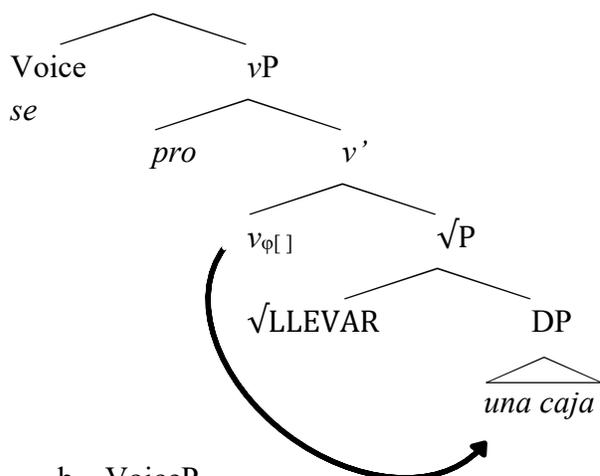
The invariant clitic *lo* (‘it’) in (49) seems to function as an agreement marker for any coreferential object DP, irrespective of the features of the latter. Béjar (2003) referred to the internal mechanisms of true object agreement of this type as being triggered by a *flat probe*. As explained more thoroughly in Béjar & Rezac (2009:44), this is a probe that will agree with any goal DP regardless of features. A byproduct of this type of agreement is the invariable and default  $\varphi$ -feature realization on the valued head. When we find this type of default agreement in combination with another clitic, we should expect it closer to the host than a moved D°. This is borne out in Argentine Spanish (50), where the default clitic *lo* appears closer to the (verb) host than the clitic *se* does.

(50) Se lo<sub>i</sub> llevó una caja<sub>i</sub>  
 CL<sub>REFL,3</sub> CL<sub>M,SG</sub> carry.PST.3SG a<sub>F,SG</sub> box<sub>F,SG</sub>

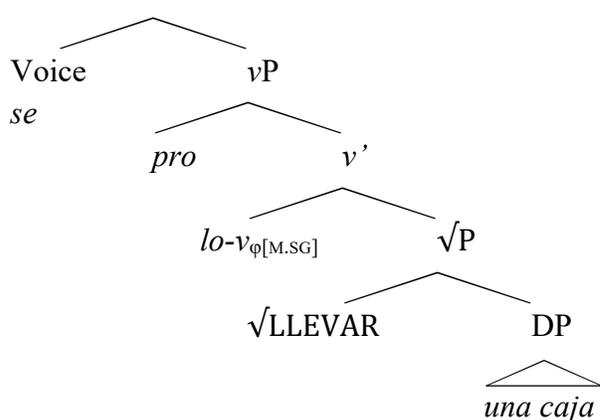
‘She took a box to him.’ (Luján 1987:115 *apud* Mayer 2017:149)

We may imagine that the probe that once triggered movement of a true doubled clitic, as in Argentine Spanish, was simplified with respect to its featural specifications and, in turn, its valuation criterion (i.e. strict valuation with no pronominal D° movement). Thus, we would expect valuation of this sort to take place before cliticization of the reflexive pronoun *se*.

(51) a. VoiceP



b. VoiceP



In (51a),  $v$  probes for some DP goal with no feature specifications set. Once it finds *una caja* ('a box') in its c-command domain, the result is  $[lo-v^0]$  (51b), which undergoes further head movement to T at which point *se* is probed.<sup>141</sup>

Applying this same idea to the Galician C-AGR phenomenon, let us imagine that the C-AGR is, in fact, an instance of true agreement and that valuation takes place on  $f$ . Although the probe for C-AGR cannot be a *flat probe* as argued for above in Spanish, it reflects the same features we saw in (25b) for accusative arguments in Galician, which I have shown is a subset of the features borne by the coreferential DP. Thus, we may assume in a Galician C-AGR clause that  $f$  would be

<sup>141</sup> Although I argued for  $f$  as the landing site for clitics in Spanish in §3.3.3, this does not change the order in which the probe on  $v$  is valued. Regardless of if  $v$ , T, or  $f$  is the functional head on which the probe is found, due to the fact that Theme arguments are probed for first (Deal 2020 and references therein) we should expect the same outcome regardless of where the clitics move.

endowed with a probe that first sought out [NUMBER], whose morphological result of Agree with the subject DP is realized as the same morpheme as accusative clitics.<sup>142</sup> The question arises, then, can the same functional head probe undergo multiple instances of Agree with the result giving rise to both movement and valuation? I claim that this is indeed what must take place if we are to assume that the Galician C-AGR morpheme arises due to true agreement. In fact, this has been proposed for agreement on T cross-linguistically (Sigurðsson & Holmberg 2008 for Icelandic, D’Alessandro 2007 for Italian, *i.a.*), although there the probing order is reversed (i.e., [PERSON] > [NUMBER]).

Piggybacking on the first point above, the second observation worth mentioning is that there is something about 3<sup>rd</sup>-person goals that seem to trigger these effects. In fact, Romero (2019) claims that the variation seen cross-linguistically is found exclusively in the 3<sup>rd</sup>-person paradigms. We have seen this in both the Andean Spanish and Galician C-AGR phenomena, and this seems to be the case for the *le*-for-*les* phenomenon in Fernández-Rubiera & Ausín (2021), as well. In this instance, we may better capture the proximity issue mentioned above. Let us imagine that for (52), there is a *flat probe* that searches for and agrees with a 3<sup>rd</sup>-person dative DP.<sup>143</sup>

- (52) No    *le*<sub>i</sub>        des            regalos    a            los        niños<sub>i</sub>  
           NEG    CL<sub>DAT.3SG</sub>    give.SUBJ.2SG    gifts        to=DAT    the<sub>M.PL</sub>    kids<sub>M.PL</sub>  
           ‘Don’t give the kids gifts.’

When the direct object *regalos* (‘gifts’) is incorporated as a clitic (*los*), combining with the dative affix *le*, the accusative clitic is closer to the verbal host than the dative. As a consequence, this

<sup>142</sup> This is revealing with respect to morphemes that undergo internal change (i.e., how they are realized in the syntax proper) but show no immediate surface-level change. I return to this point in §5.10.

<sup>143</sup> As I mentioned in fn. 140, it does not matter where the probe is located (*v*, *T*, *f*) when accounting for the cross-linguistic data presented in this chapter.

morphological number distinction of the indirect object disappears with the realization of the dative as *se*.<sup>144</sup>

- (53) No    *se*<sub>i</sub>    los    des                    a            los    niños<sub>i</sub>  
       NEG   *se*    CL<sub>M,PL</sub> give.SUBJ.2SG    to=DAT    the<sub>M,PL</sub> kids<sub>M,PL</sub>  
       ‘Don’t give them to the kids.’

While this may seem like an *ad hoc* proposal due to the ambiguity around *se* in the literature (Nevins 2007), it seems to me a concomitant of the fact that no other Romance language shows the type of phenomenon in (52) that may be undone or simply disappears with the incorporation of another argument in between the probe and its host.

To summarize, we have seen that there are morphosyntactic diagnostics that allow for more precise theoretical predictions than the observations by Zwicky & Pullum. Although these tests paint clear a picture for many languages cross-linguistically, I have shown that there are phenomena that seemingly fit into both the affix and clitic categories depending on the diagnostic. In the subsequent sections, I discuss complementizer agreement and the theoretical approaches that have been put forth in order to account for the variation found amongst natural languages (§5.8). In particular, I focus on data outside of the well-attested Germanic varieties. I apply the Galician C-AGR data from §5.7 to the three formal approaches: post-syntactic insertion, agreement, and cliticization (§5.9). Finally, I briefly discuss the diachronic change from clitic to affix based on the diagnostics and observations I have made in this chapter (§5.10).

### 5.8 Complementizer Agreement: How much cross-linguistic variability is there?

Alongside the numerous languages that show complementizer agreement, it has been shown that C-AGR phenomena show agreement marking for a plethora of different constituents. Nez

---

<sup>144</sup> When combined with an accusative clitic, both *le* (3<sup>rd</sup>-person singular) and *les* (3<sup>rd</sup>-person plural) are realized as *se*.

Perce, for example, may show agreement for both the object and the subject via more than one feature.

- (54) **ke-pe-m-ex** kaa *pro*<sub>SUBJ</sub> cewcew-tée'nix *pro*<sub>OBJ</sub>  
**C-PL-2-1** then *pro*<sub>1PL</sub> telephone-TAM *pro*<sub>2SG</sub>  
 'When we call you(SG)'

In (54), the  $\bar{A}$  particle *ke* shows agreement with three individual features from the two arguments in the clause: the [PLURAL] and [SPEAKER] from the external argument (represented here as *pro*) and [ADDRESSEE] from the internal argument (also represented as *pro*).

A unique example of this phenomenon comes from Wolof. This language boasts both matrix- (55) and subordinate-clause (56) C-AGR on a select number of C-oriented elements.<sup>145</sup> The C-AGR morpheme does not represent typical  $\phi$ -features of DP constituents but, instead, one of the fifteen noun classes found in the language.<sup>146</sup>

- (55) a. **K-u** togg ceeb ak jën?  
**CM-u** cook rice and fish  
 'Who cooked rice and fish?'
- b. **Y-u** jigéen j-i togg?  
**CM-u** woman CM-DEF.PROX cook  
 'What(PL) did the woman cook?'
- c. **F-u** jigéen j-i togg-e ceeb ak jën?  
**CM-u** woman CM-DEF.PROX cook-LOC rice and fish  
 'Where did the woman cook fish and rice?'

<sup>145</sup> I only use examples of the *u*-forms for expository purposes. I refer the interested reader to Torrence (2012) for a full list of these C-oriented functional items.

<sup>146</sup> The *u*-forms may agree with twelve of the fifteen noun classes (Torrence 2012:1151), an example of the variation with respect to how many classes may agree with each C-oriented element.

- d. **L-u** Isaa ubbé-é bunt y-i?  
**CM-u** Isaa open-INSTR door CM.PL-DEF.PROX

‘What did Isaa open the doors with?’

- (56) a. (u-j) yàmbaa **j-u** ñu tóx  
 indef-CM marijuana **CM-u** 3PL smoke

‘Some marijuana that they smoked’

- b. **S-u** fa Ayda gis-óón Dudu sax, nyu-wu-kó  
**CM-u** LOC Ayda see-PAST Dudu even greet-NEG-3SG<sub>OBJ</sub>

‘Even though Ayda saw Dudu there, she did not greet him.’

(Torrence 2012:1151-1158)

In (55a), the class marker *k-* references the (singular) human in the clause, which is interpreted as the subject in this case. In (55b), *y-* represents the plural class marker that stems from the elided direct object. In (55c), *f-* references the locative adjunct class, referencing where the action was taken place. In (55d), the single instrumental class *l-* refers to the object used by the subject in order to perform the action. C-AGR in Wolof is also found in subject relatives (56a), as well as clauses which Torrence (2012:1158) refers to as ‘even though’ and ‘instead’ clauses (56b).

Lubuksus, a Bantu language, shows possibly the most unorthodox case of C-AGR due to the fact that the goal that values the [PERSON] slot on the subordinate C is in the matrix clause.

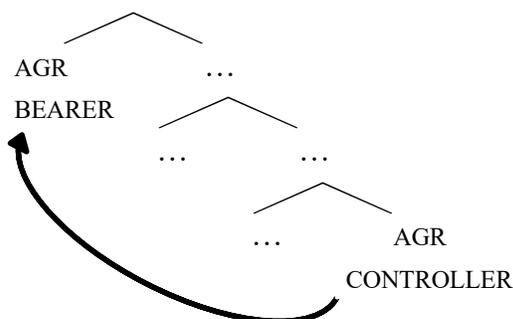
- (57) **Ba-ba-ndu** ba-bol-el-a Alfredi **ba-li** a-kha-khil-e  
**2-2-people** 2SA-say-APPL-FV 1Alfredi **2-COMP** 1SA-FUT-conquer-FV

‘The people told Alfred that he will win.’

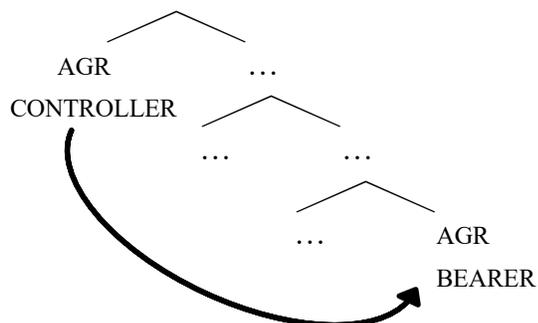
(Diercks, van Koppen & Putnam 2020:353)

Due to these data in (57), it has been proposed that Agree may probe upward (58a), the inverse of what was originally proposed for the Probe-Goal model (5b) (Chomsky 2000, 2001).

(58) a. *Downward valuation (i.e. upward probing)*



b. *Upward valuation (i.e., downward probing)*



(Preminger & Polinsky 2015:1-2)

The mechanics in (58a), supported by numerous investigators on independent grounds (cf. Wurmbrand 2011; Bjorkman & Zeijlstra 2019; Zeijlstra 2012), have the probe in a c-commanded position with respect to the goal that values it, contrary to what is typically assumed (58b).<sup>147</sup> More recent work by Diercks et al. (2020), however, has shown that the data from Bantu may be explained by assuming an anaphoric  $\phi$ -feature valuation account in which the features on C are valued after Internal Merge with the  $\nu$ P and Agree with the matrix subject.

<sup>147</sup> Although it is not the underlying reason for adopting the valuation mechanics in (58a), most of the aforementioned work has been centered around accounting for semantic notions within the syntax. As Preminger (2013) and Preminger & Polinsky (2015) have noted, this model overgenerates and makes incorrect predictions for many well-understood phenomena. Adopting a traditional model as in (58b), I leave further reasoning regarding this debate aside.



That there may be instances of C-AGR affixes stemming from a pronominal origin as well as those with a verbal genesis (van Koppen 2017) has led to numerous proposals for how these affixes arise. In §5.9, I review three main theoretical proposals in order to determine which best fits the behavior of Galician C-AGR: agreement, post-syntactic insertion, and cliticization.<sup>148</sup>

### 5.9 Theoretical approaches to C-AGR

In this section, I review the three main approaches in order to account for C-AGR cross-linguistically. I first discuss post-syntactic insertion (§5.9.1), then agreement (§5.9.2), and finally cliticization (§5.9.3).

