Not all complementisers are late: a first look at the acquisition of illocutionary complementisers in Catalan and Spanish

Núria Bosch
University of Cambridge, nb611@cam.ac.uk

This paper analyses the emergence of illocutionary complementisers (in the sense of Corr 2016, 2022) through a corpus study with Catalan and Spanish children. The production of illocutionary complementisers by ten Catalan- and Spanish-speaking children in the CHILDES database is quantified and compared to the production of finite embedding complementisers. The results indicate that illocutionary complementisers emerge early in the child production data, often well before embedding complementisers first appear. These preliminary findings, which illustrate important developmental differences between kinds of complementisers, are hard to account for in approaches that take functional categories to mature bottom-up, with left-peripheral knowledge developing last. I argue, instead, that the early emergence of illocutionary complementisers favours a view which takes the C-domain to be present early on in child grammars. I finish by considering the development of Italo-Romance complementisers as a future direction, suggesting that a deeper analysis of child 'errors' or input-divergent utterances may provide significant insights into the theoretical questions presented, as much as grammatical ones.

1 Two complementisers, two acquisition timings

The acquisition of complementisers and subordination is typically taken to be a crosslinguistically relatively late phenomenon in child language, in comparison to other very early phenomena, such as the acquisition of word-order and head-directionality (Tsimpli 2014). The earliest forms of subordination include so-called preconjunctionals, namely subordinate clauses which lack the target-language complementiser. The emergence of complementisers, such as Catalan and Spanish subordinator and relativiser *que*, is a later development during the period of early word-combinations (e.g., see Armon-Lotem 2005, Clahsen & Penke 1992, for discussion on Hebrew and German). In generative approaches to syntax, complementisers like *que* are syntactically the heads of the functional projection

Complementiser Phrase or CP (since Chomsky, 1977, 1986)[1], or of richer elaborations thereof, such as Rizzi's (1997) cartographic spine for the 'left periphery'. Converging with the observations regarding the comparatively relate acquisition of complementisers, theoretical approaches to acquisition have frequently advocated for initial unavailability of the CP domain; consider Radford's (1988) Small Clause Hypothesis, Rizzi's (1993/1994) Truncation Hypothesis and Friedmann, Belletti & Rizzi's (2021) Growing Trees Hypothesis. The latter of these, similarly to Radford, proposes a more general maturational mechanism whereby functional spines (a *v*P>TP>CP sequence, with a cartographic left periphery) develop bottom-up. These theoretical approaches are argued to help account for, i.a., the timings regarding the development of complementisers.

However, another kind of main-clause complementiser attested across Ibero-Romance and used prevalently in spoken contexts has gone unstudied in the context of the emergence of complementisers. These are so-called *illocutionary* complementisers (adopting the term in Corr 2016, 2022); that is, complementisers that do not function as a subordinator that heads a complement or relative clause in Ibero-Romance and instead introduce non-embedded matrix clauses, with several illocutionary functions. A canonical example of a complementiser which can introduce complement clauses or relative clauses is illustrated in (1), while examples of illocutionary complementisers heading a main clause are provided in (2)[2]:

(1) Subordinating complementisers
   a. Li   he   dit **que** aquesta tarda   vaig   a Barcelona   (Catalan)
      CL.IO= AUX.1SG told that this   afternoon go.1SG to Barcelona
      a veure.   un concert al   Liceu
      to see.INF a concert in.the   Liceu

      'I have told him/her that I'm going to Barcelona this afternoon to see a concert in El Liceu'

   b. No podía   creer   que hubiesen   ganado   la   (Spanish)
      not can.IMPF.3SG believe that AUX.SUBJ.IMPF.3PL   win.PTCP the
      lotería
      lottery

      'He/she couldn't believe that they'd won the lottery'

---

[1]   Note, however, that C and the precursor of the CP, namely S', are present or implicit in generative representations early on (see, e.g., Chomsky 1955, Rosenbaum 1967).
[2]   Unless otherwise noted, the Catalan and Spanish examples provided throughout this paper are mine and reflect the dialects of Central Catalan and Peninsular Spanish.

(2) Illocutionary complementisers

   a. Ai, **que**   t'atrapo!   (Catalan)
      hey that.EXCL CL.DO=catch.1SG

      'I'm coming to get you!'   (Corr 2016: 88)

   b. No hagas   esto, **que**   luego mama se   enfada   (Spanish)
      not do.SUBJ.2SG this that.CONJ then mum   CL.REFL=   get.angry.3SG

      'Don't do this, because then mum gets angry'

Illocutionary complementisers therefore represent a new and potentially productive testbed for the claim that CP-material and, particularly, complementisers emerge late in the learning path. The present paper shows that this use of *illocutionary* complementisers emerges substantially earlier than subordinating complementisers, instantiating novel evidence for early access to a CP domain in child grammars.

The relative emergence of this kind of complementiser compared to prototypical embedding complementisers could have rather far-reaching consequences for approaches to the acquisition of functional categories. Early uses of illocutionary complementisers will raise several questions: for example, how can any timing differences between these two complementiser types be explained? And, looking at the bigger picture, how does early emergence of illocutionary complementisers square with contemporary approaches to the development of functional categories, particularly those that assume a 'delayed' maturation of the CP domain?

The development of so-called illocutionary complementisers in children acquiring Ibero-Romance languages is yet to be systematically studied and contrasted to the acquisition of subordinating complementisers. In this paper, I present a preliminary corpus study on child Catalan and Spanish complementisers, showing that at least some illocutionary complementisers emerge on average earlier than subordinating complementisers and often from the earliest files. I argue, therefore, that illocutionary complementisers represent a potentially strong case study in favour of early availability of discourse and speaker/hearer-oriented material. Such a finding is at odds with many maturational approaches that assume functional categories mature bottom-up, with CP material developing at the very end of the learning path. On the other hand, I suggest that these discrepancies between complementiser types are to be expected due to (at least) two factors: firstly, given the main clause *vs* embedded clause nature of the use of these two kinds of complementisers, illocutionary complementisers require less syntactically

elaborate grammars (no knowledge of embedding) and thus should, in this respect, be accessible earlier. I contend, as well, that these patterns are coherent with approaches to acquisition and language variation that take discourse and interactional content to be acquisitionally privileged and salient and that take functional spines to initially develop 'inwardly', such as in Heim & Wiltschko's (2021) Inward Growing Spine Hypothesis (cf. also Wiltschko 2021), Biberauer (2018)'s Peripheral Speaker-Hearer Hypothesis and also Biberauer & Roberts (2015)'s emergent categorial hierarchy (see also Bosch 2023a for empirical support and further theoretical expansion).

This paper is organised as follows: section 2 introduces illocutionary complementisers in Ibero-Romance and some of their syntactic properties, focusing on Catalan and Spanish data. In section 3, I present the theoretical background and hypothesis and also provide an overview of approaches arguing for early acquisition of CP-based and speaker-hearer material. Section 4 outlines the corpus study, its methodology and, subsequently, its results and section 5 offers a discussion of the data's preliminary theoretical implications. Subsequently, section 6 investigates a possible domain of future extension of the present work, gauging the development of Italo-Romance complementisers. Section 7 concludes.

2   Illocutionary complementisers in Ibero-Romance: typology and syntactic properties

Complementisers are typically assumed to involve subordinating conjunctions that transform clauses into complements of a matrix clause (Rosenbaum 1967, Lakoff 1968, Bresnan 1972). In Spanish and Catalan, finite complement clauses are often introduced by the complementiser *que*, as exemplified below:

(3)
   a. M'adono   **que** encara no ha   arribat   (Catalan)
       CL=realise.1SG that yet   not AUX.3SG arrive.PTCP

      'I realise that he/she hasn't arrived yet'

   b. Me   dijiste   **que** se   había   arrepentido   (Spanish)
       CL.IO= tell.PST.2SG that CL.REFL= AUX.IMPF.3SG regret.PTCP

      'You told me that he/she had regretted (it)'

A conspicuous property of the Ibero-Romance language family, particularly varieties in the Iberian Peninsula, however, is their pervasive use of what seems to be the complementiser *que* introducing matrix clauses; namely, the complementiser in Ibero-Romance does not always function as a prototypical subordinator that introduces an embedded clause. These uses of *que* are attested across Ibero-Romance (though with varying degrees of frequency and robustness) and they come with a range of illocutionary functions. These illocutionary complementisers arise from main-clause constructions where the core function of *que* as a subordinator (e.g., as in 1 and 3) has been bleached in favour of utterance- and speaker-hearer-oriented functions, which are often emphatic in nature. Illocutionary *que* provides evidence of a grammatical device – a semantically-bleached subordinating complementiser – whose morpholexical material has been recycled for the purpose of encoding illocutionary information in the main clause (Corr 2016: 1-3).

Corr (2016) distinguishes at least three separate kinds of illocutionary complementisers based on both their distributional and interpretive properties. We will adopt Corr's classification and, thus, distinguish between *exclamative*, *quotative* and *conjunctive que*. These conversation- and speech-act-oriented uses of *que* will be collectively referred to as *illocutionary* complementisers. Given the child data to be considered, attention will focus on Catalan and Spanish.

Exclamative *que* introduces an indicative clause (typically, though not always, declarative) and has the illocutionary force of an exclamation:

(4)
    a. Ala, **que** ho has llençat tot al terra! (Catalan)
       hey that.EXCL CL.DO= AUX.2SG throw.PTCP everything on.the floor

       'Hey! You've thrown everything on the floor!'

    b. ¡**Que** la abuela Rosa me ha comprado una (Spanish)
       that.EXCL the grandmother Rosa CL.IO= AUX.3SG buy.PTCP a
       Nintendo nueva!
       Nintendo new

       'Grandma Rosa has bought me a new Nintendo!'

Quotative *que* constructions involve reported speech clauses introduced by *que*. These, importantly, do not require a retrievable *verbum dicendi* for the utterance to be felicitous:

(5) Context: the speaker is asked who had just phoned.
Era   la Carme. **Que**   em   trucava   per   (Catalan)
be.PST.3SG the Carme that.QUOT CL.DO=   phone.IMPF.3SG   to
felicitar-me
congratulate=CL.DO
'It was Carme. She phoned me to wish me a happy birthday'

(6) A: No entiendo   qué acabas   de decir   (Spanish)
       not understand.1SG what finish.2SG to say.INF

   'I don't understand what you've just said'

   B: ¿Qué?
      what
      'What?'

   A: **Que**   no entiendo   qué dices
      that.QUOT not understand.1SG what say.2SG

   '(I've said that) I don't understand what you're saying'

Quotative *que* can also introduce a report of a recent utterance made by the addressee, often requesting for confirmation:

(7) A: M'acompanyes   a Vic?   (Catalan)
       CL.DO=accompany.2SG   to Vic

   'Are you coming to Vic with me?'

   B: **Que**   si t'acompanyo   a Vic?
      that.QUOT if CL.DO=accompany.1SG   to Vic

   'You're asking if I'm coming to Vic with you?'

   A: Sí!
      yes

   'Yes!'

B: D'acord, ja   vinc!
  Okay   already come.1SG

  'Okay, I'm coming!'

A final third kind of illocutionary *que* involves conjunctive uses of *que*, where its function involves contextualising utterance information for the addressee. It is typically associated with causality, usually linking two root clauses in a speech-act causal relation. Often, though not necessarily, the first root clause is an imperative clause as in (8)[3].

(8)
    a. No li   diguis   això a la Paula **que**   és   un secret (Catalan)
       not CL.IO= tell.SUBJ.2SG this to the Paula that.CONJ be.3SG a   secret

       'Don't tell this to Paula because it's a secret'

    b. Dame   el diario,   por favor, **que**   siempre me   (Spanish)
       give=CL.IO the newspaper for favour that.CONJ always   CL.IO=
       gusta repasarlo   antes de ir   a   dormir
       like go.over=CL.DO before of go.INF to sleep.INF

       'Give me the newspaper, please. I always like going over it before going to bed'

The present study will also consider instances of polar interrogatives optionally headed by the complementiser *que* in Catalan, which are available in Standard Eastern Ibero-Romance. They are, thus, unavailable in Spanish[4]:

(9) **Que**   vindràs   al   final a veure   la pel·lícula?   (Catalan)
    that.INT come.FUT.2SG   in.the end to watch.INF the film

    'Are you coming to watch the film in the end?'

---

[3]   The distribution of conjunctive complementisers is not limited to Ibero-Romance; they are also observed across Romance (see Corr 2016: 226-227 and references therein for data from Romanian, Southern Italian Dialects and Rhaeto-Romance and Prins 2014 for Southern Italian Dialects specifically).

[4]   Balearic Catalan also allows the conjunction *o* 'or' to head this structure (e.g., *O vindran a Ciutadella?*, 'are they coming to Ciutadella?'; Prieto & Rigau 2007: 1). These instances are not relevant here.

