

Categorical Acquisition by Differentiation: the acquisition of the left periphery and implications for categorisation

A long-standing question in generative typology remains what is language-universal and language-variant in the categorial sequences that characterise human language(s). Cartographic work has advocated the empirical necessity of a rich hierarchy of ordered projections that is universally both *fixed* and *fine-grained* crosslinguistically (Rizzi, 1997; Cinque, 1999). This paper draws on a recent research strand arguing for crosslinguistically variable *degrees* of elaboration of the left periphery (i.a., Giorgi & Pianesi, 1997; Biberauer & Roberts, 2015; Walkden, 2017; Hsu, 2017, for morphosyntactic typology; Cournane & Klævik-Pettersen, 2023, for diachrony). We propose the working hypothesis for category acquisition in (1), following Biberauer & Roberts (2015) and Douglas (2024), and we endorse its plausibility from the perspective of the L1 acquisition of left-peripheral structure.

- (1) **Categorical Acquisition by Differentiation (CAD):** Syntactic categories *granularise* during development. Acquisition proceeds such that coarser-grained, featurally-simpler categories are acquired first, with later, finer-grained distinctions elaborating on developmentally-prior structure.

Our empirical case for (1) is two-part:

§1. We centre on the results of a **multilingual corpus study** on 10 children across 5 languages (Catalan, Italian, Spanish, German and Dutch) in CHILDES, probing the acquisition of functional categories. Specifically, the study compares the emergence of (single) CP-structures to the emergence of structures specifically showing that the child has command of a more articulated Split CP domain (henceforth, ‘Split CP-structures’). Split CP-structure was identified via utterances with co-occurring left-peripheral elements, as listed in **Table 1**.

CP-structures	V-to-C movement (Germanic only) ♦ Wh-Qs ♦ Yes/No-Qs (Germanic) ♦ Topicalisations and focalisations ♦ Illocutionary complementisers (Romance) ♦ Finite embedding
Split CP-structures	Romance: Topic > Wh-word ♦ Top > Top/Foc ♦ Complementiser > Wh/Top ♦ Topic > interrogative <i>que</i> ‘that’ (Catalan only) ♦ (Topic +) <i>Si que/sì che</i> ‘yes that’ Germanic: V3 constructions – Frame-setters, Hanging Topic/Contrastive Left-Dislocation ♦ Conditional/temporal clauses with resumptive <i>dann/dan</i> ‘then’

Table 1. Summary of structural diagnostics used

A key contrast surfaces: (i) (some) CP-structures emerge very early (average MLUw ~1.5, 21 months), including left-peripherally very high elements (illocutionary complementisers in Ibero-Romance and topics; see, i.a., the data and cartographic analyses in Rizzi, 1997; Corr, 2016). However, (ii) there is no production-evidence for further internal elaboration within CP (Split CP-structures) until a *later* stage (see **Table 2** for one child). The emergence of the latter is moreover sudden and ‘explosive’: no Split CP-structures or extremely few are detectable before average MLUw ~2.5, 31 months; after this point, they abruptly increase in quantity in a highly statistically significant manner (**Fig. 1**). Fixed-effects logistic regression also established that (ii) is unlikely to be (entirely) due to performance factors like the relative length of CP vs Split CP-structures. MLUw or age are very strong predictors ($p < .001$), while length is non-significant ($p = 0.563$).

Age	MLUw	Wh-Q	Top/Foc	Illoc	Embed	Split CP
1;07.18	1.26					
1;08.02	1.9	✓				✓
1;08.17	1.57	✓				
1;09.01	1.59					
1;10.29	1.66	✓	✓			
1;11.02	1.99	✓		✓		
1;11.20	1.99	✓	✓		✓	
2;01.12	1.86	✓	✓	✓		
2;03.01	2.55	✓		✓	✓	
2;03.22	2.64	✓		✓	✓	
2;04.13	2.69	✓	✓	✓	✓	✓
2;05.21	2.37	✓	✓	✓	✓	✓
2;07.15	2.55	✓	✓	✓	✓	✓

Table 2: Production of structures by Martina (Calambon corpus; Italian)

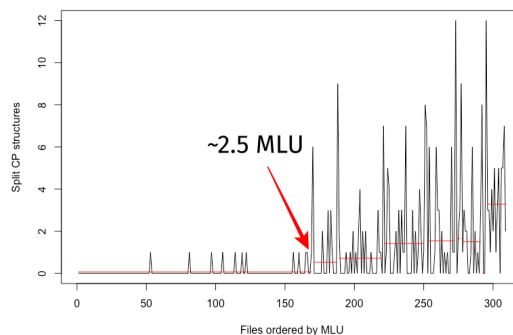


Figure 1: Production of Split CP-structures with change-point analysis (all files, 10 children)

