LONG DISTANCE AGREEMENT IN SPANISH

MA Thesis

Irene Fernández Serrano

Supervisors: Ángel J. Gallego & M. Lluïsa Hernanz

Master in Cognitive Science and Language

September 2017

UAB

Universitat Autònoma
de Barcelona

A la memoria de mi abuelo,
Pedro Serrano.

Abbreviations

Used in glosses:

1:	first person
2:	second person
3:	third person
ACC:	accusative
DAT:	dative
IND:	indicative
INF:	infinitive
NOM:	nominative
PL:	plural
REL:	relative
SBJ:	subjunctive
SG:	singular

Other abbreviations:

%:	variety with no clear dialectal tendency
ASALE:	Asociación de Academias de la Lengua Española (Association of Spanish Language Academies)
EA:	External Argument
GDLE:	*Gramática Descriptiva de la Lengua Española* (Bosque & Demonte (eds.) 1999)
H-Agr:	hyper-agreement
IA:	Internal Argument
LDA:	Long Distance Agreement
PIC:	Phase Impenetrability Condition
RAE:	Real Academia Española (Royal Academy of Spanish Language)

TABLE OF CONTENTS

1   INTRODUCTION

The nature of agreement has raised a lot of interesting questions in various generative frameworks, as it is an apparently 'non-efficient' device, given that a lot of languages have a poor inflectional paradigm (English, Chinese, Korean, to name but a few). A specific subcase of this phenomenon that has been called 'Long Distance Agreement' (LDA) (see examples in (1)) has given rise to a fairly amount of literature and discussion as it challenges the Spec-Head based approaches of early generative frameworks and seems to support the *Agree* operation proposed by Chomsky in the Probe-Goal framework of the Minimalist Program (cf. Boeckx 2008a et seq., Etxepare 2006, 2012, Preminger 2009 et seq., i.a.).

(1)   a. There seem to be likely to be three men here   [From Boeckx 2009: (2)]

   b. Vivek-ne   [**kitaab parh-nii**]   **chaah-ii**
      Vivek-Erg   book.F read-Inf.F   want-Pfv.FSg
      'Vivek wanted to read the book.'   [From Bhatt 2005: (4)]

In this thesis, we explore Spanish data that show number agreement across a clausal boundary, both finite and non-finite, as examples in (2) respectively reveal:

(2)   a. me   **molestan** [que **las cosas** no salgan como YO quiero][1]   (Argentina[2])
      me$_{DAT}$ bother$_{3PL}$ that the things no come.out$_{3PL}$ as I want
      'It upsets me when things don't come out as I'd like'

   b. Me   **duelen** [recordar   **estas cosas** de nuevo...]   (Spain)
      me$_{DAT}$ hurt$_{3PL}$ remember$_{INF}$ these things of new
      'It hurts me to remember these things again'

We will dub the relevant dependency 'hyper-agreement' (H-Agr), for it resembles 'hyper-raising' situations (Martins & Nunes 2005, 2009), attested in Brazilian Portuguese but impossible in English or Spanish:

(3)   a. As crianças$_i$ parecem [que t$_i$ gostam da babá]

                                                    [from Martins & Nunes 2009: (2)]

   b. *The children$_i$ seem [that they$_i$ like the babysitter]

   c. *Los niños$_i$ parecen [que t$_i$ adoran a la niñera]

---

[1] The agreeing elements are boldfaced in all the examples.
[2] I mark the examples from Twitter with the country of origin between brackets and I will copy them as they were written by the users (including spelling errors). The rest of examples are from the literature (indicated between squared brackets) or my judgments (without any indication).

Just as raising goes beyond what is expected in the examples in (3), agreement surpasses its standard boundaries in Spanish in the examples in (2). With this terminology, we also differentiate the cases we will study (crucially cross-clausal and not attested in the 'normative/standard' grammar of Spanish) from the ones such as (4) where the post verbal subject is far removed from the main verb:

(4) Ha   empezado al   final del año   a venir **Juan** solo a la escuela
     have$_{3SG}$ begun towards the end of the year to come John alone   to the school
     'John began coming alone to school towards the end of the year.'

[from Alexiadou et al. 2012: (64a)]

It is important to remark that this phenomenon has not been carefully studied, yet Martínez, in his chapter about agreement in the Spanish descriptive grammar (Bosque & Demonte 1999) (GDLE), briefly mentions it:

(5) a. A mí me   chifl*an*   oír   *esas canciones*
        to me me$_{DAT}$ whistle$_{3PL}$ hear$_{INF}$ those songs
        'Listen to those songs make me go nuts'

     b. Me   gust*an* más estudiar   *otras asignaturas.*
        me$_{DAT}$like$_{3PL}$ more study.inf other subjects
        'I prefer to study other subjects'

     c. Ya   sé   que te   duel*en* tener   que desmentir *esos rumores*.
        Already know that you$_{DAT}$ hurt$_{3PL}$ have$_{INF}$ that deny   those rumours
        'I know it is painful for you having to deny those rumours'

[from Martínez 1999: (206); GDLE §42.10.1.4]

The data in (5) prove that this phenomenon has been already observed in Spanish speakers; nevertheless, Martínez does not indicate the source of the data or whether it belongs to a specific Spanish variety.

Given the poor presence of the phenomenon in the literature, our first aim is empirical. We want to determine which patterns of H-Agr seem to be possible and if there is a specific dialectal distribution of the facts. In order to do so, we have carried out a search on different corpora and web resources of which Twitter has been the most useful. This social network yields a significant amount of examples, despite the rather 'rare' status of the phenomenon.

Our second aim is theoretical, we want to verify if H-Agr is indeed an instance of LDA, comparable to those attested in other languages. For this purpose, we have reviewed previous approaches to LDA and we have formulated different hypotheses that may account for Spanish H-Agr data. Our intention is not to argue for a definitive analysis but

to explore different alternatives that may conduce to a more thorough comprehension of the phenomenon and open the path for a future inquiry.

Regarding these objectives, we consider this investigation to have at least two main implications. Firstly, it presents unexplored data of different Spanish dialects that will contribute to a typology of LDA and to the description of Spanish syntax. This empirical objective also has a clear methodological impact, as it shows that new platforms such as social networks are a valuable source of information for linguistic analyses.

Secondly, LDA raises theoretical questions about the nature of agreement and its variation, a long-standing topic within formal and typological approaches to language (cf. Baker 2013 and references therein). LDA has already been explored in other languages such as Hindi-Urdu (Bhatt 2005, Boeckx 2008a, i.a.), Icelandic (Sigurðsson 1991 et seq., Boeckx 2000, 2008a, i.a.), Basque (Etxepare 2006, 2012, Preminger 2009) or different Algonquian languages (Bruening 2001, Bošković 2007, Lochbihler 2012, i.a.) Spanish H-Agr seems to share some characteristics with these languages but also presents some peculiarities of its own. These peculiarities make it necessary to revisit preceding proposals, especially in relation to theoretical notions such as the PIC (cf. Chomsky, Gallego & Ott 2017), and, in general, to rethink the status of clausal dependencies in Spanish.

The thesis is organized as follows: in section 2, we introduce H-Agr in Spanish. Firstly, we define the patterns that allow this dependency, basically psych-verbs with a DAT-NOM configuration. Secondly, we present the main characteristics of the data collected, both for non-finite and finite structures. Thirdly, we discuss the (dis)advantages of using Twitter as a corpus and describe the procedure that has been used in our study. In Section 3, we review the key theoretical notions behind LDA proposals. Then, we offer an overview of LDA literature, with special interest in Icelandic, Basque and Algonquian languages. Section 4 presents different proposals for an analysis. 5 concludes and suggests some lines for further investigation.

2   LDA IN SPANISH

To look for LDA instances in Spanish, we need to focus on structures whereby T agrees with the IA in mono-clausal structures and check if agreement holds across an embedded clause. Unlike other languages that display LDA, such as German (Polinsky 2003), Hindi (Bhatt 2005), Chamorro (Chung 2004) and Basque (person LDA[3]; Etxepare 2012), restructuring predicates are not candidates in Spanish for such dependency, as (6) shows:

(6)   a. (**Yo**) **puedo/quiero** ver   las películas
          I   can/want$_{1SG}$ see$_{INF}$ the movies

       b. *Yo **pueden/quieren** ver   **las películas**
          I   can/want$_{3PL}$ see$_{INF}$ the movies
          'I can/want to see the movies'

       c. (**Ellos**) **pueden/quieren** ver   las películas
          I   can/want$_{3PL}$ see$_{INF}$ the movies
          'They can/want to see the movies'

When the verb is inflected in plural it agrees with a plural subject, as we see in (6c), even if it is covert, as Spanish is a pro-drop language, but never with a plural object, as the ungrammaticality of (6b) reveals.

We have to turn then to non-agentive structures in which T agrees with the IA, which is typically found with unaccusative-like constructions involving passive *se* or existential/psychological predicates. Consider the following examples:

(7)   a. Se vende flores
          SE sell$_{3SG}$ flowers

       b. Se venden flores
          SE sell$_{3PL}$ flowers
          'Flowers are sold'

[From Sánchez-López 2002: (25)]

(8)   a. Había   unas mujeres en la fiesta
          there.was   some women in the party

       b. Habían   unas mujeres en la fiesta
          there.were   some women in the party
          'There where some women at the party'

[From Rodríguez-Mondoñedo 2005: (27)]

---

[3] Etxepare (2012) distinguishes two subcases of LDA in Basque: one where agreement is only in person, is possible with restructuring predicates and obeys the typical restrictions of clitic climbing; and another where agreement is only in number, not possible with restructuring predicates. We will basically focus on the latter, which is more similar to the Spanish scenario.

(9) a. A Matilda le   gusta el libro
      to Matilda her$_{DAT}$ like$_{3SG}$ the book
      'Matilda likes the book'

    b. A Matilda le   gustan los libros
      to Matilda her$_{DAT}$ like$_{3PL}$ the book
      'Matilda likes books'

The difference in agreement of the pairs in (7)-(9) does not have the same 'status'. Whereas it is compulsory with psych-verbs such as gustar 'to like' (9), it is optional in *se* impersonals (7), and it is dialectally restricted in *haber* existentials (8). Crucially, any of these structures allow a DP-agent: the *se* pronoun has suppressed that role (cf. D'Alessandro 2007, Ordóñez & Treviño 2011, i.a.) assigning it to an arbitrary or generic subject (compare it with the personal counterpart in ((10) below), the verb *haber* is impersonal (as we see in (11a) and can alternate with *estar* (11b), cf. Rodríguez-Mondoñedo 2005) and the verb *gustar* requires a dative DP that bears the experiencer role.

(10)   Marta vende flores
         Marta sell$_{3SG}$ flowers
         'Marta sells flowers'

(11) a. Hay   unas mujeres en la fiesta
        there.is some women in the party
        'There are some women at the party'

    b. Unas mujeres están en la fiesta
       some women   are in the party
       'Some women are at the party'

There are other structures in Spanish that deploy a DAT-NOM configuration without requiring ACC (departing from what Marantz's 1991 Dependent Case approach would lead us to believe) such as predicates of change (12a), (12b), existentials (12c) and inchoatives (12d):[4]

(12) a. A los chicos les   llegó   una carta
        to the kids   them$_{DAT}$ arrived$_{3SG}$ a letter
        'The kids got a letter'

---

[4] We do not include ditransitives (i), nor transitive verbs that can select a dative argument (ii) (cf. Cuervo 2010a), because in those cases the verb agrees with the the EA (subject/agent):
   (i) Pablo le   besó   la mano a la reina   / * Pablo le   besaron   las manos a la reina
       Pablo her$_{DAT}$ kissed$_{3SG}$ the hand to the queen / Pablo   her$_{DAT}$ kissed$_{3PL}$ the hands to the queen
       'Pablo kissed the queen's hand   / 'Pablo kissed the queen's hands'
   (ii) Marge le   envió   una carta a Ringo / *Marge le   enviaron unas cartas a Ringo
       Marge him$_{DAT}$ sent$_{3SG}$ a letter   to Ringo / Marge   him$_{DAT}$ sent$_{3PL}$ some letters to Ringo
       'Marge sent Ringo a letter   / 'Marge sent Ringo some letters'

    b. A los chicos les   creció rápido el pelo
       to the kids   them$_{DAT}$ grew$_{3SG}$ quickly the hair
       'The kids' hair grew quickly'

    c. Al   libro le   falta   una página
       to.the book it$_{DAT}$ lack$_{3SG}$ a page
       'The book is missing one page'

    d. A Vera se le   rompió el televisor
       to Vera SE her$_{DAT}$ broke$_{3SG}$ the TV$_{NOM}$
       'The TV broke on Vera'   [adapted from Cuervo (2010a), (2010b)]

Like psych-verbs, all these predicates obligatorily agree with their IA. Compare the examples in (12) above with a singular IA, with their plural counterparts in (13):

(13) a. A los chicos les   **llegaron** <u>unas cartas</u>
       to the kids   them$_{DAT}$ arrived$_{3PL}$ some letters
       'The kids got some letters'

    b. A los chicos les   **crecieron** rápido <u>los pelos</u>
       to the kids   them$_{DAT}$ grew$_{3SG}$ quickly the hair$_{PL}$
       'The kids' hair grew quickly'

    c. Al   libro le   **faltan** unas páginas
       to.the book it$_{DAT}$ lack$_{3PL}$ some pages
       'The book is missing some pages'

    d. A Vera se le   **rompieron** <u>los televisores</u>
       to Vera SE CL$_{DAT}$ broke$_{3PL}$ the TV$_{NOM}$
       'The TVs broke on Vera'

However, the structures in (12)-(13) are not eligible for LDA in Spanish because none of these predicates can select a clausal argument[5]. Up to this point, we can summarize the properties of LDA predicates in Spanish as the ones in (14):

(14) V {EXPERIENCER, THEME}

      EXPERIENCER = EA, dative XP

      THEME = IA, T agrees with it, can be a clause

---

[5] An apparent counterexample is the case of the verb *faltar* 'lack', which can select a clausal argument with a different meaning, namely when something is needed or required to be done. This is also valid for the predicate *quedar* 'remain', but is not that clear for *sobrar* 'be-extra', from Cuervo's (2010b) classification.
    (i) a. Aún me falta [leer cinco monografías]   (Peru)
         'I still have to read five monographies'
      b. Todavía me queda [leer como 200 páginas y hacer una síntesis]   (Chile)
         'I still have to read around 200 pages and do a summary'
      c. ??Me sobra [leer tantos capítulos]
         'It's too much for me to read so many chapters'
      d. Me basta y me sobra [leer sus tuits] para partirme de risa   (Spain) (with an idiom)
         'Reading your tweets is enough to make me crack up'
We will treat these predicates as psych-verbs solely with these interpretations.

{ 6 }

The psych-verbs of the *gustar* type match these criteria as we can see in (15), so we would expect to find instances of LDA such as the ones in (16):

(15)   a. A Matilda le   gustan los libros   REGULAR T-IA AGREEMENT
          to Matilda her$_{DAT}$ like$_{3PL}$ the books
          'Matilda likes books'

       b. A Matilda le   gusta [leer   libros]   NON-FINITE CLAUSAL ARGUMENT
          to Matilda her$_{DAT}$ like$_{3SG}$ read$_{INF}$ books
          'Matilda likes reading books'

       c. A Matilda le   gusta [que los niños lean libros]   FINITE CLAUSAL ARGUMENT
          to Matilda her$_{DAT}$ like$_{3SG}$ that the children read$_{3PL}$ books
          'Matilda likes that the children read books'

(16)   a. A Matilda le   **gustan** [leer **libros**]   LDA ACROSS NON-FINITE CLAUSE
          to Matilda her$_{DAT}$ like$_{3PL}$ read$_{INF}$ books
          'Matilda likes reading books'

       b. A Matilda le   **gustan** [que **los niños** lean libros]   LDA ACROSS FINITE CLAUSE[6]
          to Matilda her$_{DAT}$ like$_{3PL}$ that the children read$_{3PL}$ books
          'Matilda likes that the children read books'

Let us see in more detail the characteristics of psych-predicates in Spanish (cf. Marín 2015 and references therein for discussion). According to Belletti & Rizzi's (1988) classification, there are three types of psych-verbs depending on the syntactic role and case that the experiencer bears[7]:

(17)   I. *temer*-type: subject nominative experiencer
       II. *preocupar*-type: object accusative experiencer
       III. *gustar*-type: dative experiencer

We illustrate this classification in (18) below. While the ones belonging to *temer*-type behave as standard transitive verbs (such as 'like' in English), the *preocupar* and *gustar*-type display the configuration that we have been discussing, a dative experiencer and a nominative object. The particularity of the former is that it alternates with an accusative pattern, as we see in (18b) and (18b').