#### 5.9.1 C-AGR and post-syntactic $\phi$ -feature insertion

The post-syntactic insertion account as put forth by Fuß (2014) consists of two main operations, one syntactic and one post-syntactic. This is shown in (61).

(61) *Syntactic computation*

- a. AGREE: T probes for and agrees with the subject

*Post-syntactic computation*

- b. COPY: The  $\phi$ -set on T is copied
- c. FEATURE INSERTION: The copied  $\phi$ -set on T is inserted on C
- d. VOCABULARY INSERTION: The realization of the same exponent on C as on T

Fuß claims that this operation is able to account for all types of C-AGR phenomena, such as those that show adjacency effects when there is intervening material between T and C, as in Hellendoorn Dutch (62) and first conjunct agreement (FCA) as in Limburgian (63).

---

<sup>148</sup> I leave aside the anaphoric  $\phi$ -feature valuation account described for the Bantu data due to the fact that Galician C-AGR is monoclausal and the goal that values the probe never appears above the C-AGR morpheme.

(62) dat / \*darr-e **zölf**s wiej de wedstrijd wint  
 that / that-**1PL** **even** we the game win  
 ‘... that we even win the game...’ (van Koppen 2012:161)

(63) Ich denk de-**s** [toow en Marie] kump  
 I think COMP-**2SG** you<sub>SG</sub> and Marie come<sub>PL</sub>  
 ‘I think that you and Marie will come.’  
 (Haegeman & van Koppen 2012:443)

In (62), *zölf*s (‘even’) intervenes in between the complementizer *darr* and the subject *wie* (‘we’). Although *zölf*s does not have a  $\phi$ -set, an intervention effect still occurs. In (63), although the verbal agreement is plural, the complementizer only agrees with the first DP constituent *toow* (‘you’) in the conjunct phrase.

In agreement-based accounts of C-AGR (Ackema & Neeleman 2004, Carstens 2003, Gruber 2008, *i.a.*), data in (62) has been of considerable interest due to the fact that intervening material may not render the sentence ungrammatical but, instead, only blocks the probe on the complementizer element from overtly agreeing with a goal. Additionally, data in (63) showing a bifurcation between the specification of the agreement on C and the subject agreement on T. Accounting for languages like Limburgian and Tegelen Dutch (van Koppen 2005) has considerable consequences for feature inheritance (Chomsky 2008, Ouali 2008, Richards 2007, *i.a.*) due to the fact that C must not only merge with two independent sets of  $\phi$ -features but must know which  $\phi$ -set to keep and which to give to T.

Fuß' post-syntactic approach relies on a PF with proper-syntax abilities. For example, this author contends that FCA phenomena may be accounted for by an impoverishment rule that erases the second  $\varphi$ -set in the ordered pair of a conjunction constituent as in (64).<sup>149</sup>

$$(64) \quad \langle [\varphi 1], [\varphi 2] \rangle \rightarrow [\varphi 1 / \{ \_ [\varphi 1] \}]$$

The rule in (64) states that the second  $\varphi$ -set on T will be deleted on C if the minimal prosodic domain contains a  $\varphi$ -set identical to the first member of the ordered pair. Fuß derives the specifics of this rule based on data from Bavarian.

- (65) a. *dass-st* [du# und d'Maria]# an Hauptpries gwunna hoab-ts  
 COMP-2SG you and the.Mary the first.prize won have-2PL
- b. *dass-ts* #[du und d'Maria]# an Hauptpries gwunna hoab-ts  
 COMP-2PL you and the.Mary the first.prize won have-2PL
- '... that you and Mary have won first prize...'

Where # marks a prosodic boundary, Fuß claims that (65a) is only grammatical where *du* ('you') is prosodically divided from the rest of the coordinated subject. When it is not (65b), however, agreement on C bears the same  $\varphi$ -features as on T.

As highlighted recently (e.g. Preminger 2018), there is something to be said about a PF-branch that possesses the same power as the syntax proper. Primarily, it creates an impossible boundary across which to determine what must be accounted for in the syntax and what may be left to the post-syntactic branch. Although I do not endorse this view of the post-syntactic portion of the derivation, I shall show that applying a post-syntactic insertion model to Galician C-AGR falls short on numerous levels and requires post-syntactic repair strategies that go far beyond those proposed by Fuß for accounting for the C-AGR phenomena with FCA. I also show that

<sup>149</sup> Fuß (2014:73) recognizes the fact that, because Vocabulary Items in the Distributed Morphology model are not specified for ordered pairs of  $\varphi$ -features, accounting for data as in (10) is difficult without resorting to post-syntactic repairs.

while making a case for Galician C-AGR via post-syntactic insertion involves numerous repair strategies that are overly complex, there is no way in which a C-AGR in Nez Perce may be accounted for in this manner.

### 5.9.1.1 Galician C-AGR as post-syntactic insertion

Let us take the sentence in (66) as our trial example for this phenomenon.

- (66) Velaquí as estamos *pro*  
 behold CL<sub>F.PL</sub> be.PRS.1PL *pro*  
 ‘Here we are!’

I use *pro* in (13) for expository reasons, although this could be replaced with a personal pronoun or an inclusive full DP as in *os alumnos* (‘us students’, lit. ‘the students’). Let us recall that T must provide  $\phi$ -features to C after the former has been valued by the subject. As T in Romance is specified for [PERSON] and [NUMBER], the probe on T searches independently for these features. Due to the fact that the only argument in the clause is the subject, the probes on T are valued by the lone DP goal.

- (67)
- 
- ```

  graph TD
    TP --> T["T_{\phi[ ]}"]
    TP --> vP["vP"]
    vP --> pro["pro_{\phi[1, PL, M]} ..."]
  
```

Once C is merged and the derivation proceeds to the PF branch, the result is both C and T containing the ϕ -set [1, PL]. Stopping here alone would render the C-AGR morpheme equivalent to the clitic *nos* (‘us’), which I have explained is not possible due to the fact that Galician does not have clitic doubling (neither for the object or subject). How must the derivation repair this? I claim that there must be not one but two repair strategies. First, there must be a rule that states

[PERSON] must be removed from the φ -set bequeathed to C due to the fact that while T may be specified for this feature, the C-AGR morpheme never is.¹⁵⁰

(68) *Person stripping on C*: $\varphi: \langle [1, \text{PL}] \rangle \rightarrow \varphi: \langle [\text{PL}] \rangle$

The removal of [PERSON] leaves the φ -set with a lone [PLURAL] feature. As this has no individual morphosyntactic realization within a φ -feature bundle (i.e., there is no φ -set in which [NUMBER] is the only feature), a second repair strategy is required in which PF inserts a [GENDER] feature.¹⁵¹ We can immediately discredit a ‘default’ repair strategy in which the relevant feature inserted is considered the default option, as this would inherently require all Galician C-AGR marking to be [MASCULINE], contrary to what we see in (66). The primary issue with Fuß’ treatment of FCA in order to explain the Galician data is that there is no precedent for inserting missing φ -features on C independently of those already found on T. In order for the repair to work, it must search for a [GENDER] feature from some relevant φ -set as in (69).

(69) *Find φ*

- a. Search for a feature not related to [PERSON] or [PLURAL]
- b. Copy feature to C
- c. C: $\varphi \langle [\text{PL}] \rangle \rightarrow \varphi: \langle [\text{PL}, \text{FEM}] \rangle$

Considering the subject DP is the only one left, C must copy its [GENDER] feature in order to prevent ungrammaticality.

An operation such as that in (69) is contrived to such an extent that it seems highly implausible that post-syntactic insertion could be a procedure by which we can account for

¹⁵⁰ As discussed in §5.7, even in instances in which the subject is 3rd-person, the probe only searches for [NUMBER], with no [PERSON] feature specified. Thus, the morphological 3rd-person appearance of the C-AGR morpheme is a byproduct of the specification of the probe and not relevant to the specification of a 3rd-person subject.

¹⁵¹ Possibly the biggest question to answer is how PF would know this is the feature necessary in order for the derivation to converge. This may be some sort of inherent process of elimination (*Insert the only feature not remaining or having been deleted*).

Galician C-AGR. More generally, regardless of the formal mechanism by which this occurs, there must be some form of ‘search procedure’ in order to find the missing feature. Once we assume that there is a search mechanism in the post syntax, this brings the PF branch fully in line with the syntax proper, further muddying the waters of what is traditionally thought to distinguish them. The question then becomes why not assign all search, valuation, copying, etc. procedures to the post syntax? Taking a post-syntactic insertion approach, thus, seems to complicate the explanation as to how the Galician C-AGR data may be derived. I examine such an approach in the following section.

5.9.1.2 Nez Perce as post-syntactic insertion

As highlighted in example (54) of §5.8 (repeated below in (70)), C-AGR in Nez Perce may have more than one agreement feature for the same argument, whilst bearing agreement from multiple arguments.

- (70) **ke-pe-m-ex** kaa *pro*_{SUBJ} cewcew-tée'nix *pro*_{OBJ}
C-PL-2-1 then *pro*_{1PL} telephone-TAM *pro*_{2SG}
 ‘When we call you(SG)’

There are situations, however, in which the C-oriented constituent only shows inflection for 1st- and 2nd-person (71), in addition to solely the latter (72).

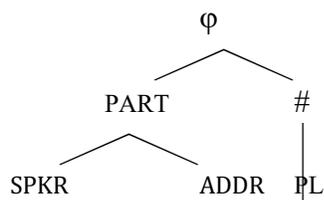
- (71) **ke-me-ex** kaa *pro*_{SUBJ} cewcew-téetu *pro*_{OBJ}
C-2-1 then *pro*_{1SG} telephone-TAM *pro*_{2SG}
 ‘When I call you.’

- (72) **ke-m** kaa *pro*_{SUBJ} cewcew-téetum *pro*_{OBJ}
C-2 then *pro*_{2SG} telephone-TAM *pro*_{1SG}
 ‘When you call me.’

(Deal 2015:6)

Deal proposes that Nez Perce has a feature geometry that accounts for [ADDRESSEE] independently of the specificity of [PARTICIPANT].

(73) *Feature geometry of C in Nez Perce* (Deal 2015:2)



In order to account for the data in (70-72), Deal proposes that C has a *satisfaction* specification for [ADDR], whilst its *interaction* behavior includes any features from (73).¹⁵² Therefore, in (71), C probes and *interacts* with the [SPKR] feature on the *pro*-subject and is then *satisfied* by probing again and agreeing with the object, specified for [ADDR]. When [ADDR] is found first by the probe, as on the structurally higher subject in (72), the probe ceases its search based on the fact that the first goal it encounters *satisfies* it.

These data immediately seem difficult to account for via a post-syntactic insertion account. First, Deal (2015:5) notes that the exponents on T need not and sometimes cannot be those specified on C, particularly when T shows agreement for 3rd-person and C never does.

(74) ke-x kaa A-nim **hi**-cwcew-téetu *pro*_{OBJ}
 C-1 then A-ERG **3SUBJ**-telephone-TAM *pro*_{1SG}

‘When A calls me.’

In (74), only the 1st-person singular object shows agreement on C, whereas T shows agreement for the 3rd-person subject. If the probe on C never agrees with 3rd-person goals and yet T must when the subject is specified for 3rd-person, a post-syntactic insertion approach to C-AGR makes

¹⁵² This may seem unorthodox considering [PLURAL] is not a [PARTICIPANT] feature and, therefore, should be sought after by a separate probe on C. However, Deal (2015:8) gives clear evidence that [PLURAL] always interacts with C as long as it is not below the argument bearing [ADDR].

poor predictions even in the most basic instances, as in (74). Moreover, the agreement shown on C is clearly based on the combination of both hierarchically positioned goals and the ϕ -features they bear, making it impossible to predict when C should show one, two, or three agreement exponents.

5.9.2 C-AGR via agreement

The literature regarding C-AGR as a true agreement phenomenon has addressed different concerns regarding syntactic agreement at different points in time in the minimalist framework. In this section, I shall focus solely on the derivational steps by which C-AGR is derived between C and the DP subject.^{153,154}

The first point to note is that C-AGR violates what has been called the *extension condition*. This has been a topic of considerable discussion for both circumventing head movement (Matushansky 2006) as well as phrasal movement that does not extend the tree (Richards 1997, *et seq.*).¹⁵⁵ The phenomenon of C-AGR, however, is an optimal case of disproving the extension condition regardless of whether by agreement or cliticization. The valuation of the ϕ -features on C may occur one of two ways: a functional head is merged bearing the probe for the relevant ϕ -features or a constituent is \bar{A} moved to a specifier of C that serves as the host for the realization of the ϕ -features. The first instance is what we see in most cases of Germanic C-AGR.

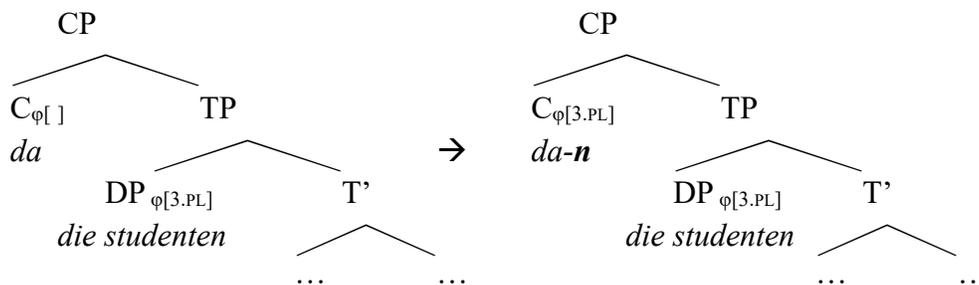
¹⁵³ There are other issues worth discussing for Germanic languages that seem not to apply to other instances of C-AGR cross-linguistically (e.g. Galician, Nez Perce, Wolof) such as the T-to-C movement account (Zwart 1993, 1997). I leave these topics aside here as they are orthogonal to my approach to C-AGR in Galician and refer the interested reader to the review in van Koppen (2017, §3).

¹⁵⁴ I have shown that there are varieties of C-AGR that do agree with DP objects such as Nez Perce and Wolof. Carstens (2003) rules this out for Germanic varieties that permit object scrambling to a position in between C and the subject due to the fact that the object has received Case as is now ‘inactive’ (cf. *Activity Condition*, Chomsky 2001), permitting C to probe further and reach the subject. As Galician C-AGR only occurs in intransitive contexts, I leave this derivational aspect aside.