These two empirical generalisations, we argue, are incompatible with extant theories of syntactic development. Early emergence of CP-structures, irrespective of their structural height, undermines the

predictions of bottom-up maturational approaches, especially *cartographic* ones, which expect universally late development of (part of) CP. Furthermore, maturational and continuity approaches (i.a., Radford, 1990, Poeppel & Wexler, 1993, and subsequent work; Westergaard, 2009) generally assume a (single) UG-given categorial sequence of *fixed*, often *cartographic*, granularity, viz. Friedmann et al.’s (2021) Growing Trees Hypothesis. All things equal, we predict we should observe evidence in child data of this fine-grained cartographic structure once CP develops. This fixed-granularity commitment renders the apparent ‘shift’ in categorial granularity in our data elusive. By contrast, we propose these results are suggestive for categorisation; they stress the analytical strengths of incorporating *changes in granularity* as part of syntactic development. We interpret the patterns under *neo-emergentist* generative approaches. Biberauer & Roberts (2015) postulate an *emergent categorial hierarchy* in which early grammars first access core ‘macroparametric’ structural properties – a ‘basic’/macro-level CP –, leading later to ‘microparametric’ refinements into finer-grained, potentially (part-)cartographic, CP-structure.

§2. We then briefly discuss how the differentiation logic in (1) also makes correct predictions for the acquisition of *parametric distinctions* in the emergence of topicalisation crosslinguistically. We summarise a **second corpus study** on **7 bilinguals** acquiring Italian/Dutch, Italian/German and Spanish/German. Its results corroborate §1, but provide further nuance regarding crosslinguistically universal and variable developmental stages: (i) (some) CP-structures again emerge early across *all* children/languages; but (ii) topic-development *varies* across languages, Germanic topics consistently emerge well before Romance ones (e.g., CLLD), with a respective average of MLUw 1.88 vs 3.11 and 27 vs 33 months. Zooming in on (ii), we compare the results against **monolingual topic-acquisition data** from **10+ typologically-diverse languages**. We establish that the development of topics (early vs late) systematically varies as a function of the *formal complexity* (e.g., operator/non-operator) of each L1’s topicalisation strategies (Table 3). We schematise the patterns as a crosslinguistic acquisition hierarchy of categories of topics; Fig. 2 shows how formally and featurally more intricate divisions correlate, without exception, to late acquisition as in Table 3.

Language	Acquisition	Source	Formal characteristics	Parametric complexity
French	Very Early	De Cat (2000, et seq.)	Adjoined or base-generated	Macroparametric
Germanic V2	Early	This talk, i.a.	Generalised V2 diacritic	Mesoparameter
Mandarin	Early	Zhu & Gavarró (2019)	Operator movement or base-generation	Mesoparameter
Japanese		Kurumada (2009)		
Korean		Lee (2001)		
European Portuguese	Early (non-CLLD topics only)	Soares (2003)	Operator movement	Mesoparameter
Catalan	Late	Bosch (2023)	Non-operator movement with CLLD	Microparametric
Spanish/Italian		This talk, i.a.		
Greek		Marinis (2001) Tsimplli (2005)		
Hebrew	Late	Friedmann et al. (2021)	Non-operator movement without CLLD	Microparametric
Brazilian Portuguese		Meira & Grolla (2023)		

Table 3. Topicalisation strategies: languages studied, earliness vs lateness of acquisition, and formal complexity.

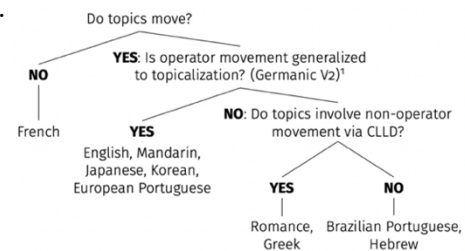


Figure 2. Formal complexity of topicalisation structures.

These findings bolster the arguments in §1, in two ways. First, we conclude that crosslinguistic topic-acquisition is *not* maturationally pre-wired as late (*pace* Radford, 1990; Rizzi, 1993/4; Friedmann et al., 2021), but rather ‘tracks’ the *formal complexity* of topicalisation strategies in each L1. Second, the crosslinguistic, *constrained* ‘flexibility’ observed in the emergence timings of topics is anticipated under CAD; coarser-grained categories, or kinds, of topics are systematically acquired first.

Overall, this paper has introduced an underexplored perspective on syntactic and categorial development – CAD. We have presented two case-studies, on functional categories and parametric distinctions, to endorse a role for differentiation in grammar construction. We furthermore argued that its adoption provides higher empirical coverage than other existing accounts. These results challenge the long-held assumption that the granularity of syntactic categorial systems remains both fixed and fine-grained throughout development, foregrounding instead the potential *hierarchical* nature of syntactic categorisation. Albeit programmatic, this work has implications for requirements on theories of functional category acquisition, the role of differentiation in human language, the crosslinguistic ‘flexibility’ of learning paths, and the ontological basis of linguistic categories more broadly.

References (selected): Biberauer & Roberts (2015) Rethinking formal hierarchies: a proposed unification. *Cambridge Occasional Papers in Linguistics*; Douglas (2024) Exploring Emergence with Substance-Free Categories. *Cambridge Occasional Papers in Linguistics*; Friedmann, Belletti & Rizzi (2021) Growing Trees: The acquisition of the left periphery. *Glossa*; Westergaard (2009) *The Acquisition of Word Order*. John Benjamins.

¹ In Germanic, operator topics fall out from its generalised V2 system, unlike the other languages considered, hence its parenthetical placement in Figure 2.