(18)   a. Diana {*ama*/*odia*/*teme*/*adora*/*detesta*/…} las tormentas
           'Diana {loves/hates/fears/adores/hates} storms'

---

[6] We will assume that even though the IA of the embedded verb is also plural, the matrix verb agrees with the EA for reasons that we will discuss later.

[7] The classification in (15) is virtually equivalent to the Italian classification (*temere, preocuppare, piacere*) (cf. Acedo-Matellán & Mateu 2015), the possible differences (pointed out by Franco 1992, among others) are not relevant for our discussion.

b. A Diana le {*preocupan/molestan/sorprenden/asustan/…*} las tormentas

b'. Las tormentas (le) {*preocupan/molestan/sorprenden/asustan/…*} a Diana
'Storms {worry/annoy/surprise/frighten} Diana'

c. A Diana le {*gustan/encantan/desagradan/…*} las tormentas
'Diana {likes/loves/dislikes} storms'

In our discussion, we will only focus on the second and third types, as they are the ones that match the criteria in (14)[8].

In relation to word order: different authors have proved that dative arguments are not left-dislocated topics in an A-bar position (Masullo 1993, Campos 1999, i.a; contra Andueza 2012); we will assume this word order[9] (cf. 19a) and consider that when the theme appears in pre-verbal position it has been topicalized (cf. 19b).

(19) a. A Pablo le encantan los idiomas
'Pablo loves languages'

b. LOS IDIOMAS le encantan a Pablo
'Pablo loves LANGUAGES'
[from Campos 1999: (139)-(140); GDLE §24.3.7]

So far little has been said about syntactic roles because there are different positions about the status of the dative argument and of the DP in IA position (cf. Cuervo 2003, Pesetsky 1995, and references therein). There is wide consensus on the fact that the DP is a theme, and that it receives nominative case (Belletti & Rizzi 1988, Franco 1992, Cuervo 2010b[10], i.a.). Regarding its syntactic status, this DP has been claimed by traditional grammars to be the subject of the clause[11], precisely because it agrees with the verb[12].

---

[8] From now on, every time we refer to psych-verbs we will be excluding *temer*-type for simplification.

[9] See Mendívil 2012: 175 for a summary of the different tests proposed in the literature.

[10] Acedo-Matellán & Mateu (2015) argue contra Cuervo (2010) that not all psych-verbs are unaccusatives, but *preocupar*-type ones are unergatives. Therefore, the IA is not always a theme. We will not enter this debate because it is not directly relevant for our discussion.

[11] Even though, as Fernández-Soriano & Táboas (1999) point out, some authors have also claimed that psych-verbs are impersonal (RAE 1973, Seco 1988) or 'pseudo-impersonal' predicates (Alcina & Blecua 1975), among other reasons because they admit adverbs, which cannot be subjects:
(i) **Así** me gusta, que te portes bien
'That's how I like it, that you behave'   [from Fernández-Soriano & Táboas 1999: (136)]

[12] As different authors have noted (Masullo 1993, Fernández-Soriano 1999, Rivero 2004, Gutiérrez-Bravo 2006, López 2007, Mendívil 2012, i.a.), these *gustar*-type dative arguments in Spanish resemble the well-known Icelandic 'quirky subjects', which are "oblique arguments that otherwise behave like surface subjects in every relevant respect" (Gutiérrez-Bravo 2006: 1). However, there is an unresolved debate about whether the analysis should be the same or not –i.e. if Spanish dative arguments are truly quirky subjects (for example, Cuervo 2010b argues in favor of dative DPs subjecthood, while Gutierrez-Bravo 2006 shows evidence against this claim). This debate does not directly affect our discussion, but interestingly relates Spanish with Icelandic, a language that has been claimed to have LDA (cf. §3.2.2).

As we are dealing with bi-clausal structures, we should point out some properties of the clausal arguments selected by psych-predicates. On the one hand, it is worth noting that the (non-)finiteness status of the clause depends on the embedded subject. The example in (20a) below shows that the subject of the infinitive clause can be recovered from the matrix clause (Hernanz 1999; GDLE §36.2.2.2) as it obligatorily corefers with the dative. By contrast, in (20b) when the clause is finite, the verb is always in subjunctive mood and the subject cannot corefer with the experiencer of the matrix clause.

(20) a. A Matilda$_i$ le$_i$ gusta [PRO$_{i/*j}$ leer   libros]
          to Matilda her$_{DAT}$ like$_{3SG}$ read$_{INF}$ books
          'Matilda likes reading books'

      b. A Matilda$_i$ le$_i$ gusta [que (ella$_{*i/j}$) lea   libros]
          to Matilda her$_{DAT}$ like$_{3PL}$ that she   read$_{3SG}$ books
          'Matilda likes that the children read books'

On the other hand, we can wonder if the clausal argument keeps the same status as the nominative DP in its mono-clausal counterpart. In other words, if the clause receives nominative Case and agrees with matrix T. Picallo (2002) argues that CPs are endowed with ϕ-features and Case. Consequently, they can agree with matrix T (or be selected by v when they are objects). She suggests that the matrix verb shows default 3rd person singular morphology because those features are negative[13].

(21)

$$[_{TP} \text{Subj}_{[[\pm C][\pm P, \pm N, \pm G]]} [T_{[P, N, G]}] [_{vP} \underline{v}_{[P, N, G]}] \text{V Obj}_{[[\pm C][\pm P, \pm N, \pm G]]}]]$$

[From Picallo 2002: (40)]
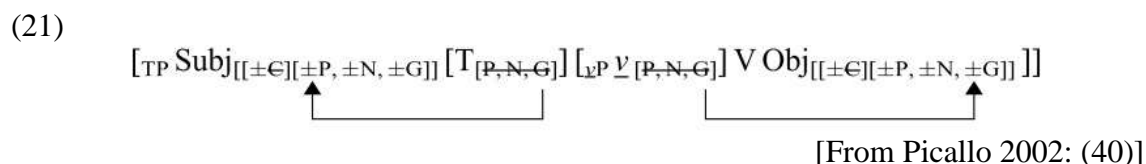
As Picallo remarks, this hypothesis would explain the (im)possibility for embedded subjects to be the goal of an agreement relation: it is impossible when there is a C (like in ECM and raising constructions). We will bear this hypothesis in mind for our proposal.

Another possibility concerning the argumental status of clauses was suggested by Torrego & Uriagereka (1992) (see Rivero (1971) for similar ideas). They claimed that indicative finite clauses are in apposition with a null pronoun, while subjunctive clauses are real arguments of the matrix verb, as it is shown in (22):

---

[13] Quer (2008), on the other hand, argues that the possibility that the default morphology is precisely result of the lack of ϕ-features on clausal arguments cannot be easily excluded. We understand then that Quer does not consider clausal arguments to agree, even though he is not explicit about it and leaves the possibility "open for further research".

(22) a. [$_{vP}$ . . . V . . . [$_{DP}$ pro] [$_{CP}$ V$_{INDICATIVE}$]]

  b. [$_{vP}$ . . . V . . . [$_{CP}$ V$_{SUBJUNTIVE}$]]

[From Gallego & Uriagereka 2011: (3)]

Taking into account that the finite argument clauses selected by psych-predicates in Spanish are obligatorily subjunctive, this proposal may also have consequences in the analysis of H-Agr, as we will see later.

In this section, we have introduced the properties of the predicates that deploy H-Agr. We have focused on non-agentive predicates, namely predicates that in Spanish require a dative argument in preverbal position and a postverbal nominative argument with which it agrees. Psych-verbs of the *gustar-* and *preocupar*-types present these characteristics given that other predicates that demand a dative argument cannot select a clausal argument. We have seen that these psych-verbs normally deploy a 3rd person singular morphology when their IA is a clause. H-Agr happens when they agree in number with the IA of the embedded clausal argument. In the following lines, we present the different patterns of this phenomenon.

2.1 The data

In this section, we show a sample of H-Agr data extracted from Twitter. Firstly, we consider instances of LDA across non-finite clauses and then we will show examples of LDA across finite clauses ('non-finite H-Agr' and 'finite H-Agr' respectively from now on). As we are dealing with unattested data, we also devote some lines to discuss the methodology and results of our search including its possible drawbacks.

2.1.1 NON-FINITE HYPER-AGREEMENT

There seems to be no lexical restriction in the psych-predicates that allow H-Agr:

(23)[14] a. Me ***agobian*** [tener tantos **deberes**...]   (Spain)
  me$_{DAT}$ overwhelm$_{3PL}$ have$_{INF}$ so.much homework
  'Having so much homework to do stresses me out…'

  b. Me ***decepcionan*** [escuchar **las voces** verdaderas de los simpsons] (Argentina)
  me$_{DAT}$ disappoint$_{3PL}$ listen.to$_{INF}$ the voices actual of the Simpsons
  'I find it disappointing to listen to the actual Simpsons' voices'

---

[14] The agreeing elements are always in bold, while we mark in italics the element we want to highlight in every example.

    c. Me ***impresionan*** [ver **imágenes** de las protestas] (Mexico)
       me$_{DAT}$ impress$_{3PL}$ see$_{INF}$ images of the protests
       'I find it impressive to see the images of the protests'

    d. Casi que me ***apetecen*** [empezar **las clases**] (Spain)
       almost that me$_{DAT}$ feel.like$_{3PL}$ start$_{INF}$ the classes
       'I almost feel like starting classes'

    e. Como me ***fascinan*** [encontrar **versiones** instrumentales de mis canciones
       how me$_{DAT}$ fascinate$_{3PL}$ find$_{INF}$ versions instrumental of my songs

       favoritas] (Mexico)
       favorite
       'It really fascinates me finding instrumental versions of my favorite songs'

We have found instances with almost every predicate fulfilling the conditions in (14). The ones with no results may be too formal for the colloquial contexts we are considering. Inversely, we have found instances with predicates that have become of the psych-type in spoken Spanish, such as the verb *chocar* 'crash' that with a dative experiencer means 'to cause astonishment or anger'[15], *matar* 'kill' or *reventar* 'blow up' that become 'be terribly annoying', as well as colloquial dialectal expressions such as *molar* 'be-cool' or *flipar* 'amaze' in Spain, *copar* 'amaze' in Argentina, etc.

(24) a. Me ***chocan*** [ver **estas noticias** en la tele] (Argentina)
       me$_{DAT}$ crash$_{3PL}$ see$_{INF}$ these news on the TV
       'It impresses me to see this news on TV'

    b. Me ***flipan*** [ver **mapas antiguos**] […] (Spain)
       me$_{DAT}$ amaze$_{3PL}$ see$_{INF}$ maps old
       'I find it amazing to see old maps'

    c. Me ***matan*** [hacer **las cosas** corriendo] (Spain)
       me$_{DAT}$ kill$_{3PL}$ do$_{INF}$ the things running
       'It kills me when I have to do things in a rush'

    d. Me ***revientan*** [tener **exámenes** por la tarde!] (Spain)
       me$_{DAT}$ blow.up$_{3PL}$ have$_{INF}$ exams for the afternoon
       'Having exams in the afternoon pisses me off'

(25) a. Me ***molan*** [leer **indirectas** por el Twitter] (Spain)
       me$_{DAT}$ be.cool$_{3PL}$ read hints for the Twitter
       'I find it cool to read hints on Twitter'

    b. Saludos desde Argentina, me ***copan*** [ver **tus videos**!!] (Argentina)
       greetings from Argentina me$_{DAT}$ amaze$_{3PL}$ watch your videos
       'Greetings from Argentina, I really love watching your videos!!'

---

[15] Definition translated from *Diccionario de la Lengua Española (DLE)* (RAE-ASALE 2014).

Some light verb + noun clusters/complexes also behave as psych-verbs (Hernanz 1999; GDLE §36.3.2.1), so they accept H-Agr. See examples of these expressions with the light verbs *dar* 'give' in (26), *hacer* 'do' in (27)[16].

(26) a. Me ***dan miedo*** [ver **sus cejas pintadas** con marcador] (Ecuador)
   me$_{DAT}$ give$_{3PL}$ fear see$_{INF}$ his/her eyebrows painted with marker
   'It frightens me to see his/her eyebrows painted with marker'

   b. me ***dan verguenza*** [ver **las propagandas** del gobierno […] (Uruguay)
   me$_{DAT}$ give$_{3PL}$ shame   see$_{INF}$ the propaganda of.the government
   'I am ashamed of seeing government's propaganda'

   c. Me ***dan coraje*** [ver **estas imágenes**] (Guatemala)
   me$_{DAT}$ give$_{3PL}$ courage see$_{INF}$ these images
   'It makes me angry to see these images'

(27) a. te haces mayor… ya no te ***hacen ilusión*** [tener **esas melenas**] (Spain)
   you$_{DAT}$ do$_{2PL}$ old anymore no you$_{DAT}$ make$_{3PL}$ illusion have$_{INF}$ these mane$_{PL}$
   'You are getting old… you aren't thrilled anymore about having such a mane'

   b. Me ***hacen gracia*** [leer **textos** de hace años] […] (Spain)
   me$_{DAT}$ do$_{3PL}$ funny read$_{INF}$ texts of do$_{3SG}$ years
   'I find it funny to read texts from some years ago'

   c. Me ***hacen falta*** [tener más **tatuajes**] (Venezuela)
   me$_{DAT}$ do$_{3PL}$ lack have$_{INF}$ more tattoos
   'I need to have more tattoos'

We also expect to find instances of H-Agr with periphrases, either as matrix (28) or embedded verb (29):

(28) a. Como te ***pueden gustar*** [ver **dientes** de una persona??] (Argentina)
   how you$_{DAT}$ can$_{3PL}$ like$_{INF}$ see$_{INF}$ teeth of a person
   'How can you like seeing people's teeth?'

   b. me ***tendrian que dar verguenza*** [hacer **muchas cosas** que hago] (Uruguay)
   me$_{DAT}$ would.have.to$_{3PL}$ that give$_{INF}$ shame do$_{INF}$ a.lot things that do$_{1SG}$
   'I would have to be ashamed of doing many of the things I do'

   c. me ***siguen aburriendo*** [hacer **introducciones y conclusiones**] (Venezuela)
   me$_{DAT}$ keep$_{3PL}$ boring   do$_{INF}$ introductions and conclusions
   'Writing introductions and conclusions still bores me'

---

[16] Causative constructions can also be collapsed with psych verbs (see Franco 1992), both with *hacer* 'make' (ia) and *poner* 'get' (ib) and in both cases they seem to accept H-Agr (ii):
   (i) [Escuchar esta canción] **me hace llorar / me pone triste**
      'Listen to this song makes me cry/feel sad'
   (ii) a. Me ***hacen reir*** [leer **esos tweets**] (Ecuador)
      'Reading those tweets make me laugh'
    b. A mí me ***ponen nervioso*** [escuchar las alarmas…] (Spain)
      'I get nervous when I here the alarms'
We leave these examples for future discussion.

(29) a. Ya sé   que te   **duelen** [*tener que desmentir* **esos rumores**]   (cf. (5))
Already know that you$_{DAT}$ hurt$_{3PL}$ have$_{INF}$ that deny   those rumours
'I know it is painful for you having to deny those rumours'

    b. Como me   **gustan** [*poder ver* **estos festejos**   tan patrios]   (Mexico)
how   me$_{DAT}$ like$_{3PL}$ can$_{INF}$ see$_{INF}$ these festivities so national
'How I like being able to see these national festivities'

Most of the examples we have seen so far deploy a 1st person singular dative clitic, but H-Agr is possible with other dative pronouns[17] (see (30) below). That is expected because the dative phrase does not interfere in the T-embedded IA agreement –in other words, it does not trigger defective intervention (Chomsky 2000, 2001).

