

Wh-word acquisition in Czech: Exploring the growing trees hypothesis*

Radek Šimík (Charles University)

Klára Matiasovitsová (Charles University)

Filip Smolík (Institute of Psychology, The Czech Academy of Sciences)

June 2024 • to appear in *Language Acquisition*

1 Introduction

Wh-words like *what*, *where*, *when*, or *which* are polyfunctional. Besides indicating the locus of ignorance in wh-interrogatives (1a) (I), they can figure in the formation of conditionals (1b) (C) and can play the role of relative pronouns in various kinds of relative clauses, including so-called free relative clauses (1c) (FR), light-headed relative clauses, i.e. relative clauses headed by pronominals (1d) (LHR), and standard relative clauses headed by a nominal phrase (1e) (HR). In addition, wh-words commonly serve as morphological bases for indefinite and quantificational expressions (1f).

- | | | | |
|-----|----|---|-----|
| (1) | a. | <i>What</i> did they discuss? | I |
| | b. | <i>When</i> I see him, (then) I'll let you know. | C |
| | c. | We spent our vacation <i>where</i> I grew up. | FR |
| | d. | This is something <i>which</i> is very easy to do. | LHR |
| | e. | I visited the hospital <i>where</i> the victims were treated. | HR |
| | f. | <i>somewhere, everywhere, wherever</i> | |

In this paper we ask how the various wh-functions are acquired by children learning Czech, a language with highly polyfunctional wh-words. The null hypothesis is that the acquisition is driven by adult input: the functions that are more frequent in adult input get acquired more quickly and easily than less frequent functions. A competing hypothesis, which we call the GROWING TREES HYPOTHESIS (after Friedmann et al. 2021), is that the acquisition proceeds along with the increasing morphosyntactic and semantic complexity implicated in the individual functions – starting from interrogatives, going through conditionals, correlatives, and free relatives, and ending with headed relatives. We approach the research question by investigating a longitudinal corpus of child and child-directed

*The research was supported by the Czech Science Foundation (GAČR) via project 21-31488J and in its earlier stages also by institutional funding (Charles University and its Faculty of Arts), specifically via PRIMUS/19/HUM/008 and Cooperaciones Linguistics. We are grateful to the following people for their feedback or technical support (in alphabetic order): Eliška Belinová, Anna Chromá, Marko Hladnik, Olga Nádvorníková, Adam Pospíšil, and Jakub Sláma. Early versions of this work were presented at Formal Approaches to Slavic Linguistics (FASL) 30 (Massachusetts Institute of Technology, 2021), at the colloquium of the Institute of the Czech National Corpus (Charles University, 2021), and at the linguistic colloquium at the University of Regensburg (2023). We are grateful to the audiences for their feedback.

speech (Chromá et al. under review). We analyze the frequencies of relevant wh-functions and address the hypotheses by investigating the effect of adult production on forward-shifted child production, relying on the concept of “Granger causality” (Granger 1969). We find no straightforward effect of adult input, in the sense that adult production of some wh-function would foster child production of that function. Instead, what we see is that adult production of syntactically simpler wh-functions foster the child production of more complex functions. We take this finding to be consistent with the growing trees hypothesis. In addition to this, we consider a few exploratory issues, such as the acquisition of different wh-sorts or the issue of the subject–object asymmetry. The results are largely in line with previous findings for other languages, while also providing hitherto unnoticed patterns; more particularly, the evidence for a subject–object asymmetry in children’s production is evident for ‘what’, but not for ‘who’. Finally, our paper reveals certain methodological limitations of a corpus-based approach to our main research question. More specifically, in order to evaluate some alternative hypotheses, the sample would need to be significantly larger than the 10+ thousand wh-words that we have examined, ideally containing more data from more children–adult parents. The value of our paper is thus not only theoretical, but also methodological.

The paper is organized as follows. Section 2 introduces the wh-complexity hierarchy, provides some grammatical evidence in its favor, and postulates our working hypothesis – the growing trees hypothesis, which is closely tied to the complexity hierarchy. Section 3 summarizes existing research about the age and order of L1-acquisition of the relevant wh-constructions. We will see that partial studies exist, but none that aims at a more comprehensive view. Our corpus methodology is described in section 4. The descriptive and inferential (statistical model-based) results of the corpus study are provided in section 5. Section 6 discusses the results and section 7 concludes.

2 The complexity of wh-functions

Research on the morphosyntax and semantics of wh-constructions has revealed many non-trivial generalizations which can be succinctly expressed by the complexity hierarchy in (2): the complexity of the wh-function grows from interrogatives (I), through conditionals/correlatives (C), free relatives (FR), to light-headed relatives (LHR) and headed relatives (HR).

- (2) **Wh-complexity hierarchy** (cf. Šimík in press)
 $I < C < FR < LHR < HR$

A morphosyntactic and semantic analysis of the wh-complexity hierarchy, as well as empirical evidence in its favor is offered in greater detail in Šimík (in press). In what follows, we only summarize the crucial points. Selected empirical evidence for the hierarchy is discussed in §2.1. Its morphosyntactic and semantic underpinnings are provided in §2.2. We finish by setting up the growing trees hypothesis (§2.3).

2.1 Evidence for the complexity hierarchy

Cross-linguistic distribution Most languages of the world have a class of “wh-words”.¹ The class is often characterized by a dedicated “wh-morpheme” (German *w-*, Slavic *k-*,

¹Exceptions to this are extremely scarce; see e.g. Arkadiev & Caponigro 2021.

Greek *p-*, etc.), but not always (cf. Japanese: *dare* ‘who’, *nani* ‘what’, *itu* ‘when’). What ultimately defines the class is the interrogative function: wh-words are primarily, and in many languages *only* interrogative (e.g. Welsh or Xhosa; see Haspelmath 1997).

In languages where wh-words have non-interrogative functions, these functions are not assigned in an accidental fashion. Lehmann (1984) noticed that there seems to be no language that uses wh-words in HRs without also using them in FRs or LHRs. Languages with wh-based FRs/LHRs without productive wh-based HRs are attested; examples include Arabic varieties (Alqurashi 2013) or Mesamerican languages (Kotek & Erlewine 2019; Caponigro et al. 2020). An analogous implication seems to hold of FRs/LHRs and Cs: If a language uses wh-words for the former, it also uses them for at least some kind of conditionals or correlatives. Most languages of the broader European region productively use correlatives (see, e.g., Lipták 2009). While English might seem as an exception, it has wh-based unconditionals (Zaefner 1991; Rawlins 2013), also called concessive conditionals (Haspelmath & König 1998), which can be considered a subtype of the broader class of conditional/correlative constructions.² Again, the implication is unidirectional: there are languages which use wh-words for the interrogative and the correlative functions without using them for free or headed relatives. Examples include Chinese (cf. so-called wh-conditionals; Liu 2018) or Turkish (Demirok 2017b). In summary, the cross-linguistic availability of wh-constructions is limited by the complexity hierarchy in (2): if a language uses wh-words for some function, it also uses them for all less complex functions, but not necessarily conversely.³

Syntactic position of wh-words Two basic syntactic positions of wh-words are distinguished: While some languages generally place their wh-words in the clause-initial (or more generally a non-canonical) position – so-called wh-ex-situ – others place their wh-words in the same position where one would expect to find their non-wh-counterparts – so-called wh-in-situ. Wh-in-situ is very common in interrogatives (Dryer 2013) and can be productive in correlatives (in Chinese and Turkish, for instance). Yet it is virtually unattested in the more complex functions – FRs and (L)HRs; see Cecchetto & Donati (2015) or Demirok (2017b) for the case of FRs and Schwartz (1971) or de Vries (2005) for HRs.⁴ It follows that exclusively wh-in-situ languages (such as Chinese) never use wh-words for forming FRs or HRs and also that in languages where wh-in-situ vs. ex-situ is optional in interrogatives (such as French), only the ex-situ strategy is available for FRs/HRs. Finally, it holds that wh-words in more complex functions tend to be structurally higher (linearly more to the left) within their clause. More precisely, it holds that an ex-situ wh-word in some construction in (2) is at least as structurally high as (and possibly higher than) the corresponding wh-word in any less complex construction.

²In some languages, e.g. in Turkish (see Demirok 2017a), correlatives and unconditionals are difficult to tell apart.

³There are reasons to believe that the generalization should be formulated not at the level of languages, but individual wh-words. For instance, the English wh-determiner *which* can be used in interrogatives, unconditionals, as well as free and (light-)headed relatives, whereas *what* can only be used up to the level of free relatives. For a discussion of this generalization, see Caponigro (2003) and Šimík (2021).

⁴Potential exceptions to this generalization include Hindi, where wh-in-situ is claimed to exist in headed relatives, but – crucially – only in appositive relatives (Lehmann 1984), Tsez, which appears to have wh-in-situ in free relatives of the ‘ever’ kind (Polinsky 2015), which, however, may turn out to be closely related to unconditionals (cf. Hirsch 2016; Šimík 2018), and Hittite (Huggard 2011, 2015; Lyutikova & Sideltsev to appear). Even if these “exceptions” were real, the generalization still remains a strong tendency.

In Hungarian or Italian, for instance, interrogative wh-words follow topics, but relative wh-words precede them; see É. Kiss (2002) and Benincà (2001), respectively.

Morphology of wh-words English is an example of a language where interrogative wh-words can be used in more complex functions without any morphological change (cf. the use of *which* in interrogatives and headed relatives). This contrasts with languages like Greek, Bulgarian, or Hungarian, in which more complex functions of wh-words are accompanied by more complex morphology. In Bulgarian, for instance, the interrogative wh-word *koj* ‘who/which’ corresponds to the relative wh-word *kojto* – obligatorily accompanied by the postfix *-to*. In line with the hierarchy in (2), it holds that a wh-word in some function is at least as morphologically complex as (and possibly more complex than) the wh-word in a less complex function. Hungarian, for instance, uses the prefix *a-* to derive relative wh-words from interrogative ones; cf. *ki* vs. *aki* ‘who’. The “bare” form is used in interrogatives, unconditionals (where it is augmented by the ‘ever’ morpheme *akár*), and to a very limited extent in correlatives (É. Kiss 2002); the prefixed form becomes obligatory from the level of FRs, i.e., it is used for FRs, LHRs, and HRs. Greek exhibits a more complex pattern, with the *p*-based wh-words (e.g. *pjos* ‘who’) receiving the *o*-prefix from the level of unconditionals and correlatives (*opjos*) and becoming even more complex in HRs (*o opíos*); see, e.g., Daskalaki (2007, 2020).

Summary The wh-complexity hierarchy is reflected in distributional, syntactic, and morphological patterns across languages. The likelihood of taking up a particular wh-function decreases with the function complexity. The complexity further correlates with the syntactic position of the wh-word – the more complex the function, the higher the position of the wh-word. And finally, the complexity is reflected in wh-morphology.⁵ We now turn to the question of what underlies the hierarchy.

2.2 Sources of the wh-complexity hierarchy

We assume that the wh-complexity hierarchy has its source in increasing semantic and morphosyntactic complexity which can be represented by standard formal-semantic and syntactic tools. This complexity is in principle compatible with cognitive complexity; by hypothesis, richer syntactic and semantic structure requires more cognitive resources to be processed and produced. Also, the source of the syntactic and semantic complexity might be cognitive at its roots and could correspond to the growing understanding of the surrounding world (cf. Cinque & Rizzi 2010).⁶

Semantic underpinnings Interrogatives are the simplest constructions because they can be represented by propositions, or more particularly open propositions, with the wh-

⁵There are indications that the hierarchy is also implicated in the diachronic development of wh-constructions; see Lehmann (1977); Mitrenina (2012); Truswell & Gisborne (2015); Meyer (2017); Belyaev & Haug (2020).