¹⁵⁵ I return to this topic regarding head movement and cliticization in §5.9.3.

(75) ... **da-n** die studenten...

COMP-**3PL** those students



(West Flemish; Haegeman 2000)

The complementizer *da* ('that') enters the derivation on C, which bears the probe to be valued by the subject. Once the probe agrees with the subject and is valued with the relevant φ -features, the agreement morpheme surfaces on the functional head in C.

For \bar{A} constituents such as phrasal PPs in East Franconian, we may imagine that the probe still merges on C and that the phrase moves to [Spec, CP], serving as the affixal host upon valuation of the probe and its morphophonological realization.

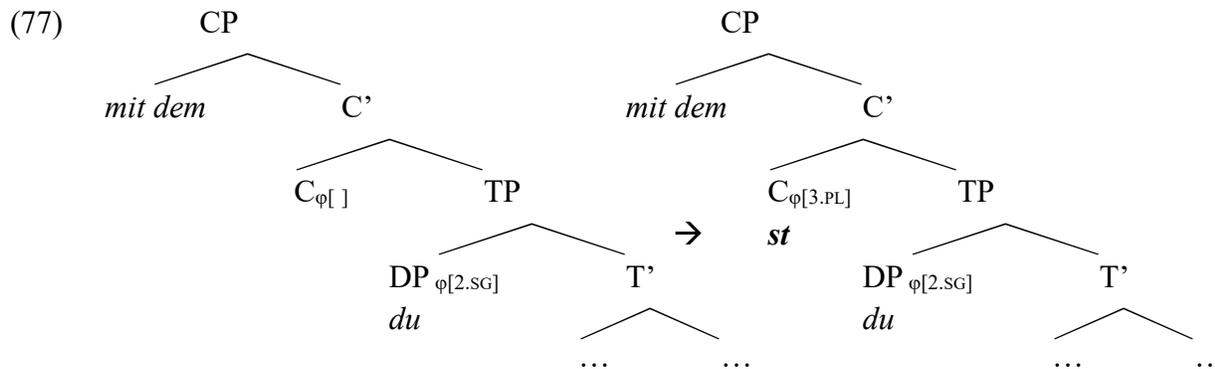
(76) Is des de Mann [mit dem]-st du gesprochen hast?

is that the man with whom_{DAT.M.SG}-2SG you speak.PRTCP have

'Is that the man with whom you have spoken?'

(Bousquette 2014:567)

The phrasal relative *mit dem* ('with whom') moves to the specifier of C and receives the agreement morpheme from C after valuation.

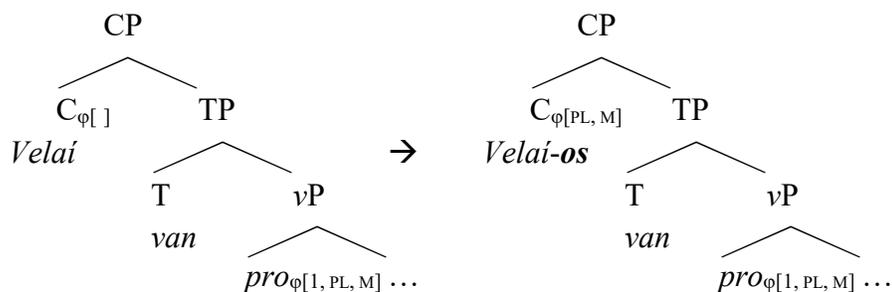


In a model without strict head movement (e.g. syntactic m-merger; Matushansky 2006), the resolution of *mit dem* as the affixal host of *-st* is solved by a lowering process that brings the phrase to the status of a head.¹⁵⁶ As I have advocated for head movement here, I claim that the affixal status of *-st* is realized as such in the post syntax. Morphological merger operations in the PF branch are widely accepted by those that reject this type of movement in the proper syntax, primarily due to the fact that it is the syntax that sets up the structure for these morphological operations as I have showed for head-to-head relations between clitics and their hosts (§5.2.1).

For Galician C-AGR, I propose a slightly different version of the merger of C endowed with φ -probe. I claim that when *velai/velaqui* is merged on C (or, more specifically, Force), it brings in with it an *f* projection and that this head is that which is endowed with the probe. First, let us imagine that the probe remained on C with *velai/velaqui*.

¹⁵⁶ This is different from the *Head Preference Principle* (van Gelderen 2004) in which a phrase is reanalyzed as a head. This operation does not take place in the proper syntax, however, but instead is realized as a form of repair in the inference engine.

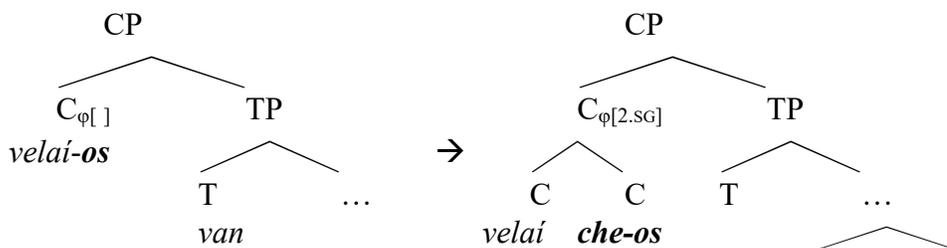
- (78) *Velaí os van*
 behold CL_{M,PL} go.PRS.3PL
 ‘Here they go!’



When C probes for the features on the *pro* subject, the relevant features (i.e., [NUMBER] and [GENDER]) are realized on the head on which the probe is found which in this case is C. This process directly mirrors the Germanic example above in (75).

The first issue with this model is the fact that we should be able to account for instances in which a dative of interest precedes the C-AGR morpheme but follows *velaí/velaquí*. Were C the functional head that bears these features, then we would expect this to be borne out, as in (79).

- (79) *Velaí chos van*
 behold CL_{DAT.2SG-CLM.PL} go.PRS.3PL
 ‘Here they go (on you)!’



The model in (79) represents two key aspects of this operation worth pointing out. First, we may now accept that the claim by Zwicky & Pullum (1983) regarding affixes being unable to attach to

clitics is at least in part incorrect.¹⁵⁷ The order in which true agreement and cliticization has taken place is the expected one (agreement > cliticization), but the spelled-out order is the reverse (clitic > agreement). Second, we may ask whether this type of multi-faceted Agree result has been attested elsewhere. Although it has never been labeled as such, possibly due to the longstanding assumptions regarding cliticization and affixation as discussed at length in §5.2, Galician-Portuguese mesocclisis may be the prime candidate to look to in order to provide a viable model.

- (80) e poder-**m-edes** defender
 and be.able.INF-**CL_{1,SG}-2PL** defend.INF
 ‘And that you will defend me’

(*Senhor fremosa, fui buscar*; ed. Arias Freixedo 2003:237)

Roberts (2007:55) claims that both the pronominal clitic and the agreement morpheme for [PERSON] are clitics, and the infinitival marker *-r* is the spell-out of T. This assumes that both clitics in (80) must attach at a higher projection, which I argue is *f*. I wish to make a case, however, that mesocclisis can be accounted for without assuming that the [PERSON] agreement inflection is a clitic.

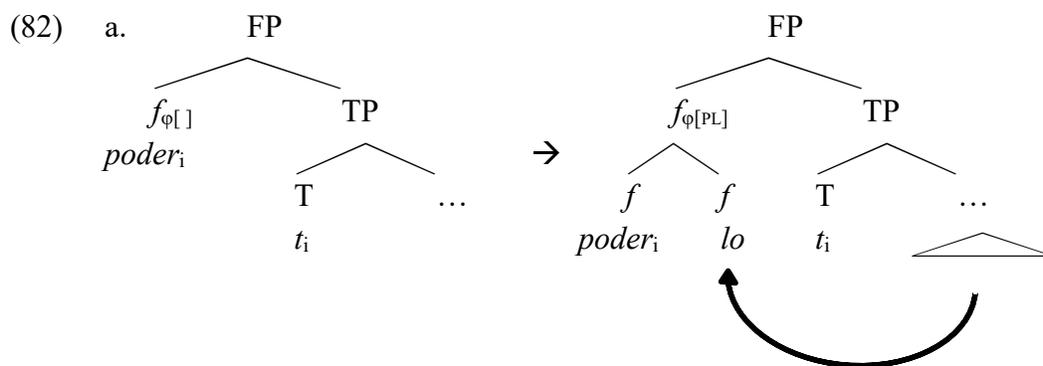
First, it should be noted that mesocclisis is coupled with another West Iberian phenomenon that Raposo & Uriagereka (2005) claim is found on *f*: inflected infinitivals. Much like mesocclisis and finite enclisis, this is a phenomenon that only survives in Western Iberian languages. The diachronic picture painted by R&U is that ‘active’ *f* in WI is the head that accounts for the different left-peripheral phenomena highlighted in §3.3.3, amongst them finite enclisis,

¹⁵⁷ That is, depending on how strict the interpretation of their claim that ‘clitics may attach to material already containing clitics but affixes may not.’ Because what is considered an affix (the C-AGR morpheme) has technically materialized on C before the clitic is attracted, this may be enough evidence for some that suggests this is actually a loophole to Z&P’s argument. Expectedly, this raises further questions and concerns that I do not pursue here but leave for further research.

mesocclisis, and inflected infinitivals. If we consider that the φ -features relevant for inflected infinitivals are on f , we may also assume that the agreement features responsible for mesocclisis are also on f . Thus, if verbal predicates must move to f in order to check the relevant agreement features for inflected infinitivals, it is feasible that the same would have been necessary in cases of mesocclisis. How would this work derivationally? Using a constructed example parallel to that in (80), a 3rd-person accusative clitic would be probed first per the model I adopt here (cf. Deal 2020).^{158,159}

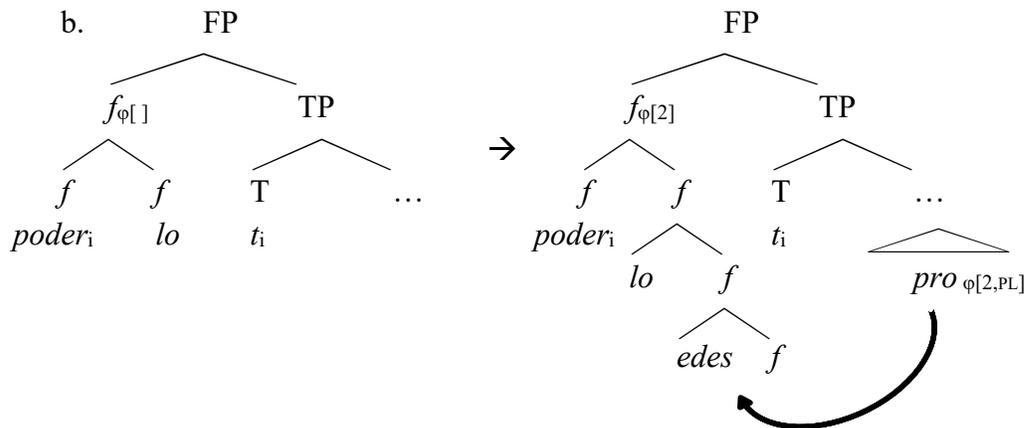
- (81) poder-**los**-edes defender
 be.able.INF-CL_{M,PL}-2PL defend.INF
 ‘You will defend them.’

Following Roberts, if we maintain that *-r* is the spell-out of T but wish to account for the [PERSON] inflection *-edes* (2.PL) as a case of agreement on f , we would expect the exact same steps as (79) but with the 2nd-person plural affix *-edes* realized after *lo* is probed. This is seen in (82).



¹⁵⁸ That the direct object is agreed with first is the common assumption in work on the PCC due to the fact that the direct object controls the outcome of PCC effects (Preminger 2019, Deal 2020, Sheehan 2020, *i.a.*). In this sense, any relevant repair strategy (Deal 2020, Rezac 2008) falls on changing the realization of the indirect object but never the direct object.

¹⁵⁹ This example is constructed based on (80) in order to show mesocclisis but with a clitic not bearing a [PERSON] feature. Mesocclisis with 3rd-person clitics was plentiful from the beginning of Galician-Portuguese with examples dating from before 1300 (Monteagudo & Boullon Agrelo 2009).



In (82a), *poder* moves to f based on its need to probe for the inflectional [PERSON] features due to the fact that T does not merge with a probe for ϕ -features.¹⁶⁰ First, however, the [NUMBER] probe on f must be satisfied, which brings the accusative clitic to f .¹⁶¹ It is then that f is able to probe for [PERSON], which is realized as agreement once the probe finds the *pro* subject (82b). It is in this way that we may account for a HOST-CLITIC-AGREEMENT pattern that has been claimed to be impossible (§5.2.1).

Although I have laid out a scenario in which cliticization may precede agreement, it begs the question of how plausible this operation is for Galician C-AGR, particularly from a learning standpoint. To be exact, we should first ask whether there is more than one cliticization site in the left periphery. We have already seen that there is robust evidence for f and the learning path along which a child comes to posit it as the functional projection that attracts clitics. Incidentally, the need for f independently of C was a question raised by R&U (2005:646): “With respect to F, these considerations raise the following two questions: (i) is F just a convenient coding for separate heads? (ii) Alternatively, could it be that F does not really exist as a separate head and is

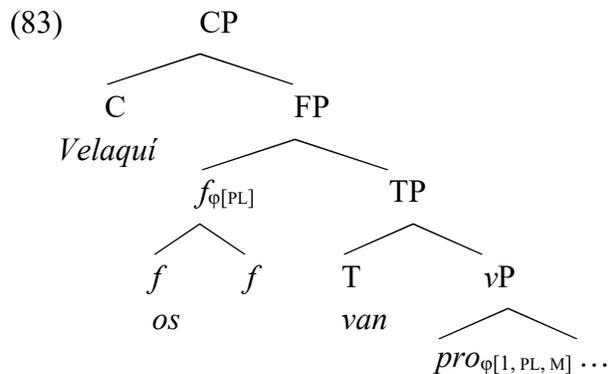
¹⁶⁰ This is different than what I showed in Chapter 3 with cliticization, where the clitic must move to f before the verb. This is further evidence for this being a true agreement phenomenon rather than cliticization-based.