The interpretative value of this interrogative *que* varies dialectally. In Balearic, Central and Northwestern Catalan, interrogative *que* introduces a neutral polar question, whereas it can only introduce counter-expectation questions in North Central Catalan, Rossellonese and Valencian (Prieto & Rigau 2007). All Catalan data to be discussed here, however, is taken from five children growing up in Catalonia, primarily in Barcelona; they therefore speak a Central Catalan variety.

Main-clause instances of illocutionary *que* such as those illustrated in (2), whilst discourse-oriented, are not easily catalogued under the umbrella of information structure that traditionally characterises the CP domain. Rather, illocutionary uses are primarily attested in spontaneous speech, charting the conversational dynamics between interlocutors, and operative in here-and-now contexts (Corr 2016). Although linguistic elements involved in the encoding of these kinds of conversational pragmatics were until relatively recently thought to lack formal syntactic properties, I will adopt neo-performative proposals (building on Ross 1970) that take such utterance- and conversation-related information to be formally represented in the syntax. These works embed these illocutionary complementisers as part of the C-domain or of a higher Speech-Act-oriented domain (following, i.a., Speas & Tenny 2003, Giorgi 2010, Haegeman & Hill 2013, Wiltschko 2014, 2021, Corr 2016, 2022, Miyagawa 2022).

Therefore, we will broadly (and simplifying grossly) assume that some structural organisation like the one in Figure 1 is largely on the right track for natural language, though I will remain agnostic as to exactly what functional projections (cartographic or otherwise) populate each of these domains. Although Figure 1 contains a speaker-hearer domain at the outer *v*P edge (cf. also Belletti 2004 on the 'low IP area'), this paper will only discuss the higher SAP domain directly above the traditional CP domain.
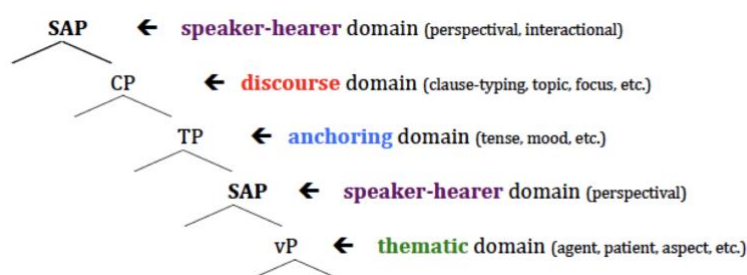

Figure 1. Clausal structure with speech-act layers (Biberauer 2018: 4).

Specifically, Corr (2016) proposes that the three types of 'illocutionary' *que* outlined above each have distinct distributional and interpretational properties (e.g., with respect to interpolation, VP deletion, variable binding, among others). On the basis

of these discrepancies, they are argued to correspond to separate functional projections at the height of the cartographic left-periphery and the so-called *Utterance Phrase* (or UP), a dedicated utterance domain above the CP (analogous to the higher SAP in Figure 1). Illocutionary complementisers are hypothesised to be 'scattered' across both the speech-act (UP) and discourse domain (CP). It is proposed that exclamative and conjunctive complementisers pattern differently to quotative complementisers, with the former being located in a higher, speech-act-oriented layer (in Corr's SALowP for exclamative *que* and SAHighP for conjunctive *que*), while quotative complementisers display the formal behaviour of a prototypical C-head (interpreted in Corr 2016 as a head at the edge of CP that heads a dedicated Evidential projection). Traditional embedding complementisers, in contrast, are assumed to be located in Rizzi's (1997) ForceP, or also in Corr's (2016) DeclarativeP (as the head encoding declarative clause-typing; see also Ledgeway 2012: 179)[5]. Interrogative complementisers have also been suggested to be C-based (e.g., Prieto & Rigau 2007 place interrogative *que* in Rizzi's 1997 Fin head; see also Rizzi 2004 on IntP, Interrogative Phrase). The precise location of these complementisers within the left periphery will not be at the forefront of our discussion; therefore, we will not survey the diagnostics adopted in this approach for each type of illocutionary complementiser (for which see Corr, 2016). What *is* worth emphasising is their placement as part of a high discourse (CP) and interactional domain (SAP or, in Corr's terms, UP). This syntactic property will become important in light of hypotheses that suggest comparatively late development of the highest layers of the clausal spine.

It is worth noting that this treatment of illocutionary complementisers was revised in Corr (2022). The revised analysis keeps the distinction between the UP and CP domains, now conceived as a phase-external utterance-oriented domain ('the edge of the edge') and a phasal propositional domain, respectively. However, it deviates from Corr (2016) by adopting a coarser-grained analysis of types of illocutionary *que* and their syntactic positioning; all illocutionary complementisers are merged at the edge of the CP domain (Rizzi's Force head), as opposed to being scattered across several heads in the UP/CP domains. The terminology for illocutionary complementisers was also adjusted (with exclamative, quotative and conjunctive complementisers receiving the terms 'affective', 'presentative' and 'dialogic' complementisers, respectively). To retain an approach which makes finer-grained distinctions between the structural positions of types of illocutionary complementisers vis-à-vis embedding complementisers within the CP/SAP domains, I will follow Corr's (2016) analysis here. As noted above, however, this

---

[5]   Other instances of non-illocutionary *que* in Ibero-Romance, such as recomplementation *que* and jussive *que* are, in some approaches, hosted by Topic and Fin, respectively (Villa-García 2012, *et seq.*).

decision will not substantively impinge on the conclusions of the study. For consistency, as well, I will adopt Corr's (2016) terminology throughout[6].

There are other structures in Catalan and Spanish that are somewhat reminiscent of the ones just introduced. Cruschina & Remberger (2018) discuss cases of complementisers being preceded by a functional element which morphologically coincides with an adjective or adverb (see 15). They suggest that the functional elements are directly merged in a speech-act-related phrase above ForceP (labelled SentienceP, and adopted from Speas & Tenny 2003), while the main-clause complementisers are located in Force. Similar constructions in Spanish and Catalan such as those introduced by *sí que* ('yes that') are also described in Batllori & Hernanz (2013) and Villa-García & González Rodríguez (2020), where they are analysed as being in projections in the left periphery. (11) illustrates some examples. These structures are formed with the emphatic polarity particle *sí* 'yes' alongside a main-clause/illocutionary *que* and they typically signal the speaker's commitment to the truth of the proposition introduced. They sometimes (though not always) have an exclamative function and can be preceded by topics, as (11b) (see also Poletto & Zanuttini 2013 on Italian *sì che*). Catalan also has available a *no que* 'no/not that' construction, where two positions for the negative marker *no* are attested (see 12). The preposed negation serves to contradict a previous affirmative assertion (Batllori & Hernanz 2013: 27) and also to confirm the polarity of the whole assertion.

(10)

    a. Evidentment **que** estudiaré   per a aquest examen!   (Catalan)
       Obviously   that study.FUT.1SG   for to this   exam

       'I'll obviously study for this exam!'

    b. ¡Claro **que** entendió!   (Spanish)
       clear   that understand.PST.3SG

       'Of course he understood!'

---

[6]   This overall discussion also raises interesting questions about the possible formal differences between 'types' of illocutionary complementisers, as noted by an anonymous reviewer; for example, whether we assume that these different types lexicalise the same or distinct syntactic heads or whether each type may have different internal compositions (e.g., in the nanosyntactic sense; cf. Villa-García 2019: 16 and Baunaz 2015: 2). Taking a stance on this matter is not imperative for the predictions and discussion of this study and it remains beyond its scope, so I will remain agnostic on the precise formal status of the subtypes of illocutionary/embedding complementisers. Nonetheless, as noted above, the exposition will proceed guided by the theoretical approaches outlined in this section, several of which do adopt a view where illocutionary *que* can lexicalise different syntactic heads.

(11)

    a. Sí **que** val   la pena, tenies   raó   (Catalan)
       yes that cost.3SG the struggle have.IMPF.2SG right

       'It certainly is worth it, you were right'

    b. Laura sí **que** ha   vuelto   de   trabajar   (Spanish)
       Laura yes that have.AUX.3SG return.PTCP   from work.INF

       'Laura has certainly come back from work'

(12)   No **que** no ha   cantat   la soprano   (Catalan)
       not that not have.3SG sing.PTCP the soprano

       'No, the soprano has NOT sung'

                              (Batllori & Hernanz, 2013: 27)

These constructions, attested across Romance, are also relevant for our purposes and will be analysed in the corpus study introduced in section 4.

   Illocutionary complementisers will be contrasted in the study to traditional *embedding* or *subordinating* complementisers in Catalan and Spanish. As described earlier, these introduce, for instance, relative clauses (13a) and complement clauses (13b):

(13)

    a. La Maria, **que** estava   al   davant de l'atri,   va   (Catalan)
       the Mary   who be.PST.3SG in.the   front   of the-atrium AUX.PST.3SG
       anunciar   els guardonats al   premi literari
       announce.INF the winners to.the prize literary

       'Mary, who was standing in front of the atrium, announced the winners of the literary prize'

    b. Ayer   nos   enteramos   de **que** van   a ser   padres (Spanish)
       Yesterday CL.REFL= find.out.PST.1PL of that go.3SG to be.INF parents

       'Yesterday we found out that they are expecting a child'

The key assumption which we will adopt from the works just introduced, then, is that illocutionary complementisers are hosted in the discourse and interactional domains (e.g., the SAP domain and CP domain in Figure 1) – a theoretical analysis

we will subsequently harness to test the predictive differences among current approaches to the acquisition of functional categories (see section 3). However, as briefly noted above, the exact hypothesised location of these complementisers in the left periphery and its empirical adequacy is going to be largely orthogonal to the aims of this paper.

Having introduced these two broad kinds of complementisers and their properties, I now turn to discussing the theoretical background of the corpus study and its hypothesis in the next section.

3   Theoretical background and hypothesis

An analysis of the acquisition of illocutionary complementisers is significant for a range of current proposals regarding the acquisition of functional categories.

A prevalent hypothesis, most recently defended in Friedmann et al. (2021) and Friedmann & Reznick (2021) under the name of the Growing Trees Hypothesis, suggests that functional categories are acquired following an innately-determined ordering of development, which specifies that the cartographic hierarchy in Universal Grammar develops bottom-up (Figure 2). The proposal finds its roots in earlier hypotheses such as Radford's (1988) Small Clause Hypothesis and Rizzi's (1993/1994) Truncation Hypothesis.

These approaches would predict delayed development of the left periphery compared to lower clausal domains such as VP and TP/IP. The predicted relative development of embedding complementisers *vs* illocutionary ones is unclear as Friedmann et al. (2021) do not discuss interactional language *per se*, but, in any event, the predictions would either expect both complementisers to become available at similar times (that is, after TP and VP knowledge have emerged) or, possibly, that some illocutionary complementisers would emerge later than embedding complementisers, by virtue of being located in a higher left-peripheral domain (SAP/UP for exclamative and conjunctive *que*). As illocutionary complementisers are located at the very top of the left periphery, neither prediction would therefore expect some illocutionary complementisers to start emerging well before embedding complementisers. In other words, the development of illocutionary and embedding complementisers is predicted to be closely linked, by virtue of bottom-up maturation and their similar hierarchical placement (for comparable approaches, see Vainikka & Young-Scholten 2011, Diercks & Bossi 2021, Diercks et al. 2023).
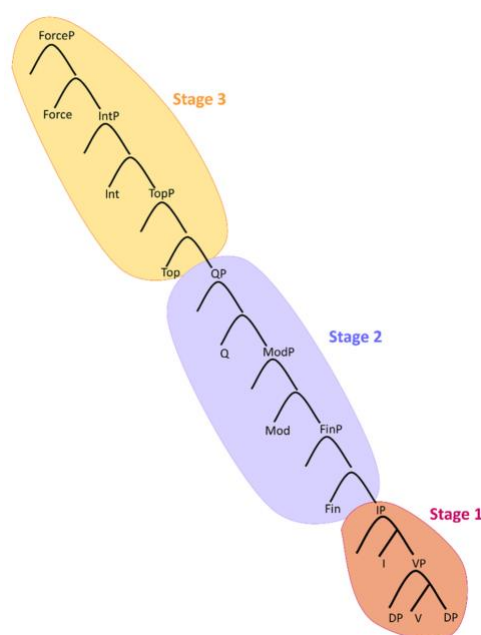
Figure 2. Stages of acquisition of the clausal domain in the Growing Trees Hypothesis (Friedmann et al. 2021: 12).