⁶An anonymous reviewer wonders whether pragmatic complexity could also be at stake. We think that there are reasons to be skeptical about this: the canonical pragmatic function of interrogatives (the lowest member of the hierarchy) is posing a question, which appears to be a fairly complex conversational move (realizing one’s ignorance and expressing a wish to reduce that ignorance by making use of the knowledge of a conversational partner); the canonical function of a headed relative clause (the highest member of the hierarchy), on the other hand, is mostly just semantic (intersective modification), without any complex pragmatics being implicated.

word corresponding to a free variable whose value is at stake in the question.⁷

- (3) Who works hard?
 \rightsquigarrow [x works hard] $\boxed{\phi[x]}$

A similar analysis is available for correlatives, with the difference that the wh-variable is no longer free, but bound by a correlative operator. This binding and coindexing with the anaphoric element in the correlative consequent (*der* in (4)) increases the complexity of the wh-construction.

- (4) Wer fleißig arbeitet, der muss keine Angst haben. (German)
 who hard works that must no fear have
 ‘Whoever works hard doesn’t need to worry.’
 \rightsquigarrow for any x , if [x works hard], x need not worry $\forall x \boxed{\phi[x]} \rightarrow \psi[x]$

This simple proposition-based analysis is not available for the more complex relative constructions, which participate in the denotation of *individuals* rather than propositions. In order for a wh-clause to (co-)determine the denotation of a noun phrase, it must not denote (open) propositions but rather properties of entities. The standard way of formally capturing a property is by a function (lambda-term) which characterizes a set of entities with that property. It is possible to think of the ex-situ wh-word in relative clauses as being involved in triggering the lambda abstraction, generating the property from the open proposition (Heim & Kratzer 1998).

- (5) Der, wer fleißig arbeitet, muss keine Angst haben. (German)
 that who hard works must no fear have
 ‘Those who work hard don’t need to worry.’
 \rightsquigarrow those [λx x work hard] need not worry DET (NP) $\boxed{\lambda x \phi[x]}$

What happens with the resulting property depends on the kind of relative construction. In FRs, the property is shifted to the maximal entity with that property (in the spirit of Jacobson’s 1995 influential proposal; see Šimík 2021). This shift is standardly considered the default operation deriving entities from properties (Partee 1987; Caponigro 2003; Dayal 2004).⁸ In LHRs, the property is used as the (quantificational) restrictor of the determiner or pronoun heading the relative clause (as illustrated below; see also Citko 2004). The additional step of using an overt determiner or pronoun corresponds to the increased complexity. The most complex situation obtains in HRs, where the relative clause not only restricts a determiner, but incorporates another restriction in the form of the relative NP head. Semantic research on this topic has shown this to be a highly non-trivial matter; see, e.g., Hulseley & Sauerland (2006).

⁷This analysis, defended e.g. by Berman (1991), can be seen as a simplified version of various other proposition-based analyses, including Hamblin’s (1973) influential analysis (assuming sets of propositions) or Beck’s (2006) (assuming an additional question operator binding the wh-variable). The semantics of interrogatives has been extremely well explored and many different proposals have been made. For recent surveys, see Krifka (2011); Cross & Roelofsen (2014); Roelofsen (2019); Hagstrom (2021).

⁸As demonstrated experimentally in Caponigro et al. (2012), it takes time for children to acquire adult-like (i.e., maximalizing) interpretation of FRs – they only do so around 6–7 years of age – despite their relatively early acquisition (see below). It is therefore possible that pre-school children apply a different type-shifting mechanism than adults, e.g. the choice function (see, e.g., von Stechow 1997 for some discussion).

Morphosyntactic complexity The increase in semantic complexity discussed above is mirrored in the morphosyntactic representation of the individual wh-constructions and wh-words. This is well-documented esp. in the so-called cartographic approach to syntax (Rizzi 1997), where scholars have demonstrated that wh-interrogatives are syntactically smaller than relative clauses. The kind of evidence supporting this is in (6), which shows that while interrogative wh-phrases obligatorily follow left-peripheral topics (6a), relative wh-phrases precede them (6b).

- (6) a. Con Mario, *di* *che cosa* volevi parlare?
with Mario about what wanted talk
‘What did you want to talk about with Mario?’
b. il ragazzo *a cui* il libro lo porterò domani
the boy to whom the book it bring tomorrow
‘the boy to whom I’ll bring the book tomorrow’ (Italian; Benincà 2001)

Bertollo & Cavallo (2012) report, in parallel to the observation above, that Italian FRs are syntactically larger than interrogatives. The Slovenian evidence in (7) (building on Hladnik 2015) suggests that HRs are structurally larger than the corresponding FRs, as only the former can accommodate an overt complementizer.

- (7) a. Živi v pokrajini, *kjer* (da) imajo najboljše vino.
lives in region where COMP have best wine
‘He lives in a region where (it is claimed) they have the best wine.’
b. Živi, *kjer* (*da) imajo najboljše vino.
lives where COMP have best wine
(Intended:) ‘He lives where (it is claimed) they have the best wine.’
(Slovenian; Marko Hladnik, p.c.)

Recently, Biskup & Šimík (2019) and Šimík et al. (2022) argued that Czech conditionals and correlatives are syntactically smaller than the corresponding free (or light-headed) relatives: while the former are clausal (CPs) and proposition-denoting, the latter are nominal (NPs/DPs) and denoting in the domain of entites. The evidence for this – stemming from rating experiments – is that only the former are transparent for long-distance extraction, as illustrated in (8).

- (8) a. To je ten chlap, kterému₁ [_C *když* zavoláš ____₁], odejdu.
this is the man who when call.2SG leave.1SG
‘This is the man who is such that when you call him, I will leave.’
b. *To je ten chlap, kterému₁ odejdu [_{FR} *když* zavoláš ____₁].
this is the man who leave.1SG when call.2SG
Intended: ‘This is the man who is such that I will leave when you call him.’
(Czech; Biskup & Šimík 2019)

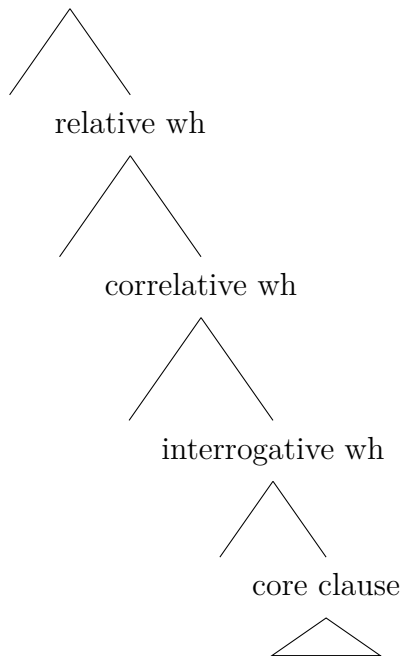
Although a detailed investigation of how all the wh-constructions in the hierarchy (2) are interrelated in terms of their syntactic size is missing (within a single language or across languages, although see Šimík in press), the available evidence is consistent with the idea that the complexity hierarchy corresponds to the relative structural size of the individual constructions.

The increasing morphological complexity of wh-words in the individual functions, as described in section 2.1, matches the increasing syntactic complexity of the wh-con-

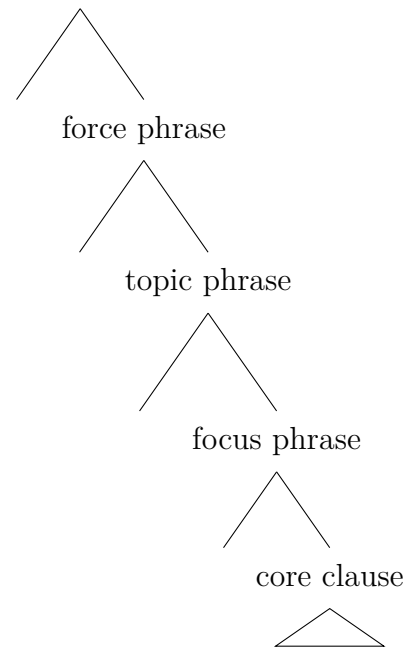
structions they appear in. In fact, many current theories of morphology routinely treat morphological phenomena in syntactic terms (Starke 2009; Siddiqi 2018). In such theories, it is natural to consider a *wh*-word used in a more complex function as a syntactic superset of *wh*-words used in less complex functions.

The growing syntactic and morphological complexity of the *wh*-functions is represented in (9), a syntactic structure which functionally matches Rizzi’s (1997) influential proposal on how the functional “peripheral” layers of clauses are structured, which is – in a simplified way – represented in (10). The interrogative function matches focus because both interrogatives and foci involve reference to alternative denotations / possible values of a free variable (Rooth 1985; Kratzer 1991; Beck 2006). The correlative function matches topic because the correlative *wh*-word can be seen as a topical referent which the correlative consequent is about (Reinhart 1981; Bittner 2001). Relative clauses correspond to Rizzi’s (1997) Force, which is implicated in the relation between a clause and its external environment, a phenomenon called syntactic embedding or subordination, which in turn necessarily involves the semantic process of abstraction, as suggested above. The incremental growth from the force phrase into the determiner phrase is implicated in the contrast between free relatives, which possibly correspond to force phrase, and (light-)headed relatives, which have an additional determiner (or, more generally, nominal) layer (Citko 2004).⁹ The difference between light-headed and nominally-headed relatives does not necessarily lie in further incremental growth, but rather in the integration – syntactic and semantic – of an additional lexical restriction provided by the nominal head of the relative clause. For relevant discussion pertaining to acquisition, see Foley (1996).

(9) *Our analysis of wh-functions*
determiner phrase



(10) *Rizzi’s (1997) analysis (adapted)*
determiner phrase



⁹There has been a debate whether free relatives are “nominal” only semantically (undergoing a semantic shift to entities) or also syntactically (being headed by covert determiner). The latter position is dominant in the literature (see, e.g., van Riemsdijk 2017; cf. Šimík 2010 for a dissenting view), but is often adopted without explicit empirical support.

Summary We have shown that the wh-complexity hierarchy has natural underpinnings in both the semantic and morphosyntactic representations of the individual wh-constructions and wh-words in the hierarchy: the more complex the wh-function, the more complex its semantic and morphosyntactic representation. The incrementally growing complexity is not always easily detectable – cf. the morphological identity of wh-words across the functions in some languages or the apparently identical syntactic shape of the individual constructions – but if an empirical contrast can be detected, then it goes in the direction predicted by the hierarchy.

2.3 Hypotheses under consideration

Growing trees Our working hypothesis is that children acquire the wh-functions in accordance with the wh-complexity hierarchy. The basic idea is that the acquisition of a wh-function is conditioned by the mastering of the simpler wh-function(s). For instance, a child is expected to acquire wh-based headed relatives only after they have mastered free relatives or conditionals, which in turn should follow the acquisition of interrogatives. The rationale behind the hypothesis is that the individual wh-functions reflect a certain degree of mental grammatical development – the semantic and syntactic structures available to the child at the moment. For instance, wh-interrogatives require the understanding of free variables and their use in expressing the locus of ignorance; wh-relatives crucially build on the notion of a variable, but in addition require the understanding of the process of abstraction, which is instrumental in turning an open proposition into a property of entities. In syntactic terms, a child can only make use of some structure, say the Force phrase (used in relative clauses) if it had already acquired the structures on which it depends, i.e., the Topic phrase and Focus phrase. To capitalize on the metaphor of the growing trees hypothesis, we can say that a child can only use wh-functions which correspond to the syntactic tree that has already “grown” in their mental grammatical representation.