¹⁶¹ Cliticization is typically seen as a left-adjunction phenomenon when the functional head to which it moves is filled. As I show in §5.9.3, however, cliticization of this pattern is also attested cross-linguistically. I leave this topic for more detailed discussion in the next section.

merely a feature of C?” Based on the parameters and left-peripheral phenomena that these authors grouped together that gave evidence for the existence (or lack thereof) of *f*/F (§3.3.3), they claim that there are two options: either C and *f* are independent heads, responsible for distinct phenomena, or *f* must be C. If the latter is true, it is not only subordinate phenomena (e.g. recomplementation) but matrix clause phenomena, as well, that must be distinguished between Romance varieties. They conclude that, as matrix C is not typically pronounced, it is tenuous to assume that the disparity between Romance languages and their licensing of different phenomena is due to a phonologically null head.

A possibly simpler way of looking at this from an acquisition standpoint is to imagine that the child has already posited a functional head (*f*) that is responsible for checking the ϕ -features of, and attracting clitic morphemes that are identical in their featural makeup to, the proposed C-AGR morpheme, argued here to be a byproduct of true agreement. Based on the evidence in her input, the child concludes that this morpheme does not surface without *velai/velaqui* heading the clause. Therefore, she assumes that, unlike in transitive sentences in which morphologically identical clitic pronouns are licensed by the verb, it is *velai/velaqui* that licenses *f*, which, in this instance, probes for agreement with the lone DP goal. Furthermore, the notion that *f* is indeed responsible for hosting this probe is further confirmed with the presence of a dative clitic as in (79). Therefore, if Galician C-AGR is indeed an agreement phenomenon, I claim that it must include *f* rather than having the ϕ -features hosted on the C head that introduces *velai/velaqui*.¹⁶²

¹⁶² Recall that my proposal for the structure in (78-79) has the probe on C as opposed to on *f* as seen in (83). The retreat from the agreement morpheme appearing on C is based on the argument for *f* by R&U as described on p. 51-52.



It seems easier, thus, to assume that *os* is a product of true agreement on f , which may further probe for a dative that couples with the agreement morpheme. Were we to assume a figure such as (78), it would be difficult to explain subsequent cliticization (79) after agreement has taken place.

With several of the same derivational mechanisms, I introduce an argument for how Galician C-AGR may be derived as a clitic phenomenon, showing that this method may be preferred due to evidence the child has seen elsewhere in her input.

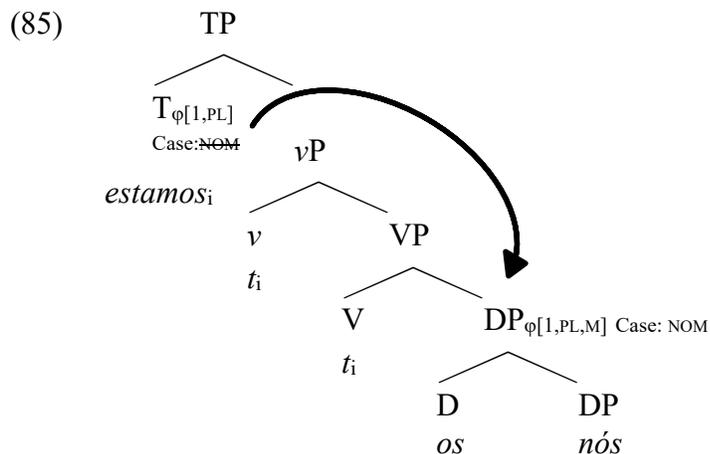
5.9.3 C-AGR via cliticization

After the earlier observations regarding cliticization in this chapter, the approach to Galician C-AGR as a clitic phenomenon should address two primary questions: i) What is the base-generated position of this morpheme?; and ii) How is the clitic brought to its landing site on f ? In §5.7, I showed that Galician C-AGR cannot be an instance of clitic doubling due to the contrast in ϕ -feature between the C-AGR morpheme and the DP subject as in (84).

- (84) Velaquí os estamos (nós)
 behold CL_{M,PL} be.PRS.1PL (we)
 ‘Here we are!’

Based on these data, I claimed that *os* could not be a product of Preminger’s (2019) *conditions on phonological chain reduction of head movement* (39), which he claims may account for the

clitic doubling being a *true* double of its coreferential DP. One option I left open, precisely to discuss here, is the idea that these data could be accounted for via the more traditional ‘Big DP’ analysis (Torrego 1992, Uriagereka 1995a, *i.a.*). As has been well documented, the advantages of this approach guarantee the same thematic relation between the clitic and its double in addition to assuring that both receive the same Case.



In (85), the probes for [PERSON] and [NUMBER] on T agree with the subject DP, assigning it Case upon valuation. The issue that this approach runs into is the fact that the ‘Big DP’ analysis is meant to capture the fact that the DP constituent and its clitic double are essentially two realizations of the same φ -set, one spelled-out as a head which moves and the other spelled-out as a phrasal constituent. As with the approach put forth by Preminger (2019), the mismatch between features of the clitic and its DP sister create an issue when accounting for the φ -set held by the mother node. In the case of (85), although the mother node is specified for 1st-person, as is the DP subject *nós*, *os* is not specified for any [PERSON] feature.

An alternative take is to assume an approach that considers the superfluous nature of the Galician C-AGR morpheme much in the same way that that superfluous datives are considered. As we have seen, the C-AGR morpheme is an optional agreement marker with a high pragmatic function. Along the same lines, high applicative arguments that surface as datives in Romance

work in a similar way. Whereas the Galician C-AGR morpheme highlights the subject in a pragmatically marked manner, high applicatives also imply the involvement of a referent not required by the verb, as in psychological predicates.

- (86) A súa opinión molesta(-me) (Galician)
 the his opinion bother.PRS.3SG(-CL_{DAT.1SG})
 ‘His opinion is bothersome (bothers me).’

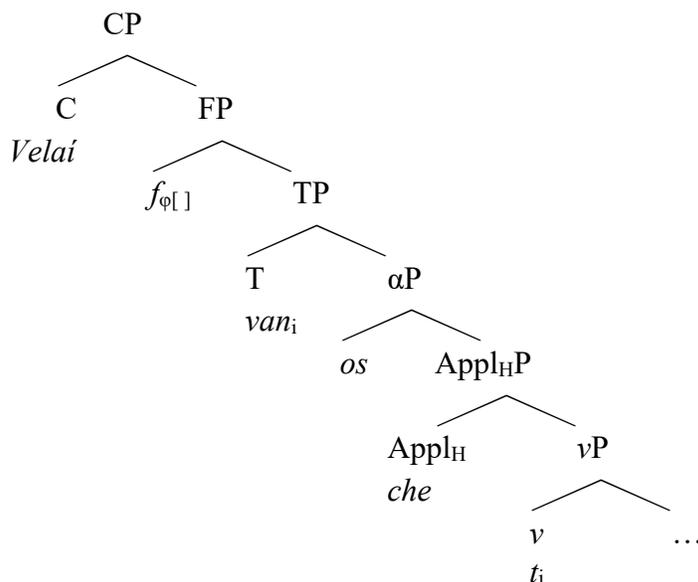
For Cuervo (2003, 2010, 2015, 2020), the nature of an applicative is determined by both its c-selectional properties as well as the head(s) that select it. For example, while low or middle applicatives are c-selected (i.e., c-commanded) by some predicating verbal head (e.g. v_{GO} or v_{BE}) high applicatives merge above it, owing to their lack of inherent participation in the event described by the syntax. Whereas low applicatives that function as true indirect objects form part of the event structure of the verbal predicate, high applicatives never do.

I wish to claim a similar approach to the base-generated position of the C-AGR morpheme. For simplicity purposes, I assume that the C-AGR clitic is base generated above the high applicative in a position I label αP .^{163,164} I claim that the probe on f agrees with both the C-AGR clitic and the dative in examples such as (79). As shown in (42) and (46), there are no person complementarity restrictions, making the probe on f *insatiable* in Deal’s (2015, 2020) terms.

¹⁶³ This may stem from Mirror Principle (Baker 1985, *i.a.*) effects. Amy Rose Deal (p.c.) claims that clitic order is considered templatic. If so, nothing hinges on the C-AGR morpheme base generating above or below the high applicative.

¹⁶⁴ The projection I claim to host the base-generated C-AGR clitic is specific to this morpheme and has no additional syntactic or semantic effects for the language as a whole.

(87) *Velai chos van* (cf. 79)



Unlike what I claimed in Chapter 3 regarding the clitic moving to *f* before a left peripheral head is merged or the need for Last Resort movement of the verb complex to *f*, I have consistently shown that the C-AGR morpheme does not surface in any other construction and does not act as normal clitics do in Western Iberian. Whereas the latter may appear in both enclitic and proclitic positions with respect to the verb, the C-AGR morpheme may only be proclitic on the verb. Moreover, whereas the surfacing of clitics is not predicated on there being material in the left periphery, this is the case of Galician C-AGR. These two distinguishing characteristics should be salient enough in the input of the learner in order for her to posit a derivational pattern distinct from that which we saw in Chapter 3. This type of head movement is technically left-adjoining, but it does instantiate a case of *tucking-in* (Richards 1997, 2001). As with the agreement-based argument in §5.9.2 with *velai/velaqui* filling the head position on C before Agree with *f* is realized, both the C-AGR morpheme and the corresponding dative clitic are probed after both of these heads are merged, and their landing site on *f* does not extend the tree.

Is there an advantage to these derivational steps over the assumption that *f* may merge and probe the clitics before *C* is merged? This would make cliticization uniform within the language, which seems conceptually advantageous for the learner. On the other hand, this structure is an outlier with respect to the cliticization phenomena observed in detail in Chapter 3. From a neutral perspective, it is not clear to me that one is truly more advantageous than another, as both options promote clear learning biases that lead the learner to different conclusions about the derivational pathway for Galician C-AGR.

5.10 Conclusion

In this chapter, I have given a thorough review of the literature regarding the longstanding question about how affixes and clitics are distinguished from one another. I have shown that Galician C-AGR has both surface-level and theoretical characteristics of both, making it a prime case for the behavior expected of a grammatical element undergoing change. I have also reviewed the three main approaches to accounting for complementizer agreement: post-syntactic insertion, true agreement, and cliticization. I have shown that post-syntactic insertion does not work cross-linguistically, showing impossibilities in both Galician and Nez Perce. While the latter is clearly an instance of true agreement, I have argued that Galician C-AGR can be derived either via true agreement or cliticization. Moreover, when derived as agreement, I have shown how clitics and affixes may interact in a way that is unexpected where an affix may be the further morpheme from the host (HOST > CLITIC > AGREEMENT). If indeed a clitic-based phenomenon, there seem to be various possibilities as to how the derivation may proceed regarding both the base-generated position of the clitic as well as the derivational steps that must be followed.

CHAPTER 6

CONCLUSIONS AND FUTURE CONSIDERATIONS

6.1 The thesis, reviewed

In the previous chapters, I have argued for a *Child Innovator Approach* to the diachronic processes that make up the outlined steps presented in this thesis. Incorporating empirical and theoretical investigation from child language acquisition, I have argued that different instances of diachronic reanalysis may be explained by understanding how a child's immature grammar and/or parsing abilities may influence her grammar in such a way that it results in deviance with respect to the adult grammar.

In Chapter 1, I outlined the inner workings of the *Language Acquisition Device* (LAD) as it is understood in the language acquisition literature. I showed that different parts of the innerworkings of this learning machinery may influence the perception of incoming data, such as the present grammar being too immature in order to accurately parse a certain string in the perceptual encoding or internal pressures of the inference engine forcing the child to commit to a structure she has never heard before. Moreover, I compared and contrasted the scenarios that are known to lead to both deviance and full convergence, highlighting how the consistent failure to accurately parse a certain string sits at the core of what we recognize as diachronic change.

In Chapter 2, I presented the data of what I have referred to as the Galician complementizer agreement (C-AGR) construction, highlighting its restrictive grammatical nature with respect to its intransitive predication coupled with the morphological accusative clitic as a subject marker. I reviewed the generative work on the topic, showing that, without accounting for the diachrony of

this construction, it is impossible to explain the idiosyncratic nature of this construction. Additionally, I reviewed a sociolinguistic account of the diachrony that attempts to attribute the synchronic behavior of the construction to two separate phenomena that I showed are geolinguistically not related. Instead, I argued for an account based on four steps, repeated below from Figure 2, that focused on the lack of full transmission and misparsing on behalf of the child from one stage to another.

Step 1: Biclausal, prosodically divided structure

Ve-lo aí vai Xan
 see.PRS.2SG-CL_{ACC.M.SG} there go.PRS.3SG Xan
 ‘You see him, there goes Xan.’

Step 2: Biclausal, prosodically unified structure

Ve-lo aí vai Xan
 see.PRS.2SG-CL_{ACC.M.SG} there go.PRS.3SG Xan
 ‘You see him there, Xan is going.’

Step 3: Monoclausal structure; reanalyzed verbal structure

Vel(o)aí vai Xan
 behold go.PRS.3SG Xan
 ‘There goes Xan.’

Step 4: Monoclausal, main clause complementizer agreement structure

Vel(o)aí o vai Xan
 behold CL_{ACC.M.SG} go.PRS.3SG Xan
 ‘There goes Xan.’

Figure 2. Proposed steps for Galician C-AGR

The subsequent chapters are centered on reconstructing how each step must have been realized in the mind of a child based on the input she received.

Chapter 3 was broken into two parts. In the first part, I compared Steps 1 & 2. In particular, I focused on the unexpected clitic placement in Step 2 where this construction may be embedded under a complementizer.

(1) Calade, diaños, calade, **que vé-lo** alí
 hush.IMP.PL devils hush.IMP.PL COMP see.PRS.2SG-CL_{M.SG} there
 e inda vos ha d’oir
 and still CL_{2PL} have.PRS.3SG COMP-hear.INF

‘Shut up, you devils, shut up, (because) you see him there and he can still hear you.’

I then reviewed data on Western Iberian cliticization patterns in child language (Costa et al. 2014), ultimately linking the patterns seen with certain phenomena to the path along which a child must go in order to master the cliticization patterns in her language. Pairing these data with the theory in Raposo & Uriagereka (2005), I showed how their theory and the necessity for the clitic to move independently of the verb is supported by the learning path a child takes. Using the data from Costa et al. (2015), I hypothesized that there is a feature-based hierarchy of groups of constituents that a child learns as she ‘undoes’ her initial enclisis-only bias. Although I showed that some constructions in the modern language have variability with respect to whether pronouns are placed enclitically or proclitically, the instance of encliticization under embedding as in (1) is a solitary instance in the history of Galician and should not be confused with other phenomena. The second part of Chapter 3 deals with category learning and how children determine the category class of a lexical item. Using findings from Cassani, et al. (2018) and Cassani et al. (2019), I showed how a main-clause functional item in the left periphery could be

posited from a multi-word string in view of the nature of reanalysis itself. In addition, I compared the case of *velai/velaquí* with the Spanish *dizque*, a functional item with multi-word origins, that is not tied to one particular projection in the left periphery. Due to the variability found in the *TILG* corpus, I concluded that Step 2 must be comprised of two micro-stages: a ‘fixed enclisis’ construction (2a) and what I referred to as ‘low *velai/velaquí*’ (2b).