(30) a. *Te*   **gustan** [ver   **películas** de terror]   (Venezuela)
you$_{DAT}$ like$_{3PL}$ watch$_{INF}$ movies   of terror
'You like watching horror movies'

    b. No *le*   **gustan** [ver   **las películas** subtituladas]   (Argentina)
No him/her$_{DAT}$ like$_{3PL}$ watch$_{INF}$ the movies   subtitled
'S/he does not like watching subtitled films.'

    c. *Nos*   **gustan** [ver   este tipo de **publicaciones**]   (Spain)
us$_{DAT}$ like$_{3PL}$ see$_{INF}$ this kind of publications
'We like seeing these kinds of publications'

    d. *Os*   **gustan** [ver   **chorradas** en la tele]   (Spain)
you$_{PL.DAT}$ like$_{3PL}$ watch$_{INF}$ rubbish   on the TV
'You like watching rubbish on TV'

    e. *Les*   **gustan** [ver   **los partidos** de fútbol][18]   (Spain)
them$_{DAT}$ like$_{3PL}$ watch$_{INF}$ the matches of football
'They like watching football matches'

Having focused on the main predicates, we turn now to the other participant in the agreement relation, the embedded IA. The examples below show that agreement can also target the embedded DP when it is introduced by a preposition, either because the object is differentially marked (31) or because the verb requires a *bona fide* prepositional object (32) (see Gallego 2016 for discussion of similar data).

(31) a. Odio los zoológicos, no me **gustan** [ver *a* **los animales** drogados y mal
hate the zoos, no me$_{DAT}$ like see$_{INF}$ to the animals drugged and badly
alimentados]   (Argentina)
fed
'I hate zoos, I don't like to see drugged and malnourished animals'

---

[17] We attribute the fact that most of examples have a 1st person singular experiencer to extra-linguistic reasons, since Twitter is mostly used to express opinions.

[18] It could also be considered that in examples such as (33c), (33d) and (33e) the verb agrees with plural the experiencer that is in plural. We leave this possibility for further research.

b. A mi me   **interesan** [conocer *a* **mis compañeros**]   (Argentina)
      to me me$_{DAT}$ interest$_{3PL}$ know$_{INF}$ to my classmates
      'I am interested in getting to know my classmates'

(32) a. me   **cuestan** [creer   *en* **todas las palabras** que me   dijo]   (Peru)
      me$_{DAT}$ cost$_{3PL}$ believe$_{INF}$ in all   the words   that me$_{DAT}$ said$_{3SG}$
      'It is difficult for me to believe all the words s/he said'

   b. me   **gustan** [reirme   *de* **las desgracias** de los demas]   (Mexico)
      me$_{DAT}$ like$_{3PL}$ laugh$_{INF}$.me of the misfortune$_{PL}$ of the others
      'I like laughing at others' misfortune'

Another interesting property of these structures is that it is possible to place an element between the matrix verb and the infinitive, such as adverbs (33) or the right-dislocated dative phrase (34) that, again, do not seem to trigger intervention effects:

(33) a. Me   **molestan** [*no* ver   **cambios**]   (Argentina)
      me$_{DAT}$ annoy$_{3PL}$ no see$_{INF}$ changes
      'It bothers me when I don't see any change'

   b. Me   **gustan** *más* [ver   **las procesiones** en Andalucia]   (Spain)
      me$_{DAT}$ like$_{3PL}$ more see$_{INF}$ the processions   in Andalucia
      'I prefer to watch processions in Andalucia'

   c. Me   **gustan** *muchisimo* [ver   **los trabajos** enmarcados.]   (Spain)
      me$_{DAT}$ like$_{3PL}$ a.lot   see$_{INF}$ the works   framed
      'I really like it when I see framed crafts'

   d. Nose   porque me   **gustan** *tanto*   [ver **esas películas**]   (Spain)
      no.know why   me$_{DAT}$ like$_{3PL}$ so.much see$_{INF}$ those movies
      'I don't know why I like so much watching those movies'

(34) a. Así   me   **gustan** *a mi* [ver   **los pasos** de palio]   (Spain)
      like.this me$_{DAT}$ like$_{3PL}$ to me see$_{INF}$ the *pasos de palio*
      'This is how I like seeing *pasos de palio*'

   b. con lo que le   **gustan**   *a esos* [hacer **documentales y series**]   (Spain)[19]
      with that them$_{3SG.DAT}$ like$_{3PL}$ to those do$_{INF}$ documentaries and series
      'they really like shooting documentaries and series'

In addition, it seems possible for some speakers to keep agreement under clausal dislocation (35) and in relative clauses[20] (36):

---

[19] This example also shows a common confusion of the singular form *le* when a plural form *les* is required (see Gallego 2017 and references therein). Speakers place the plural marker in the verb instead of in the clitic. Compare (34b) with the following example:
    (i) Con lo que **les**$_{PL}$ gusta$_{3SG}$ **a esos**$_{PL}$ [hacer documentales y series].
[20] This was already noted by Martínez (1999):
    (i) a. Tenemos que hacer *las cosas que son* necesarias hacer [«…que es necesario hacer»]
          'We have to do the things that are needed to be done'/ …'that is needed to be done'
       b. Hay *cambios que* no *son* posibles detener [«…que no es posible detener»].
          'There are changes that are not possible to stop' /… 'that is not possible to stop'.

(35) a. [*leer* **tus tweets**]$_i$ siempre me **animan a seguir** t$_i$ (Guatemala)
read$_{INF}$ your tweets always me$_{DAT}$ encourage$_{3PL}$ to carry.on
'Reading your tweets always cheers me up'

b. [*Escuchar* **sus palabras**]$_i$ me **sorprendieron** t$_i$ (Argentina)
listen.to$_{INF}$ his/her/their words me$_{DAT}$ surprised$_{3PL}$
'It found it surprising to listen to his/her/their words'

c. [*oír* **esas palabras**]$_i$ me **partieron el corazón** t$_i$ (Colombia)
hear$_{INF}$ those words me$_{DAT}$ broke$_{3PL}$ the heart
'Hearing those words broke my heart'

e. [*Ver* **cosas como estas**]$_i$ me **hacen reir** t$_i$ antes de dormir (Colombia)
see$_{INF}$ things like these me$_{DAT}$ do$_{3PL}$ laug$_{INF}$ before of sleep
'Seeing these things make me laugh before bed'

(36) a. Las **pocas cosas$_i$** [que me **gustan** comprar t$_i$] son caras en Ecuador (Ecuador)
the few things that$_{REL}$ me$_{DAT}$ like$_{3PL}$ buy$_{INF}$ are$_{3PL}$ expensive in Ecuador
'The few things I like to buy are expensive in Ecuador'

b. Estas son **las cervezas$_i$** [que me **gustan** beber t$_i$ a mi!!] (Spain)
me$_{DAT}$ are$_{3PL}$ the beers that$_{REL}$ me$_{DAT}$ like$_{3PL}$ drink$_{INF}$ to me
'These are the beers I enjoy drinking'

c. Juro que hay **cosas$_i$** [que me **cansan** estar repitiéndolas t$_i$ a cada rato] (Argentina)
swear$_{1SG}$ that there.is things that me$_{DAT}$ tire$_{3PL}$ be$_{INF}$ repeating.them to every moment
'I swear that I am tired of being constantly repeating certain things'

We have also detect that the embedded IA can be a clitic. That is crucial for the analysis of these data because it demonstrates that the embedded verb assigns accusative Case[21] – i.e. it is not defective. These clitics can be not only third person plural (masculine and feminine) (37), but also 1st and 2nd person plural (38):

(37) a. Mis videos son nulos, pero me **divierten** [hacer***los***] (Venezuela)
my videos are$_{3PL}$ useless but me$_{DAT}$ amuse$_{3PL}$ do$_{INF}$.them$_{ACC}$
'My videos are useless but I have fun making them'

b. Que lindas que son las mandalas, me **encantan** [dibujar***las*** y pintar***las***]
(Argentina)
what lovely that are$_{3PL}$ the mandalas me$_{DAT}$ love$_{3PL}$ draw$_{INF}$.them$_{ACC}$ and paint$_{INF}$.them$_{ACC}$
'Mandalas are so lovely, I love drawing and painting them'

(38) a. Al profe Felipe parece q le **encantan** [ver**nos** correr atrás de unos conitos]
(Argentina)
to.the prof Felipe seem$_{3SG}$ that him$_{DAT}$ love$_{3PL}$ see$_{INF}$.us$_{ACC}$ run$_{INF}$ behind of some little.cones
'It seems that prof Felipe loves watching us running behind some cones'

---

[21] This is contrary to Martínez's (1999) observations, he considers that there is no agreement when the embedded object is a clitic.

b. cmo m **gustan** [ver**os** tan felicesssssss] (Spain)
how me_DAT like_3PL see_INF.you_2PL.DAT so happy
'I really enjoy seeing you so happy'

(38) confirms that agreement is only in number, as only the plural of the matrix verb, which is in 3rd person plural, matches with the embedded clitic, which is in either 1st or 2nd person plural.

The above examples show that any verb fulfilling the conditions (14) (repeated below for convenience) may trigger non-finite H-Agr, including clusters of verbs such as combinations with light verbs and periphrases.

(39) V {EXPERIENCER, THEME}
EXPERIENCER = EA, dative XP
THEME = IA, T agrees with it, can be a clause

We have attested that adverbs, nor prepositions, nor dative phrases, either when they are *in situ* or right dislocated, trigger intervention effects. Moreover, the embedded verb seems to assign accusative to the embedded DP despite being the target of LDA. These two last aspects are not trivial and should be kept in mind when discussing the analysis.

2.1.2 FINITE HYPER-AGREEMENT

H-Agr across finite clauses has the same properties that we have just pointed out in non-finite clauses as for the nature of the predicates. On the other hand, the target of the agreement may vary. While in non-finite clauses the EA, PRO, could not be a candidate for agreement, this is possible in finite clauses, as the data in (40) reveals:

(40) a. me **molestan** [que **las cosas** no salgan como YO quiero] (Argentina)
me_DAT bother_3PL that the things no come.out_3PL as I want
'It upsets me when things don't come out as I'd like'

b. no me **sorprenderían** [que **ustedes dos** *fueran* novios] (Mexico)
no me_DAT surprise that you_PL two be_3PL.SBJ a.couple
'It wouldn't surprise me that you two might be boyfriend and girlfriend'

c. No me **importan** mucho [que **las personas** me *traten* mal] (Venezuela)
no me_DAT matter_3PL a.lot that the people me_DAT treat_3PL.SBJ badly
'I don't really care if people mistreat me'

On the other hand, it is also possible to find some instances where the target is the embedded IA, although this pattern seems to be far more rare:

(41) a. No me **gustan** [que ___ tenga **muchos tatuajes**] (Argentina)
no me$_{DAT}$ like$_{3PL}$ that has$_{3SG.SBJ}$ a.lot tattoos
'I don't like him/her having so many tattoos'

    b. me **molestan** [que ___ tenga **esos padres**!] (Mexico)
me$_{DAT}$ bother$_{3PL}$ that has$_{3SG.SBJ}$ those parents
'It annoy$_{3PL}$ me that s/he has such a parents'

    c. me **gustan** [que ___ hagas **videos largos**] (Spain)
me$_{DAT}$ like$_{3PL}$ that do$_{2SG.SBJ}$ videos long
'I like that you make long videos'

In the above examples, the subject or the object of the embedded clause, respectively, is the only plural DP. By contrast, in (42) below, there is any overt plural DP. The only possible target is the covert 3rd person subject (indicated with an underscore). Most of the finite H-Agr examples collected are of this sort.

(42) a. Ya me **cansan** [que ___ me *digan* q estoy gorda] (Argentina)
already me$_{DAT}$ tire$_{3PL}$ that me$_{DAT}$ say$_{3PL.SBJ}$ that am fat
'I am already tired of them telling me I am fat'

    b. No me **gustan** [que ___ me *mientan*] y más cuando ya se la verdad (Spain)
no me$_{DAT}$ like$_{3PL}$ that me$_{DAT}$ lie$_{3PL.SBJ}$ and more when already know the truth
'I don't like people lying to me, especially when I already know the truth'

In examples such as (43), it is not clear what the matrix verb is agreeing with, given that both the covert subject and the direct object of the embedded verb are plural.

(43) me **molestan** [que ___ *escriban* cosas sin sentido] (Argentina)
me$_{DAT}$ bother$_{3PL}$ that write$_{3PL.SBJ}$ things without sense
'It bothers me that they write meaningless things'

In these cases, we will assume that T agrees with the covert EA as it is the structurally highest DP available[22].

To conclude this section, we want to highlight that finite H-Agr only takes place across subjunctive clauses[23]. Many scholars have commented on the especial status of infinitive and subjunctive dependencies in Spanish in comparison to indicative (cf. Picallo 1985, Bosque 1990, 2012; Hernanz 1999; Torrego & Uriagereka 1992 (cf. (22)); Uriagereka

---

[22] This is reminiscent to Algonquian languages, as we will see in section 3.2.3
[23] We have found some instances of H-Agr across indicatives when they are embedded interrogatives:
    (i) Me **gustarían** [saber [cómo van a ser **mis hijos**]] (Spain)
    me$_{DAT}$ would.like$_{3PL}$ know$_{INF}$ how go$_{IND.3PL}$ to be my children
    'I would like to know how my children are going to be'
Although this and other similar examples could be very revealing for analyzing H-Agr we leave them aside for now since they require a comprehensive discussion that is beyond the scope of this thesis.

2006, 2008, 2015; Gallego & Uriagereka 2011, i.a.). One of the properties of subjunctive as opposed to indicatives is their more 'transparent' status, this idea is consistent with our data and a suitable analysis should account for it.

In the next section, we will discuss the origin of the data presented in these two last sections and discuss why it is pertinent to explore examples from online sources such as social networks for linguistic research.

2.2 Methodology

The almost neglected status of H-Agr in the literature does not give us a hint of where to start looking for speakers that produce these kinds of examples. In the varieties of Spanish we are in contact with, we have attested few examples in conversational language, but not enough for a detailed study: speakers reject the data when asked directly and it is not easy to make them produce it spontaneously. For this reason, we have carried out a corpus-based search that does not substitute acceptability judgments and other experimental methods (Schütze 2009, 2011), but can help us restrict the phenomenon and guide our future inquiry.

We based our search in two main questions: the first one regards the properties of the structures, namely what patterns of H-Agr exist. The second one is related to the speakers, we need to know which Spanish speakers produce/accept H-Agr and whether they come from the same region, speak the same dialect or sociolect. Once we have at least an educated guess about the answers, we will know what we are looking for and where we could find it. For these reasons, the present study has resorted to the Internet, specifically to the social network Twitter, as it has become the most advantageous method for accessing instances of spontaneous language by a wide range of speakers from different Spanish speaking areas.

De Benito & Estrada (2016a) point out that despite the increasing number of different kinds of corpora available, more and more scholars use the Internet to obtain linguistic data (see for instance, Méndez García de Paredes 2011 and Di Tullio 2011 for Spanish and Jones 2015 and Kilgariff & Grefenstette 2003 for English, apud De Benito & Estrada 2016a). One of the reasons for this growing tendency is that the more substandard a phenomenon is, the more difficulties there are to document it (De Benito & Estrada

2016a). Along the same lines, our prediction was that by searching in traditional corpora, we would not obtain barely any instance of H-Agr.

To verify this intuition, we consulted different corpora of Spanish. Firstly, we checked corpora made available by RAE-ASALE: the synchronic corpora, CREA and CORPES[24] did not seem to include any instance of H-Agr, since we have made different searches with no results. On the other hand, we found one instance of H-Agr in the diachronic corpus CORDE from the 19th Century in a novel by an Argentinian writer:

(44) - Si a la niña no le **gustan** ver **esas cosas**, yo no le he de traer la cabeza que le he
ofrecido -replicó Parra-

    If to the child no her$_{DAT}$ like$_{3PL}$ see$_{INF}$ those things […]
    'If the child does not like seeing those things […]'
                                [CORDE: *Amalia*, José Mármol, $1851 - 1855$]

We cannot know without further investigation if this example from a dialogue in a novel is an actual reflection of oral language at that time or if it is just an error. Anyhow, as expected, we cannot use these corpora for the study of H-Agr.

Secondly, we checked two corpora of oral recordings: COSER and PRESEEA. While the former, which offers data of rural European Spanish, does not collect any case of H-Agr, we found few instances in the latter:

(45) a. ¿y de qué nacionalidad le   gustan ver las películas?
         and of what nationality   you$_{DAT}$ like$_{3PL}$ see the movies
         'the movies you enjoy watching, which country are they from?'
                                [PRESEEA: Monterrey, Mexico, 05-11-2007]
      b. Me   encantan hacer   dulces
         me$_{DAT}$ love$_{3PL}$ make$_{INF}$ sweets
         'I love baking cakes and pastries'
                                [PRESEEA: Valencia, Spain, 1-12-1996]
      c. ¿y a ti   cómo te   **gustan** que ___ te   traten   normalmente?
         and to you how you$_{DAT}$ like$_{3PL}$ that   you$_{DAT}$ treat$_{3PL}$ normally
         'and how do you like to be treated?'
                                [PRESEEA: La Habana, Cuba, 09-11-2010]

This evidence does not contradict our hypothesis that H-Agr is only found in colloquial contexts[25], although these results are still very rare and lead us to use other kind of sources.