The conceptual predecessors of the growing trees hypothesis go back more than three decades (Clahsen 1990; Rizzi 1994; Friedmann & Grodzinsky 1997; Guasti 2002). For more references and discussion, see Friedmann et al. (2021), whose general approach we adopt here. Friedmann et al. (2021) use syntactic representations akin to the one in (10) to model the acquisition of the Hebrew functional syntactic structure. They argue that the acquisition proceeds in three stages. In the first stage, children acquire what we refer to as the core clause in (10). This stage includes the mastering of the basic clause structure, intransitive and transitive predicates, their argument structure and the basic word orders associated with them. In the second stage, corresponding to acquiring the focus phrase in (10), children master most kinds of wh-interrogatives, involving the clause-initial position of subject, object, and some adjunct wh-words. In the third stage, corresponding to the force phrase in (10), children master topicalization structures, relative clauses, and more generally clausal embedding.

Our own perspective partly overlaps with Friedmann et al.’s (2021). We also look at wh-interrogatives and relative clauses. In addition, however, we include an intermediate construction type, namely correlatives/conditionals. This enables a more detailed look at the acquisition stages. More particularly, we expect the acquisition of correlatives/conditionals (\approx the topic phrase level, by assumption) to precede the acquisition of relative clauses (\approx force phrase). There are additional differences related to material, method, and statistical analysis. These aspects will be described below.

Frequency The null hypothesis is that children acquire the various wh-functions in accordance with input frequency: the more a wh-function is frequent in the input, the faster the acquisition of that function. Simplifying somewhat, this null hypothesis is in line with usage-based approaches to acquisition that focus on frequency (Ambridge et al. 2015). It cannot be ruled out, of course, that the wh-complexity hierarchy is in accordance with frequency in the sense that the more complex a wh-function is, the less frequent it is. In that case, it would be impossible to decide between the two hypotheses. In order to raise the chance of finding out which hypothesis is more appropriate, we do not only look at overall frequencies, but rather at the impact of a specific adult on a specific child and look at how the impact develops over time, using a longitudinal corpus.

3 Previous work on the acquisition of wh-constructions

L1 acquisition of various types of wh-constructions has been studied from a variety of perspectives.¹⁰ Major areas of interest include the acquisition of subject–auxiliary inversion in interrogatives (since Brown 1968; see Thornton 2016 for a survey), the order of acquisition of individual wh-sorts (‘what’, ‘where’, ‘how’, ‘which’, etc.; see, e.g., Rowland et al. 2003), or subject–object asymmetries in the acquisition of both interrogatives (see Thornton 2016 for a survey) and relatives (see Lau & Tanaka 2021; Tanaka et al. online first for a survey). In our brief overview below we concentrate on the issue of greatest interest to us, namely the order of acquisition of individual wh-constructions. We are not aware of a study looking at the whole range of wh-constructions considered here. Nevertheless, it is possible to draw some inferences based on the studies of individual wh-constructions or on the comparison between selected wh-constructions.

Wh-interrogatives To start with, let us note that there is ample evidence that wh-interrogatives are the first wh-constructions to be acquired by children. Children start producing their first wh-interrogatives around (and often before) the age of 2 with the words for ‘what’ and ‘where’ being the first to get acquired (Klima & Bellugi 1966; Bellugi 1971; Bloom et al. 1982; Guasti 2000; Dąbrowska 2000; Rowland et al. 2003; Armon-Lotem 2008; Steinkrauss 2009; a.o.).

(Wh-based) headed relative clauses Headed relative clauses are acquired relatively late. As reported in Diessel (2004), who analyzed the English CHILDES (MacWhinney 1995), adult-like headed relative clauses start appearing around the age of 3 in English. Moreover, up until the age of 5, wh-based headed relatives are very rare in English child speech: out of the total of 305 finite relative clauses, there was no instance of a *which*-relative and only 6 *who*-relatives, all of which were produced by a single (the most eloquent) child in the corpus (Adam).¹¹ The results of Labelle’s (1990) study of child relative clause production in French are fully in line with Diessel’s (2004) findings for English: 37 out of 108 Labelle’s participants – children of 3 to 6 years of age – are categorized as “poor” (vs. “medium” and “good”) performers by virtue of producing 10 or fewer out of the possible total of 30 relative clauses.¹² These were likely the youngest children, indicating

¹⁰Relevant monographs or surveys include Diessel (2004); Lust et al. (2009); Thornton (2016).

¹¹Unfortunately, Diessel (2004) disregarded adverbial wh-words (as in *...the place where...*) in his analysis.

¹²Even though it is not explicitly mentioned in the article, it follows from the descriptive statistics that 7 out of the 108 children produced no relative clauses at all.

that a solid mastering of relative clause production starts significantly later than at 3 years of age.¹³ Moreover, Labelle (1990) observed no *wh*-based relative clauses at all; all relative clauses produced involved the complementizer *que/qui*.¹⁴ In their recent study of Hebrew, Friedmann & Reznick (2021) and Friedmann et al. (2021) confirm the relatively late acquisition of relative clauses. (It should be noted that Hebrew headed relatives are generally complementizer- rather than *wh*-word based.) The first relatives appear in natural production around 2;6, but “stabilize between the ages 3;3 and 4;0 years.” (Friedmann et al. 2021: 11)

Free relatives vs. headed relatives It has been observed repeatedly and across languages that free relatives are acquired earlier than headed relatives – a generalization called the “developmental primacy of free relatives” by Flynn & Foley (2004). Most of the evidence in this area stems from the elicited imitation paradigm (also called sentence repetition). This effect was first observed by Flynn & Lust (1980) for English (but see also Hamburger 1980), where children (3;6–7;7) were significantly worse at repeating headed relatives (...*the thing which*...) than free relatives (...*what*...), a difference that only starts to disappear around 6 years of age. Comparable evidence is available for Chinese (Packard 1988), Korean (Lee 1991), Tulu (Somashekar 1999), or Quechua (Courtney 2006). The fact that the primacy of free relatives is observed also for languages that do not employ *wh*-words in their relative clauses, such as Chinese or Korean, is in line with the idea that the complexity stems from the integration of the nominal head (see Foley 1996). Interestingly, no difference in the timing of mastering free and headed relatives was found for French in Foley (1996). One possible reason for the missing asymmetry could be that French free relatives in fact look a lot like light-headed relatives. See Foley (1996) for detailed discussion.

Correlatives Systematic investigations of the acquisition of correlatives are rare and the existing results are mixed. Sharma (1974) (as cited in Lakshmanan 2000: 588, fn. 2) reports the (apparently) episodic observation that “the correlative emerges early on [and abundantly, by 3 years of age; see Lust 1994] in the case of Hindi speaking children.” Lakshmi Bai (1989) reports a comparable result in the study of her two bilingual Tamil-Telugu children, who started using correlatives around the age of 3. Using elicited imitation Somashekar (1999) found that children in the youngest group (2;6–3;5) were able to repeat correlatives in 13% of the cases, a comparatively better result than the 3% success in participial headed relatives, though worse than the 20% of what could be called participial free relatives.¹⁵ In their study on Telugu, Nirmala’s (1980) found no evidence of correlatives up until the age of 3 and half. Importantly, however, Telugu correlatives are reported to be part of a formal register and are thus unlikely to appear in child-directed speech. Šimík et al. (2023) conducted a sentence imitation task to assess the relative

¹³Labelle (1990) does not provide analyses in terms of age, but she does speak about the “evolution [...] between ages 3;0 and 6;0” (p. 114) in this context. The performance measure (poor–medium–good) is thus probably a proxy of age.

¹⁴Notice that the French complementizer *que/qui* morphologically resembles *wh*-words. A similar situation obtains in Czech, as we will see. We include *wh*-complementizer-based relative clauses in our corpus analysis because it cannot be taken for granted that children are capable of telling inflected *wh*-words and invariant *wh*-complementizers apart.

¹⁵A recurring pattern in the above-mentioned studies is that participial relatives (esp. free or light-headed relatives) are acquired quite early, typically earlier than correlatives. Note, however, that participial relatives do not rely on *wh*-words and are thus only marginally relevant for our purposes.

difficulty of the acquisition of Czech correlatives (lit. ‘where policeman stands, there car waits’; ‘A policeman is standing where a car is waiting.’) and string-identical (as well as truth-conditionally indistinguishable) light-headed relatives (lit. ‘car waits there where policeman stands’). They found that preschoolers (2;10–4;3) score significantly better on correlatives than of light-headed relatives.

Conditional and temporal clauses Conditional and temporal clauses, or more generally adverbial clauses, are often considered a *sui generis* construction in L1 acquisition research. Yet there are many good reasons to treat them on a par with relatives. They can take the form of correlatives, free relatives, or (light-)headed relatives, as illustrated in (11), and in the theoretical literature, they are also treated as such (see, e.g., Bittner 2001; Caponigro 2003; Brasoveanu 2008; Caponigro & Pearl 2009; Williamson 2019).¹⁶

- (11) a. When(ever) Sue needed her most, Lisa left. ≈ correlative
 b. Lisa left when(ever) Sue needed her most. ≈ free relative
 c. Lisa left at the time when Sue needed her most. ≈ headed relative

Conditional/temporal clauses have been shown to be acquired relatively early. Reilly (1986), for instance, shows that *when*-clauses in English are naturally produced around 2;6, corroborating earlier findings (e.g., Bates 1976; Clancy et al. 1976). Also data from the English CHILDES (MacWhinney 1995) reveal that *when*-clauses are among the first clausal adverbials in child speech, nearly the only ones that productively appear – together with *because*-clauses – before the age of 3 (Diessel 2004: Chapter 7).¹⁷ Interestingly, it is the right-peripheral version (corresponding to (11b)) that gets acquired first (mean age 2;9), while the left-peripheral ones (corresponding to (11a)) start appearing later (mean age 3;2); see Diessel (2004: 168ff.). This order of acquisition is unexpected from the perspective of our wh-complexity hierarchy. Evidence from other languages than English point to an even earlier acquisition of temporals and conditionals – even before the age of 2. This holds for Japanese and Korean (Akatsuka & Clancy 1993) or Greek (Katis 1997). In Greek, in contrast to English, even the earliest instances of conditionals/temporals appear in the left-adjoined position, suggesting the need for a careful language-level analysis. Finally, let us note that none of these languages employ a wh-word in the formation of their conditionals/temporals, which makes their relation to correlatives loose.

Summary Judging by the existing findings reported above, the timing of the acquisition of the relevant (wh-)constructions is largely consistent with the wh-complexity hierarchy (2). Interrogatives are acquired first, followed by conditionals (at least in some languages) – both before or around the age of 2. Free relatives are acquired later (shortly before or around the age of 3) and consistently precede the acquisition of headed relatives. Systematic investigation of the acquisition of correlatives (grouped here together with conditionals) is scarce, but at least some findings (from Hindi; Sharma 1974) suggest early acquisition (before the age of 3). Finally, the study of Šimík et al. (2023) suggests that Czech children master correlatives earlier than string-identical light-headed relatives.

We are well aware that the inferences we have drawn from the existing findings are nothing more than estimates. The studies they are based on differ greatly in terms

¹⁶As will be detailed below, we adopt this approach in our study and include *když* ‘when’-based temporal/conditional clauses in our sample.

¹⁷*If*-clauses start appearing a bit later – shortly after the 3rd birthday.

the methodology applied as well as the language under investigation. The comparisons are drawn on a cross-child basis, which is a big problem given how greatly individual developments can differ. Lastly, most of the above-mentioned studies do not control for adult input frequencies, so it is impossible to estimate which effects can be attributed to the development of grammatical competence and which are plain frequency residues. In our study we hope to overcome at least some of these issues.