(2) a. Veña **que** **xa** **vé-los** alá están
 come.SUBJ.3SG COMP already see.PRS.2SG-CL_M.PL there be.PRS.3PL
 os carros!
 the haywains

‘Come on, because you already see them, there are the haywains!’

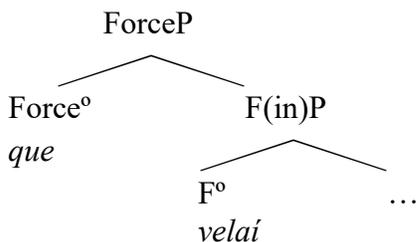
b. E cadra ben **que** **velaí** vén Alberte e
 and fit.PRS.3SG well COMP behold come.PRS.3SG Alberte and
 vou-lle dar as boas tardes
 go.PRS.1SG-CL_{DAT}.3SG give.INF the good afternoon

‘And it works out well that here comes Alberte and I’m going to tell him good afternoon.’

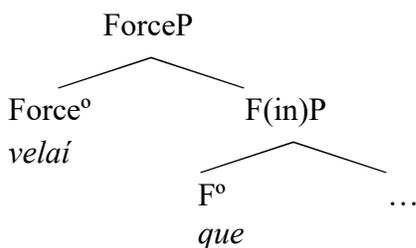
In Chapter 4, I combined the theoretical assumptions behind aspects such as frequency, diversity, and predictability in category learning with the way in which reanalysis may take place when highly variable input is presented to the learner. Presenting the data and arguments from Hudson Kam & Newport (2005), I argued that the variability between Steps 1, 2, & 3 (Figure 1) was enough in order for the learner to posit *vel(o)ai/vel(o)aquí* as a separate category, that of a main-clause complementizer. Moreover, I argued that the change from Step 2 to Step 3 (3) was

not motivated by featural movement, but instead by the pragmatic weight of the presentative nature as noted in Zanuttini (2017) and Wood et al. (2015).

(3) a. *Step 2*



b. *Step 3*



Based on the discussion in Chapter 3 regarding frequency and diversity as the most important aspects of category learning, I claimed that it was the sentence-initial frequency of *velai/velaiqui* and diminishing evidence of left-peripheral material above it that provided the child with indications to hypothesize this word as sentence-initial only. This development, I claimed, provides us with a transition into the discussion of how Step 4, the modern-day construction, was engendered. Using data from Italian child language (Belletti 2017), I showed that children also posit deviant structures based on internal pressures with no evidence for them in their PLD. Based on the hypothesis that children may create new grammatical structures based on misparsing their PLD or via internal assumptions about their language(s) not evident in their input, I claimed that an amalgamation of the two scenarios is a viable pathway toward the engenderment of new grammatical structures, as well.

In Chapter 5, I focused on the properties of the Galician C-AGR morpheme and the plausibility of its being a clitic or an affix. After reviewing much theoretical investigation done on the topic, I applied the relevant theoretical underpinnings to the Galician C-AGR morpheme, showing that it boasts characteristics of both types of morphemes. Applying these properties to the complementizer agreement phenomenon, I showed that this construction could be accounted for by both true agreement and cliticization. Furthermore, I showed that a post-syntactic insertion account, which has been argued for in several Germanic varieties (Fuß 2014), is untenable in Galician and Nez Perce, a language whose C-AGR phenomenon is agreement-based.

6.2 Theoretical importance

In this dissertation, I have labored to highlight important aspects of cross-disciplinary work that has been absent in previous approaches dealing with diachronic approaches to syntactic phenomena. Most importantly, the lack of empirical knowledge regarding child language acquisition has left large holes in the exactitude of the explanations for many diachronic phenomena. The true process of first language acquisition has been deficiently summarized and often looked over, leading to imprecise assumptions and ill-defined theoretical notions that would better be explained by the individual means by which a child learns.

By incorporating details of the internal learning mechanisms that make up the LAD, I have attempted to bridge the gap between a child's learning path and its impact on several theoretical notions, both those language-specific and those considered universal: the role of F/f and the learning path of cliticization in Western Iberian, the aspects of learning that are most prominent in the reanalysis of words/word strings, the internal creativity that children exhibit without external pressure from their input, and the diminishing importance of Universal Grammar as a factor in learning.

In Chapter 3, I emphasized the importance of a derivational mechanism that was sensitive to left-peripheral material, as the pattern between finite enclisis and proclisis is almost always dependent on the presence of a local constituent above *f*. The idea that clitics move independently of and above the verb makes Western Iberian unique amongst other Romance varieties where clitics attach low to *v* (e.g. Spanish; Roberts 2010) or to T (e.g. Romanian; Cornilescu 2020). The ability to show that children clearly must learn this process, as younger children prefer finite enclisis regardless of the context, commends the theoretical claims of Raposo & Uriagereka (2005) regarding clitic movement to *f* and Last Resort movement of the verb complex to *f* in the absence of left-peripheral material.

Additionally, using the arguments in Cassani, et al. (2018) and Cassani et al. (2019), I posited the idea that reanalysis of a word or word string is explainable by its contextual diversity and, in addition, the frequency with which it is used in said context. Thus, the cross-linguistic pattern where a functional word (or a phrase) loses ‘contextual flexibility’ and its ability to combine with certain words or form part of a certain syntactic context is diminished can be explained by the lack of salient input of certain patterns in a child’s PLD.

In order to account for the amalgamation of structures posited in order to derive the Galician C-AGR construction, I also addressed the fact that children sometimes see irregularities that are not easily explained from their input, such as cases of misparsing leading to reanalysis. When these irregularities lead to change, they are often language-specific and sometimes predicated on other changes occurring within the same grammatical structure.

6.3 Limitations and future directions

There are several aspects of this dissertation that I leave for future investigation, in addition to others that limited the thoroughness of several assumptions I was forced to make. I shall begin with the details that I leave for future studies.

Although Galician and Portuguese share very similar grammars, particularly with respect to cliticization patterns, data from Galician-speaking children should be collected and compared to the data presented from Costa et al. (2015). For the present, I have hypothesized as to how certain aspects of the Galician language that are not present in Portuguese would accelerate the learning process of certain cliticization patterns, in addition to others that may inhibit learning said patterns. In order to confirm these hypotheses, similar tests as those in Costa et al. (2015) must be done, along with others that these authors do not investigate (e.g. cliticization patterns of phrases such as *é que...* ('It's that...')). Although I have argued for several characteristics of the adult grammar that make sense of some of the more intricate aspects of cliticization patterns in Galician, tests are needed in order to determine whether a child sees these details as important cues or takes time in order to realize that these are part of her pathway to mastering cliticization in her language.

In a more universal sense, more work that incorporates the specifics of how language acquisition may affect diachronic change are needed. Work on modal verb use in children of Canadian French (Cournane 2015, 2019, 2020) has focused on deviance in a child's use of modal verbs that may eventually persist into adulthood. For Galician specifically, phrases that one would expect to induce proclisis such as *é que...* ('It's that...') may be undergoing a change to a preference for enclisis without a focus-like prosody.¹⁶⁵ If this is so, studying the clitic placement in structures such as these in child language may be immensely fruitful. Just as general work on

¹⁶⁵ I thank Miguel Giadás Quintela for much discussion on this topic.

microvariation in Romance has led to the discovery of many diachronic phenomena, identifying deviant parts of child language may lead to a prolific subfield that ties together diachrony, child language acquisition, and generative syntax.

As expected, several limitations have been unavoidable in this investigation. The fact that Galician underwent an extensive period of time without written record of its language from the 16th to the 18th centuries (*Os Séculos Escuros* ‘The Dark Ages’) puts any diachronic investigation focusing on data at this time period at a disadvantage. In particular, for this dissertation, the data available in the *TILG* corpus captures the diachronic evolution of the Galician C-AGR phenomenon at a point of noticeable overlap with respect to its development.¹⁶⁶ Although the recoverability of more revealing texts is unlikely, this places a significant portion of certain arguments on speculative grounds (e.g. the prosodic boundaries of Steps 1 & 2).

Due to the fact that deviant structures in child language are both language specific and often speculative with respect to why children use them the way they do, implementing this degree of conjecture into an already theory-heavy framework as is diachronic work proliferates the need for a certain degree of surmise. Explaining how non-universal or typologically rare phenomena surface and become part of an emerging grammar always carries some degree of uncertainty. However, doing so with minimal historical resources requires even more speculation.

6.4 Final comments

From the beginning of this dissertation, I attempted to bring a new import to the growing field that is diachronic formal syntax: incorporating empirical and theoretical data from child language acquisition in order to explain how unconventional grammatical patterns arise. I believe that my contribution will serve to further the field’s understanding of the importance of

¹⁶⁶ This is highlighted with exact dates and examples in §3.2.

looking to how a child learns in order to shed light on diachronic patterns that are both well-attested and those that have yet to be replicated in another language.

REFERENCES

- Ackema, Peter and Ad Neeleman. 2004. *Beyond Morphology. Interface Conditions on Word Formation*. Oxford: Oxford University Press.
- Adger, David. 2020. Syntax and the failure of analogical generalization: A commentary on Ambridge (2020). *First Language* 40(5-6), 560-563.
- Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21, 435-483.
- van Alem, Astrid. 2020. Complementizer is not allomorphy: A reply to Weisser (2019). *Glossa* 5(1), 1-10.
- Álvarez, Rosario. 2015. The accusative clitic in transitive verb constructions: *aí o vai, aquí as están*. *Dialectologia. Special issue, V*, 191-217.
- Álvarez Cáccamo, Celso. 1989. Variação lingüística e o factor social na Galiza. *Hispanic Linguistics* 2(2), 253-298.
- Ambridge, Ben. 2017. Syntactic Categories in Child Language Acquisition: Innate, Induced, or Illusory? In Henri Cohen & Claire Lefebvre (eds.) *Handbook of Categorization in Cognitive Science*, 567-580. Elsevier.
- Ambridge, Ben. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5-6), 509-559.
- Anagnostopoulou, Elena. 2003. *The syntax of ditransitives: Evidence from clitics*. Berlin/New York: Mouton de Gruyter.
- Anagnostopoulou, Elena. 2016. Clitic doubling and object agreement. In Susann Fischer & Mario Navarro (eds.) *Proceedings of the VII Nereus International Workshop*, 11-42.
- Arias Freixedo, Xosé Bieito. 2003. *Antoloxía da lírica galego-portuguesa*. Vigo: Edición Xerais.

- Arregi, Karlos and Andrew Nevins. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout* (Studies in Natural Language and Linguistic Theory 86). Dordrecht: Springer.
- Ausín, Adolfo and Francisco Fernández-Rubiera. 2021. *Leísmo: One le or two?* In Juan J. Colomina-Almiñana & Sandro Sessarego (eds.) *Language Patterns in Spanish and Beyond: Structure, Context and Development*, 122-143. London: Routledge.
- Baier, Nico. 2017. Antilocality and Antiagreement. *Linguistic Inquiry* 48(2): 367-377.
- Baker, Mark. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.
- Baker, Mark. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.) *The Limits of Syntactic Variation*, 351-373.
- Baker, Mark and Ruth Kramer. 2018. Doubled clitics are pronouns. *Natural Language & Linguistic Theory* 36(4), 1035-1088.
- Bar-Sever, Galia and Lisa Pearl. 2016. Syntactic categories derived from frequent frames benefit early language processing in English and ASL. *Proceedings of the 40th Annual Boston University Conference on Child Language Development*, 32-46.
- Barbosa, Pilar. 1995. *Null subjects*. Doctoral dissertation, Massachusetts Institute of Technology.
- Barbosa, Pilar. 2000. A Window into the Null Subject Property. Clitics: A window into the null subject property. In João Costa (ed.) *Portuguese Syntax: New Comparative Studies*, 31-93. New York/Oxford: Oxford University Press.
- Béjar, Susana. 2003. *Phi-syntax: A theory of agreement*. Doctoral dissertation, University of Toronto.
- Béjar, Susana & Milan Rezac. 2003. Person licensing and the derivation of PCC effects. In Ana Teresa Perez-Leroux & Yves Roberge (eds.), *Romance linguistics: Theory and acquisition*, 49-62. Amsterdam: John Benjamins.

- Béjar, Susana, and Milan Rezac. 2009. Cyclic Agree. *Linguistic Inquiry* 40, 35–73.
- Belletti, A. 2004. Aspects of the Low IP Area. In L. Rizzi (ed.), *The structure of CP and IP. The Cartography of Syntactic Structures, vol. 2*, 16-51. Oxford: Oxford University Press.
- Belletti, Adrianna. 2017. Internal grammar and children's grammatical creativity against poor inputs. *Frontiers in Psychology* 8, 1-7.
- Belletti, Adriana 2018. On *a*-marking of object topics in the Italian left periphery. In Roberto Petrosino, Pietro Cerrone & Harry van der Hulst (eds.) *From Sounds to Structures*, 445-466. Berlin/Boston: de Gruyter/Mouton.
- Belletti, Adrianna and Luigi Rizzi. 1988. Psych-verbs and ϕ -theory. *Natural Language & Linguistic Theory* 6(3), 291-352.
- Belloro, Valeria. 2007. *Spanish clitic doubling: A study into the syntax-pragmatics interface*. Doctoral dissertation, State University of New York at Buffalo.
- Benincà, Paola, Adam Ledgeway, and Nigel Vincent (eds.). 2014. *Diachrony and dialects: Grammatical change in the dialects of Italy*. Oxford: Oxford University Press.
- Berwick, Robert. 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, M.A.: MIT Press.
- Berwick, Robert. and Noam Chomsky. 2016. *Why only us: Language and evolution*. Cambridge, M.A.: MIT Press.
- Biberauer, Theresa. 2017. Factors 2 and 3: a principled approach. *Cambridge Occasional Papers in Linguistics* 10, 38–65.
- Biberauer, Theresa. 2019a. Children always go beyond the input: The Maximise Minimal Means perspective. *Theoretical Linguistics* 45(3/4), 211-224. Berlin/Boston: de Gruyter Mouton.
- Biberauer, Theresa. 2019b. Some thoughts on the complexity of syntactic complexity. *Theoretical Linguistics* 45(3/4), 259-274. Berlin/Boston: de Gruyter Mouton.