---

[24] These corpora consist of data from different sources, mostly written (journalistic, academic, literary, etc.) from different Spanish speaking countries gathered from 1974 to 2004 and from 2001 to 2004 respectively.
[25] Interestingly, the three examples from PRESEEA were uttered by people with higher education, the interviewer in (45a) and (45c), and the interviewee in (45b), who also had a degree in Spanish studies. This

Thirdly, we have consulted two corpora made up of examples from the web, *Corpus del Español* (CdE) and EsTenTen. Data from the web could overcome the lack of evidence from the traditional format of corpora that we have just reviewed, given that the sample is much broader. These two corpora allow us to use regular expressions[26], which facilitate the search for syntactic patterns[27]. In Figure 1 we can see the results of one of these searches in CdE, which can give us an overview of some of the possible combinations, the frequency of the patterns[28] and their origin by country.
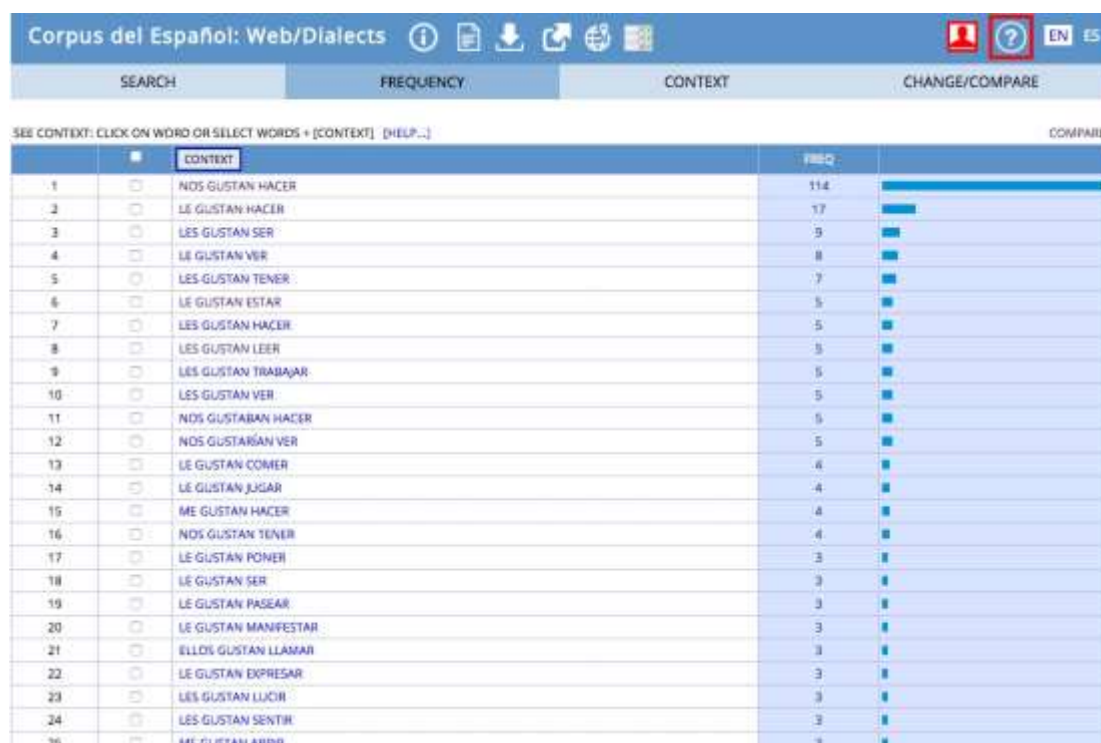
Figure 1. Search verb of personal pronoun + *gustar* 'to like' in plural + infinitive in CdE (regular expression: _PP*gust*n _VR*).

However, we are not going to use the examples from the web, neither from these corpora nor from a manual search (Google search) because of one main drawback: the difficulty of being sure of the author/source of a given text. On the one hand, the author is often not known or, if known, the information we need is not available (basically if the person is a Spanish native speaker and from which region). On the other hand, web texts are often

---

fact would suggest that these cases are a processing error or that it is a tendency of oral language, regardless of the level of education (i.e. greater normative pressure).

[26] A regular expression is "a search pattern used for matching one or more characters within a string" (Christensson, Per. 2013. https://techterms.com/definition/regular_expression).

[27] Typical web searching engines such as Google do not offer the option of searching with regular expressions. See Kilgarriff & Grefenstette (2003 §6) for a detailed list of the drawbacks of this type of web search engines in linguistic research.

[28] The results shown in Figure 1 do not seem to be totally reliable, at least as for the first one most of the 114 supposed examples are repetitions of the same one.

copied, edited or translated from other sources (including automatic generation/translation) (see Schütze 2009, 2011).

In light of the unsatisfactory results obtained in all corpora mentioned in the preceding lines, we opted for Twitter as our main source of data. In the next section, we outline the advantages of this procedure.

2.2.1 TWITTER

There is an increasing interest in linguistic data from online social networks for the study of linguistic variation. Some scholars have already employed this new source of information for defining dialectal tendencies (see Donoso & Sánchez 2017 and references therein). These platforms can be characterized by their near-synchronous communication (Eisenstein 2014) –that differs, for example, from email communication– and their interactive purpose (De Benito & Estrada 2016a). These factors favor the imitation of oral/colloquial language in a writing medium, something that is especially appealing for language researchers (although some authors do not consider them instances of real language, see Kilgarriff & Grefenstette 2003 for discussion).

Twitter has become one of the most useful platforms of this sort because of its microtext format (messages of maximum 140 characters), the fact that most of the users keep their profiles public and the relatively easy procedures for extracting data. In this regard, different techniques are being developed in the field of computational linguistics to extract data and metadata, such as the geolocalization of the microtexts. These methods seem to be fruitful for the study of lexical variation (see Gonçalves & Sánchez 2014, Jones 2015, Kulkarni et al. 2016, i.a.), but can be implemented in syntactic variation studies as well (see for instance Ruiz Tinoco 2013).

De Benito & Estrada (2016a) distinguish three options when using Twitter as a data-mining source: i) search directly on the platform; ii) create a corpus with data of a specific phenomenon; and iii) create a general corpus for linguistic research. Let us start with the last option: there are already research groups working on the construction of Twitter corpora for linguistic uses, such as Variaciones (Ruiz Tinoco 2014) and ASinEs (Gallego 2014) for Spanish. We could look for H-Agr using regular expressions in these corpora, but we got almost the same results as in traditional corpora, noted above. This may be due to the rarity of the phenomenon in comparison to the totality of tweets, even though these corpora are rather large.

The second option, creating a small corpus of a specific phenomenon, is typically carried out using specific computational techniques that restrict the extraction of data (by coordenates, specific time, etc.; see for instance De Benito & Estrada 2016b or Gonçalves & Sánchez 2014, 2016). We expect, though, this method to produce a lot of noise (undesired results) when looking for a phenomenon such as H-Agr, that involves searching for complex strings of words.

Considering these difficulties, we have carried out a manual search in Twitter advanced search (Figure 2) by looking for strings such as the ones in (46) below.

Figure 2. Twitter advanced search (https://twitter.com/search-advanced)

(46)   a. dative pronoun (*me, te, le(s), nos, os*) + psych-verb[29] in 3rd person plural (*-n*) + infinitive

    a'. Example: "te sorprenden leer"

    b. dative pronoun (*me, te, le(s), nos, os*) + psych-verb in 3rd person plural (*-n*) + subordinate conjunction (*que*)

    b'. Example: "te sorprenden que"

Since the start of this research we have collected around 1.500 tweets of H-Agr, not on a daily basis, but extracting newly published tweets from the last weeks every time. This data is not exhaustive, meaning that, in as much as we are not carrying out a quantitative

---

[29] Vanhoe's (2002) list of Spanish psych-verbs has been very useful for this purpose.

analysis we do not collect all the instances found, but only the ones that add something new (linguistically or geographically) to the sample.

In relation to the results, there are two factors that need to be underlined. Firstly, tweets are, in general, full of spelling and grammatical mistakes. These are attributable to a lack of attention or to rapid self-editing (Schütze 2011) or sometimes purportedly made, as part of the Internet jargon or just for fun (De Benito & Estrada 2016a). In both cases, they should not be taken as genuine dialectal forms. H-Agr data must be then interpreted with caution, given that it could easily result in typing or editing errors (often only one letter changes between the non-agreeing/agreeing versions, e.g. *gusta/gustan*).

Secondly, we have tried to be sure that all examples come from native speakers. In this sense, Twitter permits us to be more certain about the identity of the author of the text, considering that every tweet is linked to a personal profile (we will avoid corporative accounts). Moreover, we have used the information that the user provides in his/her profile about their origin. This information is not always available (the user can choose whether to make it public) or it can be deceitful, but still we find it more reliable than the coordenates retrieved by Twitter-corpora such as Variaciones (Ruiz Tinoco 2014). These coordenates are more precise but indicate the location where the tweet was sent, which does not necessarily match the place of origin of the speaker.

We have seen that Twitter as a data-mining source has significant advantages in comparison to other corpora. In relation to our first research question, the search for different strings has allowed us to organize the results and have an overall idea of the different patterns of H-Agr (presented in section 2.1.). We will try to answer our second question in the following lines.

2.3 A 'substandard' phenomenon?

As pointed out earlier, one of the main questions that we need to answer for having a comprehensive view of the phenomenon we are studying is which Spanish speakers produce H-Agr. The impossibility of performing a quantitative analysis makes it difficult to know how rare is the data we are dealing with. Even if we collected thousands of tweets the results would not be representative in the sense that the use of Twitter is not the same in all Spanish speaking countries and by all speakers. In the future, we expect to contrast the found data with native speakers in order to shed light on this issue; until then, the following ideas are educated guesses about what the present study can reveal.

We have already commented that we are inclined to think that H-Agr is more common in spoken language and that all these tweets are a reflect of that. On the other hand, it is also difficult to assess which patterns of the ones presented earlier are more generalized among H-Agr speakers. We would like to know if they all accept both the non-finite and finite version of the phenomenon and with all the variables that we have discussed or if it depends on other factors.

The reader may have noted that the examples presented in earlier sections are from very different Spanish speaking countries. In the following maps, we have pinned the places where at least one example of H-Agr has been found:
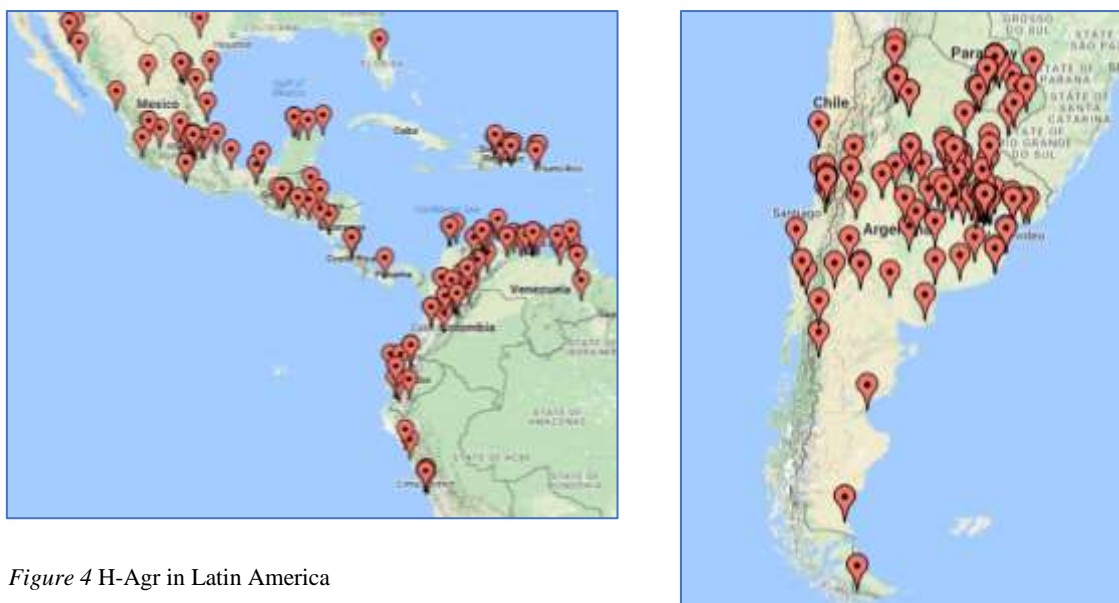

Figure 3 H-Agr in Spain

Figure 4 H-Agr in Latin America

As shown in the maps, we have found tweets with H-Agr by speakers of every Spanish speaking country (with exception of Bolivia[30]), both from European and Latin-American varieties. If we are talking about a dialect (or dialects), it does not seem to be subject to a geographic tendency. Therefore, as we have already suggested, and like what Etxepare (2006, 2012) has observed for LDA in Basque, we could consider H-Agr a 'substandard' phenomenon.

This leads us to the issue of optionality. H-Agr could be freely optional (as it is implausible that this version would have substituted the canonical one), or speakers could be bi-dialectal, as Rodríguez-Mondoñedo (2005) suggests for the agreeing version of the existential *haber* (cf. (8)):

> At best, speakers who accept both are bi-dialectal. Being a native speaker of SII [agreement version] myself, remember my surprise when I was first introduced to SI [non-agreeing version] as the "correct" dialect. Now, however, I can accept SI *haber*-sentences. (Rodríguez-Mondoñedo 2005: fn.18)

This also reminds us of other cases of unexpected agreement in Spanish varieties, such as *haber* existentials in different varieties (cf. Rodríguez-Mondoñedo 2005, Aloy 2016, i.a.), agreement across PP in *se* impersonals in American varieties (cf. Gallego 2016) and agreeing plural markers in singular clitics in Argentinian Spanish (cf. Mare 2016). These data, along with H-Agr, could suggest that there is a growing preference for agreement in Spanish; we leave this hypothesis for future research.

We have discussed the advantages of using new platforms as a source for linguistic data for overcoming some of the limitations we are facing when collecting data. Twitter has allowed us to face two main methodological problems: to access spontaneous language without speakers noticing it and accessing Spanish speaking areas that could not be reached by our means. At the same time, it presents some of the pitfalls of corpus-based searches, essentially, we cannot contrast data directly with speakers. For this reason, we have labeled it 'substandard' without being more specific, insofar that we only know that it is ubiquitous, regardless of the dialectal area.

---

[30] If a location is not pinned in these maps, it does not necessarily mean that there are no H-Agr speakers there, but that we have not found any instance on Twitter.

3   LDA IN CONTEXT

3.1 *Agree* and the PIC

Before we discuss the different proposals for analyzing LDA in different languages it is necessary to establish some key theoretical notions discussed in those proposals, mainly the concepts *Agree* and the PIC (*Phase Impenetrability Condition*).

Agree, postulated by Chomsky (2000), is basically agreement at a distance and it dispenses with the technical assumptions of some previous approaches, such the need to create a Spec-H dependency. Agree takes place as follows: some functional heads are taken from the lexicon with their ɸ-features unvalued. Therefore, they act as a Probe that looks for a Goal with matching features in its c-command domain (Gallego 2009). In short, "Agree is a (long distance) context-sensitive mechanism to check features – where 'check' means, to be precise, 'match and provide a value'" (Gallego 2009: 164). When the Goal (DP) values the Probe (T or *v*), it is assigned structural Case. This process is illustrated in (1):

(47)   a. $[_{vP}\ v[u\varphi]\ [_{VP}\ V\ DP[\varphi]\ ]]$ Agree (v, DP)

           b. $[_{vP}\ v[\checkmark\varphi]\ [_{VP}\ V\ DP[\varphi]\ ]]$ v's φ-features are valued

           c. $[_{vP}\ v[\checkmark\varphi]\ [_{VP}\ V\ DP[\varphi]ACC\ ]]$ v's φ-features are deleted

           $\rightarrow$ DP receives Case

[From Gallego 2009: (2)]

Consider next the conditions for Agree (Gallego 2010: 35; adapted from Chomsky 2001):

(48)   a. Probe and Goal must be active for Agree to apply

      b. Agree divides into Match and Valuation

      c. Probes must contain a full set of features (it must be complete) to delete the uninterpretable FF of matched Goals

In order to understand these conditions, it is necessary to clarify some notions:

(49)   *Activity Condition*

       Uninterpretable (unvalued) morphology renders syntactic objects 'active'.