However, whilst bottom-up maturation has arguably been one of the dominant perspectives in the acquisition literature, it is not obvious from other work in language acquisition and variation that expressive, illocutionary and discourse content should necessarily be developmentally 'delayed' in comparison to material located at other layers of the clausal domain, such as vP/VP or TP. In other words, it is not self-evident that making developmental predictions primarily on the basis of structural height (in the adult grammar or the grammar specified by Universal Grammar, UG) should be the preferred mode of reasoning. There may be independent reasons for certain structural elements (whether located high in the clausal domain or not) to emerge early, particularly in emergentist approaches that do not assume a UG-given functional spine (i.a., Biberauer & Roberts 2015, Ramchand & Svenonius 2014, Biberauer 2019a, Ramchand 2023). For one, Roeper (2007, *et seq*.) already noted the importance of a 'parallel' language possibly being in play in early word combinations alongside simple referential forms – that is, a language that reflects primarily attitudes and feelings, besides propositional and referential language. A host of other terms, denoting seemingly 'vague' but nonetheless highly expressive meaning (such as *uh-oh*, *well*, *huh*, *oops*) abound in early utterances. As Roeper notes (p. 40), these expressive words tend to occur before and after sentences, sometimes even in co-occurring constructions (e.g., *Well, gee, yes, maybe I can*, in response to *Can you sled down that hill?*). Albeit remarkably hard to characterise semantically, they are among the first words a child uses and, he suggests, "may give the biggest clue to the quality of children's thoughts" (Roeper 2007: 39). Joint work with Christopher Potts also underscored the

possibility that early child combinations might primarily denote purely expressive meaning, besides any conceivable propositional content (Potts & Roeper 2006).

Proceeding with a convergent line of reasoning, there is a small, but nonetheless growing, range of theoretical approaches to acquisition that argue for an acquisitionally privileged role of some speaker-hearer and discourse content, not just lexically, but also syntactically – the approaches introduced below, then, do anticipate that at least some illocutionary complementisers might emerge early on. This, notably, will underscore a predictive difference of these approaches compared to bottom-up maturation: such a prediction is *ipso facto* ruled out in a bottom-up approach. Since these discourse/perspectival elements are located at the CP or SAP domain, they should emerge late and closely approximate the emergence of subordinating complementisers. It is worth emphasising head-on, however, that neither Roeper (2007) nor any of the approaches outlined in what follows necessarily predict that *all* relevant discourse and interactional material in a language will emerge early and/or simultaneously. Importantly, in this context, they expect early production and acquisition of *some* illocutionary complementisers, but critically without any presuppositions that the complete inventory of illocutionary complementisers (and other discourse elements) should emerge in one fell swoop.

Roeper & Rohrbacher (1994) and Galasso (2003), for instance, argue that the production data in English shows evidence of early knowledge of CP, in particular wh-questions; viz. Galasso's 'Empty Middle' approach, whereby the earliest stages project CP>VP before developing into CP>IP>VP[7]. Tsimpli (2005) similarly provides data from child Greek that supports early acquisition of the left-periphery and draws on the difference between LF-interpretable and LF-uninterpretable features: LF-interpretability is taken to be responsible for a distinction between (i) discourse-related features (e.g., focusing and topicalisation/dislocation), and (ii) between peripheral LF-interpretable features (focusing or wh-questioning) on the one hand, and non-peripheral, uninterpretable features such as inflectional features, on the other (Tsimpli 2005: 185-186). In the two children studied, phenomena such as focalisation and wh-questioning (involving LF-interpretable features) emerge early on. This is followed by Clitic-Left Dislocation and clitic doubling (involving discourse-related and uninterpretable features). The inflectional domain (which requires non-peripheral uninterpretable features) only clearly emerges subsequently. These results are taken to suggest that 'peripheral' positions are grammatically represented from the earliest stages.

---

[7]   See also Perkins & Lidz (2021) and Perkins, Ying, Williams & Lidz (2021) for behavioural evidence that 18- and 20-month-olds can represent abstract non-local dependencies like wh-questions and, according to preliminary results, that they can subsequently harness them to constrain hypotheses during verb learning. Thanks to Laurel Perkins for providing me with a copy of the slides of the latter reference.

Other works have also underscored the early emergence of some CP-based knowledge. Van Kampen (2010), notably, has argued for a 'typological guidance' approach to acquisition, whereby the major typological properties of a language are invariably the first to be acquired and are the "system's bootstrap for learnability" (van Kampen 2010: 264). Based on Dutch data, she makes the case for early acquisition of the V2 rule, where early topic-comment structures and operator-comment structures with [+finite] verbs with surface V2/V1 order (e.g., moet*operator* papa*topic* doen*comment* 'daddy has to do (it)'; Sarah week 120) setting the stage for the full acquisition of V-to-C shortly after. Additionally, the acquisition data suggests that finiteness, not tense, is likely the trigger for verb movement to the C-domain, as tense and agreement markings are established after verb movement is already apparent (see also van Kampen 2009ab, on early topic-comment structures and the acquisition of V2 and wh-questions). In this approach, the early acquisition of some CP-related knowledge is a consequence of major typological characteristics of the child's L1 guiding the learner, and, in Dutch, some of these are encoded precisely in the C-domain (as is the case in other V2 and V1 languages). As these approaches indicate, then, the possibility that some discourse knowledge emerges early representationally, potentially even before TP-based material, is not unlikely. More generally, however, they argue, in different forms, against a strictly bottom-up approach to development. To the extent that illocutionary complementisers are located at either the CP or SAP domain, these results could anticipate a potential early emergence of illocutionary complementisers.

In a more recent proposal, Heim & Wiltschko (2021) propose an inward maturational pathway of the universal and interactional spines in Wiltschko (2014, 2021), hypothesising they develop inwardly, starting from the edges and then developing towards the centre of the spine – this proposal is dubbed the Inward Growing Spine Hypothesis in Heim & Wiltschko (2021) (cf. also Wiltschko 2021: 220). This hypothesis rests on two interrelated models of clausal organisation, namely the Universal Spine Hypothesis and the Interactional Spine Hypothesis, outlined in Wiltschko (2014, 2021).

Being argued to be phylogenetically prior (see Hinzen & Wiltschko, 2023), the approach takes the thematic and categorisation-based domain (Wiltschko's 2014 Classification head, or *v*P) and the interactional and speech-act-based domain (Wiltschko's 2021 Responding head) to develop first. Subsequently, Linking (the CP domain, hosting, e.g., wh-questions or embedding complementisers) matures. The development of the Anchoring domain (or TP, in a language grammaticalising tense) occurs at a subsequent step; Wiltschko's (2021) fully fleshed-out interactional spine also matures in its entirety at later stages (see Heim & Wiltschko, 2021, for greater detail). Data on the acquisition of English *huh* is presented as preliminary evidence

for part of Heim & Wiltschko's (2021) Inward Growing Spine Hypothesis. As a result, their Inward Growing Spine would accordingly predict child production data to present evidence of at least some illocutionary complementisers early on, given the proposed inward maturational timeline, where some representation of the speaker-hearer domain is present from the start.

In a proposal predating Heim & Wiltschko (2021), Biberauer (2018) also converged on the hypothesis that some interactional and speaker-hearer-oriented material may be early-acquired, albeit for somewhat different reasons. Specifically, Biberauer (2018, *et seq.*) argues that structural edges (in particular, *phasal* edges) will be of particular significance in acquisition, as these constitute the locus of here-and-now and speaker-hearer-oriented material. This is captured in her Peripheral Speaker-Hearer Hypothesis, or PSHH, defined below (Biberauer 2018: 4):

(14)   **The Peripheral Speaker-Hearer Hypothesis (PSHH**)
        Speaker-hearer perspective is formally encoded at the edges of phasal domains, where phasal domains are independently signalled, realizationally (PF) and interpretively (LF) privileged structural domains, the precise identity of which differs from language to language, and the 'size' of which may also differ from derivation to derivation language-internally (i.e. the 'dynamic'[8] perspective on phases).

If perspectival, speaker-hearer elements – such as modal particles, emphatic polarity particles, illocutionary complementisers, among others – are necessarily phase peripheral in the PSHH approach, these edge-based elements would serve a key bootstrapping heuristic in acquisition (namely, signalling domain edges) and so, at least some of these perspectival elements, may be expected to be acquired early (see the discussion in Biberauer 2018 for some synchronic and diachronic consequences, also in the context of spoken, vernacular varieties).

Therefore, the PSHH makes important predictions in our context, insofar as illocutionary complementisers are located at phase edges, encoding, i.a., perspectival and speaker-hearer information (see section 2). In particular, PSHH predicts an acquisitionally significant role for illocutionary information by virtue of the role of structural *edges* in the system and their associated perspectival nature, as outlined above. Crucially, and as already discussed in section 2, illocutionary complementisers cannot be readily analysed in the same way as subordinating C-elements. Following Corr (2016), several illocutionary complementisers are located

---

[8]   See so-called dynamic or contextual approaches to phases (i.a., den Dikken 2007, Gallego 2006, 2010, Harwood 2013, Wurmbrand 2013, Bošković 2014).

higher up in the clause than embedding complementisers, in the SAP domain, and they are also peripheral. As a consequence, PSHH likewise foresees a possible developmental mismatch between the emergence of illocutionary *vs* embedding complementisers – the peripheral and speaker-hearer facet of illocutionary complementisers is something acquirers are expected to be particularly drawn to.

Unlike the Inward Growing Spine Hypothesis, however, the prediction that illocutionary complementisers should be early-acquired importantly does not find its aetiology in some biological mechanism that trails through an innate clausal hierarchy and dictates acquisition orderings, as would be the case in maturational approaches. Instead, it is phasal edges and their distinctive role in the system that facilitate crucial syntactic domain-size learning in a language acquisition context and comprise productive 'ways in' for elements that have not yet been fully formally integrated into the structure (see Biberauer 2018 for further details on this latter point, and also Biberauer 2019a; for further detail on the Inward Growing Spine Hypothesis and the PSHH see the more expanded exposition in Bosch 2023b).

Finally, a similar theoretical proposal is made in Biberauer & Roberts (2015). This approach represents an attempt to unify independently suggested formal hierarchies and encodes these as a single formal hierarchy that is not UG-given, but instead an emergent property of the interaction of the three factors of language design (namely, UG, the Primary Linguistic Data and third-factor, general-cognitive principles, see Chomsky 2005; cf. also Biberauer 2019a). Crosslinguistically, according to Biberauer & Roberts (2015), clauses can be analysed at different levels of 'magnification' or 'granularity' (adapted from Biberauer & Roberts 2015: 6):

(15)   **Extended Projection** (V) (Grimshaw 1991) > **phase** (C, v) (Chomsky 2001, 2008) > **Core Functional Category** or **CFC** (C, T, v) (Chomsky, 2000, 2001) > '**cartographic field**' (e.g., Tense, Mood, Aspect, Topic, Focus) > **semantically distinct head** (as in Cinque 1999; Speas & Tenny 2003; Frascarelli & Hinterhölzl 2007, Haegeman & Hill 2013, among others).

(15) can be reformulated as a hierarchy, which identifies overall similarities between independently proposed hierarchies and functional sequences in the literature (see the references in 16)[9]. The hierarchy thus begins at an initially 'undivided' categorial space and gradually increments the level of grain with which the system is analysed:

---

[9]   There may still be a level of nanosyntactic organisation, as Biberauer & Roberts (2015: 4, fn. 1) note, but we leave this aside as it is orthogonal to the present discussion.

(16)
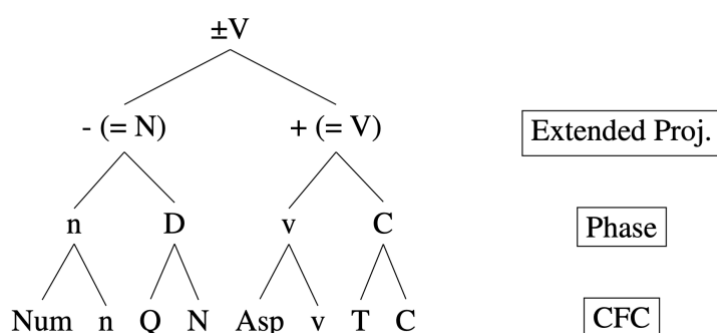
Importantly for present purposes, the hierarchy in (14) represents a learning path, in which acquirers successively divide the categorial space into ever more fine-grained units (cf. also the Successive Division Algorithm in Dresher 2009). After initially establishing a basic distinction between predicates/arguments (possibly aided by prosodic cues; see the discussion in Biberauer 2019a) and subsequently one between nominal/verbal Extended Projections (the Extended Projection layer in 14), the earliest developmental stages display evidence of a further subdivision between an underspecified discourse (CP-internal) domain (hosting, e.g., early wh-words) and a thematic (vP-internal) domain (the Phasal stage in 14). These early CP-structures can also include some early interactional language (of the sort in, i.a., Speas and Tenny 2003, Haegeman and Hill 2013, Wiltschko 2021, Heim and Wiltschko 2021), encoded at this stage as part of this coarse-grained CP and not distinguished as a separate SAP domain until a later developmental stage (see 13). The TP domain would be expected to emerge after basic discourse/interactional and thematic material is established (the Core Functional Category or CFC stage). In Bosch (2023a) I provide corpus data from five languages that supports such a developmental progression, at least partially. Like the Inward Growing Spine Hypothesis and the PSHH, therefore, this neo-emergentist[10] approach to syntactic categories argues too for early availability of discourse and illocutionary information in developing grammars.