- Biberauer, Theresa and Ian Roberts. 2012. Towards a parameter hierarchy for auxiliaries: Diachronic considerations. *Cambridge occasional papers in linguistics* 6(9), 267-294.
- Biberauer, Theresa and Ian Roberts. 2016. Parameter typology from a diachronic perspective: the case of Conditional Inversion. In E. Bidese, F. Cognola & M. Moroni (eds.), *Theoretical approaches to linguistic variation*, 259–291. Amsterdam: John Benjamins.
- Biberauer, Theresa and Ian Roberts. 2017. Conditional inversion and types of parametric change. In Bettelou Los & Pieter de Haan (eds.) *Word Order Change in Acquisition and Language Contact: Essays in honour of Ans van Kemenade*, 57-77. Amsterdam: John Benjamins.
- Biberauer, Theresa, Ian Roberts & Michelle Sheehan. 2014. No-choice parameters and the limits of syntactic variation. *Proceedings of the 31st West Coast Conference on Formal Linguistics*, 45-55. Somerville, MA: Cascadilla Proceedings Project.
- Bjorkman, Bronwyn and Hedde Zeijlstra. 2019. Checking up on (φ-)Agree. *Linguistic Inquiry* 50, 527-569.
- Boeckx, Cedric and Evelina Leivada. 2013. Entangled Parametric Hierarchies: Problems for an Overspecified Universal Grammar. *PLoS ONE* 8(9), 1-11.
- Bonet, Eulàlia. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Doctoral dissertation, Massachusetts Institute of Technology.
- Bošković, Željko and Jairo Nunes. 2007. The Copy Theory of Movement: A view from PF. In Jairo Nunes & Norbert Corver (eds.) *The Copy Theory of Movement*, 13-74. Oxford: Blackwell Publishers.
- Bousquette, Joshua. 2014. Complementizer Agreement in eastern Wisconsin: (Central) Franconian features in an American heritage community. *STUF-Language and Typology and Universals* 67(4), 561-588.

- Brandi, Luciana and Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In Osvaldo Jaeggli & Ken Safir (eds.) *The Null Subject Parameter*, 111-142. Dordrecht: Springer.
- Bybee, Joan. 2003a. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2003b. 19 Mechanisms of Change in Grammaticalization: The Role of Frequency. In Brian Joseph & Richard Janda (eds.) *The Handbook of Historical Linguistics*, 602-623. New York: John Wiley & Sons.
- Cardinaletti, Anna and Uri Shlonsky. 2004. Clitic positions and restructuring in Italian. *Linguistic Inquiry* 35(4), 519-557.
- Carstens, Vicki. 2003. Rethinking complementizer agreement: Agree with a Case-checked goal. *Linguistic Inquiry* 34(3): 393-412.
- Cassani, Giovanni. 2019. *The Bootstrapping Toolbox: Which Cues are More Useful to Learn Lexical Categories and Why*. Doctoral dissertation, University of Antwerp.
- Cassani, Giovanni, Robert Grimm, Walter Daelemans & Steven Gillis. 2018. Lexical category acquisition is facilitated by uncertainty in distributional co-occurrences. *PLoS ONE* 13(12), 1-36.
- Cassani, Giovanni, Robert Grimm, Steven Gillis, and Walter Daelemans. 2019. What makes a distributional context useful? Lexical diversity is more important than frequency. Ms.
- Chemla, Emmanuel, Toben Mintz, Savita Bernal and Anne Christophe. 2009. Computational principles of language acquisition. Categorizing words using ‘frequent frames’: what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science* 12(3), 396-406.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, M.A.: MIT Press.

- Chomsky, Noam. 1981. *Lectures in Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, M.A.: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89-155. Cambridge, M.A.: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, 1-50. Cambridge, M.A.: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero, & Maria-Luisa Zubizarreta (eds.) *Foundational issues in linguistic theory*, 133–166. Cambridge, M.A.: MIT Press.
- Chomsky, Noam, Ángel Gallego and Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics Special Issue*, 2019, 229-261.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Colasanti, Valentina and Guiseppina Silvestri. 2018. Matrix Complementizers in Italo-Romance. In Silvio Cruschina, Adam Ledgeway & Eva-Maria Remberger (eds.) *Italian dialectology at the interfaces*, 155-183. Amsterdam: John Benjamins.
- Compton, Richard. 2016. Mutually conditioned mood and object agreement in Inuit. In Christopher Hammerly & Brandon Prickett (eds.) *Proceedings of NELS 46(1)*, 241-250.
- Cornilescu, Alexandra. 2020. Ditransitive constructions with differentially marked direct objects in Romanian. In Anna Pineda & Jaume Mateu (eds.) *Dative constructions in Romance and beyond*, 117-142. Berlin: Language Science Press.

- Costa, João, Alexandra Fiéis, and Maria Lobo. 2015. Input variability and late acquisition: clitic misplacement in European Portuguese. *Lingua* 161: 10-26.
- Cournane, Ailís. 2014. In Search of L1 Evidence for Diachronic Reanalysis: Mapping Modal Verbs. *Language Acquisition* 21(1), 103-117.
- Cournane, Ailís. 2015. *Modal Development: Input-Divergent L1 Acquisition in the Direction of Diachronic Reanalysis*. Doctoral dissertation, University of Toronto.
- Cournane, Ailís. 2017. In defence of the child innovator. In Éric Mathieu and Robert Truswell (eds.) *Micro-change and Macro-change in Diachronic Syntax*, 10-24 Oxford: Oxford University Press.
- Cournane, Ailís. 2019a. A developmental view on incrementation in language change. *Theoretical Linguistics* 45(3/4), 127-150. Berlin/Boston: de Gruyter Mouton.
- Cournane, Ailís. 2019b. Grammatical representations versus productive patterns in change theories. *Theoretical Linguistics* 45(3/4), 287-297. Berlin/Boston: de Gruyter Mouton.
- Cournane, Ailís. 2020. Learning modals: A grammatical perspective. *Language and Linguistic Compass* 14(10), 1-22. Wiley-Blackwell.
- Crain, Stephen, Takuya Goro and Rosalind Thorton. 2006. Language acquisition is language change. *Journal of Psycholinguistic Research* 35(1), 31-49.
- Cuervo, María Cristina. 2003. *Datives at large*. Doctoral dissertation, Massachusetts Institute of Technology.
- Cuervo, María Cristina. 2010. Against ditransitivity. *Probus* 22, 151-180.
- Cuervo, María Cristina. 2020. Datives as applicatives. In Anna Pineda & Jaume Mateu (eds.) *Dative constructions in Romance and beyond*, 1-39. Berlin: Language Science Press.

- D'Alessandro, Roberta. 2008. *Impersonal si constructions: Agreement and Interpretation*. Berlin/New York: Mouton de Gruyter.
- Deal, Amy Rose. 2015. Interaction and satisfaction in ϕ -agreement. In Thuy Bui & Deniz Ozyildiz (eds.) *Proceedings of NELS 45(1)*, 179-192.
- Deal, Amy Rose. 2020. Interaction, satisfaction, and the PCC. Ms.
- Demonte, Violeta, and Olga Fernández Soriano. 2009. Force and finiteness in the Spanish complementizer system. *Probus* 21, 23-49.
- Demonte, Violeta, and Olga Fernández Soriano. 2014. Evidentiality and illocutionary force: Spanish matrix *que* at the syntax-pragmatics interface. In Andreas Dufter & Álvaro S. Octavio de Toledo y Huerta (eds.) *Left sentence peripheries in Spanish: Diachronic, variationist and comparative perspectives*: 217-252.
- Diercks, Michael, Marjo van Koppen and Michael Putnam. 2020. Agree probes down: Anaphoric feature valuation and phase reference. In Peter W. Smith, Johannes Mursell & Katharina Hartmann (eds.) *Agree to Agree: Agreement in the Minimalist Programme*, 347-389. Berlin: Language Science Press.
- Emmorey, Karen, Stephen McCullough, Sonya Mehta and Thomas J. Grabowski. 2014. How sensory-motor systems impact the neural organization for language: direct contrasts between spoken and signed language. *Frontiers in Psychology* 5(484), 1-13.
- Emmorey, Karen, Stephen McCullough, Sonya Mehta and Thomas J. Grabowski. 2016. The neural circuits recruited for the production of signs and fingerspelled words. *Brian and Language* 160, 30-41.

- Etxepare, Ricardo. 2007. Aspects of quotative constructions in Iberian Spanish. *Anuario del Seminario de Filología Vasca Julio de Urquijo: International Journal of Basque Linguistics and Philology* 41(2): 25-58.
- Etxepare, Ricardo. 2010. From hearsay evidentiality to samesaying relations. *Lingua* 120(3): 604-627.
- Fasanella, Adriana. 2014. *On how Learning Mechanisms shape Natural Languages*. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Fasanella, Adriana and Jordi Fortuny. 2016. Deriving linguistic variation from learnability conditions: The Chunking Procedure. In Luis Eguren, Olga Fernández-Soriano & Amaya Mendikoetxea (eds.), *Rethinking Parameters*, 105-132. Oxford: Oxford University Press
- Fernández Rei, Francisco. 1990. *Dialectoloxía da lingua galega*. Vigo: Edición Xerais.
- Fernández Rei, Francisco, Manuel González González and Rosario Álvarez. 1995. *Atlas Lingüístico Galego. vol. II Morfoloxía non verbal*. Santiago de Compostela: Instituto da Lingua Galega.
- Fernández-Rubiera, Francisco. 2009. *Clitics at the edge: clitic placement in Western Iberian Romance languages*. Doctoral dissertation, Georgetown University.
- Fernández-Rubiera, Francisco. 2013. Clisis revisited: Root and embedded contexts in Western Iberian. In C.M. Salvesen & H.P. Helland (eds.), *Challenging Clitics*, 55-86. Amsterdam: John Benjamins.
- Ferreira, Fernanda and Nikole Patson. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1(1-2), 71-83.
- Ferreiro, Manuel. 1999. *Gramática histórica galega. Vol. I: Fonética e Morfosintaxe*. Santiago de Compostela: Laiovento.

- Fodor, Janet. 1998. Parsing to learn. *Journal of Psycholinguistic Research* 27(3): 339-374.
- Fodor, Janet. 2017. Ambiguity, parsing, and the evaluative measure. *Language Acquisition*, 24(2): 85-99.
- Fodor, Janet and William Sakas. 2012. Disambiguating syntactic triggers. *Language Acquisition*, 19(2): 83-143.
- Fodor, Janet and William Sakas. 2017. Learnability. In Ian Roberts (ed.) *The Oxford Handbook of Universal Grammar*, 249-269.
- Friederici, Angela. 2009. Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences* 13(4), 175-181.
- Freixeiro Mato, Xosé Ramón. 2006. *Gramática da lingua galega. Vol. II: Morfoloxía*. Vigo: A Nosa Terra.
- Fuß, Eric. 2014. Complementizer agreement (in Bavarian): Feature inheritance or feature insertion? In Günter Grewendorf & Helmut Weiß (eds.) *Bavarian syntax. Contributions to the Theory of Syntax*, 51-82. Amsterdam: John Benjamins.
- Fuß, Eric. 2017. Language Change. In Ian Roberts (ed.) *The Oxford Handbook of Universal Grammar*, 459-485.
- Galgiardi, Annie and Jeffrey Lidz. 2014. Statistical insensitivity and the acquisition of Tsez noun classes. *Language* 90(1), 58-89.
- Galgiardi, Annie, Naomi Feldman, and Jeffrey Lidz. 2017. Modeling statistical insensitivity: Sources of suboptimal behavior. *Cognitive science* 41(1), 188-217.
- Gallego, Ángel. 2016. Lexical items and feature bundling: Consequences for microparametric approaches to variation. In Luis Eguren, Olga Fernández-Soriano & Amaya Mendikoetxea (eds.), *Rethinking Parameters*, 133-169. Oxford: Oxford University Press

- Gallego, Ángel and Juan Uriagereka. 2016. Estar=Ser+X. *Borealis: An International Journal of Hispanic Linguistics*, 5(1), 123-156.
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.
- van Gelderen, Elly. 2009. *Feature economy in the linguistic cycle. Historical syntax and linguistic theory*. Oxford: Oxford University Press.
- van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.
- Gibson, Edward and Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25(3), 407-454.
- Gleitman, Lila. 1990. The structural sources of verb meanings. *Language Acquisition*, 1(1), 3-55.
- Goldowsky, Boris Neill. 1995. *Learning structured systems from imperfect information*. Doctoral dissertation, University of Rochester.
- Graham, Lamar. 2015. *The history of the future: morphophonology, syntax, and grammaticalization*. Doctoral dissertation, University of Georgia.
- Gravely, Brian. 2020. Only syntactic movement will do: Morphosyntactic variation in Galician and Icelandic. Talk given at 9th Cambridge Comparative Syntax Workshop (CamCos9). 19-20 January, 2021 (online).
- Gravely, Brian. 2021. DOM and non-canonical word order in Romance: The case of Galician. *Buscharest Working Papers in Linguistics. Special Issue: The Syntax and Semantics of Internal Arguments*, 5-31.
- Gravely, Brian and Timothy Gup-ton. 2021. Verbless DP interrogative constructions and enclisis in Galician. In Juan J. Colomina-Almiñana & Sandro Sessarego (eds.) *Language Patterns in Spanish and Beyond: Structure, Context and Development*, 97-121. London: Routledge.