It is assumed that the unvalued features for C, T and *v* are ɸ-features, while nominals have an unvalued Case-feature. As we have seen in (47), the Probe values its ɸ-features with the Goal and it receives Case in exchange. When this operation has taken place, the elements become 'frozen' (=inactive) for further operations.

(50) *Match*

F[eature] and F match if they belong to the same attribute class (e.g., [number], [Case], etc.), independently of value (e.g., singular vs. plural, nominative vs. accusative, etc.).

[From Gallego 2010: 36]

For example, a Probe with [uN] looks for a Goal with [iN] and they match, the next step is Valuation: if the N features have the same value (e.g. *plural*) the Probe with receive this feature (will be valuated). Nevertheless, not every Match is followed by Valuation, in other words, there can be Match without Agree. Gallego (2010) comments on two cases in which there is Match without Valuation:

(51) [$_{CP}$ C [$_{TP}$ T$_{[\phi]}$ [$_{v*P}$ v$_{[\phi]}$ arrived John$_{[3.SG]}$]]]   [From Gallego 2010: 37]

In an accusative structure such as the one in (51), *v* is cannot assign Case because it is defective. Consequently, *v* matches the ϕ-features of the DP *John* without valuation. As *John* has not received Case, it remains active until it receives nominative Case from T (non-defective).

The second case is known as *defective intervention*: Probe cannot reach Goal$_2$ because there is another inactive element interfering, Goal$_1$.

(52)   [From Gallego 2010: 38]

Having discussed Agree, let us turn to PIC, to which the operation is closely connected. Chomsky assumes that the derivation is carried out in chunks or *phases*[31], which are cyclic domains. One of the functions of these domains is to regulate displacement operations, Move and Agree. In other words, these operations cannot freely cross certain boundaries. The PIC establishes these limits:

(53) *Phase Impenetrability Condition (PIC)*
In phase α with head H, the domain of H is not accessible to operations outside α; only H and its edge are accessible to such operations.

[Chomsky 2000: 108, apud Gallego 2010: 57]

The PIC states that only the edge of each phase (Spec-C and Spec-*v*) is accessible for Agree. LDA poses a problem because it is non-local in the PIC's sense: as Richards (2009) indicates, any Agree operation holding between T and V's complement, or

---

[31] We will assume that only CPs and *v*Ps are phases (cf. Chomsky 2000, Gallego 2010, Chomsky, Ott & Gallego 2017), even though there is a long debate about whether other phrases can be phases as well.

between *v* and an item inside an embedded CP exceeds the PIC and crosses a phase boundary. As a result, different scholars have tried to accommodate their analyses to the PIC in different ways (see Richards 2009).

Another option, pursued by Bošković (2003, 2007) is to accommodate the PIC so that it does not clash with LDA data. He argues that, in fact, the PIC is redundant with intervention effects and proposes "Agree closest" as a more suitable operation. According to this proposal, Agree is not restricted by the PIC, but only by intervention effects.

This proposal is reminiscent of Chomsky's (2008) version of PIC (also in Chomsky, Gallego & Ott 2017), which we have dubbed PIC$_2$:

(54) *PIC$_2$*
     Transferred phases remain accessible, but they cannot be modified at later cycles.

In other words, PIC$_2$ "permits Probe-Goal relations across phase boundaries, as long as these only affect properties of the Probe" (Chomsky, Gallego & Ott 2017: 10). In the next section, we review some proposals that account for LDA assuming Agree. We will see how the PIC affects such analyses and how the new conception of the PIC may have an impact on future proposals.

3.2 Previous approaches to LDA

LDA is one of the phenomena that has provided empirical support for *Agree*. It has been proved in different languages that in LDA configurations the DP does not move, overtly nor covertly, to the matrix clause. Thus, agreement has to take place at a distance (cf. Boeckx 2008a, 2009, Etxepare 2006, 2012, i.a.):

(55)   [… V$_{\phi i}$ … [$_{XP}$ … DP$_{\phi i}$ …]]
       |_____|
              AGREE

This is a shift from the Spec-H relation postulated for agreement to a Head-complement domain relation, which is involved in LDA configurations[32]. We review some approaches to LDA in different languages classifying them among these options[33]:

---

[32] Some authors have claimed that it is possible to maintain a Spec-H or a local analysis for some 'apparent' LDA configurations, Chandra (2007) for Hindi is an example of the former and Polinsky (2003) for different languages for the latter.

[33] This list is only for Algonquian languages. However, it is also quite significant for the different approaches to LDA present in the literature. See for example Polinsky (2003) for other analyses.

(56) a. to relax the locality constraints on Agree such that the Phase Impenetrability Condition (PIC) does not apply;

    b. to analyze the embedded clause as deficient (e.g., a restructuring configuration or an ECM construction); or

    c. to propose that a copy of the phrase participating in long-distance agreement has a copy in the left-periphery of the embedded clause.

<div align="right">[From Fry 2016: 8; cf. Richards 2009]</div>

We basically focus on the analysis of Icelandic, Basque and Algonquian languages[34] as they seem to be the closest counterparts to Spanish vis-à-vis LDA properties, as we will discuss in turn. Firstly, we look at option (56b), specifically to restructuring approaches to LDA. Secondly, we outline the proposals for Icelandic (Boeckx 2009) and Basque (Etxepare 2012), which also posit defective embedded clauses, although not of a restructuring nor of an ECM sort. Finally, we turn to Algonquian languages, which have been analyzed under both option (56c) (Fry and Hamilton to appear) and option (56a) (Bošković 2003, 2007).

3.2.1 THE RESTRUCTURING APPROACH

Restructuring (cf. Rizzi 1978, i.a.) consists, roughly, on a biclausal structure becoming monoclausal by erasing functional structure dividing the two clauses so that a modal and an infinitive end up forming a head unit (cf. Rizzi 1982, apud Ordóñez 2012). We need to comment on the possibility of Spanish H-Agr being a result of such process at least for two reasons. On the one hand, because other languages with LDA such as Hindi (Bhatt 2005) or person LDA in Basque (Etxepare 2012) seem to fit in this kind of approach. On the other hand, because Martínez (1999), the only reference to this phenomenon in Spanish (cf. (5)), considers that it is the result of some sort of restructuring process:

> Although less frequently, speakers interpret too as a verbal periphrasis –inappropriately, as different grammars insist– some personal verb with infinitive subject constructions. Consequently, it is formulated as the subject, not the infinitive (singular) but its direct object (if it is plural).

<div align="right">[Martínez 1999 (GDLE §42.10.1.4); translation mine IFS]</div>

Martínez equates the phenomenon of H-Agr to another case of attested optionality in agreement in Spanish, the case of *se* passives. We have already mentioned mono-clausal

---

[34] For simplification, we will talk about Algonquian languages in general, but it is necessary to keep in mind that the discussion may not be valid for all languages from that family.

se constructions in (7), repeated below as (57), but we have not talked about *se* passives with non-finite arguments such as the ones in (58) and (59):

(57) a. Se vende   flores
       SE sell$_{3SG}$ flowers

    b. Se venden flores   [From Sánchez-López 2002: (25)]
       SE sell$_{3PL}$ flowers
       'Flowers are sold'

(58)   Se {pudo/**pudieron**} vender **todas las flores**
        SE could$_{3SG}$/could$_{3PL}$ sell$_{INF}$ all   the flowers
        'All the flowers could be sold'

(59) a. Se {consiguió/**consiguieron**} vender **todas las flores**
       SE achieved$_{3SG}$/achieved$_{3PL}$ sell$_{INF}$ all   the flowers
       'They managed to sell all the flowers'

    b. Se {intentará/**intentarán**} vender **todas las flores**
       SE will.try$_{3SG}$/will.try$_{3PL}$ sell$_{INF}$ all   the flowers
       'They will try to sell all the flowers'

    c. Se {quiere/**quieren**} vender **todas las flores**
       SE want$_{3SG}$/want$_{3PL}$ sell$_{INF}$ all   the flowers
       'They want to sell all the flowers'

(58) is a typical case of restructuring where the auxiliary *poder* 'can' shows agreement because it forms a periphrasis with the infinitive. The cases in (59), although not being auxiliaries (modals, aspectuals, and verbs of motion) have been claimed to be some sort of restructuring/semi-auxiliar predicates (see Hernanz & Rigau 1984, Herbeck 2013, Paradís 2016 among many others)[35].

However, contrary to Martínez's suggestion, predicates that allow H-Agr cannot be considered of the restructuring sort in Spanish (but of the *gustar* and *preocupar*-type, following Belletti & Rizzi's 1988 classification; cf. §2). This is also pointed out by Etxepare (2012) for number LDA in Basque, possible with verbs such as *gustatu* 'like'. Among other evidence, clitic climbing is not possible:

(60) a. Me   gusta   comprar**los**
       me$_{DAT}$ like$_{3SG}$ buy$_{INF}$.them$_{ACC}$

    b. *Me **los**   gusta comprar
       me$_{DAT}$ them$_{ACC}$ like$_{3SG}$ buy$_{INF}$
       'I like to buy them'   [From Etxepare (2012: (84))]

---

[35] Even though *se* constructions are relevant for our purposes, especially because they resemble DAT-NOM constructions in their lack of an agent, we will not discuss them in this paper (see Gallego 2016, Ordóñez & Treviño 2016, i.a.).

Etxepare further argues that no restructuring analysis, be it Cinque's (1999, 2001, et seq.) or Wurmbrand's (2001, 2004, et seq.), can be adopted for a *gustar*-type verb. On the one hand, if we followed Cinque's analysis, we would need to claim that *gustar* can be a functional or semi-lexical item, something that does not seem very plausible. For example, Etxepare indicates that this type of verbs is not in Heine & Kuteva's lexicon of grammaticalization (2002).

On the other hand, according to Wurmbrand, in order to maintain the lexical content of the restructuring verb, clausal complements are bare VPs and there is some sort of event structure unification. As Etxepare indicates, it is not clear what the semantic result of such reduction would be in cases such as the one in (60), since *comprar* requires an agentive subevent in its conceptual structure.

It is remarkable, though, that some infinitives seem to be more 'transparent' than others. Examples of non-finite H-Agr with infinitives such as such as *hacer* 'do/make', *tener* 'have' or *ver* 'see' abound, while combinations with semantically fuller verbs such as *comprar* 'buy' or *encontrar* 'find' seem to be more restricted. Compare (60) with (61):

(61) a. Me **asustan** [*tener* **pesadillas**]   (Mexico)
me$_{DAT}$ frighten$_{3PL}$ have$_{INF}$ nightmares
'It frightens me to have nightmares'
            (=what frightens me are the nightmares)

 b. Me **molestan** [*tener* **las uñas largas**]   (Argentina)
me$_{DAT}$ annoy$_{3PL}$ have$_{INF}$ the nails long
'It annoys me to have long nails'
            (=what bothers me are the long nails)

 c. Te **gustan** [*ver* **películas** de terror]   (Venezuela)
you$_{DAT.SG}$ like$_{3PL}$ watch$_{INF}$ movies   of terror
'You like to watch horror movies'
            (=what I like are horror movies)

(62) a. No me **gustan** [*comprar* **zapatos**]   (Peru)
no me$_{DAT}$ like$_{3PL}$ buy$_{INF}$ shoes
'I don't like to buy shoes'
            (=what I don't like is to buy shoes, *but I like to wear them*)

 b. Ya   me **cansaron** y me **aburrieron** [*hacer* **cupcakes**] (Argentina)
already me$_{DAT}$ tire$_{3PL}$ and me$_{DAT}$ bore$_{3PL}$ make$_{INF}$ cupcakes
'I am already bored and tired of baking cupcakes'
            (=I am tired of baking cupcakes, *not of eating them*)

c. Les **encantan** [*leer* **cartas** de amor] (Peru)
them$_{DAT}$ love$_{3PL}$ read$_{INF}$ letters of love
'They love reading love letters'
(=what I like is to read love letters, *but not to write them*)

The infinitives in (61) seem to provide little lexical content, inasmuch as the same sentence without the infinitive would be virtually equivalent in meaning (the one in brackets), maybe because the meaning of the infinitive is quite redundant − i.e. to be afraid of nightmares is more or less equivalent to be afraid of *having* them. Contrarily, the examples in (62) have a different interpretation. Here, the event signaled by the infinitive is more relevant (consider a possible contrast in meaning in italics). Therefore, the examples in (61) would be more plausible candidates for a restructuring analysis à la Wurmbrand.

Yet, as we saw in (37) (see (63) below) the embedded DP can be replaced by an accusative clitic. If the embedded verb assigns accusative Case, the embedded clause should be at least a vP, but in no case a bare VP.

(63) Me **gustan** [comprar*los*]
me$_{DAT}$ like$_{3PL}$ buy$_{INF}$.them$_{ACC}$
'I like to buy them'

Finally, if H-Agr were possible with complex verb constructions (see (64) below), it would trigger undesired consequences under a restructuring approach. It would force us to postulate big verb clusters such as the ones underlined in the examples:

(64) a. Como te <u>*pueden gustar* ver</u> **dientes** de una persona?? = (28a)
how you$_{DAT}$ can$_{3PL}$ like$_{INF}$ see$_{INF}$ teeth of a person
'How can you like seeing people's teeth?'

b. Como me <u>**gustan** *poder ver*</u> **estos festejos** tan patrios = (29a)
how me$_{DAT}$ like$_{3PL}$ can$_{INF}$ see$_{INF}$ these festivities so national
'How I love watching these national festivities'

c. Me <u>*hacen gracia* oír</u> hablar de **leyes electorales** injustas (Spain)
me$_{DAT}$ do$_{3PL}$ funny hear.$_{INF}$ talk$_{INF}$ of laws electoral unfair
'I find it funny to hear them talking about unfair electoral laws'

Having discussed why the restructuring approach does not seem suitable for H-Agr Spanish data, we now turn to other proposals that can be contemplated for languages that display non-restructuring LDA.

Icelandic and Basque have both been claimed to display long distance *Agree* (Boeckx 2009) and share some properties with the structures that we have seen to be compatible with non-finite H-Agr. Consider first these examples from Icelandic:

(65) a. Henni **leiddust strákarnir**.
        her_DAT bored_3PL the.boys_NOM

    b. ??*Henni leiddist   strákarnir.
          her_DAT bored_3SG the.boys_NOM
       'She found the boys boring.'

    c. Mér   virðist/**virðast þeir**   vera skemmtilegir.
       me_DAT seem_3SG/3PL   they_NOM be   interesting
       'It seems to me that they are interesting.'

                                        [From Boeckx 2009: (59), (60)]

LDA in Icelandic (65c) is only possible with 'quirky' structures, similar to what we have reported in Spanish. Even if we do not assume that Spanish oblique arguments are truly quirky subjects (cf. fn. 12), these structures resemble in both languages in that they involve a non-agentive predicate that selects a dative argument and a nominative object. In addition, in both languages this nominative object triggers obligatory agreement in mono-clausal structures. Compare (65) with (66):

(66) a. Le   **aburrieron los chicos**
        her_DAT bored_3PL   the.boys_NOM

    b. *Le   aburrió   los chicos
        her_DAT bored_3SG the boys_NOM
       'She found the boys boring.'

    c. a mi me   **gustan** [que ___ **sean** interesantes][36]   (Colombia)
       to me me_DAT like_3PL that (they) are_3PL interesting
       'I like them to be interesting'

Two differences between Spanish and Icelandic LDA should be noted: unlike Icelandic, Spanish does not seem to be subject to intervention effects, at least with datives (cf. (34)), and 1st and 2nd person agreement is also possible in quirky constructions[37]. Although very interesting, these differences are not relevant for our purposes.

---

[36] We cannot directly equate the bi-clausal structure in (66c) to the Spanish version because of the nature of the predicate, a direct translation in Spanish would involve a raising predicate *parecer* 'seem' that would lead to a different discussion.

[37] While Icelandic quirky subjects can only inflect for 3rd person singular, Spanish psych-verbs do show 1st and 2nd person morphology in mono-clausal structures (e.g. *me gustas_2SG tú* 'I like you'). Rivero (2004), though, claims that this agreement is also banned for some Spanish psych-verbs (see Gutiérrez-Bravo 2006 for counterarguments).