Therefore, several separate and recent proposals in the literature of acquisition and variation converge on the proposal that some discourse, interactional and 'here-and-now' linguistic material is expected to be acquisitionally most accessible.

---

[10]   The prefix *neo-* is used to set the approach apart from traditional emergentist approaches, such as those in the usage-based or Construction Grammar tradition. The approaches are fundamentally distinct. Neo-emergentism (e.g., as originally outlined in the Maximise Minimal Means model in Biberauer 2019a), still retains the essence of a generativist model, with formal variation being consistently encoded via formal features and still regulated by Universal Grammar (albeit poor in incarnation).

I will adopt here the neo-emergentist perspective in Biberauer & Roberts (2015), Biberauer (2018, 2019ab) and the associated findings in Bosch (2023a), according to which development proceeds in the neo-emergentist and edge-centred manner just outlined. According to Biberauer & Roberts (2015), the child initially makes a basic predicate/argument (or 'archi' N/V) distinction (see Douglas 2018, Song 2019 and Biberauer 2019a on 'archi-V' and 'archi-N' features and categories), which then aids them with making a first representational division into the verbal and nominal Extended Projections. Subsequently, the learning path proceeds by making a coarse-grained subdivision between discourse/interactional (CP- and SAP-internal) material and thematic (*v*P-internal) material[11]. In contrast to bottom-up perspectives, this approach *does* expect at least some illocutionary complementisers to be harnessed early on because of their interactional and discourse nature. These are also predicted to surface before embedding complementisers are first detectable, owing to the fact that illocutionary complementisers are attested in main-clause contexts and thus do not require knowledge of subordination. The predictions made by this approach, as far as illocutionary complementisers are concerned, are identical to those made by Heim & Wiltschko's (2021) Inward Growing proposal; namely, that at least some illocutionary complementisers will emerge early and before embedding complementisers. The present work will not discuss their predictive differences beyond the domain of complementiser types (though this latter point receives some attention in Bosch 2023a).

To reiterate, however, none of these approaches are committed to the absolute claim that *all* illocutionary complementisers (in their complete inventory of subtypes) will necessarily emerge before any embedding complementisers; the vital prediction that matters in the present context is that they do expect early production of *some* illocutionary complementisers, a scenario which would be disallowed in bottom-up maturational approaches. This work will not establish the relative development of each subtype of illocutionary complementiser (conjunctive, quotative, exclamative, interrogative, etc.), compared to embedding ones, and so will leave open the possibility that some kinds of illocutionary complementisers may develop later than others. This is left for future work.

---

[11]   Therefore, I will not aim to tease apart here whether CP emerges before/after the Speech-Act domain. This would entail a larger project and, necessarily, this paper is more modest in scope. For now, I take these two layers to be potentially encoded as a single underspecified or coarse-grained domain at the earliest stage, in line with Biberauer & Roberts's (2015) hierarchy. This working and speculative assumption inevitably requires further work, so it is perfectly possible that, as more developmental data becomes available, it will have to be dropped or reshaped.

With this in mind, we have paved the way for the presentation of the corpus study in the next section, which will discuss its methodology and results.

4   Corpus study

4.1 Summary of the structures analysed

To summarise the discussion in section 2, the following constructions were analysed and quantified in every corpus. As discussed earlier, these structures are hypothesised to be located in a variety of positions spanning both the utterance-related layer and the traditional CP (proposition-oriented) domain. Both (1) and (2), however, are treated as groups. In other words, we did not quantify and collect the precise number of utterances in which every type of illocutionary complementiser was used. All subtypes were instead quantified as a single group. The same holds for subordinating complementisers, as far as their subtypes (2a) and (2b) are concerned. This is expanded upon in the following section (4.2). The subtypes (1a-f) and (2a-b) are nonetheless listed below for clarity and precision, to note the specific structures that fall under the umbrella of 'illocutionary' and 'subordinating' complementisers in the current study.

1. **Illocutionary complementisers**
   a. Exclamative
   b. Conjunctive
   c. Quotative
   d. Interrogative (in Catalan only)
   e. Adverb/adjective + *que*
   f. (Topic) *sí/no que* ('yes/no that')
2. **Subordinating complementisers**
   a. Complement clauses
   b. Relative clauses introduced by *que* (including free relatives[12])

4.2 Methodology

Using the CLAN programme and its *kwal* command (MacWhinney 2000), we automatically extracted all occurrences of *que* and their conversational contexts for

---

[12]   Free relatives are included in the analysis on the grounds that these often feature the complementiser *que* in Spanish and Catalan. For example, Sp. *Prefiero el **que** comí ayer* 'I prefer the one I ate yesterday'. Just like with relative clauses more generally, any free relatives that do not include *que* are excluded from the analysis (e.g., Cat. *M'agrada on vius* 'I like where you live').

10 Catalan and Spanish children in the CHILDES corpora. After manually excluding utterances where *que* did not correspond to a complementiser, this yielded N = 1318 utterances from children aged 0;11 to 4;08 that contained a complementiser. This is summarised in Table 1. The utterances were annotated for type of complementiser, either illocutionary or subordinating. The number of instances in which each kind of complementiser was produced per file was also counted. As the data is very often ambiguous and it is hard to discern the kind of illocutionary complementiser produced (e.g., discriminating between quotative and exclamative complementisers), I did not quantify each subtype of illocutionary complementiser and I treated them as a heterogeneous group, aiming simply to compare the (un)availability of illocutionary *vs* embedded-clause complementisers (more broadly, diagnosing different types of speech acts in child language is known to be a non-trivial task, see Cameron-Faulkner 2014 and references therein). Detailed coding of these subtypes would be a welcome avenue of future work, however.

Table 1. CHILDES corpora studied (Serra & Solé 1989, Bel 2001, Llinàs-Grau & Ojea 2000, Aguado Orea & Pine 2015, Aguirre 2000, Vila 1990) and utterances with complementisers.

Language	Corpus	Children	N
Catalan	Serra/Solé	Laura	202
		Gisela	201
		Àlvar	15
		Guillem	105
	Júlia	Júlia	4
Spanish	Llinàs/Ojea	Irene	90
		Yasmin	42
	Aguado-Orea/Pine	Juan	244
	Aguirre	Magín	295
	Vila	Emilio	120
Total		10	1318

Cases of complementisers followed by an adjective only (such as in Cat. *Que bonic!* 'how beautiful!') were disregarded, on the grounds that these do not feature a finite verb (and so may not be reliable diagnostics for CP and Speech-Act-related projections). Importantly, they can also be conflated with other structurally and interpretively similar exclamative constructions, such as some kinds of wh-exclamatives. Therefore, only instances of sentences with a complementiser and a verb were counted. Any ambiguous, unclear or unintelligible utterances were excluded. Consecutive repetitions of the same construction were also excluded, as were identical imitations of an utterance by an adult.

Additionally, I adopt Mean Length of Utterance (MLU) as the guiding measure for syntactic development, to be used to compare the stages across all children. As several authors have noted (Clahsen, Penke & Parodi 1993, Paradis & Genesee 1997, Caprin & Guasti 2009, Friedmann et al. 2021), age is to be avoided in both intralinguistic and crosslinguistic comparisons, to evade the great variability in linguistic development that is observed across 2-year-olds and to provide a developmental metric that, in later studies, will make it possible to match the children studied with other children acquiring the same or other languages. We calculated the word-based MLU (hereafter, MLUw) for each file and child. MLUw was calculated via the *mlu* program in CLAN, by running the command 'mlu +t*CHI -t%MOR *.cha'. The lowest and highest MLUw values in each child's dataset were used to determine the MLUw range for their production.

The choice of word-based MLU metrics over morpheme-based MLUs follows from two considerations: firstly, not all CHILDES corpora (including most of the corpora studied here) are morphosyntactically tagged (namely, have a %mor tier), entailing morpheme-based metric are not available in practice. Word-based MLUs are not just the only alternative in this instance, but they also facilitate uniform comparison and the replicability of the present data for other CHILDES corpora beyond the ones studied for which morphosyntactic tagging is also unavailable. Furthermore, as noted in Ezeizabarrena & Garcia Fernandez (2017), the calculation of morpheme-based MLUs is highly susceptible to annotation errors and subjectivity: the researcher or annotator must determine which morphemes the children harness productively, which may in turn lead to arbitrary and error-prone decisions. Therefore, I adopt word-based MLUs throughout (MLUw) and take them to be a more uniform and objective developmental metric.

The details for each child studied, their MLUw range and the number of files analysed are reported in Table 2.

Table 2. Children studied in the CHILDES database and summary information.

Language	Corpus	Children	Age range	Files analysed	MLUw range
Catalan	Serra/Solé	Laura	1;07-4;00	19	1.03-3.47
		Gisela	1;04-4;02	20	1.02-3.51
		Àlvar	1;02-3;01	21	1.0-3.37
		Guillem	1;01-4;00	34	1.01-3.88
	Júlia	Júlia	1;07-2;06	17	1.15-2.74
Spanish	Llinàs/Ojea	Irene	0;11-3;02	40	1.0-4.94
		Yasmin	1;10-2;09	47	1.29-3.21
	Aguado-Orea/Pine	Juan	1;10-2;05	65	1.34-3.39
	Aguirre	Magín	1;07-2;10	29	1.24-3.07
	Vila	Emilio	0;11-4;08	35	1.0-3.23
Total		10			

4.3 Results

Table 3 reports the breakdown of productions by type of complementiser for each child. Overall, among the 1318 total utterances with complementisers, 1009 of them (76.6%) corresponded to examples with illocutionary and interrogative complementisers and 309 corresponded to subordinating complementisers (23.4%). As expected, given the main-clause nature of illocutionary complementisers, these are much more frequent in the child production data.

Table 3. Proportion of use by type of complementiser.

Language	Children	Illocutionary	Embedding
Catalan	Laura	154 (76.2%)	48 (23.8%)
	Gisela	148 (73.6%)	53 (26.4%)
	Àlvar	9 (60%)	6 (40%)
	Guillem	85 (81%)	20 (19%)
	Júlia	3 (75%)	1 (25%)
Spanish	Irene	58 (64.4%)	32 (35.6%)
	Yasmin	36 (85.7%)	6 (14.3%)
	Juan	164 (67.2%)	80 (32.8%)
	Magín	248 (84.1%)	47 (15.9%)
	Emilio	104 (86.7%)	16 (13.3%)
Total		1009 (76.6%)	309 (23.4%)

We provide examples of illocutionary complementiser productions in (17), for Catalan children, and (18), for Spanish. At the earliest stages, utterances with illocutionary complementisers are syntactically simple, typically involving an illocutionary complementiser alongside a verb and a (null/overt) subject.

(17)

    a. **Que**   ja   no fa   mal?   (Guillem, MLUw 1.99)
       that.INT already not make.3SG pain

       'Does it not hurt anymore?'

    b. Ai,   **que**   crema!   (Laura, MLUw 1.35)
       ouch that.EXCL burn.3SG

       'Ouch, it's burning!'

    c. Espera't,   **que**   estic fent   això   (Gisela, MLUw 2.61)
       wait=CL.REFL that.CONJ   am   doing this

       'Wait, I'm doing this'

    d. **Que**   no hi   és   (Àlvar, MLUw 1.84)
       that.QUOT not CL.LOC= is

       '(I've already said) it's not there'

(18)

    a. **Que**   no qu(i)ero   (Juan, MLUw 1.58)
       that.QUOT not want.1SG

       '(I said) I don't want to'

    b. ¡**Que**   viene!   (Magín, MLUw 1.78)
       that.EXCL come.3SG

       'He/she is coming!'

c. Ay, no, **que**   me   harán   daño   (Emilio, MLUw 2.2)
ouch no that.CONJ CL.IO= do.FUT.3PL   harm
a la barriga
to the tummy

'Ouch, no, they'll hurt my tummy'

In contrast, the first instantiations of subordinating complementisers are already syntactically more sophisticated, by virtue of being attested at a later developmental stage (see discussion below). Some examples illustrating early relative clauses and complement clauses are given in (19). Note that several of the English translations below feature a null *that*-complementiser. Whilst we would not be able to diagnose availability of complementation in child English exclusively on the basis of production of overt complementisers, this problem does not arise with the Catalan and Spanish data. This is because null complementisers are generally ungrammatical in embedding contexts in both Catalan and Spanish (except in an infrequent set of typically formal contexts with subjunctive mood in Spanish and, to a lesser extent, in Catalan; see, i.a., Etxepare 1996, Antonelli 2013, Llinàs-Grau & Fernández-Sánchez 2013[13]).