4 Corpus study

4.1 Corpus

For our corpus analysis, we used the transcripts of the speech of 6 monolingual, typically developing Czech children (1;7–3;9) that constitute the Chromá corpus in CHILDES (Chromá et al. under review). They were recorded by their caregivers at home during routines and play about twice a month. Besides the children’s speech, we also analyzed child-directed speech. The whole corpus consists of 75 237 children’s words and 187 423 words of adults (and in one case also an 8-year old brother).

4.2 Data extraction

Using CLAN (MacWhinney 2000) we extracted all utterances containing wh-words, e.g. *kdo* ‘who’, *co* ‘what’, *kam/kde* ‘where’, *kdy/když* ‘when’, *jak* ‘how’, *proč* ‘why’, *jaký/kerý* ‘which’, *čí* ‘whose’, *kolik* ‘how much/many’, and various variants thereof. Most of Czech wh-words can be used in most of the wh-constructions under investigation. The sample included utterances with *když* ‘when’, which is the conditional/free relative correlate of the interrogative *kdy* ‘when’. We have decided to also include *co* ‘what’ and *jak* ‘how’ in their relative complementizer function, mainly because one cannot be sure that children are capable of telling wh-words (inflected proforms) and wh-complementizers apart. Finally, we included wh-words with the postfix *-koli(v)* ‘ever’ in case these participated in the formation of one of the constructions under investigation.

We have eliminated repetitions within one utterance and exact repetition of an utterance within two utterances immediately following each other, either by the same speaker or by different speakers. We also excluded revised parts. Utterances that were read (from books) or sung were excluded, too. On the other hand, we did not exclude all incomplete, interrupted, and partly unintelligible utterances (see below). The final data consists of 1786 children’s utterances containing wh-words and 9028 adults’ utterances. For purposes of our statistical analyses we have excluded instances of embedded interrogatives (due to their complicated status with respect to the wh-complexity hierarchy; see our remarks in section 4.3) and were thus left with 1674 children’s utterances and 8256 adults’ utterances.

4.3 Data annotation and examples of constructions

Each wh-word containing utterance was annotated for the following variables (setting aside technical annotations): child recorded (6 values), child vs. adult (2 values), child age at the time of recording, mean length of utterance/MLU (computed per recording session), wh-word (particular form used), wh-lemma (dictionary form), wh-phrase (relevant if the wh-word was a proper part of a larger wh-phrase, like *which dog*), and – most importantly for us – the wh-construction which the wh-word participates in forming. The

wh-construction annotation was accompanied by the degree of confidence in the annotation, from 1 = high confidence, 2 = moderate confidence (often cases of ellipsis that can be reconstructed), 3 = low confidence.¹⁸

The wh-construction was annotated by the second author and the result was checked by the first author (both native speakers of Czech). The final annotation is the result of mutual agreement between the two annotators. The included categories of wh-functions are listed below. We include basic information about the annotation procedure, but also about the properties of the individual constructions in the target adult grammar.¹⁹

- **Interrogatives/I:** A wh-construction was annotated as I if the wh-word was part of main/matrix (not embedded) clause that ended with a question mark. Interrogative wh-phrases in Czech matrix interrogatives mostly occupy the initial clausal position (Veselovská 2021), see example (12a), but can also be placed elsewhere if there is information structural motivation for it – for instance the need to place a contrastive topic in the initial position (cf. Biezma 2020), in which case the wh-phrase appears in the clause-final position, characteristic of foci in Czech (Jasinskaja & Šimík forthcoming). An example of this is provided in (12b), where the referent of *tam* ‘there’ contrasts with places and *jaká* ‘what [color of a lego piece]’ (involving NP ellipsis) ranges over colors, yielding a canonical contrastive topic–focus (CT–F) structure (Büring 2003).

- (12) a. kde je máma? [Anna 1;09.30]
 where is mummy
 ‘where’s mummy?’
 b. tam patří jaká? [Julie 3;03.04]
 there belongs what.kind
 ‘what [color of a lego piece] belongs there?’

- **Embedded interrogatives/EI:** The EI annotation applied to wh-clauses syntactically and semantically dependent on a superordinate element (typically an attitude predicate like ‘know’, ‘say’, ‘ask’, etc.) conveying propositional contents (rather than referring to an entity). They can often be distinguished from free relatives by being asked about by ‘what’, independently of the wh-word used (Caponigro 2003). E.g. in (13a) one would ask ‘What will you tell daddy?’ rather than ‘Where will you tell daddy?’. As opposed to matrix interrogatives, EIs always place their wh-words in the initial position; we have not found a case where this would not be respected in our corpus. Czech interrogative wh-phrases are routinely discontinuous (see Veselovská 2021, a phenomenon called left-branch extraction), which is illustrated in (13b), where *jak* ‘how’ is fronted and *dlouhý vlasy* ‘long hair’ stranded.

- (13) a. já jdu tátovi říct, kde má kočičku [Anna 2;04.26]
 I go.1SG daddy tell.INF where have.3SG kitty
 ‘I’m going to tell daddy where his kitty is’

¹⁸The annotated sample as well as the annotation manual can be accessed here: https://osf.io/semxc/?view_only=8108bf938aac43dea2347cdc95729871.

¹⁹Instances of rare wh-construction types like the modal existential construction (Šimík 2011) were excluded from the sample.

- b. podívej, jak mám dlouhý vlasy [Klára 3;00.24]
 look.IMP how have.1SG long hair
 ‘look how long my hair is’

In this paper, we provide descriptive statistics about EIs, but do not include them in our inferential analyses. The reason for this is that the status of EIs in the wh-complexity hierarchy is not straightforward, which in turn complicates the predictions about their acquisition and about the effect of adult EI input on children’s output of wh-constructions. More specifically, EIs are like root interrogatives in that they denote propositions, but like relatives in that they are embedded. This ambivalence is reflected in their cross-linguistic and cross-constructural morphosyntactic behavior: in some respects they behave on a par with interrogatives (typically interrogative rather than relative wh-morphology), in others with relatives (often a ban on wh-in-situ, see Bobaljik & Wurmbrand 2015, sometimes relative morphosyntax, see Zimmermann 2018; Kellert 2018).

- **Correlatives or conditionals/C:** The C annotation was used for wh-clauses that did not convey the propositional contents of an attitude predicate and were positioned to the left of the superordinate clause they modify. Typically, the wh-word had a demonstrative correlate in the superordinate clause (cf. *když–tak* ‘when–so/then’ in (14a)). Most of the C clauses were temporal/conditional adverbials introduced by *když* ‘when’, as in (14a); only rarely were they other types of correlatives (an example is provided in (14b)). The wh-words or phrases must appear in the clause-initial position in Cs, as well as in all the construction types discussed below; as with EIs, we have not encountered any cases where this would be violated.

- (14) a. když má ty vočka, tak to je chobotnice [Viktor; 2;06.23]
 when has those eyes so it is octopus
 ‘when it has those eyes, then it’s an octopus’
 b. co je uklizený, tak to můžem mít takhle [Julie 3;05.24]
 what is tidied.away so that can.1PL have.INF this.way
 ‘What’s tidied away can be (placed) like this.’

- **Free relatives/FR:** A FR is very similar to C except that it is placed to the right of its superordinate clause. They can be told apart from EIs by being asked about using the wh-word that occurs in the wh-clause (e.g. in (15a) would be asked about by ‘When will you wash it?’ rather than ‘What will you wash (it)?’). Like Cs, most FRs are temporal/conditional clauses introduced by *když* with only rare occurrence of other wh-sorts of the kind in (15b).

- (15) a. ale opláchnu si to jenom, když si lehnu [Viktor 2;10.24]
 but wash REFL it only when REFL lie.down
 ‘But I’ll only wash it when I lie down.’
 b. já stavím, co jsi už dělala [Viktor 2;07.02]
 I build.1SG what AUX.2SG already made
 ‘I’m building what you’ve already made.’

- **Light-headed relatives/LHR:** LHRs are relative clauses headed by pronominal (rather than nominal) expressions, typically demonstratives (e.g. *ten* ‘that (one)’),

tam ‘there’ etc.), illustrated in (16a), but also universals (*všechno* ‘everything’), as in (16b). LHRs are usually introduced by wh-pronouns, but can also be introduced by complementizers (like headed relatives), possibly combined with resumptive pronouns. A structure like this (rather exceptional in our data) is illustrated in (16c); the structure *co + na tom* ‘what.COMP + on it’ could in principle be replaced by *na čem* ‘on what’.

- (16) a. *my jsme to přivezli tam, kde bydlel ten Petr.* [Viktor 3;05.29]
 we AUX.1PL it brought there where lived DEM Petr
 ‘We brought it where Petr lived.’
- b. *všechno, co bylo v pořádku, už jsem vyndal* [~ 3;06.19]
 everything what was in order already AUX.1SG put.out
 ‘I’ve put out everything that was all right’
- c. *tohle to, co na tom jezdí* [~ 2;06.23]
 this that what.COMP on that rides
 ‘this [thing] that (s)he’s riding on’

- **Headed relatives/HR:** HRs are standard nominally headed relative clauses. They can be introduced by relative pronouns like *ktorej* ‘which’ in (17a), pronominal adverbials like *kde* ‘where’ in (17b), or wh-complementizers like *co* ‘what’ or *jak* ‘how’; see (17c).²⁰

- (17) a. *bůh je takovej táta ktorej je zlobivej* [Viktor 2;09.25]
 god is a.kind.of dad which is mean
 ‘God is a kind of dad who is mean.’
- b. *tys šlápla na špinavou díru, kde bydlely* [Klára 3;04.01]
 you.AUX.2SG stepped on dirty hole where lived
ty liščata
 DEM little.foxes
 ‘You stepped on the dirty whole where the little foxes lived.’
- c. *⟨dárky⟩ jak jsme je rozbalili* [Julie 3;03.04]
 gifts how.COMP AUX.1PL them unpacked
 ‘gifts that we unpacked’

The reliability of the construction type annotation was checked by inter-annotator agreement. A student annotator (Czech native speaker) was trained in the annotation and subsequently annotated 1000 randomly selected tokens from the analyzed sample. Her annotation was compared to the authors’ annotation. Cohen’s κ was 0.88. After the exclusion of the interrogatives – the most frequent and easiest-to-annotate construction type, Cohen’s (1960) κ was 0.78. Subsequent analysis of the annotations revealed that most disagreements were due to C vs. FR annotations, which is understandable given that most of both C and FR were adverbial clauses introduced by *když* ‘when’ and the fragmentary character of many utterances made it difficult to decide on the linear order of the adverbial and its superordinate clause.²¹

²⁰We use ⟨ ⟩ to indicate ellipsis.

²¹For purposes of our statistical analysis, we have collapsed C+FR and LHR+HR (see below). With these construction pairs collapsed, Cohen’s κ was 0.94 (with interrogatives) and 0.91 (without interrogatives).

4.4 Analytic approach

The goal of our paper is to describe the use of different wh-word categories in children, test their acquisition sequence and evaluate the effect of adult input. We address these goals using cumulative link mixed models with the category of the wh-word as the ordinal dependent variable. This way, we can test whether the acquisition sequence implied by the theoretical assumptions is in line with the data, i.e. whether it is related to age. Subsequently, we tested whether wh-word production in children could be explained by the frequencies of these structures in the input language they were exposed to. To do this, we used the lagged regression approach, examining the effects of parental wh-word use on children’s use of these elements in subsequent transcripts. The use of cross-lagged regression may be understood as a test of Granger causality (Granger 1969), i.e., of regular precedence of one type of events before another type of events. Observing Granger causality does not necessarily imply direct causality between the preceding and following event, but it shows that the two phenomena are in potentially causal relations.