Boeckx's (2008b) analysis proposes that agreement is triggered by v°, not by T°, which then would assign nominative case. This is explained by the following hypothesis:

(67)   a verbal head *v*° is endowed with the option of nominative Case licensing only if it assigns a theta role realized as Quirky Case to an NP in its specifier (Boeckx 2008b: 95)

Boeckx suggests that (67) is reminiscent of Burzio's Generalization (1986) because nominative (structural Case) is only available if Quirky Case (thematic information) is present. He further argues that, since Quirky DPs are never agents (but experiencers, goals or beneficiaries), the *v°* that selects them is a *v°[non-agentive]*.

He assumes that there is no embedded C, so matrix *v°* can establish an Agree relation with the embedded DP that has raised to the edge of the embedded *v*P to fulfill PIC requirements. In other words, the embedded DP does not need to move to the matrix clause but can establish an agreement relationship from the embedded clause.

Regarding Basque LDA, we can list two main similarities with Spanish H-Agr. As we have already noted in the previous section, Basque number LDA can never occur with restructuring predicates, but it is possible with psych-predicates[38] (compare (68) with (69) below). Furthermore, the status of LDA in both languages resemble in that their optionality is due to sociolinguistic factors –i.e. it is a 'substandard' phenomenon (cf. §2.3).

(68)   a. [Nobela erromantikoak irakurtzea]   gustatzen zaio
        novel   romantic$_{PL-ABS}$ read$_{N-DET-ABS}$ like$_{HAB}$ Aux$_{(3SG.ABS-3SG.DAT)}$ [39]

       b. [**Nobela erromantikoak** irakurtzea]   gustatzen **zaizkio**
        novel   romantic$_{PL-ABS}$ read$_{N-DET-ABS}$ like$_{HAB}$ Aux$_{(3PL.ABS-3SG.DAT)}$
        'He/she likes to read romantic novels'
                                                        [From Etxepare 2006: (1)]

(69)   a. Le   gusta [leer novelas románticas]   (Spanish)
          him/her$_{DAT}$ like$_{3SG}$ read novels romantic

       b. Le   **gustan** [leer **novelas románticas**]   (%Spanish = H-Agr)
          him/her$_{DAT}$ like$_{3PL}$ read novels romantic
          'He/she likes to read romantic novels'

---

[38] Mendívil (2012) has proposed an analysis of psych-verbs in Spanish as an ergative-absolutive pattern. If we take this proposal seriously, Spanish would be closer to Basque than it seemed.
[39] ABS = absolutive, N = nominalized, DET = determiner, HAB = habitual

For the analysis of number LDA in Basque, Etxepare (2012) argues for a one-step operation[40] by postulating an unvalued number feature in the auxiliary:

(70)   [$_{CP}$ C… Aux$_{[uNumber:]}$… [$_{DP1}$ –a… [$_{vP1}$ DP$_{[iNumber]}$ v…

[Etxepare 2012: (113)]

The relevant structure in (70) shows that the matrix auxiliary finds the embedded DP with a number feature as its Goal, given that the latter has moved to the vP edge. Etxepare claims that here agreement seems to go in two directions: the controller of number is the embedded object, while the nominalized clause[41] controls Case. At the same time, he proves that the embedded DP receives Case from the embedded v. This would be consistent with Bhatt's (2005) claim that the Goal being active is not a requirement for Agree (against Chomsky's *Activity Condition*). In other words, the embedded DP can be a Goal for Agree even though it has already checked its Case feature.

We can conclude from this section that the gist of the analyses for Icelandic (Boeckx 2009) and Basque (Etxepare 2012) LDA is the same. Setting aside the differences, both proposals assume that there is no CP layer. In addition, both respect the PIC, as the embedded DP (the Goal) moves to the edge of the vP in order to be accessible by the Probe from the matrix clause.

3.2.3 LDA in Algonquian languages

Algonquian languages[42] have also been claimed to show LDA (known in the literature as cross-clausal agreement). In this case, agreement can take place across declarative embedded clauses and even across embedded interrogatives in some languages. In (71) we can see examples from Ojibwe, (71a) shows no agreement, (71b) and (71c) show agreement with the embedded agent and the embedded patient respectively. (72) shows an example of LDA across an embedded question in Blackfoot.

---

[40] This proposal is an updated version of the one put forward by Etxepare (2006) and Preminger (2009). They proposed that the agreement takes place in two steps: firstly, between the nominal element and the head of the clause and, secondly, between the head of the clause and the finite auxiliary. This approach, though, could not predict that Basque LDA cannot cross more than one clause (Etxepare 2012).

[41] We are simplifying a rather complex issue. For more details about Basque nominalized clauses see Etxepare 2012 and references therein.

[42] Algonquian languages are a subfamily of Native American Languages. They were formerly spoken along the east coast of North America and west to the Rocky Mountains, but currently, most of the original 30 languages are no longer spoken. Typologically, some of their main features are being polysynthetic, hierarchical, nonconfigurational head-marking languages with discontinuous constituents and relatively free word order (Pentland 2006).

(71) a. Ngikendaan gii-bashkizwaadj. (Ojibwe)
   ni-giken-daan   gii-bashkizaw-aa-d
   1-know-INTR(INDEP) PAST-shoot-3OBJ-2(CONJ)
   'I know that you shot him.'

  b. Ggikenimin gii-bashkizwaadj.
   gi-giken-im-in   gii-bashkizaw-aa-d
   2-know-TR-1>2(INDEP) PAST-shoot-3OBJ-2(CONJ)
   'I know that **you** shot him.'

  c. Ngikenimaa gii-bashkizwaadj.
   ni-giken-im-aa   gii-bashkizaw-aa-d
   1-know-TR-1>3(INDEP) PAST-shoot-3OBJ-2(CONJ)
   'I know that you shot **him**.

           [From Lochbihler & Mathieu 2016: (50)][43]

(72) nít-ssksinoa-**wa** **m**-aníst-sskonata'psspi.   (Blackfoot)
   1-know-3   3-manner-strong
   'I know how strong he is.
        [Bošković 2003: (5), from Frantz (1978)]

Let us come back to the three possibilities of analysis listed in (56) (repeated as (73) for convenience):

(73) a. to relax the locality constraints on Agree such that the Phase Impenetrability Condition (PIC) does not apply;

  b. to analyze the embedded clause as deficient (e.g., a restructuring configuration or an ECM construction); or

  c. to propose that a copy of the phrase participating in long-distance agreement has a copy in the left-periphery of the embedded clause.

We have already pointed out that option (73a) is supported by Bošković's (2003, 2007) analysis. He claims that data from Blackfoot (see (72) above) and other Algonquian languages such as Chukchee provide empirical support to argue that Agree must be able to cross strong-CP phases. Therefore, the PIC in Chomsky 2000, 2001's version does not hold. If this is so, nothing prevents a Probe from searching for a Goal inside a CP domain, unless there is a higher intervener. This is precisely what we would claim for our analysis, following PIC$_2$. We will come back to this in section 4.

On the other hand, Fry (2016) and Fry & Hamilton (to appear) adopt option (73c) (similarly to what has been proposed for Tsez by Polinsky & Potsdam 2001). They posit

---

[43] Glosses from the authors: CONJ =conjunct; INDEP =independent; INTR =intransitive; OBJ =object; PAST =past tense; X > Y= X agent, Y patient.

that the embedded DP leaves a copy in the left-periphery of the embedded CP[44] in order to participate in Agree:

(74)
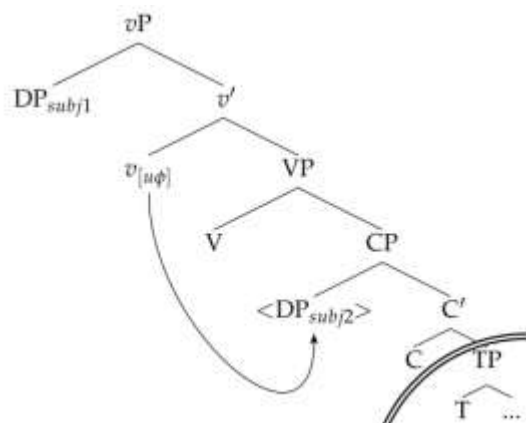
A special interpretation of the DP at the left periphery (in Tsez it is a topic, cf. Polinsky & Potsdam 2001 while in Algonquian languages such as Ojibwe it is a source of direct evidence, cf. Fry & Mathieu in press) justifies this analysis. The variation among languages would be then specified by the featural content of C, either δ or ϕ-features (Lochbihler & Mathieu (2016)). In Spanish, the embedded DP does not seem to have a particular interpretation, even though it is a possibility that must be carefully studied in the future.

According to this hypothesis, "following Oxford (2014), if v does not find a potential Goal for Agree within its search domain, the result is not a derivational crash (Chomsky 1995) but rather the appearance of default morphology (Preminger 2011)" (Fry 2016: 8). In other words, if the DP does not move to Spec-C the matrix verb shows default morphology. The underlying assumption seems to be that this happens when C does not have any features (nor δ or ϕ). Another reason could be that when the embedded clause possesses ϕ-features, LDA is banned (Bošković 2003).

In sum, in accounting for LDA across embedded clauses in Algonquian there seem to be two options, the embedded DP could either stay *in situ*, if we assume that Agree has access to the lower part of the phase, or either leave a copy in the left periphery. It should be noted, as already put forward by Bošković (2003, 2007), that these analyses challenge

---

[44] Boeckx (2009) claims something similar, although the DP moves overtly to the edge of embedded CP.

the Activity Condition again. Data from Algonquian languages reveal that the same DP can be the Goal of two Probes, the embedded and the matrix verb.

4   ANALYSIS

In what follows we discuss two strategies for analyzing the H-Agr data in Spanish. The former collects different aspects of the approaches we have just seen for Icelandic, Basque and Algonquian languages. Following Chomsky's ideas on Agree and the updated version of the PIC, we propose two hypotheses based on defective domains. The latter is based on a quite different approach without leaving the Minimalist framework, namely Uriagereka & Castillo's (2002) proposal for cyclic movement to which we will refer as *Tucking in* (Richards 1997).

4.1   Defective domains

The main concern for a suitable analysis of H-Agr is to account for the variation between the 'canonical' non-agreeing version and the agreeing one. We offer two possibilities:

A. *TP-analysis:*
      i. non-agreeing: the embedded clause is a CP
      ii. H-Agr: the embedded clause is a TP

B. *pro-analysis:*
      i. non-agreeing: the embedded clause is in apposition with a nominal (*pro*)
      ii. H-Agr version: the embedded clause is a *bona fide* argument

Both hypotheses are underpinned by the same intuition: the embedded clause in the non-agreeing version must be more complex in order to bar agreement. In other words, the embedded clause loses some structure, becomes defective, when there is agreement. Let us consider each option in detail, for non-finite and for finite H-Agr respectively.

4.1.1 TP-ANALYSIS

In (75) we can see that the embedded clause has lost some of its functional structure: it is a TP, instead of a CP.

(75) $[_{CP}$ C $[_{TP}$ T$_{[uN\_]}$ $[_{vP}$ $v_{[defective]}$ $[_{VP}$ **V$_{3PL}$** $[_{TP}$ PRO $[_{vP}$ v $[_{VP}$ V$_{INF}$ DP$_{[iN: PL]}$]]]]]]]]

ACC

AGREE
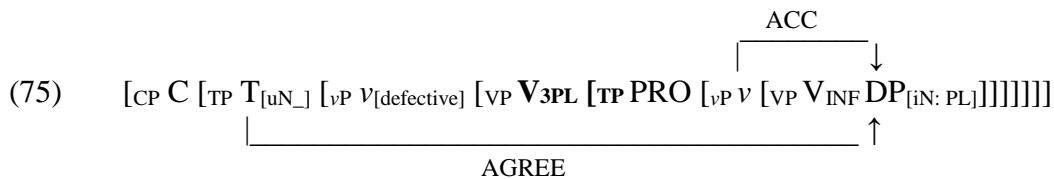
Matrix *v* is defective (does not assign Case)[45], while the embedded *v*P does assign ACC to the embedded DP (as data in (37) revealed). The novelty in respect of similar approaches is that, following PIC$_2$, the embedded DP remains *in situ*, it does not need to move to the edge of the phase to be accessible. Matrix T, as the Probe with an unvalued number feature can search inside the domain of the lower phases until it matches with the valued number feature of the DP, assuming that PRO is not an intervener[46]. Crucially, the embedded DP receives Case from embedded *v*, but not from T. The DP then becomes frozen but it is still visible for T as the Goal of Agree.

Compare H-Agr analysis in (75) with its non-agreeing counterpart in (76):

(76)   [$_{CP}$ C [$_{TP}$ T [$_{vP}$ *v*$_{[defective]}$ [$_{VP}$ **V3SG** [$_{CP}$ [$_{TP}$ PRO [$_{vP}$ *v* [$_{VP}$ V$_{INF}$ DP]]]]]]]]
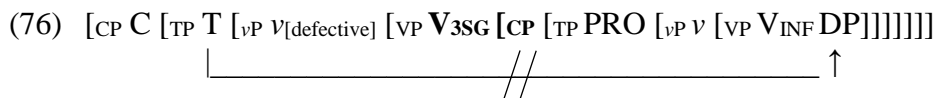
The relevant difference, in this case, is that the embedded clause does have its whole CP structure. An immediate counterargument is that if T is able to search inside a lower phase (*v*P) in (75), nothing prevents it from doing the same across the CP phase. A plausible hypothesis could be that C bears some kind of nominal features that act as defective interveners blocking agreement so that it cannot go further down. One possibility would be to assume Picallo's (2002) proposal (cf. (21)) according to which the clausal argument would possess negative ϕ-features. T would then find the clause as a Goal and trigger default 3rd singular morphology. This would support the above-mentioned idea that LDA is not possible when the embedded clause possesses ϕ-features (Bošković 2003).

4.1.2 PRO-ANALYSIS

In the second scenario that we propose, the non-agreeing version has some additional structure: V selects a nominal *pro* as its complement to which the CP is an apposition. Consistently, T values its ϕ-features with *pro* and is not able to extend its search further on until the embedded DP.
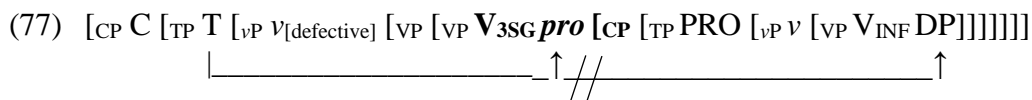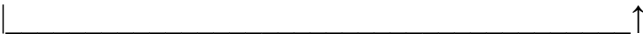
(77)   [$_{CP}$ C [$_{TP}$ T [$_{vP}$ *v*$_{[defective]}$ [$_{VP}$ [$_{VP}$ **V3SG** *pro* [$_{CP}$ [$_{TP}$ PRO [$_{vP}$ *v* [$_{VP}$ V$_{INF}$ DP]]]]]]]]]

H-Agr clauses would lack this *pro*, allowing agreement to cross the CP boundary:

(78) [$_{CP}$ C [$_{TP}$ T [$_{vP}$ $v_{[defective]}$ [$_{VP}$ **V3PL** [$_{CP}$ [$_{TP}$ PRO [$_{vP}$ v [$_{VP}$ V$_{INF}$ DP]]]]]]]]
   |———————————————————————————↑

The rest of the elements would behave as we have postulated for the TP-analysis: matrix T and *v* cannot assign Case, the embedded DP receives ACC from embedded *v* and remains visible for Agree, according to PIC$_2$.

This is a variation of Torrego & Uriagereka's (1992) analysis (cf. (22), repeated as (79) for convenience):

(79) a. [$_{vP}$ . . . V . . . [$_{DP}$ pro] [$_{CP}$ V$_{INDICATIVE}$]]

  b. [$_{vP}$ . . . V . . . [$_{CP}$ V$_{SUBJUNTIVE}$]]

Since we consider that all clausal arguments have the structure in (79a), we need a different account for the indicative/subjunctive distinction. This is not the only consequence of this analysis, the assumption that clausal arguments in Spanish are appositions is not trivial and requires careful study that lies beyond the scope of this thesis. Despite the possible pitfalls, in the next section we show that this analysis may be more plausible for finite H-Agr and, even if it should be discarded in the future, it is at least thought-provoking.