(19)
a. Saps   **que** no vindrà   ningú?   (Gisela, MLUw 2.61)
know.2SG that not come.FUT.3SG no.one

'Do you know no one is going to come?'

b. Una vegada hi   havia   un nen **que**   (Júlia, MLUw 2.74)
one time   CL.LOC= AUX.IMPF.3SG   a boy that
es   diu   Andreu
cl.REFL=   say.3SG Andreu

'Once upon a time, there was a boy named Andreu'

c. En una capsa **que** hi   ha   aquí   (Àlvar, MLUw 2.82)
in a   box   that CL.LOC= AUX.3SG   here

'In a box that's here'

_____

[13]   Llinàs-Grau & Fernández-Sánchez (2013) additionally argue that, although these constructions are superficially similar in that a complementiser is missing, the constructions that allow a null *que* in Catalan and Spanish do not involve the same process as the ones that may underlie *that*-deletion in English.

(20)

    a. Quiero   **que** sea   un zapato   (Yasmin, MLUw 2.47)
        want.1SG that be.SUBJ.3SG a   shoe

        'I want it to be a shoe'

    b. ¿No ves   **que** estaba con la raqueta?   (Irene, MLUw 3.23)
        not see.2SG that was   with the racquet

        'Don't you see it was next to racquet?'

    c. Había   un señor muy malo **que** tiraba   (Juan, MLUw 2.73)
        aux.IMPF.3SG   a   man   very bad who throw.IMPF.3SG
        flechas
        arrows

        'There was a very bad man who was throwing arrows'

We turn now to the predictions made by the different accounts outlined in section 3. As per a bottom-up maturational account, both kinds of complementisers are predicted to be late acquisitions and to surface at roughly similar times, on the grounds that these are hosted at the very top of the functional spine and that a rigid, biologically-timed developmental pathway specifies the acquisition ordering of functional categories. In contrast, this is not borne out in the ten children studied. What we find instead is that the first illocutionary complementisers typically appear well before embedding complementisers and never later. In a couple of children only (Júlia and Emilio), both types of complementisers emerge simultaneously in the production data. Although the ten children differ in the number of complementisers they produce, these patterns recur in all of them.

    The MLUw stage at which illocutionary complementisers emerge for every child is reported in Table 4[14]:

---

[14]   The calculation of the average MLUw excludes outlier values, which were Júlia's (see also discussion below).

Table 4. Emergence of illocutionary and embedding complementisers.

Language	Children	Illocutionary	Embedding
Catalan	Laura	1;10.22	3;00.02
		1.15 MLUw	2.42 MLUw
	Gisela	1;08.24	2;08.00
		1.13 MLUw	2.61 MLUw
	Àlvar	2;02.06	2;06.25
		1.84 MLUw	1.91 MLUw
	Guillem	2;02.28	2;11.25
		1.54 MLUw	2.44 MLUw
	Júlia	2;06.25	2;06.25
		2.74 MLUw	2.74 MLUw
Spanish	Irene	1;08.09	1;09.10
		1.88 MLUw	3.28 MLUw
	Yasmin	1;10.08	2;05.18
		1.93 MLUw	2.47 MLUw
	Juan	1;11.11	2;01.21
		1.58 MLUw	1.77 MLUw
	Magín	1;09.01	1;10.00
		1.78 MLUw	2.73 MLUw
	Emilio	2;04.17	2;04.17
		2.18 MLUw	2.18 MLUw
Total		1.67 MLUw	2.42 MLUw

We ran a paired-samples t-test in R (R Core Team 2022) to confirm that the MLUw at the point of emergence (as the dependent variable) is different for illocutionary and subordinating complementisers (the independent variable, encoded as a binary variable). A dataset containing the MLU when illocutionary *vs* embedding complementisers first emerge in every child was compared. Any outliers in the dataset were excluded. This only involved Júlia's MLUw value for the emergence of illocutionary complementisers, which was radically different compared to the other children. As a result, I excluded her data for both illocutionary and embedding complementisers from the t-test. A Shapiro-Wilk test for normality confirmed that the dataset is normally distributed (p = 0.8347), rendering it suitable for a paired-samples t-test. The results of the t-test showed that MLUw values for illocutionary complementisers were lower (M = 1.67, SD = 0.35) compared to embedding complementisers (M = 2.42, SD = 0.45) with respect to their point of emergence. There was a highly statistically significant difference between the MLU value of emergence of the two complementiser types of 0.5456 ($t(17)$ = 5.6201, p < .001), with a 95% confidence interval that ranged from 0.3408 to 0.7504. This justifies rejection of the

null hypothesis that there is no difference between the point of emergence (measured with MLUw) of illocutionary *vs* embedding complementisers: the results confirmed that illocutionary complementisers were significantly more likely to emerge earlier than embedding complementisers.

On average, the two kinds of complementisers emerged at the following MLUw values:

Table 5. Average and range of MLUw values across language groups for the emergence of illocutionary and embedding complementisers.

	Illocutionary	**Embedding**
Catalan	1.41 MLUw (range 1.13-1.84)	2.35 MLUw (range 1.91-2.61)
Spanish	1.87 MLUw (range 1.58-2.18)	2.49 MLUw (range 1.77-3.28)
Combined	1.67 MLUw (range 1.13-2.18)	2.42 MLUw (range 1.77-3.28)

Crucially, before embedding complementisers first emerge, these structures with early illocutionary complementisers are neither infrequent nor lexically specific, plausibly indicating that they represent syntactically productive knowledge. This is shown in Table 6, which displays the range of verb types with which these early illocutionary complementisers are found, prior to the emergence of illocutionary complementisers, and their frequency (number of utterances attested). Recall that Júlia and Emilio are the only two children that show simultaneous emergence of illocutionary and embedding complementisers; therefore, their rows are shaded in grew and are left blank.

Table 6. Types of verbs with illocutionary complementisers before the emergence of embedding complementisers and overall frequency of illocutionary complementisers at this point.

	Unacc	**Unerg**	**Trans**	**Modal**	**Copula**	**Impers**	**Freq**
Laura	✓		✓	✓	✓	✓	20
Gisela	✓	✓	✓	✓	✓		7
Àlvar					✓		1
Guillem	✓	✓	✓	✓		✓	11
Júlia							
Irene			✓				1
Yasmin	✓	✓	✓		✓	✓	18
Juan	✓	✓	✓		✓		10
Magín	✓		✓				10
Emilio							

Abbreviations: Unacc = unaccusative, Unerg = unergative, Trans = transitive, Impers = impersonal, Freq = frequency.

Table 6 illustrates, then, two further empirical contributions: (i) for most children, the earliest illocutionary complementisers can be found with a wide range of verb classes, making it unlikely these complementisers form part of lexically specific or rote-learned formulae. And (ii), illocutionary complementisers do not just emerge early in the dataset, they are also considerably frequent from the very start, in turn strengthening two of our arguments: the possibility that they are acquired and represented early, well before embedding complementisers, and combined with (i) above, that their early emergence *cannot* feasibly be a function of lexical specificity.

Finally, the children's acquisition of both kinds of complementisers over time is shown in Figure 3, which plots their development according to the children's MLUw (with the y-axis indicating number of occurrences attested for each complementiser type). Each child's individual development and their graphs are collated in Appendix A.
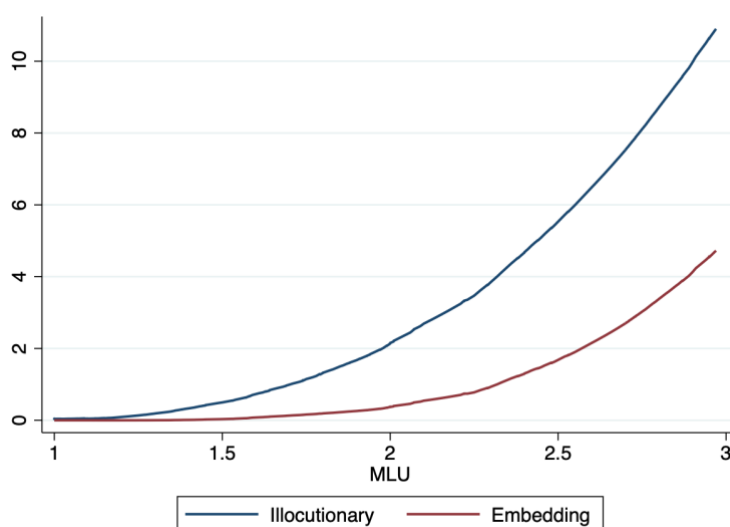
Figure 3. The development of complementisers in the Catalan and Spanish children.

As can be seen from Figure 3 and the slopes of the curves, illocutionary complementisers both emerge earlier and develop faster in frequency than their subordinating counterparts. A one-sample Kolmogorov-Smirnov test was carried out to check that the two distributions are statistically significantly different from each other, which rejected the null hypothesis that the two curves are equal (D = 2.0000, *p* < .001).

In summary, illocutionary complementisers are accessible early on in most of the children studied and are almost always produced several files before embedding complementisers emerge. In four out of the ten children studied (Laura, Gisela, Guillem and Juan), these emerge from the earliest multiword utterances, by around 1.5 MLUw or even earlier. Furthermore, an analysis of their frequency and accompanying verb types before the emergence of embedding complementisers reveals that they are generally very frequent from the earliest recordings and furthermore not restricted to specific verb classes. Importantly, this ruled out the possibility that the earliest illocutionary complementisers are merely rote-learned formulae, not indicative of productive syntactic knowledge. In light of all this, we conclude, therefore, that the acquisition of illocutionary complementisers and embedding complementisers is developmentally dissociable and unrelated, unlike what the predictions of a bottom-up maturational account would lead us to expect.

5   Discussion

The results of the corpus study show that illocutionary complementisers emerge in our data significantly earlier than embedding complementisers, supporting the common predictions made by approaches to the acquisition of functional categories that anticipate early emergence of the CP and the SAP domains (e.g., Galasso 2003, Tsimpli 2005, Roeper 2007, van Kampen 2010, Biberauer & Roberts 2015, Biberauer 2018, Heim & Wiltschko 2021; recall section 3). As expected by these approaches, illocutionary complementisers often emerge well before embedding complementisers, and never later than the latter in the ten children studied. This is hypothesised to be due to both their speaker-hearer and main-clause nature. We have also established that, for some children at least, some illocutionary complementisers appear available from the earliest MLU stages (around 1.5 MLUw or sometimes even earlier), indicating that a CP domain is possibly accessible at this stage for these children. In several children, too, the earliest illocutionary complementisers, preceding the emergence of embedding complementisers, are both frequent in the production data and are used in a wide variety of verb types, which we took to lend credence to the conclusion that some productive syntactic knowledge of illocutionary complementisers has emerged by this point. These findings support multiple recent approaches that propose children make efficient use of interactional/discourse, edge-based and 'here-and-now' cues in building their incipient grammars. As noted earlier, the present study does not establish whether all types of illocutionary complementisers emerge before embedding complementiser (a separate empirical question). The important point, nonetheless, is that the first instances of illocutionary complementisers are never attested after the emergence of embedding complementisers.

This observation, although necessarily preliminary, could have significant consequences for developmental theories that pursue a bottom-up maturational track, and which would predict CP knowledge to emerge substantially late, with illocutionary and embedding complementisers developing at comparatively 'delayed' stages. The early availability of illocutionary complementisers contradicts such proposals and suggests instead that, whilst embedding complementisers may indeed be later phenomena compared to, for instance, argument structure and basic tense/agreement marking (as bottom-up proposals predict), not all complementisers are equally stagnant at early stages. I preliminarily suggest, therefore, that bottom-up approaches do not seem well-suited to account for the emergence of this grammatical knowledge, at least in Catalan and Spanish, where these illocutionary complementisers abound in both child production, spoken language and, thus, carer input[15]. Instead, they point, in a novel way, to an acquisitionally advantaged role of peripheries and edges, interactional language and the CP domain more broadly, as variously entertained by Biberauer's (2018a) PSHH, Heim & Wiltschko's (2021) Inward Growing Spine Hypothesis and Biberauer & Roberts's (2015) emergent categorial hierarchy (section 3). This paper supplements the findings in Bosch (2023a), where it is suggested that the children studied largely abide by the predictions made by Biberauer & Roberts's (2015) emergent categorial hierarchy and that their data does not support bottom-up structural development as far as emergence (first use of a given structure) is concerned.