- Gruber, Bettina. 2008. *Complementizer Agreement: New Evidence from the Upper Austrian Variant of Gmunden*. Master's thesis, University of Vienna.
- Guimarães, Maximiliano. 2004. *Derivation and representation of syntactic amalgams*. Doctoral dissertation, University of Maryland, College Park.
- Gupton, Timothy. 2010. *The syntax-information structure interface: Subjects and clausal word order in Galician*. Doctoral dissertation, University of Iowa.
- Gupton, Timothy. 2012. Object clitics in Galician and complications for clausal analyses. In Kimberly Geeslin and Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, 272-284. Somerville, MA: Cascadilla Proceedings Project.
- Gupton, Timothy. 2014. *The syntax-information structure interface: Subjects and clausal word order in Galician*. Berlin/Boston: de Gruyter/Mouton.
- Hacquard & Jeffrey Lidz 2019. Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language* 34(1), 73-96.
- Haegeman, Liliane. 1992. *Theory and Description in Generative Syntax: A Case Study in West-Flemish*. Cambridge: Cambridge University Press.
- Haegeman, Liliane. 2000. The External Possessor Construction in West Flemish. *GG@G* 1, 1-20.
- Haegeman, Liliane and Marjo van Koppen. 2012. Complementizer agreement and the relation between C° and T°. *Linguistic Inquiry* 43(3), 441-454.
- Hale, John. 2011. What a rational parser would do. *Cognitive Science* 35(3), 399-443.
- Hale, Mark. 1998. Diachronic syntax. *Syntax* 1(1), 1-18.
- Hale, Mark. 2007. *Historical Linguistics: Theory and method*. Wiley-Blackwell.

- Hale, Kenneth and Samuel Jay Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In Richard Kayne, Raffaella Zanuttini & Thomas Leu (eds.) *An Annotated Syntax Reader: Lasting Insights and Questions*, 53-109. Blackwell Publishing.
- Han, Chung-hye, Julien Musolino, and Jeffrey Lidz. 2016. Endogenous sources of variation in language acquisition. *Proceedings of the National Academy of Sciences* 113(4), 942-947.
- Harbour, Daniel. 2007. *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features*. Dordrecht: Springer.
- Harizanov, Boris. 2014. Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian. *Natural Language & Linguistic Theory* 32(4), 1033–1088
- Harley, Heidi and Elizabeth Ritter. 2002. Person and number in pronouns: A feature geometric analysis. *Language* 78(3), 482–526
- Harrigan, Kaitlyn, Valentine Hacquard and Jeffrey Lidz. 2016. Syntactic bootstrapping in the acquisition of attitude verbs: think, want, and hope. *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, 196-206. Somerville, MA: Cascadilla Proceedings Project.
- Harrigan, Kaitlyn, Valentine Hacquard and Jeffrey Lidz. 2018. Three-year-olds' understanding of desire reports is robust to conflict. *Frontiers in Psychology* 9, 199.
- Harrigan, Kaitlyn, Valentine Hacquard and Jeffrey Lidz. 2019. Hope for syntactic bootstrapping. *Language* 95(4), 642-682.
- Haspelmath, Martin. 2004. Explaining the Ditransitive Person-Role Constraint: A usage-based approach. *Constructions* 2, 1-71.

- Hauser, Marc, Noam Chomsky and W. Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298(5598), 1569-1579.
- Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd and Tania Kuteva. 2012. Grammaticalization theory as a tool for reconstructing language evolution. In Maggie Tallerman & Kathleen Gibson (eds.) *The Oxford Handbook of Language Evolution*, 511-527. Oxford: Oxford University Press.
- Hill, Virginia. 2007. Vocatives and the pragmatics-syntax interface. *Lingua* 117(12), 2077-2105.
- Hill, Virginia. 2012. A main clause complementizer. In Lobke Aelbrecht, Liliane Haegeman & Rachel Nye (eds.) *Main Clause Phenomena: New Horizons*, 279-295. Amsterdam: John Benjamins.
- Hill, Virginia. 2013. *Vocatives: How syntax meets with pragmatics*. Leiden/Boston: Brill.
- Hopper, Paul and Elizabeth Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.
- Hornstein, Norbert, Jairo Nunes and Kleantes Grohmann. 2005. *Understanding Minimalism*. Cambridge: Cambridge University Press.
- Hudson Kam, Carla and Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and children learners in language formation and change. *Language learning and development* 1(2), 151-195.
- Hudson Kam, Carla and Elissa Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59(1): 30-66.

- Joseph, Brian. 2011. Historical linguistics and sociolinguistics: Strange bedfellows or natural friends? In Nils Langer, Steffan Davies & Wim Vandebussche (eds.) *Language and history, linguistics and historiography. Interdisciplinary approaches*, 67-88. Oxford & Bern: Peter Lang.
- Kalin, Laura. 2018. Licensing and Differential Object Marking: The view from Neo-Aramaic. *Syntax* 21 (2): 112-159.
- Kalin, Laura. 2019. Nominal licensing is driven by valued (ϕ -)features. *Nordlyd* 43(1): 15-29.
- Kato and Eduardo Raposo. 2007. Topicalization and European and Brazilian Portuguese. In José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez & María José Cabrera (eds.) *Romance Linguistics 2006: Selected Papers from the 36th Linguistics Symposium on Romance Languages (LSRL XXXVI), New Brunswick, March-April 2006*, 199-212. Amsterdam: John Benjamins.
- Kayne, Richard. 1991. Romance clitics, verb movement, and PRO. *Linguistic Inquiry* 22(4), 647-686.
- Kayne, Richard. 2005. *Movement and silence*. Oxford: Oxford University Press.
- Kemp, Charles, Andrew Perfors, and Joshua Tenenbaum. 2007. Learning overhypotheses with hierarchical Bayesian models. *Developmental science* 10(3), 307-321.
- Kidd, Celeste, Steven Piantadosi and Richard Aslin. 2012. The Goldilocks effects: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7(5).
- Kidd, Celeste, Steven Piantadosi and Richard Aslin. 2014. The Goldilocks Effect in Infant Auditory Attention. *Child Development* 85(5), 1795-1804.

- Klausenburger, Jürgen. 2000. *Grammaticalization: studies in Latin and Romance morphosyntax*. Amsterdam/Philadelphia: John Benjamins.
- van Koppen, Marjo. 2005. *One Probe, Two Goals: Aspects of Agreement in Dutch Dialects*. Doctoral dissertation, Leiden University.
- van Koppen, Marjo. 2012. The distribution of phi-Features in Pronouns. *Natural Language & Linguistic Theory* 30, 135-177.
- van Koppen, Marjo. 2017. Complementizer Agreement. In Martin Everaert & Henk van Riemsdijk (eds.) *The Wiley Blackwell Companion to Syntax, 2nd ed.*, 1-40.
- Koring, Loes, Iain Giblin, Rosalind Thornton and Stephen Crain. 2020. Like dishwashing detergents, all analogies are not the same: A commentary on Ambridge (2020). *First Language* 40(5-6), 596-599.
- Kramer, Ruth. 2014. Clitic doubling or object agreement: the view from Amharic. *Natural Language & Linguistic Theory* 32(2): 593-634.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.
- Laka, Itziar. 1990. *Negation in syntax—on the nature of functional categories and projections*. Doctoral dissertation, Massachusetts Institute of Technology.
- Laka, Itziar. 1996. *A brief grammar of Euskara, the Basque language*. Vitoria-Gasteiz: Euskal Herriko.
- Lasnik, Howard and Jeffrey Lidz. 2017. The argument from the poverty of the stimulus. In Ian Roberts (ed.) *The Oxford Handbook of Universal Grammar*, 221-248.
- Ledgeway, Adam, Norma Schifano, and Guiseppina Silvestri. 2019. Differential Object Marking and the properties of D in the dialects of the extreme south of Italy. *Glossa* 4(1), 1-25.

- Lidz, Jeffrey. 2019. Learning, Memory, and Syntactic Bootstrapping: A Meditation. *Topics in Cognitive Science*, 12(1), 1-13.
- Lidz, Jeffrey and Annie Galgiardi. 2015. How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics* 1, 333-353.
- Lidz, Jeffrey, Henry Gleitman and Lila Gleitman. 2003. Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition* 87(3), 151-178.
- Lidz, Jeffrey and Laurel Perkins. 2018. Language Acquisition. In John T. Wixted (ed.) *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 4th Ed*, 83-131. New York: John Wiley & Sons.
- Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Wiley-Blackwell.
- Lightfoot, David. 2002. Explaining language change: An evolutionary approach. *Journal of Linguistics* 38(2), 410-414.
- Lightfoot, David. 2006. *How new languages emerge*. Cambridge: Cambridge University Press.
- Lightfoot, David. 2020. *Borne to parse: How children select their languages*. Cambridge, M.A.: MIT Press.
- Lightfoot, David and Marit Westergaard. 2007. Language acquisition and language change: Interrelationships. *Language and Linguistics Compass* 1(5), 396-415.
- Lim, Dongsik. 2010. *Evidentials and Interrogatives: A Case Study from Korean*. Doctoral dissertation, University of Southern California.

- Lohndal, Terje. 2014. *Phrase structure and argument structure: A case study of the syntax-semantics interface*. Oxford: Oxford University Press.
- Longa, Victor, Guillermo Lorenzo and Gemma Rigau. 1996. Expressing modality by recycling clitics. *Catalan working papers in linguistics* 5(1), 67-79.
- Longa, Victor, Guillermo Lorenzo and Gemma Rigau. 1998. Subject clitics and clitic recycling: locative sentences in some Iberian Romance languages. *Journal of Linguistics* 34(1), 125-164.
- López, Luis. 2012. *Indefinite objects: Scrambling, choice functions, and differential marking*. Cambridge, M.A.: MIT Press.
- Lorenzo, Ramón. 1977. *La traducción gallega de la Crónica General y de la Crónica de Castilla*. Orense: Instituto de Estudios Orensanos Padre Feijoo.
- Lugrís Freire, Manuel. 1931. *Gramática do idioma galego: esta obra leva como aditamentos os nomes de persoas, toponimia, meses e días, denominación dos ventos, táboa de castelanismos, mostras literarias e vocabulario galego-castelán*. Moret.
- Manetti, Claudia and Adrianna Belletti. 2017. The Production of Clitic Left Dislocations by Italian-Speaking Children and the Role of Intervention. In Maria LaMendola & Jennifer Scott (eds.) *Proceedings of the 41st annual Boston University Conference on Language Development*, 443-453. Somerville, MA: Cascadilla Proceedings Project.
- Martín-González, Javier. 2002. *The syntax of sentential negation in Spanish*. Doctoral dissertation, Harvard University.
- Martins, Ana Maria. 1993. Clitic placement from Old to Modern European Portuguese. In Henning Andersen (ed.), *Historical linguistics 1993*, 295-308. Amsterdam: John Benjamins.

- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37(1), 69-109.
- Mayer, Elisabeth. 2017. *Spanish clitics on the move: Variation in time and space*. Berlin/Boston: Walter de Gruyter.
- McGinnis, Martha. 2005. On markedness asymmetries in person and number. *Language* 81, 699–718.
- Mintz, Toben. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90(1), 91-117.
- Mintz, Toben. 2006. Finding the verbs: Distributional cues to categories available to young learners. In Kathryn Hirsh-Pasek & Roberta Golinkoff (eds.) *Action meets word: How children learn verbs*, 31-63.
- Mintz, Toben, Elissa Newport and Thomas Bever. 2002. The distributional structure of grammatical categories in speech to young children. *Cognitive science* 25(4), 393-424. Wiley-Blackwell.
- Monteagudo, Xosé Henrique and Ana Isabel Boullon Agrelo. 2009. *De verbo a verbo: documentos en galego anteriores a 1260*. Santiago de Compostela: Universidade de Santiago de Compostela.
- Moraes de Castilho, Célia Maria. 2005. *O processo de redobramento sintático no português medieval: A formação das perífrases com estar*. Doctoral dissertation, Universidade Estadual de Campinas.
- Morin, Yves Charles. 1985. On the two French subjectless verbs *voici* and *voilà*. *Language* 61(4), 777-820.
- Musolino, Julien, Stephen Crain and Rosalind Thornton. 2000. Navigating negative quantificational space. *Linguistics* 38(1), 1-32.

- Musolino, Julien and Jeffrey Lidz. 2006. Why children aren't universally successful with quantification. *Linguistics* 44(4), 817-852.
- Naigles, Letitia. 1996. The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58(2), 221-251.
- Neokleous, Theoni. 2014. *The L1 acquisition of clitic placement in Cypriot Greek*. Doctoral dissertation, University of Cambridge.
- Neokleous, Theoni. 2015. The L1 acquisition of clitic placement in Cypriot Greek. *Lingua* 161, 27-47.
- Nevins, Andrew Ira. 2011. Multiple Agree with clitics: Person complementarity vs. omnivorous number. *Natural Language & Linguistic Theory* 29(4), 939–971.
- Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, M.A.: MIT Press.
- Omaki, Akira. 2010. *Commitment and flexibility in the developing parser*. Doctoral dissertation, University of Maryland, College Park.
- Omaki, Akira and Jeffrey Lidz. 2015. Linking Parser Development to Acquisition Syntactic Knowledge. *Language Acquisition* 22, 158-192.
- Ormazabal, Javier and Juan Romero. 2007. The Object Agreement Constraint. *Natural Language & Linguistic Theory* 25(2): 315-347.
- Ouali, Hamid. On C-to-T ϕ -feature transfer: The nature of Agreement and Anti-Agreement in Berber. In Roberta D'Alessandro, Susann Fischer & Gunnar Hrafn Hrafnbjargson (eds.) *Agreement Restrictions*, 159-180. Berlin/Boston: de Gruyter Mouton.
- Oxford, Will. 2014. *Microparameters of agreement: A diachronic perspective on Algonquian verb inflection*. Doctoral dissertation, University of Toronto.

- Pearl, Lisa. 2007. *Necessary bias in natural language learning*. Doctoral dissertation, University of Maryland, College Park.
- Pearl, Lisa. 2020a. Modeling syntactic acquisition. To appear in: *The Oxford Handbook of Experimental Syntax*.
- Pearl, Lisa. 2020b. Poverty of the stimulus without tears. Ms.
- Pearl, Lisa and Amy Weinberg. 2007. Input filtering in syntactic acquisition: Answers from language change modeling. *Language learning and development* 3(1), 43-72.
- Pearl, Lisa and Jeffrey Lidz. 2013. Parameters in language acquisition. In Cedric Boeckx & Kleanthes Grohmann (eds.) *The Cambridge Handbook of Bilingualism*, 129-159.
- Perfors, Amy, Joshua B. Tenenbaum and Terry Regier. 2011. The learnability of abstract syntactic principles. *Cognition* 188, 306-338.
- Pesetsky, David and Esther Torrego. 2007. The syntax and valuation and the interpretability of features. In Simin Karimi, Vida Samilan & Wendy Wilkins (eds.) *Phrasal and clausal architecture: Syntactic derivation and interpretation*: 262-294.
- Phillips, Colin. 2012. Individual variation and constraints on language learning. *Linguistic Approaches to Bilingualism* 2(3), 281-286.
- Phillips, Colin and Lara Ehrenhofer. 2015. *Linguistic Approaches to Bilingualism* 5(4), 409-453.
- Pine, Julian, Daniel Freudenthal, Grzegorz Krajewski and Fernand Gobet. 2013. Do young children have adult-like syntactic categories? Zipf's law and the cases of the determiner. *Cognition* 127(3), 345-360.
- Platzack & Rosegren 1997. On the subject of imperatives: A minimalist account of the imperative clause. *Journal of Comparative Germanic Linguistics* 1(3): 177-224.