4.1.3 WHAT ABOUT FINITE CLAUSES?
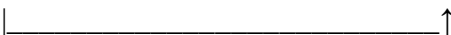
The analysis of finite H-Agr is a challenge, not only because its intrinsic complexity, but also because our data are limited in that, because of practical constraints, we could not observe different patterns in the same speaker. Consequently, we cannot assure that all speakers that produce H-Agr sentences accept both finite and non-finite versions. This information would help us to determine if both versions should be explained under the same analysis or if they require different accounts.

For the moment, we can discuss if the *TP-* and *pro*-analyses operate for non-finite H-Agr as well. Consider them respectively in (80):

(80) a. [$_{CP}$ C [$_{TP}$ T$_{[uN\_]}$ [$_{vP}$ $v_{[defective]}$ [$_{VP}$ **V3PL** *que* [$_{TP}$ DP$_{[iN:\ PL]}$ [$_{vP}$ v [$_{VP}$ V$_{INF}$ DP]]]]]]]]
     |———————————————————————↑

  b. [$_{CP}$ C [$_{TP}$ T [$_{vP}$ $v_{[defective]}$ [$_{VP}$ **V3PL** [$_{CP}$ *que* [$_{TP}$ DP$_{[iN:\ PL]}$ [$_{vP}$ v [$_{VP}$ V$_{INF}$ DP]]]]]]]]
     |———————————————————↑

Whereas in (80b) the complementizer *que* occupies its position in Cº, such space is not available in (80a), because the CP layer is not projected. If we want to keep the *TP-analysis* for finite clauses we should then postulate a new status for *que* that crucially cannot be C.

Regarding the *pro-analysis*, apart from its best suitability concerning the complementizer, it seems on the right track if we compare it to English:

(81) a. Me gusta (***pro***) [CP que canten]
      b. I like **it** [CP that they sing]

We are postulating then that bi-clausal psych-verb constructions behave equally in English and in Spanish, both select a pronoun, overt in English and covert in Spanish, being the finite clause dependent on that pronoun. Because of the nature of the predicates, the difference remains on Case assignment: while *it* receives ACC, Spanish *pro* receives NOM.

To explain the contrast between indicative and subjunctive clauses in the *pro*-analysis we can contemplate Uriagereka's (2006, 2008) proposal. He argues that, assuming that C receives Case, V assigns oblique case to subjunctive C (as it would be the counterpart to partitive NP) and direct Case to indicative C.

Another option would be to assume Gallego's (2010) claim that subjunctive constructions in Spanish behave as ECM constructions in English:

(82) a. Jon wants [TP Sansa to rule the North]   ENGLISH ECM
          |_____↑
                ACC

      b. *Jon quiere [TP Sansa gobernar$_{INF}$ el Norte]   SPANISH ECM (IMPOSSIBLE)

      c. Jon quiere [CP C$_{def}$ que Sansa *gobierne$_{SBJ}$* el Norte] SPANISH SUBJUNCTIVE
          |_____↑
                ACC

Matrix *v* assigns abstract accusative to the EA of the subjunctive clause because C is defective (see Gallego 2010: §3.5 for technical details). This option is especially suitable for our *TP-analysis*: finite H-Agr would be then even closer to English ECM since we postulate that subjunctive clauses in H-Agr cases are TPs.

We have not addressed the issue of why H-Agr involves number but not person agreement. Etxepare (2012) claims that person and number split in agreement. He highlights that this conclusion is not only valid for Basque LDA but also explains a vast

amount of crosslinguistic data (Baker 2008, apud Etxepare 2012). This is a possibility that should also be explored for Spanish.

Our proposals resemble to the ones for Basque and Icelandic LDA in that they posit less structure in the subordinate clause in order to agreement to cross its boundary. Under the PIC$_2$ the embedded DP is accessible in its original position. For finite clauses, we have not weigh the possibility of having a copy in the left-periphery, as the DP does not seem to have a special interpretation. However, the idea that features in C may be involved in the alternation between the agreeing and non-agreeing version seems to be in the right track, at least if we assume a *TP-* or similar analysis.

Furthermore, these approaches are consistent with the claim that the Activity Condition does not hold for Agree (Bhatt 2005, Bošković's 2007, Etxepare 2012). In the non-finite version, Agree targets a DP that has received Case. More significantly, in the finite version, it targets a DP that has already been targeted before by the embedded verb. The fact that the matrix verb only agrees in number may be crucial in explaining this fact. In any case, this evidence lead us to rethink conditions on Agree.

4.2 Tucking-in (Uriagereka & Castillo 2002)

In this section, we study a not-so-canonical type of analysis, suggested by Uriagereka & Castillo (2002) (U&C from now on) for long-distance *wh*-movement. U&C proposed to adapt the so-called *Tree Adjoining Grammar* (Kroch and Joshi 1985) (TAG) to the Minimalist framework, following Richards (1997) *tucking-in.*

Long-distance *wh*-movement, such as the one in (83) and other long-distance dependencies have been explained by 'successive cyclicity' roughly, the movement proceeds in a step-by-step fashion. This idea comes from Chomsky's (1973) claim that movement must obey a *Subjacency Condition*, whereby movement cannot cross more than one of the bounding nodes (S = CP and NP = DP) (Gallego 2011), which in the Minimalist Program have been substituted by phases which are not accessible but their edges (U&C).

(83)   a. Who John thinks that Mary loves?

b. [CP Who [TP John T [vP *v* [VP thinks [CP that [TP Mary T [vP *v* [VP loves t$_i$]]]]]]]]
     ↑_____↑_____↑_____↑_____|

U&C question the status of phase edges and of the *wh*-feature that the embedded Comp is claimed to possess to trigger movement. In this context, they propose an analysis inspired in the TAG, whose main idea is very simple: *bona fide* movement is always local and what moves the elements further is tree splitting by insertion of other elements in between, as we can see in (84) below. As U&G point out, this approach captures the intuition that the simpler question *"who does Mary love?"* underlies the more complex one.

(84)

```
                    CP
                  /    \
              who_j     C'
                      /    \
      merge → W           C'
           /‾‾‾‾\        /    \
    John thinks that C      [_IP Mary loves t_j]
```

[From U&C: (9)]
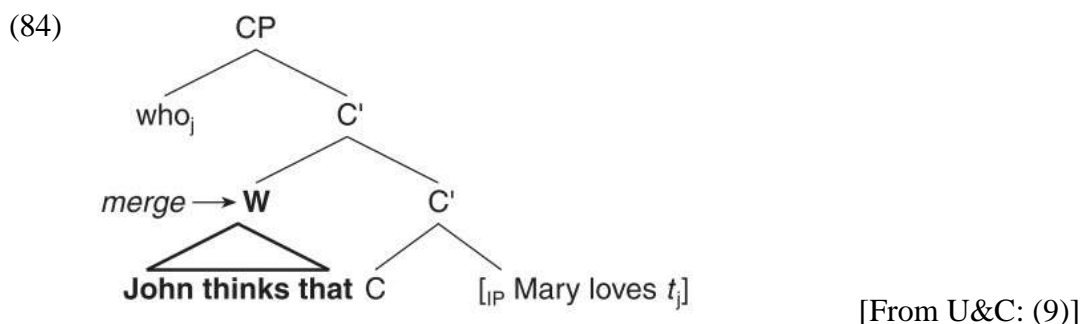
The idea in (84) is improved by Richards' tucking in, which U&G claim that it is tacitly assumed in Chomsky 2000. The gist of this proposal is that the lexical items are merged below the matrix CP one by one, instead of inserting a whole subtree:
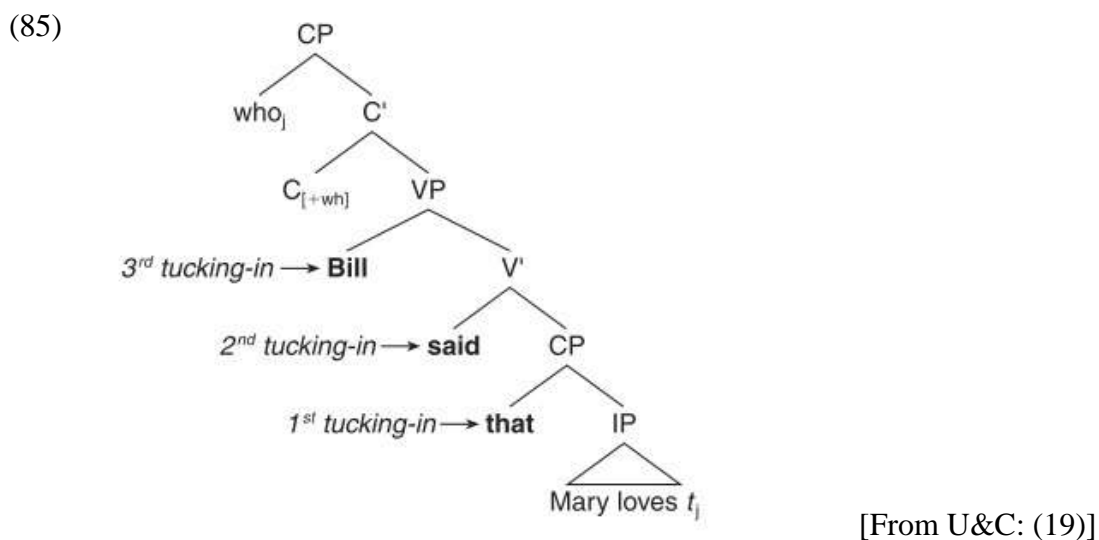
(85)

```
                        CP
                      /    \
                  who_j     C'
                          /    \
                     C_[+wh]    VP
                              /    \
       3rd tucking-in → Bill      V'
                                /    \
         2nd tucking-in → said      CP
                                  /    \
             1st tucking-in → that    IP
                                     /‾‾\
                               Mary loves t_j
```

[From U&C: (19)]

According to U&C, this proposal does not violate the requirements in *Bare Phrase Structure* (Chomsky 1995), it can explain island violations (when the computational system cannot distinguish two Comps close to each other during the derivation) and it

dispenses with *ad hoc* features required for the derivation to proceed[47] (see U&C for technical details).

We find this proposal very appealing for the kind of long distance dependency we are dealing with. Following this approach, the relevant elements would agree within a regular adjacent context and become further separated, since the rest of the elements are being tucked in. Examples in which there is only an infinitive separating the main verb from the DP are the most basic patterns. As illustrated in (86), the infinitive would be tucked in, forcing the agreeing elements to part.

(86)   Me **asustan (las) pesadillas** >> Me **asustan** *tener* **pesadillas**
                                                                    ↑

Whether the DP in examples such as (86) requires the determiner or not depends on the selection conditions of the predicate, which would take place in FL. According to this idea, agreement would be strictly syntactic while the elements would 'accommodate' to semantic requirements post-syntactically.

Finite H-Agr data could also be explained by these means:

 (87) a. Me **sorprenden** tanto [que **las cosas** sean así]   (from Twitter, n.p.)
         me$_{DAT}$ surprise$_{3PL}$ so.much that the things are$_{3PL.SBJ}$ this.way
         'I am very surprised by things being this way'

     b. Me **sorprenden** tanto **las cosas** >> Me **sorprenden** tanto *que sean así* **las cosas**
                                                                    1st↑ 2nd↑   3rd↑
     c. Me **sorprenden** tanto *que* **las cosas$_i$** *sean así* **t$_i$**

Taking into account that the order in (87b) is also possible in Spanish, we assume that tuck-in would take place with the subject *in situ* (in postverbal position) and it will move (or not) to its preverbal position afterwards (87c).

We are aware that the technical details of this proposal should be discussed thoroughly. However, given the apparently oral/spontaneous nature of the phenomenon, a non-that-canonical approach could work. A proposal such as the one we have discussed embraces the intuition that the sentence may start as a mono-clausal psych-verb construction, where T agrees with the IA argument and then insert the rest of the material resulting in a canonical bi-clausal sentence with deviant agreement.

---

[47] Even it is not explicitly said in their squib, U&C's proposal violates the *No Tampering Condition* (Chomsky 2007, 2008) according to which, once X, Y have been merged, "Merge cannot break up X or Y, or add new features to them. Therefore Merge is invariably ''to the edge''" (Chomsky 2008: 138).

5   CONCLUSIONS

In this thesis, we have presented a panoramic view of Spanish hyper-agreement phenomenon from an empirical and theoretical perspective. We have looked at structures whereby agreement goes beyond it is expected: T agrees with a deeply embedded argument, either the IA of an infinitive or the EA of a subjunctive predicate.

We have shown that H-Agr is possible in DAT-NOM configurations, where the dative phrase is an experiencer and the object is a clause. The predicates that allow this configuration are basically psych-verbs of the *gustar* and *preocupar*-types (according to Belletti & Rizzi's 1988 classification).

We departed from three isolated examples from the GDLE, which had barely any replication in standard corpora. Nevertheless, Twitter gave us access to spontaneous language from heterogeneous Spanish varieties, both from America and Europe. Our search on this platform has resulted in hundreds of instances of H-Agr. This substantial number of results, as well as its consistency with other phenomena that seem to indicate a tendency for agreement in Spanish, casts doubts on the idea that these examples are only lapsus.

An important matter for future studies is, precisely, determining which of our examples belong to the paradigm and which are mere errors. We expect to solve this puzzle by finding informants who could provide insight about the optionality of the phenomenon and other unresolved questions. Along the same lines, these questions may open the door for psycholinguistic research about agreement in Spanish and other Romance languages.

Notwithstanding these empirical limitations, we have been able to systematize and organize the data so we could extract the properties of H-Agr in order to determine if it was a genuine instance of LDA. On the one hand, we have seen that despite their outstanding differences, H-Agr makes Spanish comparable to other languages that display LDA, such as Icelandic, Basque and Algonquian languages. On the other hand, we have shown that H-Agr can be analyzed by means of Agree.

Because of the complexity of the phenomenon and the empirical constraints, we have not defended a specific analysis, but we have explored two different strategies: defective domains and tucking-in. The latter suggests a different perspective that captures the intuition behind H-Agr data: there is a simple sentence (where local T-IA takes place regularly) underlying the complex bi-clausal structure with deviant agreement. Although very appealing the technical implementation of this hypothesis needs of careful study.

The other strategy builds on the intuition that there is variation in the boundaries between matrix and subordinate clauses that interact with agreement. A *TP*-analysis assumes that this variation is due to the lack of the CP-layer, while a *pro*-analysis suggests that the presence of a nominal element is what regulates the variation. While the former is more intuitive, the latter better unifies finite and non-finite H-Agr, even though it forces us to rethink the status of clausal dependencies. Both hypotheses implement updated versions of key theoretical notions such as the *Phase Impenetrability Condition* and the *Activity Condition* (Chomsky 2000, 2001 et seq.).

This thesis may be another instance of how the study of long distance agreement provides challenging data that requires revisiting non-trivial theoretical assumptions. Our proposals may contribute to the already fruitful research about long distance dependencies, clausal complementation, the relationship among indicative, subjunctive and infinitive clauses, and the nature of agreement in Spanish in relation to other languages.

6 REFERENCES

Alcina, Juan & Blecua, José Manuel. 1975. *Gramática española.* Barcelona: Ariel.

Acedo-Matellán, Victor & Jaume Mateu. 2015. Los verbos psicológicos: raíces especiales en estructuras corrientes. In Rafael Marín (ed.), *Los predicados psicológicos*, 81-104. Madrid: Visor.

Alexiadou, Artemis, Elena Anagnostopoulou, Gianina Iordachioaia & Mihaela Marchis. 2012. In Support of Long Distance Agree. In Alexiadou, Artemis, Tibor Kiss & Gereon Müller (eds.), *Local Modelling of Nonlocal Dependencies in Syntax*, 55–81. Berlin: Walter de Gruyter.

Aloy, Núria. 2016. La concordancia en el verbo existencial *haber* en español. Barcelona: Universitat de Barcelona MA Thesis.

Andueza, Patricia. 2012. The Subject of Psych Verbs in Spanish. In González-Rivera, Melvin & Sandro Sessarego (eds.), *Current Formal Aspects of Spanish Syntax and Semantics,* 95-120. Newcastel upon Tyne, UK: Cambridge Scholars.

Baker, Mark C. 2008. *The Syntax of Agreement and Concord.* Cambridge: CUP.

Baker, Mark C. 2013. Agreement and Case. In Marcel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax,* 607-654. Cambridge: CUP.

Belletti, Adriana & Luigi Rizzi. 1988. Psych Verbs and Theta Theory. *Natural Language and Linguistic Theory* 6. 291-352.

Bhatt, Rajesh. 2005. Long-distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23. 757-807.