These findings must be treated as preliminary, due to the limited and exploratory nature of the corpus study presented here and the relatively small sample (N = 10). Nonetheless, the patterns reported in this study had thus far gone unnoticed and represent novel evidence that makes the case for early availability of a discourse/interactional domain in child grammars (cf. Roeper 2007 on the importance of early expressive language). More attention should also be devoted to the role of maturation in language development: while it is clear that embedding complementisers emerge later than the first illocutionary complementisers, it remains to be determined whether this is simply because embedding complementisers require more sophisticated syntactic knowledge accommodating subordination or whether this is due to some maturational timeline (the Inward Growing Spine Hypothesis would predict Linking to be accessible later due to maturational constraints). If it is correct, however, that some illocutionary complementisers such as interrogative and quotative complementisers are located in the CP or Wiltschko's Linking domain (and the child encodes them as such), their early emergence might be at odds with the Inward Growing Spine Hypothesis. A

---

[15]   Although this study did not quantify the adult data, a cursory examination of the adult utterances in CHILDES indicates that illocutionary complementisers are very frequent in the input, at least impressionistically.

separate corpus study analysing the relative development of the different subtypes of illocutionary complementisers would be needed to test for this. Generally, the data presented here does not have direct implications regarding the cause behind the relative delay of embedding complementisers. A theoretically more parsimonious approach would be one that foregoes appeal to innate developmental mechanisms such as maturation and resorts, for instance, to the differences between main and embedded contexts and their syntactic complexity, as well as to the salience of interactional/discourse content in acquisition. Notwithstanding the advantage of these 'early-CP' and 'inward-growing' approaches in accounting for the presented data, it remains to be seen whether they make the correct predictions for acquisition more generally and, importantly, whether development of syntactic categories necessarily needs to be modelled via a maturational mechanism, as suggested by Heim & Wiltschko (see Bosch 2023a for arguments that a non-maturational, neo-emergentist approach may fare better).

Additionally, further research is required to determine whether these findings recur in other Spanish and Catalan children and, particularly, in other Ibero-Romance languages, notably Portuguese. An analysis of the acquisition data of children acquiring languages beyond Ibero-Romance also appears to be a potentially productive avenue; I discuss one potential extension of the present work in relation to Italo-Romance, and some preliminary results based on shallow empirical investigation, in the following section 7. Examination of *cross*linguistic differences in the acquisition of illocutionary complementisers could also be helpful. As seen in Table 4 and Table 5, Catalan children appear faster than the Spanish children studied in producing illocutionary complementisers. Conceivably, this could be due to the relatively higher prevalence of illocutionary complementisers in Catalan, given the frequent use of interrogative complementisers in spoken language, which are unavailable in Spanish. Alternatively, it could simply be an accidental fact of this dataset. The sample per language (5 children) is likely too small to make any conclusive statements, but future work should aim to probe for these timing differences further. One should also pay due attention to dialectal differences; for instance, whether developmental timings differ between children speaking different dialects, such as different varieties of Latin American *vs* Peninsular Spanish, or even Catalonian Spanish *vs* Spanish spoken in other areas of the Iberian Peninsula. Finally, this paper mainly employs naturalistic corpus data, but experimental tests, such as comprehension studies, would also be revealing as regards the state of children's knowledge.

6 Future extensions: first impressions of Italo-Romance complementiser development

Many Italo-Romance varieties, particularly in central and northern Italy, generally present a more restricted system of illocutionary complementisers compared to Ibero-Romance. Broadly, they only allow some conjunctive uses of *che*, exclamative *che* with subjunctive mood and also interrogative uses of *che* (the latter only in some varieties, such as Roman or other Tuscan dialects; see Giurgea & Remberger 2016). In contrast, several Southern Italian Dialects (SIDs), such as Calabrian, present a more extensive system of illocutionary complementisers, allowing main-clause *che* to surface with interrogative, imperative, optative, conjunctive and exclamative functions (or a subset of these; see Ledgeway 2010, 2015: 120-121, 2020: 33-35, Prins 2014, D'Alessandro & Felice 2015). Besides Italo-Romance, Romanian and some Rhaeto-Romance varieties (e.g., Badiotto) allow use of the complementiser *că* and *che* in conjunctive contexts (Corr 2016: 226-227). The use of illocutionary complementisers in Italo-Romance, with various degrees of robustness, indicates, therefore, that a detailed analysis of their development would be a welcome addition to the discussion surveyed in this paper.

Indeed, a preliminary look at the data of Italian children in the CHILDES database already suggests that they are also attested reasonably early on, and, interestingly, often in creative configurations that are ungrammatical in many Italian varieties (but crucially attested in other Romance varieties or SIDs). Some examples of main-clause uses of *che* by Italian children at relatively early stages (before 2.0 MLUw) are given in (21). All data cited in (21-22) is taken from the following CHILDES corpora: the Calambrone corpus (Martina and Diana; Cipriani et al. 1989) and the Tonelli corpus (Elisa and Marco; Tonelli, Dressler & Romano 1995).

(21)
    a. **Che** gira   (Martina; 1;08.02, MLUw 1.9)
        that turn.3SG

        'He/she/it stirs (it)'

    b. **Che** legge   (Martina; 1;08.02, MLUw 1.9)
        that read.3SG

        'She is reading' (in response to *Diglielo alla mamma cosa fa la bimba*, 'tell mum what the child is doing').

    c. **Che**   ride!   (Martina; 1;11.02, MLUw 1.99)
       that.EXCL laugh.3SG

       'He/she is laughing!'

    d. **Che**   piove   (Martina; 2;01.12, MLUw 1.86)
       that.CONJ rain.3SG

       'It's raining' (in response to *l'ombrello*?, 'the umbrella?', asking what someone was doing with an umbrella)

(21b) involves the use of *che* as a response to an imperative (generally disallowed in Italian). (21c) is a fairly transparent case of an exclamative complementiser and (21d) is arguably a case of a conjunctive complementiser, where the child is implying that the person is using an umbrella *because* it is raining (note also that *che* in Italian cannot generally replace *perché* 'because' as a response to an open question). The function of (21a) is unclear based on the surrounding context[16]; it appears to indicate either surprise or exclamation, but it could also have an imperative reading (see footnote 16 for conversational context)[17].

Importantly, three Italian informants rated all the sentences in (21) as ungrammatical and disallowed in their respective varieties (the Pisa and Roma regions for two of the informants, and the Veneto region for the other; the former share the linguistic background of Martina). Neither Martina nor the children cited below were raised in the south of Italy (where some of these structures would be tolerated), at least according to the information in CHILDES. An examination of the

---

[16]   Example (21a) was uttered in the following conversational context (my own translations):
       *MOT: devo girare il caffé ('I have to stir the coffee')
       *CHI: **che gira** ('It/he/she stirs')
       *MOT: giri te? ('Are you stirring (it)?')
       *MOT: gira ancora, gira ('Stir (it) again, stir (it)')
[17]   Note that Moscati & Rizzi (2021) also study the development of *che* in Martina (among other children). However, the reported time of emergence is later than that suggested by the examples in (21) above, which date from 1;08 or 20 months. In Moscati & Rizzi (2021), *che* is reported to emerge in Martina at 21 months (Table 2, p. 6). This plausibly results from including different data: for matrix uses of *che,* Moscati & Rizzi only consider polar exclamatives (e.g., **sì che** *lo sai!* lit. 'yes that it know!', 'yes you know it!', p. 9), wh-exclamatives (e.g., **che** *bella!* 'so nice!', p. 9) and wh-questions. None of the examples in (21) would fall into these categories, as they are target-deviant, potentially explaining their omission from their dataset. Significantly, the discussion above suggests that examining cases of input-divergent or ungrammatical productions such as (21) also sheds potentially important light on the formal status of children's representations. Thanks to an anonymous reviewer for drawing my attention to Martina's data in Moscati & Rizzi.

utterances by Martina's parents (who speak a Tuscan variety) indicates that only conjunctive and interrogative complementisers are present in the recordings. No instances of exclamative, quotative or other illocutionary complementisers are attested. Only the utterance *che c'ha la tazza, sì, brava!* ('there's a cup, yes, good!'; file 1;11.02, line 518) instantiates an exclamative complementiser in the parental input. As noted by a Tuscan informant, this specific construction is possible in Tuscan varieties, but otherwise exclamative *che* faces several restrictions (as noted earlier, Giurgea & Remberger 2017 comment that exclamatives with *che* in Italian are typically found with subjunctive mood). Therefore, since these forms are not generally sanctioned in these varieties of Italian and probably cannot have been learned directly from the input, this innovative use of main-clause complementisers is remarkable. Given that these input-deviant structures are not scarce in some children (see the four examples in (21), all by Martina), they must instead reflect active 'reuse' of the complementiser and conjunction *che* and its generalisation to a broader range of speaker-hearer functions than those present in the children's input. An aspect worth highlighting, however, is that these wider uses of *che* in (21) nonetheless reflect attested instantiations of illocutionary complementisers in other Romance varieties, like Ibero-Romance and some SIDs.

This apparent generalisation of the interactionally-oriented functions of Italian *che* also extends to later developmental stages. (22) presents some instances of main-clause *che* being harnessed for seemingly exclamative, interrogative and quotative purposes, most of which are generally disallowed in Italian (and particularly in northern regions of Italy, where Elisa and Marco grew up).

(22)

    a. Oh, **che**   c'ha   un   (Diana; 2;06.00, MLUw 5.53)
        oh   that.EXCL CL.LOC=have.3SG   a
        lunghi   pelosi!
        long.PL hairy.PL

        (lit.) 'Oh, there's a long hairy!' (possibly meaning 'There's (a) long hair(s)!')

    b. **Che**   io ti   chiudo   la   (Diana; 2;06.00, MLUw 5.53)
        that.QUOT I   CL.IO= close.1SG the
        bocca, sai?
        mouth know.2SG

        '(I've said) I'll shut your mouth, you know?'

> c. **Che**   lo   metto   qui   (Elisa; 2;01.06, MLUw 4.47)
>   that.QUOT CL.DO= put.1SG here
>   '(I've said) I'm putting this here'

> d. E **che**   vuoi   un posto tu?   (Marco; 2;01.27, MLUw 2.16)
>   and that.INT want.2SG a   place you
>
>   'And do you want a place?'

Similarly to (21), the constructions in (22) with a main-clause *che* are very frequently judged to be ungrammatical in Italian (with the exception of some SIDs, as noted above). An exclamative complementiser is exemplified in (22a); as noted earlier, this is often disallowed in Italian, though one informant did accept the structure[18]. (22b-22c) could instantiate examples of quotative complementisers, as these examples were produced right after similar utterances, a characteristic of quotative *que* in Ibero-Romance (recall section 2): these are *chiudi, la bocca no chiudi bene?* ('shut up, don't you close your mouth properly?') and *lo metto qui* ('I'm putting this here'), respectively. Both (22b) and (22c) are generally classed as borderline ungrammatical by my informants (though with some inter-speaker variation), who additionally report that *guarda* 'look/watch' preceding *che* would form a more natural construction in their varieties (e.g., *guarda che lo metto qui*, 'Look, I'm putting it here')[19]. (22d), on the other hand, is a case of an interrogative *che* heading a neutral polar question (analogous to the Catalan example in 9). The informant from the Veneto region rejected (22d) as a possible utterance in their north-eastern variety. This may again be significant, given that Marco was raised in the Northeast of Italy. In contrast, interrogative *che* is possible in Roman and Tuscan varieties (as expected, my Tuscan informants accepted these examples). The parental input in Diana, Elisa and Marco's recordings also did not contain any instances of quotative, exclamative or interrogative complementisers. It is transparent from (21-22), then, that the linguistic creativity evidenced by many of these 'errors' may far exceed the constructions allowed in the Italo-Romance varieties that these children were exposed to, indicating that there is a possible developmental stage in which Italian children 'maximise' the use of illocutionary *che* (cf. also Hudson-Kam & Newport 2005 on 'maximisation' and regularisation in acquisition).

---

[18]   Setting aside here the orthogonal error in the noun phrase (with a missing noun), where the child likely intended *capelli lunghi* ('long hairs'), as opposed to *un lunghi pelosi* (literally, 'a long hairy').

[19]   Note that there is no prosodic break between *guarda* and *che* above (at least according to the judgements of my three informants). Crucially, then, *che* here would be an embedding complementiser ('be warned/observe that…'), *not* an illocutionary complementiser.

These creative structures that outstrip the input the children are presented with certainly warrant further scrutiny. Greater attention to these seemingly 'generalised' complementisers may turn out to be theoretically illuminating, not just developmentally, but also from a diachronic and language variation perspective. For one, all the equivalent sentences in (21-22) represent grammatical configurations in Catalan[20] and most of them are also productive in Spanish (with the main exception being interrogative complementisers). The fact that the 'errors' by the Italian children in (21-22) align with the speaker-hearer functions that the complementiser *che/que* can fulfil in other Romance varieties (such as the ones analysed here for Catalan and Spanish) is highly suggestive.