- Poletto, Cecilia. 1993. *La sintassi del soggetto nei dialetti italiani settentrionali*. Doctoral dissertation, Università di Padova, Trieste, Udine e Venezia.
- Preminger, Omer. 2009. Breaking agreements: distinguishing agreement and clitic doubling by their failures. *Linguistic Inquiry* 40(4), 619-666.
- Preminger, Omer. 2011. *Agreement as a fallible operation*. Doctoral dissertation, Massachusetts Institute of Technology.
- Preminger, Omer. 2013. That's not how you agree: a reply to Zeijlstra. *The Linguistic Review* 30(3), 419-500.
- Preminger, Omer. 2014. *Agreement and its failures*. Cambridge, M.A.: MIT Press.
- Preminger, Omer. 2018. How to tell a syntactic phenomenon when you see it. Talk given at *Groningen Syntax Workshop*. 17 June, 2018. University of Groningen.
- Preminger, Omer. 2019. What the PCC tells us about “abstract” agreement, head movement, and locality. *Glossa* 4(1), 1-42.
- Preminger, Omer & Maria Polinsky. 2015. Agreement and semantic concord: a spurious unification. Ms.
- Putnam, Michael. 2020. One Feature—One Head: Features as Functional Heads in Language Acquisition and Attrition. In Pedro Guijarro-Fuentes & Cristina Suárez-Gómez (eds.) *New Trends in Language Acquisition Within the Generative Perspective*: 3-25.
- Putnam, Michael and Marjo van Koppen. 2009. C-agreement or something close to it. In Artemis Alexiadou, Jorge Hankamer, Thomas McFadden, Justin Nuger & Florian Schäfer (eds.), *Advances in comparative Germanic syntax*, 41-58. Amsterdam: John Benjamins.
- Raposo, Eduardo. 1986. On the null object in European Portuguese. In Osvaldo Jaeggli & Carmen Silva-Corvalán (eds.) *Studies in Romance Linguistics*, 373-390. Dordrecht: Foris.

- Raposo, Eduardo. 1998. Definite/zero alternations in Portuguese: Towards a unification of topic constructions. In Armin Schwegler, Bernard Tranel, & Myriam Uribe-Extbarria (eds.) *Romance Linguistics: Theoretical Perspectives. Selected papers from the 27th Linguistic Symposium on Romance Languages (LSRL XXVII), Irvine, February 1997*, 197-212. Amsterdam: John Benjamins.
- Raposo, Eduardo. 2000. Clitic positions and verb movement. In João Costa (ed.), *Portuguese syntax: new comparative studies*, 266-297. New York: Oxford University Press.
- Raposo, Eduardo and Juan Uriagereka. 2005. Clitic placement in Western Iberian: a Minimalist view. In Guglielmo Cinque & Richard Kayne (eds.), *The Oxford handbook of comparative syntax*, 639-697. Oxford: Oxford University Press.
- Ramus, Franck and Simon Fisher. 2009. In M.S. Gazzaniga (ed.) *The cognitive neurosciences 4th ed.*, 855-871. Cambridge, M.A.: MIT Press.
- Rezac, Milan. 2008. *The syntax of eccentric agreement: The Person Case Constraint and absolutive displacement in Basque*. *Natural Language & Linguistic Theory* 26(1), 61-106.
- Richards, Marc. 2007. On feature inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38(3), 563-572.
- Richards, Norvin. 1997. *What moves where in which languages?* Doctoral dissertation, Massachusetts Institute of Technology.
- Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.
- Richards, Norvin. 2010. *Uttering Trees*. Cambridge, M.A.: MIT Press.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, M.A.: MIT Press.

- Rizzi, Luigi. 1994. Some Notes on Linguistic Theory and Language Development: The case of root infinitives. *Language Acquisition* 3, 371-393.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: handbook in generative syntax*, 281-337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2013. Core computational principles in natural language syntax. *Lingua* 130, 1-13.
- Rizzi, Luigi and Ian Roberts. 1989. Complex inversion in French. *Probus* 1(1), 1-30.
- Roberts, Ian. 1993. A formal account of grammaticalisation in the history of Romance futures. *Folia Linguistica Historica* 133:1-2.219-258.
- Roberts, Ian. 2004. The C-system in Brythonic Celtic. In Luigi Rizzi (ed.), *The structure of IP and CP (The Cartography of Syntactic Structures 2)*, 297–328. New York, NY: Oxford University Press.
- Roberts, Ian. 2007. *Diachronic Syntax*. Oxford: Oxford University Press
- Roberts, Ian. 2010a. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge, M.A.: MIT Press.
- Roberts, Ian. 2010b. Grammaticalization, the clausal hierarchy and semantic bleaching. In Elizabeth Traugott & Graeme Trousdale (eds.) *Gradience, Gradualness and Grammaticalization*, 45-73.
- Roberts, Ian. 2012. *Verbs and diachronic syntax: A comparative history of English and French*. Dordrecht: Springer.
- Roberts, Ian. 2014. Subject clitics and macroparameters. In Paola Benincà, Adam Ledgeway, and Nigel Vincent (eds.) *Diachrony and dialects: Grammatical change in the dialects of Italy*, 177-201. Oxford: Oxford University Press.

- Roberts, Ian and Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization (No. 100)*. Cambridge: University Press.
- Rodríguez-Modoñedo, Miguel. 2007. *The syntax of objects: Agree and differential object marking*. Doctoral dissertation, University of Connecticut.
- Romero, Juan. 2019. Empty Categories and Clitics. In Ángel Gallego (ed.) *The Syntactic Variation of Spanish Dialects*, 108-126. Oxford: Oxford University Press.
- Rothman, Jason and Noam Chomsky. 2018. Towards eliminating arbitrary stipulations related to parameters: linguistic innateness and that variational model. *Linguistic Approaches to Bilingualism* 8(6), 764-769.
- Saco y Arce, Juan Antonio. 1886. *Gramática gallega*. Lugo: Imprenta de Soto Freire.
- Sheehan, Michelle, Theresa Biberauer, Ian Roberts & Anders Holmberg. 2017. *The Final-Over-Final Condition: A Syntactic Universal*. Cambridge: Cambridge University Press.
- Santamarina, Antón (dir.). 2018. *Tesouro informatizado da lingua galega*. Santiago de Compostela: Instituto da Linga Galega. <http://ilg.usc.gal/TILG/>
- Sheehan, Michelle. 2020. The Romance Person Case Constraint is not about clitic clusters. In Anna Pineda & Jaume Mateu (eds.) *Dative constructions in Romance and beyond*: 143-171. Berlin: Language Science Press.
- Sigurðsson, Halldór Ármann and Anders Holmberg. 2008. Icelandic dative intervention: Person and number are separate probes. In Roberta D'Alessandro, Susann Fischer & Gunnar Hrafn Hrafnbjargarson (eds.) *Agreement restrictions*, 251-280. Berlin: De Gruyter Mouton.
- Speas, Margaret and Carol Tenny. 2003. Configurational properties of point of view roles. In Anna Maria Di Sciullo (ed.) *Asymmetry in Grammar, Vol. I: Syntax and Semantics*. Amsterdam: John Benjamins.

- St. Clair, Michelle, Padraic Monaghan and Morten Christiansen. 2010. Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition* 116(3), 341-360.
- Starke, Michael. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1): 1-6.
- Starke, Michael. 2011. Towards an election solution of language variation. Ms.
- Stevenson, Harold and Morton Weir. 1959. Variables affecting children's performance in a probability learning task. *Journal of Experimental Psychology* 57(6), 403.
- Suñer, Margarita. 1988. The Role of Agreement in Clitic-Doubled Constructions. *Natural Language & Linguistic Theory*, 6(3): 391-434.
- Torrego, Esther. 1992. Case and agreement structure. Ms.
- Torrence, Harold. 2012. The morpho-syntax of silent wh-expression in Wolof. *Natural Language & Linguistic Theory* 30: 1147-1184.
- Traugott, Elizabeth Closs. 2012. Pragmatics and language change. In Keith Allan & Kasia M. Jaszczolt (eds.) *The Cambridge Handbook of Pragmatics*, 549-566. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs and Richard B. Dasher. 2001. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Travis, Catherine. 2006. *Dizque*: A Colombian evidentiality strategy. *Linguistics* 44(6): 1269-1297.
- Tyler, Matthew. 2019. Absolute Promotion and the Condition on Clitic Hosts in Choctaw. *Natural Language & Linguistic Theory* 37: 1145-1203.
- Uriagereka, Juan. 1988. *On Government*. Doctoral dissertation, University of Connecticut.

- Uriagereka, Juan. 1995a. An F position in Western Romance. *Discourse configurational languages* 26.79-123.
- Uriagereka, Juan. 1995b. Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26: Winter, 79-123.
- Uriagereka, Juan. 1996. Determiner clitic placement. In R. Freiden (ed.) *Current Issues in Comparative Grammar: Studies in Natural Language and Linguistic Theory* 35, 257-295. Dordrecht: Springer.
- Uriagereka, Juan. 1999. Multiple spell-out. In Samuel David Epstein & Norbert Hornstein (eds.) *Working Minimalism*: 251-282. Cambridge, M.A.: MIT Press.
- Uriagereka, Juan. 2007. Clarifying the notion “parameter”. *Biolinguistics* 1, 99-113.
- Walkden, George. 2011. Abduction or Inertia? The logic of syntactic change. *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research*, 230-239.
- Walkden, George. 2012. Against inertia. *Lingua* 122, 891-901.
- Weinreich, Uriel, William Labov & Malvin Herzog. 1968. Empirical Foundations for a Theory of Language Change. In Winfred Lehmann & Yakov Malkiel (eds.) *Directions for Historical Linguistics*, 95-195. Austin: University of Texas Press.
- Weisleder, Adrianna and Sandra Waxman. 2010. What’s in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of child language* 37(5), 1089-1108.
- Westergaard, Marit. 2006. Triggering V2: On the frequency of the input cue and possible consequences for language change. Presented at *Workshop on “Inversion and Verb Movement”*, University of Tromsø 30-31 January, 2006.

- Westergaard, Marit. 2008. Acquisition and change: On the robustness of the triggering experience for word order cues. *Lingua* 118(12): 1841-1863.
- Westergaard, Marit. 2009a. Microvariation as diachrony: A view from acquisition. *The Journal of Comparative Germanic Linguistics* 12(1): 49-79.
- Westergaard, Marit. 2009b. *The acquisition of word order: Micro-cues, information structure, an economy*. Amsterdam: John Benjamins.
- Westergaard, Marit. 2013. The acquisition of linguistic variation. In Terje Lohndal (ed.) *In Search of Universal Grammar: From Old Norse to Zoque*: 275-298. Amsterdam: John Benjamins.
- Westergaard, Marit. 2014. Linguistic variation and micro-cues in first language acquisition. *Linguistic Variation* 14(1): 26-45.
- Wexler, Kenneth. 2004. Theory of Phasal Development: Perfection in Child Grammar. In Aniko Csirmaz, Andrea Gualmini & Andrew Nevins (eds.) *MITWPL 48: Plato's Problems: Papers on Language Acquisition*, 159-209. Cambridge, M.A.: MIT Working Papers in Linguistics.
- Wexler, Kenneth. 2014. Tough-Movement Developmental Delay: Another Effect on Phasal Computation. In Massimo Piattelli-Palmarini & Robert Berwick (eds.) *Rich Languages From Poor Inputs*, 146-167.
- Wimmer, Heinz & Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1), 103-128.
- Wolfe, Sam. 2018. *Verb Second in Medieval Romance*. Oxford: Oxford University Press.

- Wood, Jim, Laurence Horn, Raffaella Zanuttini and Luke Lindemann. 2015. The Southern Dative Presentative Meets Mechanical Turk. *American Speech* 90(3), 291-320.
- Woolford, Ellen. 2010. Active-stative agreement in Choctaw and Lakota. *Revista Virtual de Estudos da Linguagem* 8: 6-46.
- Wurmbrand, Susanne. 2011. Reverse Agree. Ms.
- Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
- Yang, Charles. 2005. On productivity. *Linguistic variation yearbook* 5(1), 265-302.
- Yang, Charles. 2016. *The price of linguistic productivity. How children learn to break the rules of language*. Cambridge, M.A.: MIT Press.
- Yang, Charles. 2018. A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 8(6), 665-706.
- Yuan, Michelle. 2020. Dependent case and clitic dissimilation in Yimas. *Natural Language & Linguistic Theory* 38: 937-985.
- Yuan, Michelle. 2021. Diagnosing Object Agreement vs. Clitic Doubling: An Inuit Case Study. *Linguistic Inquiry* 52(1): 153-179.
- Zanuttini, Raffaella. 2017. Presentatives and the syntactic encoding of contextual information. In Enoch Aboh, Eric Haeberli, Genoveva Puskás & Manuela Schönenberger (eds.) *Elements of Comparative Syntax: Theory and Description*, 221-255. Boston/Berlin: de Gruyter Mouton.
- Zdrojewski, Pablo and Liliana Sánchez. 2014. Variation in accusative clitic doubling across three Spanish dialects. *Lingua* 151: 162-176.
- Zeijlstra, Hedde. 2012. There Is Only One Way to Agree. *Linguistic Review* 29, 491-553.

Zwart, Jan-Wouter. 1993. *Dutch Syntax: A Minimalist Approach*. Doctoral dissertation, University of Groningen.

Zwart, Jan-Wouter. 1997. *Morphosyntax of verb movement: A minimalist approach to the syntax of Dutch*. Vol. 39. Springer.

Zwicky, Arnold and Geoffrey Pullum. 1983. Cliticization vs. Inflection: The Case of English *n't*. *Language* 59(3): 502-513