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


Abstract

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.

**Keywords:** syntax, acquisition, illocutionary complementisers, Catalan, Spanish, Ibero-Romance.

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 preconjunctinals, 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), 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 vP>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. **Catalan**

\[ \text{Li} \quad \text{he} \quad \text{dit} \quad \text{que} \quad \text{aquesta tarda} \quad \text{vaig} \quad \text{a} \quad \text{Barcelona} \]

\[ \text{CL.IO= AUX.1SG told that this afternoon go.1SG to Barcelona} \]

\(^1\) 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.
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. Spanish

No podia creer que hubiesen ganado la lotería

not can.IMPF.3SG believe that AUX.SUBJ.IMPF.3PL win.PTCP the lottery

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

(2) Illocutionary complementisers

a. Catalan, Corr 2016: 88

Ai, que t’atrapo!

hey that.EXCL CL.DO=catch.1SG

‘I’m coming to get you!’

b. Spanish

No hagas esto, que luego mama se enfada

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. Catalan
M’adono que encara no ha arribat
CL=realise.1SG that yet not AUX.3SG arrive.PTCP
‘I realise that he/she hasn’t arrived yet’

b. Spanish
Me dijiste que se había arrepentido
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. Catalan
   Ala, *que* ho has llençat tot al terra!
   hey that.EXCL.CL.DO= AUX.2SG throw.PTCP everything on.the floor
   ‘Hey! You’ve thrown everything on the floor!’

b. Spanish
   ¡*Que* la abuela Rosa me ha comprado una
   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)

Catalan
   Context: the speaker is asked who had just phoned.
   Era la Carme. *Que* em trucava per
   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’

Spanish
   A: No entiendo qué acabas de decir
   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:
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. Catalan
No li diguis això a la Paula que és un secret
‘Don’t tell this to Paula because it’s a secret’

b. Spanish
Dame el diario, por favor, que siempre me gusta repasarlo antes de ir a dormir
‘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 Spanish⁴:

(9) Catalan
Que vindrás al final a veure la pel·lícula?
‘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 vP 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

![Diagram](source)

**Source:** 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)\. 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 decision will not substantively impinge on the conclusions of the study. For consistency, as well, I will adopt Corr’s (2016) terminology throughout\. 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

---

\[5\] 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*.).

\[6\] 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.
Not all complementisers are late

Isogloss 2023, 9(1)/12

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 (labelledSentienceP, 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. Catalan
   Evidentment **que** estudiaré per a aquest examen!
   ‘I’ll obviously study for this exam!’

b. Spanish
   ¡Claro **que** entendió!
   ‘Of course he understood!’

(11)

a. Catalan
   Sí **que** val la pena, tenies raó
   ‘It certainly is worth it, you were right’

b. Spanish
   Laura sí **que** ha vuelto de trabajar
   ‘Laura has certainly come back from work’

(12) Catalan, Batllori & Hernanz (2013: 27)
   No **que** no ha cantat la soprano
   ‘No, the soprano has NOT sung’

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):
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
Not all complementisers are late, 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

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. 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.

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 vP) 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, realisationally (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-externally (i.e. the ‘dynamic’ 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 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) >

---

8 See so-called dynamic or contextual approaches to phases (i.a., den Dikken 2007, Gallego 2006, 2010, Harwood 2013, Wurmbrand 2013, Bošković 2014).
semantically distinct head (as in Cinqu 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:

\[
(16)
\]

\[
\begin{array}{c}
\pm V \\
- (\equiv N) \\
+ (\equiv V) \\
\hline \\
n \\
D \\
v \\
C \\
\hline \\
\text{Num} \\
\text{n} \\
\text{Q} \\
\text{N} \\
\text{Asp} \\
v \\
\text{T} \\
\text{C} \\
\hline \\
\text{Extended Proj.} \\
\text{Phase} \\
\text{CFC}
\end{array}
\]

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.

\(^9\) 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.

\(^{10}\) 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).
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.

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 (vP-internal) material. 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 & Wittschko’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.

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.

---

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.
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) *si/no que* (*yes/no that*)

2. **Subordinating complementisers**
   a. Complement clauses
   b. Relative clauses introduced by *que* (including free relatives12)

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 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.

---

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’).
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

---

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

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

2014 and references therein). Detailed coding of these subtypes would be a welcome avenue of future work, however.
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

<table>
<thead>
<tr>
<th>Language</th>
<th>Corpus</th>
<th>Children</th>
<th>Age range</th>
<th>Files analysed</th>
<th>MLUw range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>Serra/Solé</td>
<td>Laura</td>
<td>1;07-4:00</td>
<td>19</td>
<td>1.03-3.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gisela</td>
<td>1;04-4:02</td>
<td>20</td>
<td>1.02-3.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Àlvar</td>
<td>1;02-3:01</td>
<td>21</td>
<td>1.0-3.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guillem</td>
<td>1;01-4:00</td>
<td>34</td>
<td>1.01-3.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Júlia</td>
<td>1;07-2:06</td>
<td>17</td>
<td>1.15-2.74</td>
</tr>
<tr>
<td>Spanish</td>
<td>Llinàs/Ojea</td>
<td>Irene</td>
<td>0;11-3:02</td>
<td>40</td>
<td>1.0-4.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yasmine</td>
<td>1;10-2:09</td>
<td>47</td>
<td>1.29-3.21</td>
</tr>
<tr>
<td></td>
<td>Aguado-Orea/Pine</td>
<td>Juan</td>
<td>1;10-2:05</td>
<td>65</td>
<td>1.34-3.39</td>
</tr>
<tr>
<td></td>
<td>Aguirre</td>
<td>Magín</td>
<td>1;07-2:10</td>
<td>29</td>
<td>1.24-3.07</td>
</tr>
<tr>
<td></td>
<td>Vila</td>
<td>Emilio</td>
<td>0;11-4:08</td>
<td>35</td>
<td>1.0-3.23</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1009 (76.6%)</td>
</tr>
</tbody>
</table>