Boeckx, Cedric. 2000. Quirky agreement. *Studia linguistica* 54(3), 354-380.

Boeckx, Cedric. 2008a. *Aspects of the syntax of agreement*. London: Routledge.

Boeckx, Cedric. 2008b. The PPC and patterns of exclusivity. In D'Alessandro, Roberta, Susann Fischer & Gunnar H. Hrafnbjargarson (eds.), *Agreement restriction,* 87-101. Berlin: Walter de Gruyter.

Boeckx, Cedric. 2009. On Long-Distance Agree. *An International Journal of Theoretical Linguistics* 111(1). 1-32.

Bošković, Željko. 2003. Agree, Phases, and Intervention Effects. *Linguistic Analysis* 33. 54–96.

Bošković, Željko. 2007. On the locality and motivation of move and Agree: An even more minimal theory. *Linguistic Inquiry* 38, 589-644.

Bosque, Ignacio (ed.). 1990. *Indicativo y subjuntivo.* Madrid: Taurus.

Bosque, Ignacio. 2012. Mood: Indicative vs. Subjunctive. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The Handbook of Hispanic Linguistics,* 373-394. London: Willey-Blackwell.

Bruening, Benjamin. 2001. *Syntax at the Edge*: *cross-clausal phenomena and the syntax of Passamaquoddy.* Cambridge, MA: MIT dissertation.

Campos, Héctor. 1999. Transitividad e intransitividad. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1519-1574. Madrid: Espasa Calpe.

Cuervo, María Cristina. 2003. *Datives at Large.* Cambridge, MA: MIT dissertation.

Cuervo, María Cristina. 2010a. Against Ditransitivity. *Probus* 22. 151-180.

Cuervo, María Cristina. 2010b. Some Dative Subjects Are Born, Some Are Made. In C. Borgonovo et al. (eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 26–37. Somerville, MA: Cascadilla Proceedings Project.

Chandra, Pritha. 2007. *(Dis)agree: movement and agreement reconsidered.* College Park, MD: University of Maryland dissertation.

Chomsky, Noam. 1995. Bare Phrase Structure. In G. Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, Oxford: Blackwell.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Robert Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step*, 89-155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud,* 134–166. Cambridge, MA: MIT Press.

Chomsky, Noam, Ángel J. Gallego & Dennis Ott. 2017. Generative Grammar and the Faculty of Language: Insights, Questions and Challenges. Ms.

Chung, Sandra. 2004. Restructuring and Verb-Initial Order in Chamorro. *Syntax* 7. 199-233.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective.* New York, Oxford University Press.

Cinque, Guglielmo. 2001. Restructuring and Functional Structure. Ms., University of Venice.

D'Alessandro, Roberta. 2007. *Impersonal* si *Constructions.* Amsterdam: Walter de Gruyter.

De Benito, Carlota & Ana Estrada. 2016a. Si Alvar levantara la cabeza: cómo tratar con informantes virtuales. *Revista Internacional de Lingüística iberoamericana* 2(28). 7–17.

De Benito, Carlota & Ana Estrada. 2016b. Variación en las redes sociales: datos twilectales. *Revista Internacional de Lingüística iberoamericana* 2(28). 77–114.

Di Tullio, Ángela. 2011. Infinitivos introducidos por de. *Cuadernos de la ALFAL* 3. 176-187.

Donoso, Gonzalo & David Sánchez. 2017. Dialectometric analysis of language variation in Twitter. *Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial'2017).* 16-25.

Eisenstein, Jacob. 2014. Identifying regional dialects in online social media. Ms., Georgia Institute of Technology.

Etxepare, Ricardo. 2006. "Number long distance agreement in (substandard) Basque". In J. A. Lakarra & J. I. Hualde (eds.), *Studies in Basque and historical linguistics in memory of R. L. Trask. Supplements of the Anuario del Seminario de Filologia Vasca "Julio de Urquijo"* XL(1-2): 303-350.

Etxepare, Ricardo. 2012. *Agreement Configurations in Basque: Lessons from Distance*. Bayonne, France: CNRS-IKER habilitation thesis.

Fernández Soriano, Olga. 1999. Two types of impersonal sentences in Spanish: Locative and dative subjects. *Syntax* 2.101-140.

Fernández-Soriano, Olga & Susana Táboas. 1999. Construcciones impersonales no reflejas. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1723-1778. Madrid: Espasa Calpe.

Franco, Jon. 1992. Towards a typology of psych verbs: Evidence from Spanish. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"* 27. 119-134.

Frantz, Donald. 1978. Copying from complements in Blackfoot. In Eung-Do Cook and Jonathan Kaye (eds.), *Linguistic studies of native Canada*, 89-110. Vancouver: University of British Columbia Press.

Fry, Brandon J. 2016. Cross-clausal agreement in Algonquian: its wide implications for syntactic theory. Invited student talk at the 21st Workshop on Structure and Constituency in the Languages of the Americas (WSCLA 21), April 1-3, 2015, Université du Québec à Montréal, Montréal, Québec, Canada.

Fry, Brandon J. & Michael David Hamilton. To appear. Long-distance agreement in Algonquian: accounting for syntactic variation. In *Proceedings of the 50th annual meeting of the Chicago Linguistics Society*. Chicago: Chicago Linguistics Society.

Fry, Brandon J. & Éric Mathieu. In press. La sémantique de l'accord à longue distance en ojibwé algonquin. In *Papers of the 46th Algonquian Conference*. SUNY Press.

Gallego, Ángel J. 2009. Defective C-T in Romance. *Probus* 21. 163–216.

Gallego, Ángel J. 2010. *Phase Theory*. Amsterdam: John Benjamins.

Gallego, Ángel J. 2016. Three Types of Prepositions in Spanish *se* Sentences. Consequences for cross-dialectal studies. *Dialectologia* 17. 51-70.

Gallego, Ángel J. 2017. Agreement asymmetries under dislocation: Evidence for a biclausal analysis. Ms., Universitat Autònoma de Barcelona.

Gallego, Ángel J. & Juan Uriagereka. 2011. Dos tipos de argumentos y la distinción indicativo vs. subjuntivo. *Cuadernos de la ALFAL* 3.188-199.

Gonçalves, Bruno & David Sánchez. 2014. Crowdsourcing Dialect Characterization through Twitter. *PLoS ONE* 9(11). 1-6.

Gonçalves, Bruno & David Sánchez. 2016. Learning about Spanish Dialects through Twitter. *Revista Internacional de Lingüística iberoamericana* 2(28). 65-75.

Gutiérrez-Bravo, Rodrigo. 2006. A reinterpretation of quirky subjects and related phenomena in Spanish. Ms., CIESAS-Mexico City.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization.* Cambridge University Press.

Herbeck, Peter. 2013. (Backward) Control and Clitic Climbing On the Deficiency of Non-finite Domains in Spanish and Catalan. *Proceedings of ConSOLE XXI*, 123–145. University of Leiden.

Hernanz, M. Lluïsa. 1999. El infinitive. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 2197-2356. Madrid: Espasa Calpe.

Hernanz, M. Lluïsa & Gemma RIGAU. 1984. Auxiliaritat i reestructuració. *Els marges* 31. 29-51.

Jones, Taylor. 2015. Toward a Description of African American Vernacular English Dialect Regions Using 'Black Twitter'. *American Speech* 90(4): 403-440.

Kilgarriff, Adam & Gregory Grefenstette. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3). 333–347.

Kroch, Anthony & Aravind K. Joshi.1985. The Linguistic Evidence of Tree Adjoining Grammar. Philadelphia: University of Pennsylvania Department of Computer and Information Science Technical Report MS-CIS-85-16.

Kulkarni, Vivek, Bryan Perozzi & Steven Skiena. 2016. "Freshman or Fresher? Quantifying the Geographic Variation of Language in Online Social Media". *Procceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016).* 615-618.

Lochbihler, Bethany C. 2012. *Aspects of argument licensing*. Montreal, Québec: McGill University Dissertation.

Lochbihler, Bethany C. & Eric Mathieu. 2016. Clause Typing and Feature Inheritance of Discourse Features. *Syntax* 19(4). 354–391.

López, Luis. 2007. *Locality and the Architecture of Syntactic Dependencies*. New York, NY: Palgrave Macmillian

Marantz, Alec. 1991. Case and licensing. In G. Westphal, B. Ao & H.-R. Chae (eds.), *Proceedings of the 8th Eastern States Conference on Linguistics (ESCOL 8).* Ithaca, NY: CLC Publications, 234–253.

Mare, María. 2016. Concordancias duplicadas en español. Handout. Seminari del CLT, Universitat Autònoma de Barcelona.

Marín, Rafael. 2015. *Los predicados psicológicos*. Madrid: Visor.

Martínez, José Antonio. 1999. La concordancia. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 2695-2786. Madrid: Espasa Calpe.

Martins, Ana Maria & Jairo Nunes. 2005. Raising Issues in Brazilian and European Portuguese. *Journal of Portuguese Linguistics* 4. 53–77.

Martins, Ana Maria & Jairo Nunes. 2009. Syntactic change as chain reaction: the emergence of hyper-raising in Brazilian Portuguese. In Paola Crisma & Giuseppe. Longobardi (eds.), *Historical Syntax and Linguistic Theory*. New York: OUP.

Masullo, Pascual J. 1993. Two types of quirky subjects: Spanish versus Icelandic. *Proceedings of Nels* 23. 303-317.

Méndez García de Paredes, Elena. 2011. Si yo fuera estado allí, no fuera pasado eso. Pervivencia de un aparente arcaísmo en la lengua de Internet. In José J. Bustos Tovar (ed.), *Homenaje a Antonio Narbona,* vol. 2, 1009-1032. Sevilla: Universidad de Sevilla.

Mendívil, José Luis. 2012. Ergativity as transitive unaccusativity. *Folia Linguistica* 46(1). 171–210.

Ordóñez, Francisco. 2012. Clitics in Spanish. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The Handbook of Hispanic Linguistics,* 423-452. London: Willey-Blackwell.

Ordóñez, Francisco & Esthela Treviño. 2011. Impersonals with Passive Morphology. In Luis A. Ortiz-López, *Selected Proceedings of the 13th Hispanic Linguistics Symposium,* 314-324. Somerville, MA: Cascadilla Proceedings Project.

Ordóñez, Francisco & Esthela Treviño. 2016. Agreement and DOM with impersonal se. In Mary A. Kato & Francisco Ordóñez (eds.), *The Morphosyntax of Portuguese and Spanish in Latin America*, 236-258. Oxford: OUP.

Oxford, Will. 2014. *Microparameters of agreement: a diachronic perspective on Algonquian verb inflection*. Toronto: University of Toronto Dissertation.

Pentland, David H. 2006. Algonquian and Ritwan Languages. In Keith Brown, (ed.), *Encyclopedia of Languages and Linguistics,* 2nd edn., 161–6. Amsterdam: Elsevier.

Pesetsky, David. 1995. *Zero Syntax: Experiencers and Cascades.* MIT Press.

Picallo, M. Carme. 1985. *Opaque Domains.* New York, NY: CUNY Dissertation.

Picallo, M. Carme. 2002. Abstract Agreement and Clausal Arguments. *Syntax* 5(2). 116–147.

Polinsky, Maria. 2003. Non-canonical agreement is canonical. *Transactions of the philological society* 101, 279-312.

Polinsky, Maria & Eric Potsdam. 2001. Long-distance agreement and topic in Tsez. *Natural Language and Linguistic Theory* 19. 583-646.

Paradís, Anna. 2016. Més enllà dels verbs de reestructuració, *Miscel·lània homenatge a Jordi Bruguera, vol. IV. Estudis de Llengua i Literatura Catalanes*, 209-229. Barcelona: Publicacions de l'Abadia de Montserrat.

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.* Cambridge, MA: MIT dissertation.

Quer, Josep. 2008. Argumental Clauses and Nominal Features. *Lingue e linguaggio* 1. 93–110.

Real Academia Española.1973. *Esbozo de una nueva gramática de la lengua española.* Madrid: Espasa Calpe.

Richards, Mark. 2009. Probing the past: On reconciling long-distance agreement with the PIC. Ms., University of Leipzig.

Richards, Norvin. 1997. *What moves where when in which language*. Cambridge, MA: MIT dissertation.

Rivero, María Luisa. 1971. Mood and presupposition in Spanish. *Foundations of Language* 7. 302-336.

Rivero, María Luisa. 2004. Spanish quirky subjects: person restrictions and the Person-Case constraint. *Linguistic Inquiry* 35. 494-502.

Rizzi, Luigi. 1978. A restructuring rule in Italian syntax. *Transformational Studies in European Languages*, 113-158. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1982. *Issues in Italian syntax.* Dordrecht: Foris.

Rodríguez-Mondoñedo, Miguel. 2005. Case and Agreement in Spanish Existential Constructions (and beyond)". Ms., U. Conn.

Ruiz Tinoco, Antonio. 2013. Variación léxica y sintáctica del español en las redes sociales. *Actas del I Congreso Internacional sobre el español y la cultura hispánica del Instituto Cervantes de Tokio*, 231–251. Centro Virtual Cervantes.

Ruiz Tinoco, Antonio. 2016. "Geocorpus del español de las redes sociales y cartografía automática". *IX Congreso Internacional Asociación Asiática de Hispanistas*.

Sánchez-López, Cristina. 2002. Las construcciones con *se*. Estado de la cuestión. In Cristina Sánchez López (ed.), *Las construcciones con se*, 18-142. Madrid: Visor.

Schütze, Carson T. 2009. Web searches should supplement judgements, not supplant them. *Zeitschrift fur Sprachwissenschaft* 28(1). 151–156.

Schütze, Carson T. 2011. Linguistic evidence and grammatical theory. *WIREs Cognitive Science* 2. 206–221.

Seco, Ramón. 1988. *Manual de Gramática Española.* Madrid: Aguilar.

Sigurðsson, Halldór Ármann. 1991. Icelandic Case-marked PRO and the licensing of lexical arguments. *Natural Language and Linguistic Theory* 9. 327–364.

Torrego, Esther & Juan Uriagereka. 1992. Indicative Dependents. Ms., U. Mass Boston/UMD.

Uriagereka, Juan. 2006. Complete and Partial Infl. In Cedric Boeckx (ed.), *Agreement Systems*, 267–298.Amsterdam: John Benjamins.

Uriagereka, Juan. 2008. 2008. *Syntactic Anchors. On Semantic Structuring*. Cambridge: CUP.

Uriagereka, Juan. 2015. Subordinación: indicativos y subjuntivos. In Ángel J. Gallego (ed.), *Perspectivas de sintaxis formal*, 265-298. Madrid: Akal.

Uriagereka, Juan & Juan Carlos Castillo. 2002. A note on successive cyclicity. In J. Uriagereka, *Derivations*, 136–46. London: Routledge.

Vanhoe, Henk. 2002. *Aspectos de la sintaxis de los verbos psicológicos en español: un análisis léxico funcional.* Ghent, the Netherlands: Ghent University Dissertation.

Wurmbrand, Susi. 2001. *Infinitives: Restructuring and clause structure*. Berlin: Mouton de Gruyter.

Wurmbrand, Susi. 2004. Two types of restructuring: Lexical vs. functional. *Lingua* 114(8). 991-1014.

Corpora

ASinEs: Gallego, Ángel. 2014. *Atlas Sintáctico del Español.* <http://asines.org>

CdE: Davies, Mark. 2002-. *Corpus del Español: 100 million words, 1200s-1900s.Web/Dialects.* <http://www.corpusdelespanol.org>

COSER: Fernández-Ordóñez, Inés (dir.). 2005-. *Corpus Oral y Sonoro del Español Rural.* <http://www.corpusrural.es> ISBN 978-84-616-4937-2

CORPES: Real Academia Española. *Corpus del Español del Siglo XXI.* <http://web.frl.es/CORPES/>

CORDE: Real Academia Española. *Corpus Diacrónico del Español.* <http://corpus.rae.es/cordenet.html>

CREA: Real Academia Española. *Corpus de Referencia del Español Actual.* <http://corpus.rae.es/creanet.html>

PRESEEA: 2014-. *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. <http://preseea.linguas.net>

EsTenTen: Sketch Engine. 2011. *Spanish Web 2011. EsTenTen11 Eu+Am.* <https://www.the.sketchengine.co.uk/estenten-corpus/>.

Variaciones: Ruiz Tinoco, Antonio. 2014. *Variación léxica y sintáctica del español en las redes sociales.* <http://variaciones.org/>