To perform our statistical analyses, we relied on cumulative link mixed models as implemented in the R package *ordinal* (Christensen 2013). This type of models is used for ordinal dependent variables, and tests whether predictor variables are associated with increased probability of observing higher ordinal values of the dependent variable. Unlike standard linear models, there is no assumption of equal distance between the different values of the dependent variable, say between 1 and 2 vs. 2 and 3. The ability to fit the mixed version of the cumulative link makes it possible to account for the fact that our corpus contains recordings from 6 different children and their caregivers. All models were fit with random intercept for each child. More complex random structure was tested but could not be estimated, which is not surprising given the low number of children.

5 Results

5.1 Overall descriptive results

Table 1 provides basic information regarding the absolute frequencies of individual construction occurrences per child, including the associated adult input. Almost half of all children’s datapoints come from Viktor (834 out of 1786) and Viktor is the author of 70 % of all child non-interrogative instances. The other extreme is Jan, who only produced interrogatives (including one embedded interrogative) and even that in very low frequencies. Another source of asymmetry in the data is the predominance of interrogatives, whose frequency is around 80 % in both adults’ and children’s speech. On the other hand, light-headed relatives are represented by only 24 instances in child speech (22 of which come from Viktor).

The above-mentioned imbalances in the data as well as the overall scarcity of some construction types clearly limit the options of further analyzing our data. It seems clear that our working hypothesis cannot be addressed fully. In order to allow for any meaningful statistics – descriptive or inferential – we have decided to collapse the data from two pairs of adjacent wh-constructions – C+FR and LHR+HR. Collapsing C and FR is motivated not just by the adjacency of the two functions, but also by the fact that the majority of both C and FR occurrences are cases of *když* ‘when’ clauses (92 % in the case of C and 70 % in the case of FR). Thus there is a clear sense in which the two construction types form a natural class in our data. Collapsing LHR and HR is motivated by

Aneta 2;02–3;03 11.5 h in 38 sessions				Anna 1;09–2;07 6.5 h in 14 sessions			
	Adults	Child	Total		Adults	Child	Total
I	1482	327	1809	I	505	174	679
C	47	5	52	C	36	7	43
FR	41	5	46	FR	39	0	39
LHR	21	0	21	LHR	14	2	16
HR	59	2	61	HR	49	2	51
EI	169	12	181	EI	76	6	82
Total	1819	351	2170	Total	719	191	910

Jan 1;07–2;09 7 h in 11 sessions				Julie 1;07–3;09 14 h in 32 sessions			
	Adults	Child	Total		Adults	Child	Total
I	824	66	890	I	1615	173	1788
C	9	0	9	C	105	19	124
FR	19	0	19	FR	168	10	178
LHR	1	0	1	LHR	30	0	30
HR	12	0	12	HR	113	3	116
EI	59	1	60	EI	188	6	194
Total	924	67	991	Total	2219	211	2430

Klára 2;04–3;04 4.5 h in 38 sessions				Viktor 2;06–3;08 14.5 h in 32 sessions			
	Adults	Child	Total		Adults	Child	Total
I	1036	99	1135	I	1675	561	2236
C	17	14	31	C	87	56	143
FR	19	3	22	FR	87	39	126
LHR	6	0	6	LHR	20	22	42
HR	5	3	8	HR	115	82	197
EI	36	13	49	EI	244	74	318
Total	1119	132	1251	Total	2228	834	3062

Table 1: Absolute frequencies per child

category	wh-word		Children		Adults		Total
human	<i>kdo</i>	‘who’	37	91	519	465	556
thing	<i>co</i>	‘what’	762	855	4453	4360	5215
location	<i>kde</i>	‘where’	240	167	776	849	1016
time	<i>kdy</i>	‘when’	6	6	29	29	35
manner	<i>jak</i>	‘how’	49	96	538	491	587
reason	<i>proč</i>	‘why’	233	84	281	430	514
amount	<i>kolik</i>	‘how many’	11	13	69	67	80
identity	<i>který</i>	‘which’	3	8	47	42	50
quality	<i>jaký</i>	‘what kind of’	59	76	403	386	462
possession	<i>čí</i>	‘whose’	0	4	22	18	22
Total			1400		7137		8537

Table 2: Frequencies of interrogative wh-word categories (wh-words represent prototypical lemmas; gray = expected frequencies)

both construction types being headed by (pro)nominal material – something that clearly singles them out in the dataset.

In what follows, we thus work with the simplified wh-complexity hierarchy in (18).

- (18) **Wh-complexity hierarchy** (simplified for purposes of statistical analysis)
 $I < C+FR < (L)HR$

Table 2 provides the frequencies of individual ontological categories of wh-words in the production of interrogatives in children and adults. Expected frequencies are provided for a better orientation in where the children under/overperform (relative to adults).²² The categories that children in our sample ask less about than expected are humans and manners. They also use relatively few determiner-like wh-words (*který*, *jaký*, *čí*). What children ask more about than expected are locations and reasons or causes.²³ While examining the frequencies, it is good to keep in mind that 40% of all interrogative wh-words were produced by a single child (Viktor) and 63% by two children (Viktor and Aneta). One should therefore be cautious in generalizing the findings.

Table 3 provides the frequencies of the types of conditionals and free relatives. As we have already noted, the absolute majority of the C+FR category is represented by temporal/conditional clauses introduced by *když* ‘when’. The relativization of individuals (‘who’/‘what’), on the other hand, is rather rare. The frequencies in child speech match the frequencies in adult speech almost perfectly ($\chi^2(2) = 0.472, p = 0.79$). We should keep in mind that 60% of all C+FR instances among children are due to a single child

²²By expected frequencies we mean frequencies standardly used for computing the χ^2 test of independence; an expected frequency = sum of column \times sum of row divided by the grand sum. For instance, the expected frequency of children’s production of ‘who’ equals $(1400 \times 556)/8537$.

²³An anonymous reviewer suggests to treat ‘why’ interrogatives as a separate category – more on a par with relatives than interrogatives. This strategy is adopted in Friedmann et al. (2021), who notice that in Hebrew, as in Italian (Rizzi 2001), ‘why’ appears higher in the interrogative structure than other wh-words (namely above topics). We are not aware of clear evidence that ‘why’ would be special in this way in Czech. Also, more than half of the ‘why’ interrogatives produced by children in our sample involve clausal ellipsis (aka sluicing), i.e., simply a one-word utterance ‘Why?’. In the absence of deeper evidence, we cannot be sure that the ellipsis site involves any structure in children’s grammar or whether it is just an unstructured propositional variable (see Merchant 2001 for background discussion).

category	wh-word		Children		Adults		Total
individual	<i>kdo/co</i>	‘who’/‘what’	6	6	26	26	32
temp/cond	<i>když</i>	‘when’	133	130	553	556	686
rest			19	22	95	92	114
Total			158		674		832

Table 3: Frequencies of conditional/free relative types (gray = expected frequencies)

category	Children		Adults		Total
relative pronoun/proadverb	97	68	231	260	328
relative complementizer	19	48	213	184	232
Total	116		444		560

Table 4: Frequencies of relative clause types (gray = expected frequencies)

(Viktor).

Table 4 provides the frequencies of relative clause introduction types. Czech possesses two main strategies of introducing a relative clause – either by using a relative pronoun (typically *který* ‘which’, sometimes *jaký* ‘what kind of’) / proadverb (*kde* ‘where’, *jak* ‘how’, etc.), or an invariant (uninflected) relative complementizer (*co* ‘what’ or *jak* ‘how’), often combined with a resumptive pronoun. In our sample, the relative pronoun/proadverb strategy clearly dominates in children as compared to the adults ($\chi^2(1) = 37.833, p < 0.001$). But again, it is important to keep in mind that most relative clause data, namely 71 %, come from a single child (Viktor).

We continue by briefly examining the issue of the subject–object asymmetry – the recurrent finding that wh-subjects are easier to produce and comprehend than wh-objects. Due to the observed absolute frequency differences among ‘what’, ‘who’, and complex wh-phrases (with determiner-like wh-words) in both child and adult speech (see Tables 2 for the case of interrogatives), we have divided the data so that any potential effects in the ‘who’ subset and complex wh-phrase subset do not get lost in the ‘what’-subset. For purposes of this analysis, we have aggregated the data from all types of wh-constructions. In Table 5 we can see a clear subject–object asymmetry in the use of ‘what’: the subject function clearly dominates in children ($\chi^2(2) = 120.221, p < 0.001$).²⁴ This kind of asymmetry is not evident with ‘who’ ($\chi^2(2) = 2.605, p = 0.27$) and complex wh-phrases ($\chi^2(2) = 2.199, p = 0.33$), as shown in Tables 6 and 7.

We close this section by briefly inspecting the mismatches between child production

²⁴Note that this is not only due to cases of subjects of copular clauses (‘What is X?’), whose proportion within the subject function is comparable in both child and adult speech.

function	Children		Adults		Total
subject	339	222	1448	1565	1787
object	198	296	2186	2088	2384
other	16	36	272	252	288
Total	553		3906		4459

Table 5: Frequencies of syntactic functions of ‘what’ (gray = expected frequencies)

function	Children		Adults		Total
subject	32	34	431	429	463
object	6	6	79	79	85
other	5	3	30	32	35
Total	43		540		583

Table 6: Frequencies of syntactic functions of ‘who’ (gray = expected frequencies)

function	Children		Adults		Total
subject	63	56	255	262	318
object	47	53	255	249	302
other	15	17	81	79	96
Total	125		591		716

Table 7: Frequencies of syntactic functions of complex wh-phrases (gray = expected frequencies)

and the target grammatical properties of the wh-constructions under consideration. Overall, deviations from the target grammar (sometimes called “errors”) were relatively rare and do not seem to point to any specific pattern that would be related to our hypotheses. The error annotation provided by the authors of the Czech CHILDES (Chromá et al. under review) reveals only 8 deviations from the adult norm in the use of wh-words – mostly an unsuitable wh-word choice (such as *kolik* ‘how much/many’ instead of *kdy* ‘when’ or *čím* ‘what.INSTR’ instead of *proč* ‘why’). As already hinted at in section 4.3, we have not encountered any obvious mismatches in the clausal position of wh-words: in interrogatives, they were sometimes in situ (in line with adult grammar); in other constructions, they were always in the clause-initial position (ex-situ). Czech is a language that does not allow for much ordering flexibility of wh-words in the fronted position – wh-words cannot be preceded by topics, for instance – and the child data are adult-like also in this respect.

5.2 Age effects

We start by an informal description of the frequencies of individual wh-constructions per child and age; see Table 8.²⁵ Despite the low absolute frequencies, the development in four of the children is consistent with the simplified wh-complexity hierarchy. Unsurprisingly, interrogatives are clearly the first wh-constructions to be used by all children. What is more important is the development of the further types, where we consistently observe the primacy of C+FR over (L)HR: Julie’s C+FR start appearing shortly before her third birthday, while the frequencies of (L)HR remain very low until the end of observation (3;09). Klára’s C+FR start appearing before her third birthday, while (L)HR remain nearly unattested until the end of observation (3;04). An analogous development is visible in Aneta, where C+FR start appearing around her third birthday, without there being a similar tendency in (L)HRs. Finally, while Viktor produced all construction types from the onset of observation (2;06), cases of Cs and FRs are more frequent until around 3;03, where (L)HR start gaining momentum.