Looking at the bigger picture, though, that language acquisition always involves going beyond a finite input is an indisputable fact. It is much less clear, however, *how* this extension is carried out and what forms and linguistic content are more or less likely to be subject to overgeneralisations when a (perceived) grammatical regularity is detected[21] (Biberauer, 2019b: 211-212). It is in this respect that careful analysis of child overgeneralisation errors like the ones in (21-22) could prove most fruitful. Notably, the potential 'going beyond the input' scenario illustrated in the Italian children above appears telling from a PSHH standpoint (recall section 3), given that these complementisers with perspectival functions are structurally peripheral. From this perspective, then, these superficially innocent and 'anecdotal' examples could potentially elucidate children's keenness to innovatively generalise already-acquired forms or forms present in the input (but with a different distribution) by endowing them with novel speaker-hearer meaning, as expected from a PSHH and Maximise Minimal Means perspective (see also Biberauer 2019a: 71-79, 2019b). As they stand, these comments remain largely speculative, but further exploration of the acquisition of these complementisers in Italian (as well as their synchronic dialectal/regional variation, which is far from fully documented) is clearly a desirable future direction. I leave these strands of work to future research.

---

[20]   Although the equivalent of (21b) is only marginally acceptable in Catalan in that context (?*Que llegeix*, 'he/she is reading'; my own judgements).

[21]   These overgeneralisations may later, if re-enforced in peer-to-peer contexts, lead to diachronic change (see Cournane 2019, Biberauer 2019b). In this context, it also worth remarking the misleading nature of the term 'errors', which are imbued with a judgement according to which there is a failure to reach a given (here, adult-like) 'target' or 'goal'. This conflicts with a non-teleological view of acquisition and acquisition-driven diachronic change (as noted by Cournane 2017: 10); therefore, terms such as 'input-divergent' (following Cournane 2017) are to be preferred.

7   Conclusion

In this paper, I investigated the emergence of illocutionary complementisers in child Catalan and Spanish and compared it to the use of embedding complementisers. I established that illocutionary complementisers consistently emerge earlier in child speech than embedding complementisers. Since at least some illocutionary complementisers are often attested at early MLU values, I argued that this finding is hard to accommodate in bottom-up maturational approaches to the acquisition of functional categories, which propose the CP domain is acquired last. This developmental pattern can instead be understood with approaches which suggest an interactional/discourse domain alongside *v*P are first to develop in the learning path. The developmental differences between the two kinds of complementisers feasibly also come down to the distinction between main/embedded clauses, with embedding complementisers appearing later owing to the increased syntactic sophistication of the subordination structures of which they form part. Setting out an avenue for future expansion beyond Catalan and Spanish, I have also offered preliminary evidence suggesting illocutionary complementisers are attested relatively early in Italian child speech, often in configurations that illustrate generalisation of *che* to a wider range of speaker-hearer functions than those evident in the children's input. More broadly, the early emergence of illocutionary complementisers lends additional support to the salience of speech-act and discourse material in acquisition. It suggests that further study of their acquisition can likely enhance our understanding of the make-up of early child grammars and the heuristics used to leverage and generalise formal knowledge from systematicities in the input. Echoing the sentiment in, i.a., Roeper (2007), van Kampen (2010), Biberauer (2018) and Heim & Wiltschko (2021), their development helps underscore the possibility that *peripheral* and thus far underdiscussed elements may nonetheless be *core* stepping stones in the process of grammar construction.

Acknowledgments

References

Aguado Orea, Javier & Juan Pine. 2015. Comparing different models of the development of verb inflection in early child Spanish. *PloS ONE* 10(3). e0119613. https://doi.org/10.1371/journal.pone.0119613.

Aguirre, Carmen. 2000. *La adquisición de las categorías gramaticales en español*. Madrid: Ediciones de la Universidad Autónoma de Madrid.

Antonelli, André. 2013. On the Left Periphery of Spanish Complementizerless Clauses. In Jennifer Cabrelli Amaro & et al. (eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, 15–26. Somerville, MA: Cascadilla Proceedings Project.

Armon-Lotem, Sharon. 2005. The Acquisition of Subordination: From Preconjunctionals to Later Use. In Dorit Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on Language and Language Development*, 192–204. Kluwer. https://doi.org/10.1007/1-4020-7911-7_15.

Baunaz, Lena. 2015. On the various sizes of complementizers. *Probus: International Journal of Romance Linguistics* 27(2). 193-236. https://doi.org/10.1515/probus-2014-0001.

Bel, Aurora. 2001. *Teoria lingüística i adquisició del llenguatge: anàlisi comparada dels trets morfològics en català i en castellà*. Barcelona: Institut d'Estudis Catalans.

Belletti, Adriana. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.), *The structure of CP and IP: The Cartography of syntactic structures*, vol. 2, 16–51. Oxford: Oxford University Press.

Biberauer, Theresa. 2018. Peripheral significance: a phasal perspective on the grammaticalisation of speaker perspective. Talk presented at *DiGS 20* (York), June. https://www.york.ac.uk/media/languageandlinguistics/documents/conferences/digs20/Biberauer handout DiGS20.pdf.

Biberauer, Theresa. 2019a. Factors 2 and 3: Towards a principled approach. *Catalan Journal of Linguistics, Special Issue*, 45–88. https://doi.org/10.5565/rev/catjl.219.

Biberauer, Theresa. 2019b. Children always go beyond the input: The Maximise Minimal Means perspective. *Theoretical Linguistics* 45(3-4). 211-224. https://doi.org/10.1515/tl-2019-0013.

Biberauer, Theresa & Ian Roberts. 2015. Rethinking formal hierarchies: a proposed unification. *Cambridge Occasional Papers in Linguistics* 7. 1–31.

Bosch, Núria. 2023a. *Emergent Syntax and Maturation: a neo-emergentist approach to syntactic development*. MPhil thesis, University of Cambridge.

Bosch, Núria. 2023b. Not All Complementisers Are Late: A first look at the acquisition of illocutionary complementisers in Catalan and Spanish. *Cambridge Occasional Papers in Linguistics* 15. 1-38.

Bošković, Željko. 2014. Now I'm a Phase, Now I'm Not a Phase: On the Variability of Phases with Extraction and Ellipsis. *Linguistic Inquiry* 45(1). 27–89. https://doi.org/10.1162/ling_a_00148.

Bresnan, J. W. 1972. Theory of complementation in English syntax. Ph.D. thesis, Massachusetts Institute of Technology. https://doi.org/10.4324/9781315458731.

Cameron-Faulkner, Thea. 2014. The development of speech acts. In Danielle Matthews (ed.), *Pragmatic development: Trends in Language Acquisition Research*, 37-52. Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.10.03cam.

Caprin, Claudia & Maria Teresa Guasti. 2009. The acquisition of morphosyntax in Italian: A cross-sectional study. *Applied Psycholinguistics* 30(1). 23–52. https://doi.org/10.1017/s0142716408090024.

Chomsky, Noam. 1955. *The logical structure of linguistic theory*. Chicago: University of Chicago Press.

Chomsky, Noam. 1977. On Wh-Movement. In Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71-132. New York: Academic Press.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 85–156. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1–50. Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three Factors in Language Design. *Linguistic Inquiry*, 36(1). 1–22. https://doi.org/10.1162/0024389052993655.

Chomsky, Noam. 2008. On Phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133–166. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262062787.003.0007.

Cinque, Guglielmo. 1999. A*dverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., Pantano, N., Pfanner, L., Poli, P., Sarno, S., Bottari, P., Cappelli, G., Colombo, C., & Veneziano, E. 1989. *Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio* (1/84 Italian Ministry of Health). Stella Maris Foundation

Clahsen, Harald & Martina Penke. 1992. The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone-corpus. In Jürgen Meisel (ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition*, 181–224. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-2803-2_7.

Clahsen, Harald, Martina Penke & Teresa Parodi. 1993. Functional Categories in Early Child German. *Language Acquisition* 3(4). 395–429. https://doi.org/10.1207/s15327817la0304_3.

Corr, Alice. 2016. *Illocutionary complementisers and utterance syntax*. Ph.D. thesis, University of Cambridge.

Corr, Alice. 2022. *The Grammar of the Utterance: How to Do Things with Ibero-Romance*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198856597.001.0001.

Cournane, Ailís. 2017. In defence of the child innovator. In Éric Mathieu & Robert Truswell (eds.), *Micro-change and Macro-change in Diachronic Syntax*, 10-24. Amsterdam: John Benjamins. https://doi.org/10.1093/oso/9780198747840.003.0002.

Cournane, Ailís. 2019. A developmental view on incrementation in language change. *Theoretical Linguistics* 45(3-4). 127-150. https://doi.org/10.1515/tl-2019-0010.

Cruschina, Silvio & Eva-Maria Remberger. 2018. Speaker-oriented syntax and root clause complementizers. *Linguistic Variation* 18(2). 336–358. https://doi.org/10.1075/lv.16009.cru.

D'Alessandro, Roberta & Claudio di Felice. 2015. The diachrony of Abruzzese complementation. *Revue Roumaine de Linguistique* 40(2-3). 129-145.

Diercks, Michael & Madeleine Bossi. 2021. Minimalist syntax is psychologically real: Lessons from (counter)cyclicity. https://ling.auf.net/lingbuzz/005671.

Diercks, Michael, Katherine Johnson, Galia Bar-Sever & Madeline Bossi. 2023. Developmental Minimalist Syntax. https://ling.auf.net/lingbuzz/007134.

den Dikken, Marcel. 2007. Phase Extension Contours of a theory of the role of head movement in phrasal extraction. *Theoretical Linguistics* 33(1). 1–41. https://doi.org/10.1515/tl.2007.001.

Douglas, Jamie. 2018. Maōri subject extraction. *Glossa: a journal of general linguistics* 3(1). 110. https://doi.org/10.5334/gjgl.566.

Dresher, B. Elan. 2009. *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642005.

Etxepare, Ricardo. 1996. Null Complementizers in Spanish. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"* 32(2). 469–496. https://doi.org/10.1387/asju.8657.

Ezeizabarrena, Maria-José and Garcia Fernandez, Iñaki. 2017. Length of Utterance, in Morphemes or in Words?: MLU3-w, a Reliable Measure of Language Development in Early Basque. *Frontiers in Psychology* 8. 2265. https://doi.org/10.3389/fpsyg.2017.02265.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form: Generalizations across languages*, 87–116. Amsterdam: John Benjamins. https://doi.org/10.1075/la.100.07fra.

Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2021. Growing Trees: The acquisition of the left periphery. *Glossa: a journal of general linguistics* 6(1). 31. https://doi.org/10.16995/glossa.5877.

Friedmann, Naama & Julia Reznick. 2021. Stages rather than ages in the acquisition of movement structures: Data from sentence repetition and 27696 spontaneous

clauses. *Glossa: a journal of general linguistics* 39(1). 143. https://doi.org/10.16995/glossa.5716.

Galasso, Joseph. 2003. *The Acquisition of Functional Categories: A Case Study*. Indiana University, IUCL Publications.

Gallego, Ángel. 2006. Phase sliding. Unpublished Ms., Universitat Autònoma de Barcelona.

Gallego, Ángel. 2010. *Phase Theory*. Amsterdam: John Benjamins. https://doi.org/10.1075/la.152.

Giorgi, Alessandra. 2010. *About the Speaker: Towards a Syntax of Indexicality*. Oxford: Oxford University Press. https://doi.org/10.1080/07268602.2013.838743.

Giurgea, Ion & Eva-Maria Remberger. 2016. Illocutionary force. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance Languages*, 863-878. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0053.

Grimshaw, Jane. 1991. Extended projection. Unpublished Ms., Brandeis University.

Haegeman, Liliane & Virginia Hill. 2013. The syntacticization of discourse. In Raffaella Folli, Christina Sevdali & Robert Truswell (eds.), *Syntax and its limits*, 370–390. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199683239.003.0018.

Harwood, William. 2013. *Being progressive is just a phase: dividing the functional hierarchy.* Ph.D. thesis, Ghent University.

Heim, Johannes & Martina Wiltschko. 2021. Acquiring the form and function of interaction: a comparison of the acquisition of sentence-final particles and tag questions in the Brown corpus. Talk presented at the *LAGB Annual Meeting* 2021, 8 September.

Hinzen, Wolfram & Martina Wiltschko. 2023. Modelling non-specific linguistic variation in cognitive disorders. *Journal of Linguistics* 59(1). 1–27. https://doi.org/10.1017/S0022226722000123.

Hudson-Kam, Carla L. & Elissa L. Newport. 2005. Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development* 1(2). 151-195. https://doi.org/10.1080/15475441.2005.9684215.

Lakoff, Robin. 1968. *Abstract Syntax and Latin Complementation*. Cambridge, MA: MIT Press.

Ledgeway, Adam. 2010. The clausal domain: CP structure and the left periphery. In Roberta D'Alessandro, Adam Ledgeway & Ian Roberts (eds.), *Syntactic Variation: The Dialects of Italy*, 38-51. Cambridge: Cambridge University Press.

Ledgeway, Adam. 2012. *From Latin to Romance: Morphosyntactic typology and change.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199584376.001.0001.