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

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

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?**
   that.INT already not make.3SG pain
   ‘Does it not hurt anymore?’

b. **Ai, que crema!**
   ouch that.EXCL burn.3SG
   ‘Ouch, it’s burning!’

c. **Espera’t, que estic fent això**
   wait=CL.REFL that.CONJ am doing this
   ‘Wait, I’m doing this’

d. **Que no hi és**
   that.QUOT not CL.LOC= is
   ‘(I’ve already said) it’s not there’

(18)

a. **Que no qu(i)ero**
   that.QUOT not want.1SG
   ‘(I said) I don’t want to’

b. **¡Que viene!**
   that.EXCL come.3SG
   ‘He/she is coming!’

c. **Ay, no, que me harán daño**
   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\(^\text{13}\)).

\(^{13}\) 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.
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\textsuperscript{14}:

\begin{tabular}{|l|}
\hline
\textbf{MLUw} & \textbf{Children} & \\
\hline
2.61 & Gisela & \\
2.74 & Júlia & \\
2.82 & Álvar & \\
2.47 & Yasmin & \\
3.23 & Irene & \\
2.73 & Juan & \\
\hline
\end{tabular}

\textsuperscript{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

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

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Illocutionary</th>
<th>Embedding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>1.41 MLUw (range 1.13-1.84)</td>
<td>2.35 MLUw (range 1.91-2.61)</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.87 MLUw (range 1.58-2.18)</td>
<td>2.49 MLUw (range 1.77-3.28)</td>
</tr>
<tr>
<td>Combined</td>
<td>1.67 MLUw (range 1.13-2.18)</td>
<td>2.42 MLUw (range 1.77-3.28)</td>
</tr>
</tbody>
</table>

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 grey 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

<table>
<thead>
<tr>
<th>Unacc</th>
<th>Unerg</th>
<th>Trans</th>
<th>Modal</th>
<th>Copula</th>
<th>Impers</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laura</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>20</td>
</tr>
<tr>
<td>Gisela</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>7</td>
</tr>
<tr>
<td>Àlvar</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1</td>
</tr>
<tr>
<td>Guillem</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11</td>
</tr>
<tr>
<td>Júlia</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Irene</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1</td>
</tr>
<tr>
<td>Yasmin</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>18</td>
</tr>
<tr>
<td>Juan</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>10</td>
</tr>
<tr>
<td>Magín</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>10</td>
</tr>
<tr>
<td>Emilio</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

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 1.
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, Tsimpili 2005, Roepen 2007, van Kampen 2010, Biberauer & Roberts 2015, Biberauer 2018, Heim & Witschko 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 MLU 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. 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

---

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.
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 Wiltshko’s Linking domain (and the child encodes them as such), their early emergence might be at odds with the Inward Growing Spine Hypothesis. A 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 & Wiltshko (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 crosslinguistic 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
   that.turn.3SG
   ‘He/she/it stirs (it)’

b. **Che** legge
   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!
   that.EXCL laugh.3SG
   ‘He/she is laughing!’

d. **Che** piove
   that.CONJ rain.3SG
   ‘It’s raining’ (in response to *l’ombrell*, ‘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; it appears to indicate either surprise or exclamation, but it could also have an imperative reading (see footnote 16 for conversational context).

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 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.

Example (21a) was uttered in the following conversational context (my own translations):

*MOT: devo girare il caffè (‘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)’)

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., si che lo sai! lit. ‘yes that it know!’, ‘yes you know it!’), wh-exclamatives (e.g., che bella! ‘so nice!’), 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.
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
   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
   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. (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’). 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

---

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.
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).

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 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 (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.

Although the equivalent of (21b) is only marginally acceptable in Catalan in that context (?Que llegex, ‘he/she is reading’; my own judgements).

These overgeneralisations may later, if re-enforced in peer-to-peer contexts, lead to diachronic change (see Courmane 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 Courmane 2017: 10); therefore, terms such as ‘input-divergent’ (following Courmane 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 vP 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

Immense thanks go to Theresa Biberauer for supervising my MPhil thesis and the small side-project from which this work stems, and for her significant aid with this paper. Thanks too to Adam Ledgeway for answering some of my questions and recommending useful references. I am also grateful to Sara Cardullo, Marco Fioratti, Elena Isolani, Leonardo Russo Cardona and Giusy Truncellito for providing native speaker judgements at various points. Finally, thank you to reviewers of Isogloss, 2023 LAGB Annual Meeting and BCGL 16 for very helpful comments. Any remaining errors are entirely my own. This research was generously supported by a Benefactors’ Scholarship (Hayes Scholarship) from St John’s College and a Cambridge Opportunity Master’s Studentship from the Cambridge Trust.

References


Clahsen, Harald & Martina Penke. 1992. The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the


Heim, Johannes & Martina Wiltschko. 2021. Acquiring the form and function of


van Kampen, Jacqueline. 2009a. The Non-Biological Evolution of Grammar: Wh-
Not all complementisers are late.