²⁵This data presentation is inspired by Friedmann et al. (2021), who rely on the so-called Guttman scales (Guttman 1944), revealing implicational hierarchies of the kind discussed here.

age days	Aneta			Anna			Jan			Julie			Klára			Viktor			Y	
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3		
610							3	0	0											
640							3	0	0											
670				8	0	0	15	0	0											1
700				6	0	0	na	na	na											
730				21	0	1	7	0	0	2	0	0								
760				12	0	0	18	0	0	4	0	0								
790	2	0	0	9	0	0	2	0	0	5	0	0								
820	4	0	0	27	4	1	na	na	na	3	0	0								
850	29	0	0	na	na	na	na	na	na	3	0	0								
880	9	0	0	13	0	0	9	0	0	na	na	na	23	2	0					
910	33	0	0	23	0	1	8	0	0	1	0	0	12	0	0					2
940	19	1	0	13	3	0	na	na	na	1	0	0	9	0	0	24	9	7		
970	31	0	0	42	0	1	na	na	na	1	0	0	0	0	1	46	3	8		
1000	na	na	na				na	na	na	3	1	0	5	0	0	39	6	5		
1030	5	0	0				1	0	0	na	na	na	na	na	na	51	2	2		
1060	27	0	0							4	1	0	7	0	0	46	9	5		
1090	21	3	1							na	na	na	5	4	0	21	3	1		
1120	38	1	0							16	3	0	7	2	0	30	7	3		
1150	55	4	0							8	2	0	13	2	0	16	0	2		
1180	5	1	0							46	2	1	2	2	0	70	11	2		
1210	49	0	1							na	na	na	6	0	1	59	8	14		
1240										21	5	0	10	5	1	31	2	7		
1270										29	7	0				67	13	27		3
1300										9	5	2				57	3	10		
1330										na	na	na				2	0	1		
1360										na	na	na				2	19	10		
1390										17	3	0								
sum	327	10	2	174	7	4	66	0	0	173	29	3	99	17	3	561	95	104		

Table 8: Number of constructions (1 = I, 2 = C+FR, 3 = (L)HR) documented in the corpus per child and age, divided by 30-day periods (e.g. 640 = age between 611 and 640 days) and years (horizontal lines); empty slots and “na” = no data collected

We examined the effects of age on the likelihood using the more complex wh-categories by the means of a statistical model (see section 4.4 for more detailed information). These effects were robust and statistically significant ($z = 5.90, p < 0.001$, see Model 1 in Table 9). This lends support to the assumed complexity hierarchy of wh-elements. However, as our data indicate, this might reflect the overwhelming majority of interrogative wh-words in children’s data and not the whole hierarchy. We thus repeated the analysis omitting the interrogative wh-words and examined the relative order of the two more complex categories – C+FR and LHR+HR. In this analysis, there was no effect of age ($z = 0.30, p = 0.77$, see Model 2 in Table 9), so despite the suggestive tendencies described above, there is no statistical support for a developmental ordering between these categories.

5.3 Input effects

The cross-lagged analyses examined the effects of age and the previous proportion of interrogative wh-words among all wh-words in the adult language in the transcripts. Proportion of interrogative wh-words was chosen as the relevant measure of parental input because interrogatives are clearly acquired before the remaining categories, while there is no significant effect of age on the relative ordering of the two remaining categories. The effect of previous adult interrogative wh-words was significant and positive ($z = 2.21, p = 0.03$; Model 3 in Table 9), indicating that the previous use of interrogative wh-words by adults is related to later use of more complex categories by children. We also fit a model that included an interaction between age and the previous adult use of interrogative wh-words. The interaction was significant ($z = 2.05, p = 0.04$; Model 4 in Table 9), and its shape is best visible in Figure 1, which shows the estimated probability of producing each wh-category (each category in a separate panel) for different proportions of interrogative wh-use in parental language.²⁶ The panel for interrogative usages in children indicates that low interrogative use in the input (red line) initially relates to a slightly decreased use of interrogatives in children. However, with increasing age, this tendency turns to the opposite: the decrease in interrogative use in children – and a proportional increase of C+FR and (L)HR use – is slowest when parents show low proportion of interrogatives. In turn, high percentage of interrogatives in parents is related to the fastest decrease of interrogative use and a proportional increase of C+FR and (L)HR use in children.

In addition to the analysis of complete data, we again fit models that excluded interrogatives from child and adult data. In a model without an interaction, no significant effect of age was found ($z = 0.05, p = 0.96$) but a significant effect of the previous use of C+FR wh-words ($z = 3.22, p = 0.001$; Model 5 in Table 9). When an interaction between the two variables was included, it was not significant ($z = 1.51, p = 0.13$; Model 6 in Table 9). This means that higher proportion of C+FR in parents is regularly followed by a decreased C+FR use and a proportional increase of (L)HR production in children, regardless of age. This effect of C+FR parental input on C+FR / (L)HR child production,

²⁶The three particular cutoff points in the proportions of I input (0.73, 0.83, 0.94) are a means to visualize the effect of the continuous variable (adult frequency) on the likelihood of using the target structures; they are not part of the corresponding statistical model. They are determined automatically by the graphing algorithms to highlight the effects and interactions of the frequency variable, while staying in realistic ranges for the actual data. They represent the proportion of interrogative words in the productions of all the adults. The raw frequencies for individual structures can be found, for each child and their parents, in Table 1.

Predictors	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Odds Ratios	p	Odds Ratios	p	Odds Ratios	p	Odds Ratios	p	Odds Ratios	p	Odds Ratios	p
Age (years)	3.49	<0.001	1.12	0.768	4.39	<0.001	0.13	0.234	1.02	0.958	0.30	0.188
Past I input (adult)					5.65	0.027	0.00	0.077				
Age (years) * Past I input (adult)							55.42	0.040				
Past C+FR input (adult)									7.26	0.001	0.00	0.255
Age (years) * Past C+FR input (adult)											13.94	0.131
Random effects												
σ^2	3.29		3.29		3.29		3.29		3.29		3.29	
N	6 _{name}		5 _{name}		6 _{name}		6 _{name}		5 _{name}		5 _{name}	
Observations	1671		258		1611		1611		248		248	

Table 9: Statistical models



Figure 1: Probability of producing a given wh-category (panel) as a function of age (x-axis) and proportion of interrogative input (color)

along with the absence of age effects, is visually represented in Figure 2.²⁷

6 Discussion

Our hypothesis (Friedmann et al. 2021) that L1-acquisition of wh-functions proceeds along the wh-complexity hierarchy, repeated in (19), which finds multi-faceted support in the syntax and semantics of wh-constructions, could not be fully addressed by our dataset. Despite the relatively large corpus (1786 wh-words in the children’s production), a majority of the wh-words were used in the interrogative function, leaving only a few hundred for the more complex functions. Moreover, the performance of the individual children – 6 in total – varies dramatically; about a half of all the data come from a single child (Viktor 2;06–3;08), while another child (Jan 1;07–2;09) produced only 67 wh-words, all of which were interrogative.

- (19) **Wh-complexity hierarchy**
 $I < C < FR < LHR < HR$

In order for any statistical analysis to be feasible, we have decided to simplify the hierarchy, collapsing, on the one hand, correlatives/conditionals with free relatives, and on the other hand light-headed relatives and headed relatives. The simplified hierarchy is repeated in (20).

²⁷The models included C+FR as a predictor, but nothing would change if we included (L)HR instead. Since they are complementary and thus a linear function of one another, the results would be the same, just with different signs and scaling.

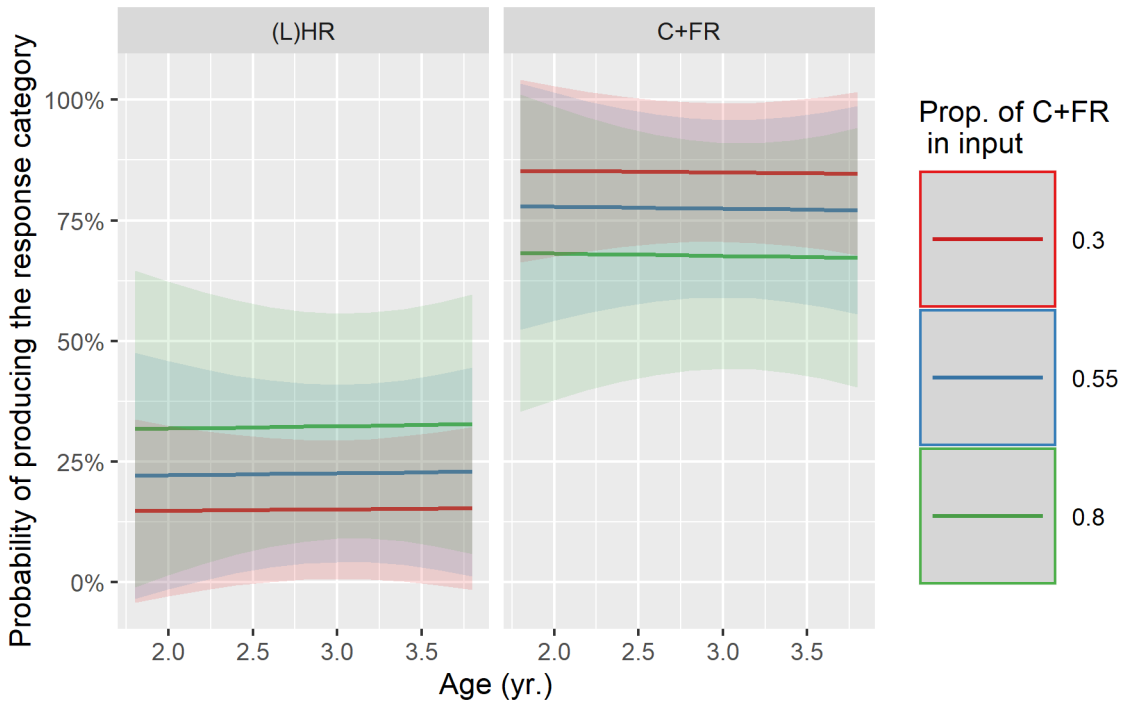


Figure 2: Probability of producing a given wh-category (panel) as a function of age (x-axis) and proportion of C+FR input (color)

- (20) **Wh-complexity hierarchy** (simplified for purposes of statistical analysis)
 $I < C+FR < (L)HR$

As expected, we see a clear primacy of interrogative wh-words, which are not only the most frequent ones, but also the earliest to be acquired. This is confirmed by our ordinal regression analysis, which revealed a strong effect of age of acquisition (interrogatives used earlier than the more complex categories). When interrogatives are dropped from the analysis, however, we see no statistical evidence for the ordering between C+FR on the one hand and (L)HR on the other. The numerical tendencies, which can be inspected in Table 8, are consistent with the simplified hierarchy (20).