Ledgeway, Adam. 2015. Parallels in Romance nominal and clausal microvariation. *Revue roumaine de linguistique* LX(2-3). 105-127.

Ledgeway, Adam. 2020. The north-south divide: Parameters of variation in the clausal domain. *L'Italia Dialettale* 81. 29-77. https://doi.org/10.17863/CAM.45096.

Llinàs-Grau, Mireia & Javier Fernández-Sánchez. 2013. Reflexiones en torno a la supresión del complementante en inglés, español y catalán. *Revista de la Sociedad Española de Lingüística* 43. 55–88.

Llinàs-Grau, Mireia & Ana Isabel Ojea. 2000. PhonBank Spanish Llinàs/Ojea Corpus. Doi:10.21415/T5K01Q.

MacWhinney, Brian. 2000. The CHILDES project: The database. Mahwah, New Jersey: Lawren Erlbaum Associates, Inc. https://doi.org/10.4324/9781315805641.

Miyagawa, Shigeru. 2022. *Syntax in the Treetops*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/14421.001.0001.

Moscati, Vincenzo & Luigi Rizzi. 2021. The Layered Syntactic Structure of the Complementizer System: Functional Heads and Multiple Movements in the Early Left-Periphery. A Corpus Study on Italian. *Frontiers in Psychology*, 12. 627841. https://doi.org/10.3389/fpsyg.2021.627841.

Paradis, Johanne & Fred Genesee. 1997. On continuity and the emergence of functional categories in bilingual first-language acquisition. *Language Acquisition* 6(2). 91–124. https://doi.org/10.1207/s15327817la0602_1.

Perkins, Laurel & Jeffrey Lidz. 2021. Eighteen-month-old infants represent nonlocal syntactic dependencies. *PNAS* 118(41). https://doi.org/10.1073/pnas.2026469118.

Perkins, Laurel, Yuanfan Ying, Alexander Williams & Jeffrey Lidz. 2021. Object wh-questions with unknown verbs are transitive for 20-month-olds. Talk presented at *Boston University Conference on Language Development* 46 (Boston).

Poletto, Cecilia & Raffaella Zanuttini. Emphasis as reduplication: Evidence from *sì che/no che* sentences. *Lingua* 128. 124-141. https://doi.org/10.1016/j.lingua.2012.10.016.

Potts, Christopher & Tom Roeper. 2006. The narrowing acquisition path: From expressive small clauses to declaratives. In Ljiljana Progovac, Kate Paesani, Eugenia Casielles-Suárez & Ellen Barton (eds.), *The Syntax of Nonsententials: Multidisciplinary perspectives*, 183–201. Amsterdam: John Benjamins. https://doi.org/10.1075/la.93.09pot.

Prieto, Pilar & Gemma Rigau. 2007. The Syntax-Prosody Interface: Catalan interrogative sentences headed by *que. Journal of Portuguese Linguistics* 6(2). 29–59. https://doi.org/10.5334/jpl.139.

Prins, Karine. 2014. *Overt complementizers in root clauses in southern Italian dialects*. Master's thesis, Universiteit Leiden.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Radford, Andrew. 1988. Small children's small clauses. *Transactions of the Philological Society* 86(1). 1–43. https://doi.org/10.1111/j.1467-968x.1988.tb00391.x.

Ramchand, Gillian. 2023. The Symbolic Domain. Talk presented at *V-NYI 6 Winter School*, 23 January.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 46. 152–174. https://doi.org/10.1016/j.langsci.2014.06.013.

Rizzi, Luigi. 1993/1994. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3(4). 371–393. https://doi.org/10.1207/s15327817la0304_2.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-5420-8_7.

Rizzi, Luigi. 2004. Locality and Left Periphery. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, Volume 3, 104–131. Oxford: Oxford University Press.

Roeper, Tom. 2007. *The Prism of Grammar: How Child Language Illuminates Humanism*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5569.001.0001.

Roeper, Tom & Bernhard Rohrbacher. 1994. Null Subjects in Early Child English and the Theory of Economy of Projection. *IRCS Technical Reports Series* 162. https://doi.org/10.1007/978-94-017-3232-1_14.

Rosenbaum, P. 1967. Phrase structure principles of English complex sentence formation. *Journal of Linguistics* 3(1). 103–118. https://doi.org/10.1017/s0022226700012986.

Ross, John R. 1970. On declarative sentences. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 166–183. Waltham, MA: Ginn.

Serra, Miquel & Rosa Solé. 1989. CHILDES Catalan Serra/Solé Corpus. https://doi.org/10.21415/T58W3N.

Song, Chenchen. 2019. *On the Formal Flexibility of Syntactic Categories*. Ph.D. thesis, University of Cambridge.

Speas, Peggy & Carole Tenny. 2003. Configurational properties of point of view roles. In Anna Maria Di Sciullo (ed.), *Asymmetry in grammar*, 315–344. Amsterdam: John Benjamins. https://doi.org/10.1075/la.57.15spe.

Tonelli, Livia, Wolfgang Dressler & R. Romano. 1995. Frühstufen des Erwerbs der italienischen Konjugation. *Suvremena Linguistika* 21. 3-15.

Tsimpli, Ianthi-Maria. 2005. Peripheral positions in early Greek. In Melita Stavrou & Arhonto Terzi (eds.), *Advances in Greek generative syntax: In honor of Dimitra Theophanopoulou-Kontou*, 179–216. Amsterdam: John Benjamins. https://doi.org/10.1075/la.76.08tsi.

Tsimpli, Ianthi-Maria. 2014. Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. https://doi.org/10.1075/lab.4.3.01tsi.

Vainikka, Anne & Martha Young-Scholten. 2011. *The Acquisition of German: Introducing Organic Grammar*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110263848.

van Kampen, Jacqueline. 2009a. The Non-Biological Evolution of Grammar: Wh-Question Formation in Germanic. *Biolinguistics* 3(2-3). 154–185. https://doi.org/10.5964/bioling.8719.

van Kampen, Jacqueline. 2009b. What's so difficult about a wh-pronoun? Poster presented at *Relating Asymmetries between Speech & Comprehension in the Acquisition of Language* (Groningen), 24 January.

van Kampen, Jacqueline. 2010. Typological guidance in the acquisition of V2 Dutch. *Lingua* 120(2). 264–283. https://doi.org/10.1016/j.lingua.2008.07.005.

Vila, I. 1990. Adquisición y desarrollo del lenguaje. Barcelona: Graó.

Villa-García, Julio. 2012. Recomplementation and locality of movement in Spanish. *Probus: International Journal of Romance Linguistics* 24. 257–314. https://doi.org/10.1515/probus-2012-0011.

Villa-García, Julio. 2019. Recomplementation in English and Spanish: Spanish: Delineating the CP space". *Glossa: a journal of general linguistics* 4(1). 56. https://doi.org/10.5334/gjgl.845.

Wiltschko, Martina. 2014. *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139833899.

Wiltschko, Martina. 2021. *The Grammar of Interactional Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108693707.

Wurmbrand, Susi. 2013. Tagalog infinitives: Consequences for the theory of phases, voice marking, and extraction. Unpublished Ms., University of Connecticut, Storrs.
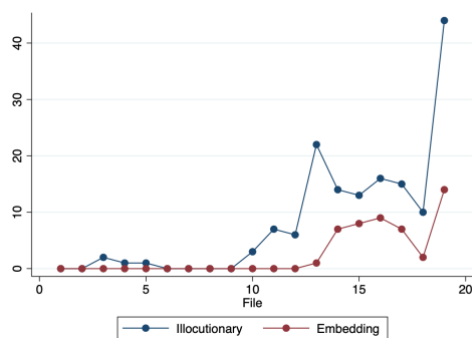
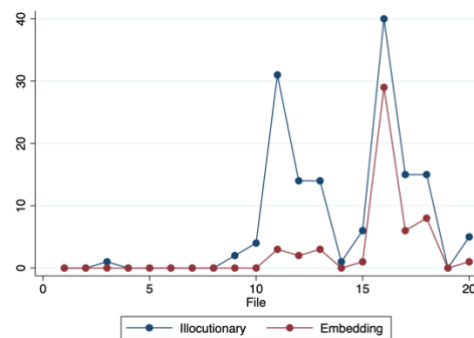## Appendix A: Additional graphs

Catalan

Figure A.1. Laura's development.
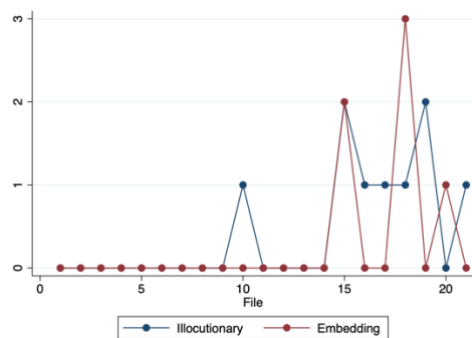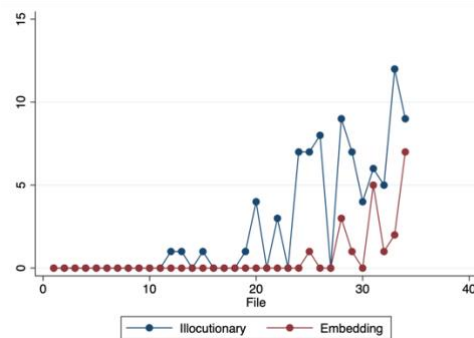

Figure A.2. Gisela's development.

Figure A.3. Àlvar's development.

Figure A.4. Guillem's development.
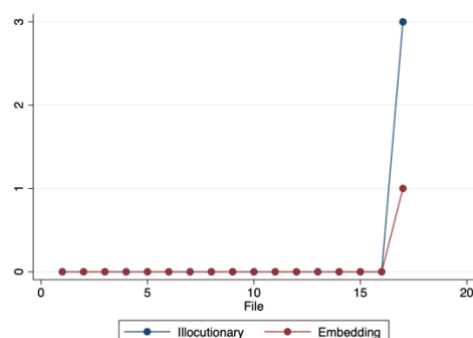
Figure A.5. Júlia's development.
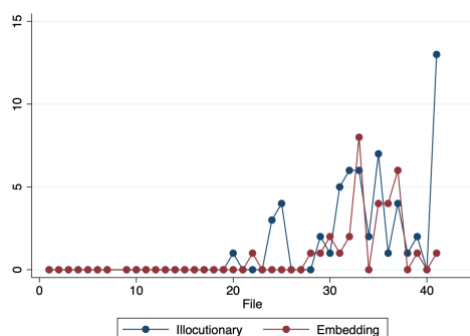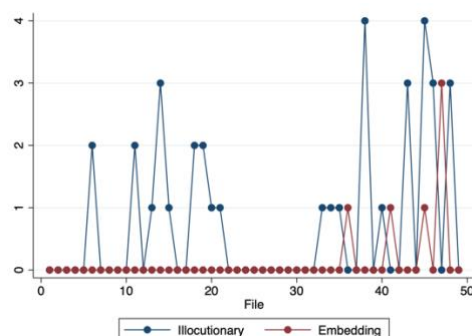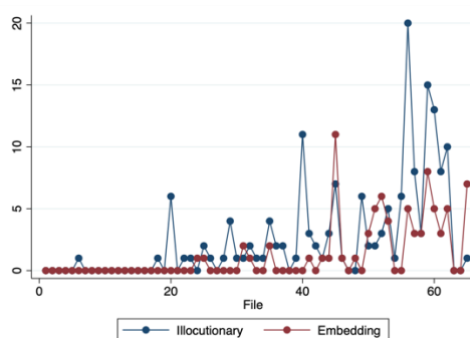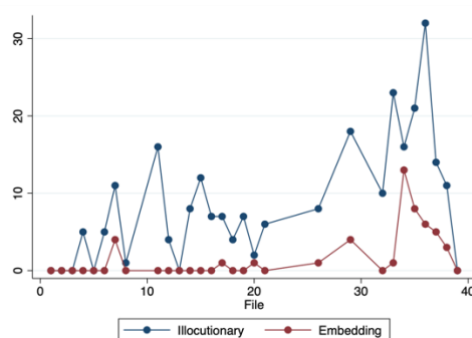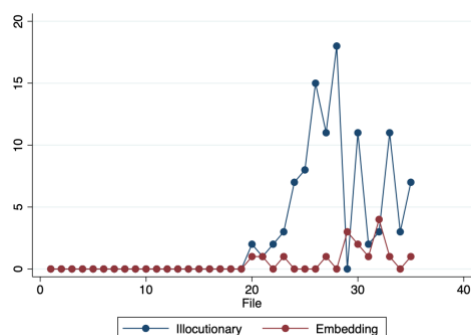
Spanish

Figure A.6. Irene's development.

Figure A.7. Yasmin's development.

Figure A.8. Juan's development.

Figure A.9. Magín's development.

Figure A.10. Emilio's development.