An important part of assessing our hypothesis was examining the effect of the parental input: if children’s behavior mirrors parental input, there is little left to explain for any alternative hypothesis. It could be, for instance, that children start producing more complex wh-functions after their frequency rises in the parental input. Relying on lagged correlation analyses (in the spirit of Granger’s 1969 causality), we demonstrated that there is no such simple correlation. What our models show instead is that the increase of children’s production of more complex wh-functions (C+FR and (L)HR in particular) is positively correlated with the proportion of parental interrogative input (and negatively with C+FR/(L)HR input) – the more interrogatives the parents use (and the less C+FR/(L)HR), the more likely the children are to produce wh-words in the C+FR/(L)HR functions – an effect that gets more robust with the children’s age. A parallel correlation – albeit without the interaction with age – is observed when the interrogatives are dropped from the analysis: the likelihood of producing (L)HR wh-words is positively correlated with the proportion of C+FR in the input (and negatively with (L)HR input). To put

it more generally – a higher proportion of less complex functions in the input seems to foster the production of more complex functions. This observation does not seem to be easily reconciled with a simple frequency-based account along the lines of Ambridge et al. (2015), while it is consistent with the entertained growing trees hypothesis, where the assumption is that children gradually mature to acquire new – incrementally more complex – syntactic and semantic structures. Note that this assumption does not exclude effects of input in general. Children might acquire the wh-hierarchy independently of input language from some endogenous sources, such as the conceptual system, but it can also be implied by complex relations between the input frequency of wh-words and other word classes, or by other phenomena in the input. This is consistent with different types of emergentist approaches to language acquisition (MacWhinney 2004; Steedman 2017).²⁸

Overall, our findings are consistent with those reported in the literature. Interrogative wh-words are the first to be used, with more complex wh-functions only appearing later (see Section 3). We were not able to statistically support the reported developmental primacy of free relatives (vs. headed relatives; see Flynn & Foley 2004), although our data are consistent with it and the descriptive results point in the expected direction. It must be emphasized, however, that the majority of free relative (and correlative) instances in our corpus were cases of adverbial ‘when’-clauses, rather than “genuine” entity-denoting free relatives (introduced, e.g., by ‘what’). Corpus thus does not seem to be the optimal tool for studying free relatives or correlatives (in Czech), as they are simply too infrequent.

Our findings corroborate and complement the recent findings of Friedmann & Reznick (2021); Friedmann et al. (2021), based on Hebrew. More particularly, our data support the finding that interrogatives are mastered earlier than other wh-constructions, but also suggest the developmental primacy of conditionals (including correlatives and free relatives) over (light-)headed relatives.

What is worth noting is the early acquisition and use of wh-based (light-)headed relatives in Czech. At first sight, Czech is similar to both French (Labelle 1990) or English (Diessel 2004) in that it has both wh-(pronoun-)based relative clauses (...*the man who*...) and complementizer-based relatives (...*the man that*...). But while wh-based relatives start appearing quite late in these two languages, showing a clear preference for complementizer-based relatives, the situation in Czech appears to be reversed, with wh-based relatives being acquired very early (around the age of 3). Two factors come to mind as possible sources of this cross-linguistic asymmetry. First, the English and French complementizers (*that* and *que*, respectively) have a broader distribution: they can be used as relative, but also as declarative complementizers (cf. *he said that*.../ *the claim that*...); in Czech, the relative and the declarative complementizers are different (*co* lit. ‘what’ and *že*, respectively). Second, English and French wh-based relative pronouns dominate in registers to which children are typically not exposed (see, e.g., Biber et al. 1999; Blanche-Benveniste 1990), while in Czech they are more frequent than relative complementizers even in everyday conversations (as evidenced by an auxiliary search in the oral v1 corpus; Kopřivová et al. 2017).

As correctly pointed out by an anonymous reviewer, there are many predictions made by the growing trees approach that are not easy to test relying just on naturalistic data.

²⁸An anonymous reviewer suggests that input might play a more nuanced role than we suggest; in particular, maybe children “need the frequency of non-interrogative wh to reach a tipping point for acquisition to get started.” While we find this hypothesis plausible and worth exploring, it would require a much richer dataset (possibly richer by orders of magnitude) – both in terms of the number of children and the amount of recorded material – to be able to estimate the envisaged tipping point, or rather tipping points (different for each child).

One such prediction concerns the structural position of *wh*-words in the constructions under investigation: Is it really the case that fronted *wh*-words in interrogatives are structurally lower than those in headed relatives? Based on our data, we cannot tell. Clearly, our argument from the order of acquisition and the influence of parental input is just one of many that could be offered in support of the growing trees hypothesis. We hope that future research can complement our argument with more specific ones.

7 Conclusion

Wh-words are at heart of many languages' grammatical systems. They are primarily used for marking the locus of ignorance in interrogative constructions, but often they participate in the formation of other lexical items (indefinites) and constructions, including various types of relative clauses, correlatives, or conditionals. Grammatical and typological studies have contributed much to our understanding of how the individual uses of *wh*-words relate to one another – which ones are more basic and which likely to be derived. The literature on L1-acquisition of *wh*-words has, for the most part, concentrated on the development within a single construction type, yielding important results related to subject–object asymmetries or to the acquisition of the individual *wh*-sorts (thing, human, place, etc.).

The present work contributes to a less voluminous tradition – one concerned with the speed and order of acquisition across different *wh*-construction types. It does so by looking at Czech, a Slavic language with a broad distribution of *wh*-words and one whose L1-acquisition is understudied. In our research we have aimed to address the GROWING TREES hypothesis (Friedmann et al. 2021), a recent incarnation of an approach to acquisition according to which children acquire the syntax and semantics of complex constructions gradually – by incrementally extending the functional sequence of clauses, whereby new production and comprehension options open up as new functional heads are added on the top of the already acquired syntactic tree. The growth of the syntactic tree is matched by the growth of the semantic competence. Building on previous (grammatical) research, we have postulated a *wh*-complexity hierarchy, according to which the interrogative *wh*-construction is the most basic one, being followed by correlatives/conditionals, free relatives, light-headed relatives, and finally the most complex (nominally) headed relatives.

Even though the Czech corpus we relied on turned out not to be large enough for evaluating the validity of the whole hierarchy, our results (descriptive and inferential) are consistent with a proper subpart of the hierarchy – with interrogatives being the most basic, (light-)headed relatives the most complex, and correlatives/conditionals and free relatives being in the middle. At the same time, we have seen that the pattern of *wh*-construction development is not easily accounted for by parental input, in the sense that increased input of a particular construction would be matched by increased production of that construction by children. What we find instead is that parental input of simpler constructions boosts the production of more complex ones. A possible interpretation of this finding is that children use parental input to solidify attained knowledge, while constructing a new layer of structure and meaning enabling them to produce new and more complex constructions.

While the present work delivers some promising results in the study of L1-acquisition of *wh*-constructions and syntax/semantics more generally, it also uncovers some important limitations of the method applied. Despite the relatively large sample (over 10 thousand

wh-words, all that the Czech CHILDES currently makes available), it is still not large enough to allow for drawing more detailed inferences about the acquisition of the selected wh-constructions. A chief reason for this is the high frequency of interrogatives, leaving only a fraction of occurrences to other wh-constructions. A much larger corpus would be needed to fully address our questions as well as alternative hypotheses about how the acquisition of wh-words and construction proceeds. We hope that the present study provides a useful methodological guide for similar ones in the future.

References

- Akatsuka, Noriko & Patricia M. Clancy. 1993. Conditionality and deontic modality in Japanese and Korean: Evidence from the emergence of conditionals. In Patricia M. Clancy (ed.), *Japanese/Korean linguistics*, vol. 2, 177–192. Stanford, CA: CSLI Publications.
- Alqurashi, Abdulrahman A. 2013. *Arabic free and restrictive relatives: A minimalist and an HPSG approach*. Colchester: University of Essex dissertation.
- Ambridge, Ben, Evan Kidd, Caroline F. Rowland & Anna L. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(2). 239–273. <https://doi.org/10.1017/S030500091400049X>.
- Arkadiev, Peter & Ivano Caponigro. 2021. Conveying content questions without wh-words: Evidence from Abaza. In Patrick Grosz, Luisa Martí, Hazel Pearson, Yasutada Sudo & Sarah Zobel (eds.), *Proceedings of Sinn und Bedeutung 25*, London: University College London and Queen Mary University of London. <https://doi.org/10.18148/sub/2021.v25i0.925>.
- Armon-Lotem, Sharon. 2008. The interaction between question formation and verbal morphology in the acquisition of Hebrew: A minimalist perspective. In Sharon Armon-Lotem, Gabi Danon & Susan Rothstein (eds.), *Current Issues in Generative Hebrew Linguistics*, 223–244. Amsterdam: John Benjamins.
- Bates, Elizabeth. 1976. *Language and context: The acquisition of pragmatics*. New York, NY: Academic Press.
- Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14(1). 1–56. <https://doi.org/10.1007/s11050-005-4532-y>.
- Bellugi, Ursula. 1971. Simplification in children’s language. In Renira Huxley & Elisabeth Ingram (eds.), *Language acquisition: Models and methods*, 95–119. New York: Academic Press.
- Belyaev, Oleg & Dag Haug. 2020. The genesis and typology of correlatives. *Language* 96(4). 874–907. <https://doi.org/10.1353/lan.2020.0065>.
- Benincà, Paola. 2001. The position of topic and focus in the left periphery. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 39–64. Amsterdam: North Holland.

- Berman, Stephen. 1991. *On the semantics and logical form of wh-clauses*. Amherst, MA: University of Massachusetts dissertation.
- Bertollo, Sabrina & Guido Cavallo. 2012. The syntax of Italian free relative clauses: An analysis. In Lena Baunaz & Tomislav Socanac (eds.), *Generative Grammar in Geneva 8*, 59–76. http://www.unige.ch/lettres/linge/syntaxe/journal/volume_huit_2012.html.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar and spoken and written English*. London: Longman.
- Biezma, María. 2020. Non-informative assertions: The case of non-optional wh-in-situ. *Semantics & Pragmatics* 13(18). 1–54. <https://doi.org/10.3765/sp.13.18>.
- Biskup, Petr & Radek Šimík. 2019. Structure of conditional and (cor)relative clauses: New evidence from locality. In Maggie Baird & Jonathan Pesetsky (eds.), *NELS 49: Proceedings of the 49th Annual Meeting of the North East Linguistic Society, Volume 1*, 135–144. Amherst, MA: GLSA Publications.
- Bittner, Maria. 2001. Topical referents for individuals and possibilities. In Rachel Hastings, Brandon Jackson & Zsofia Zvolenszky (eds.), *SALT 11: Proceedings from the 11th Conference on Semantics and Linguistic Theory*, 36–55. Ithaca, NY: Cornell University. <https://doi.org/10.3765/salt.v11i0.2854>.
- Blanche-Benveniste, Claire. 1990. Usages normatifs et non normatifs dans les relatives en français, en espagnol et en portugais. In Johannes Bechert, Giuliano Bernini & Claude Buridant (eds.), *Toward a typology of European languages* (Empirical Approaches to Language Typology [EALT] 8), 317–335. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110863178.317>.
- Bloom, Lois, Susan Merkin & Janet Wootten. 1982. Wh-questions: Linguistic factors that contribute to the sequence of acquisition. *Child Development* 53(4). 1084–1092. <https://doi.org/10.2307/1129150>.
- Bobaljik, Jonathan & Susi Wurmbrand. 2015. Questions with declarative syntax tell us what about selection? In Ángel Gallego & Dennis Ott (eds.), *50 years later: Reflections on Chomsky's Aspects* (MIT Working Papers in Linguistics 77), Cambridge, MA: MITWPL.
- Brasoveanu, Adrian. 2008. Uniqueness effects in correlatives. In Atle Grønn (ed.), *Proceedings of Sinn und Bedeutung 12*, 47–65. Oslo: ILOS.
- Brown, Roger. 1968. The development of wh questions in child speech. *Journal of Verbal Learning and Behavior* 7(2). 279–290. [https://doi.org/10.1016/S0022-5371\(68\)80002-7](https://doi.org/10.1016/S0022-5371(68)80002-7).
- Büring, Daniel. 2003. On D-trees, beans, and B-accent. *Linguistics and Philosophy* 26(5). 511–545. <http://www.jstor.org/stable/25001898>.
- Caponigro, Ivano. 2003. *Free not to ask: On the semantics of free relatives and wh-words cross-linguistically*. Los Angeles: University of California dissertation.
- Caponigro, Ivano & Lisa Pearl. 2009. The nominal nature of *where*, *when*, and *how*: evidence from free relatives. *Linguistic Inquiry* 40(1). 155–164. <https://doi.org/10.1162/ling.2009.40.1.155>.

- Caponigro, Ivano, Lisa Pearl, Neon Brooks & David Barner. 2012. Acquiring the meaning of free relative clauses and plural definite descriptions. *Journal of Semantics* 29(2). 261–293. <https://doi.org/10.1093/jos/ffr014>.
- Caponigro, Ivano, Harold Torrence & Roberto Zavala (eds.). 2020. *Headless relative clauses in Mesoamerican languages*. New York: Oxford University Press.
- Cecchetto, Carlo & Caterina Donati. 2015. *(Re)labeling*. Cambridge, MA: MIT Press.
- Christensen, Rune Haubo Bojesen. 2013. Ordinal: Regression models for ordinal data. R package version 2013.9-30. <http://www.cran.r-project.org/package=ordinal/>.
- Chromá, Anna, Jakub Sláma, Klára Matiasovitsová & Jolana Treichelová. under review. A morphologically annotated longitudinal corpus of spoken czech child-adult interactions. *Language Resources and Evaluation* <https://doi.org/10.21415/3ZNE-HX03>.
- Cinque, Guglielmo & Luigi Rizzi. 2010. The cartography of syntactic structures. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 51–65. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199544004.013.0003>.
- Citko, Barbara. 2004. On headed, headless, and light-headed relatives. *Natural Language & Linguistic Theory* 22(1). 95–126. <https://doi.org/10.1023/B:NALA.0000005564.33961.e0>.
- Clahsen, Harald. 1990. Constraints on parameter setting: A grammatical analysis of some acquisition in stages in German child language. *Language Acquisition* 1(4). 361–391. https://doi.org/10.1207/s15327817la0104_3.
- Clancy, Patricia, Terry Jacobsen & Marilyn Silva. 1976. The acquisition of conjunction: A cross-linguistic study. In *Papers and Reports on Child Language Development 12*, 71–80. Stanford, CA: Stanford University, Department of Linguistics.
- Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1). 37–46. <https://doi.org/10.1177/001316446002000104>.
- Courtney, Ellen H. 2006. Adult and child production of Quechua relative clauses. *First Language* 26(3). 317–338. <https://doi.org/10.1177/0142723706062677>.
- Cross, Charles & Floris Roelofsen. 2014. Questions. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*, Stanford, CA: Metaphysics Research Lab, Stanford University. <http://plato.stanford.edu/archives/sum2015/entries/questions/>.
- Dąbrowska, Ewa. 2000. From formula to schema: The acquisition of English questions. *Cognitive Linguistics* 11(1–2). 83–102. <https://doi.org/10.1515/cogl.2001.013>.
- Daskalaki, Evangelia. 2007. Free relatives in Greek: The interaction between case (mis)matches and relativization strategies. In Artemis Alexiadou (ed.), *Studies in the morphosyntax of Greek*, 251–293. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Daskalaki, Evangelia. 2020. Types of relative pronouns. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.), *Syntactic architecture and its consequences II: Between syntax and morphology*, 277–296. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.4280653>.

- Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4). 393–450. <https://doi.org/10.1023/B:LING.0000024420.80324.67>.
- Demirok, Ömer. 2017a. A compositional semantics for Turkish correlatives. In Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel & Edward J. Rubin (eds.), *WCCFL 34: Proceedings of the 34th West Coast Conference on Formal Linguistics*, 159–166. Somerville, MA: Cascadilla Proceedings Project. <http://lingref.com/cpp/wccfl/34/paper3308.pdf>.
- Demirok, Ömer. 2017b. Free relatives and correlatives in wh-in-situ. In Andrew Lamont & Katerina Tetzloff (eds.), *NELS 47: Proceedings of the Forty-Seventh Annual Meeting of the North East Linguistic Society*, Amherst, MA: GLSA Publications.
- Diessel, Holger. 2004. *The acquisition of complex sentences*. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2013. Position of interrogative phrases in content questions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/93>.
- É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press.
- Flynn, Suzanne & Claire Foley. 2004. On the developmental primacy of free relatives. In Aniko Csirmaz, Andrea Gualmini & Andrew Nevins (eds.), *MIT Working Papers in Linguistics 48: Plato's problem: Problems in language acquisition*, 59–69. Cambridge, MA: MITWPL.
- Flynn, Suzanne & Barbara Lust. 1980. Acquisition of relative clauses: Developmental changes and their heads. In *Cornell Working Papers in Linguistics 1*, 33–45. Ithaca, NY: Cornell University.
- Foley, Claire. 1996. *Knowledge of the syntax of operators in the initial state: The acquisition of relative clauses in French and English*. Ithaca, NY: Cornell University dissertation.
- Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2021. Growing trees: The acquisition of the left periphery. *Glossa: a journal of general linguistics* 6(1). 131. 1–38. <https://doi.org/10.16995/glossa.5877>.
- Friedmann, Naama & Yosef Grodzinsky. 1997. Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language* 56(3). 397–425. <https://doi.org/10.1006/brln.1997.1795>.
- Friedmann, Naama & Julia Reznick. 2021. Stages rather than ages in the acquisition of movement structures: Data from sentence repetition and 27696 spontaneous clauses. *Glossa: a journal of general linguistics* 6(1). 143. 1–69. <https://doi.org/10.16995/glossa.5716>.
- Granger, Clive W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3). 424–438. <https://doi.org/10.2307/1912791>.

- Guasti, Maria Teresa. 2000. An excursion into interrogatives in early English and Italian. In Marc-Ariel Friedemann & Luigi Rizzi (eds.), *The acquisition of syntax*, 105–128. Harlow: Longman. <https://doi.org/10.4324/9781315839899-5>.
- Guasti, Maria Teresa. 2002. *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.
- Guttman, Louis. 1944. A basis for scaling quantitative data. *American Sociological Review* 9(2). 139–150. <https://doi.org/10.2307/2086306>.
- Hagstrom, Paul. 2021. Questions and interrogatives: Exhaustivity vs quantificational variability. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas E. Zimmermann (eds.), *The Wiley Blackwell companion to semantics*, vol. 4, 2579–2606. Oxford: John Wiley & Sons. <https://doi.org/10.1002/9781118788516.sem044>.
- Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10(1). 41–53. <https://www.jstor.org/stable/25000703>.
- Hamburger, Henry. 1980. A deletion ahead of its time. *Cognition* 8(4). 389–416. [https://doi.org/10.1016/0010-0277\(80\)90002-5](https://doi.org/10.1016/0010-0277(80)90002-5).
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198235606.001.0001>.
- Haspelmath, Martin & Ekkehard König. 1998. Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 563–640. Berlin: Mouton de Gruyter.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.
- von Heusinger, Klaus. 1997. *Salienz und Referenz: Der Epsilonoperator in der Semantik der Nominalphrase und anaphorischer Pronomen*. Berlin: Akademie Verlag.
- Hirsch, Aron. 2016. A compositional semantics for *wh-ever* free relatives. In Nadine Bade, Polina Berezovskaya & Anthea Schöller (eds.), *Proceedings of Sinn und Bedeutung 20*, 341–358. <http://semanticsarchive.net/sub2015/SeparateArticles/Hirsch-SuB20.pdf>.
- Hladnik, Marko. 2015. *Mind the gap: Resumption in Slavic relative clauses*. Utrecht University dissertation. http://www.lotpublications.nl/Documents/390_fulltext.pdf.
- Huggard, Mattyas. 2011. On *wh*-(non-)movement and internal structures of the Hittite preposed relative clause. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd Annual UCLA Indo-European Conference*, 83–104. Bremen: Hempen.
- Huggard, Mattyas. 2015. *Wh-words in Hittite: A study in syntax-semantics and syntax-phonology interfaces*. Los Angeles, CA: University of California dissertation. <https://escholarship.org/uc/item/1gb9f3dg>.
- Hulsey, Sarah & Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics* 14(2). 111–137. <https://doi.org/10.1007/s11050-005-3799-3>.

- Jacobson, Pauline. 1995. On the quantificational force of English free relatives. In Emon Bach, Eloise Jelinek, Angelika Kratzer & Barbara Partee (eds.), *Quantification in natural languages, Vol II*, 451–486. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-2817-1_15.
- Jasinskaja, Katja & Radek Šimík. forthcoming. Slavonic free word order. In Jan Fellerer & Neil Bermel (eds.), *The Oxford guide to the Slavonic languages*, Oxford: Oxford University Press. <https://ling.auf.net/lingbuzz/004919>.
- Katis, Demetra. 1997. The emergence of conditionals in child language: Are they really so late? In Angeliki Athanasiadou & René Dirven (eds.), *On conditionals again*, 355–386. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/cilt.143.20kat>.
- Kellert, Olga. 2018. Questions with definite markers in (Old) Romance, with focus on Old Spanish. *Isogloss* 4(1). 55–84. <https://doi.org/10.5565/rev/isogloss.56>.
- Klima, Edward S. & Ursula Bellugi. 1966. Syntactic regularities in the speech of children. In John Lyons & Roger J. Wales (eds.), *Psycholinguistics Papers: The Proceedings of the 1966 Edinburgh Conference*, 183–208. Edinburgh: Edinburgh University Press.
- Kopřivová, Marie, David Lukeš, Zuzana Komrsková, Petra Poukarová, Martina Waclawíčová, Lucie Benešová & Michal Křen. 2017. *ORAL: Korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017*. Praha: Ústav českého národního korpusu FF UK. <https://www.korpus.cz>.
- Kotek, Hadas & Michael Yoshitaka Erlewine. 2019. Wh-indeterminates in Chuj (Mayan). *Canadian Journal of Linguistics* 64(1). 62–101. <https://doi.org/10.1017/cnj.2018.25>.
- Kratzer, Angelika. 1991. The representation of focus. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 825–834. Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110126969.10.825>.
- Krifka, Manfred. 2011. Questions. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, chap. 77, 1742–1785. Berlin: de Gruyter. <https://doi.org/10.1515/9783110255072.1742>.
- Labelle, Marie. 1990. Predication, wh-movement, and the development of relative clauses. *Language Acquisition* 1(1). 95–119. https://doi.org/10.1207/s15327817la0101_4.
- Lakshmanan, Usha. 2000. The acquisition of relative clauses by Tamil children. *Journal of Child Language* 27(3). 587–617. <https://doi.org/10.1017/S0305000900004293>.
- Lakshmi Bai, B. 1989. Relativization in child language and history. In Aditi Mukherjee (ed.), *Osmania University Publications in Linguistics 5: Language variation and language change*, 141–152. Hyderabad: Osmania University.
- Lau, Elaine & Nozomi Tanaka. 2021. The subject advantage in relative clauses: A review. *Glossa: a journal of general linguistics* 6(1). 34. 1–34. <https://doi.org/10.5334/gjgl.1343>.
- Lee, Kwee-Ock. 1991. *On the first language acquisition of relative clauses in Korean: The universal structure of COMP*. Ithaca, NY: Cornell University dissertation.

- Lehmann, Christian. 1984. *Der Relativsatz*. Tübingen: Gunter Narr Verlag.
- Lehmann, Winfred P. 1977. *Proto-Indo-European syntax*. Austin, TX: University of Texas Press.
- Lipták, Anikó (ed.). 2009. *Correlatives cross-linguistically*. Amsterdam: John Benjamins. <https://doi.org/10.1075/lfab.1>.
- Liu, Mingming. 2018. *Varieties of alternatives: Focus particles and wh-expressions in Mandarin*. Springer. <https://doi.org/10.1007/978-981-10-6208-7>.
- Lust, Barbara. 1994. Functional projection of CP and phrase structure parametrization: An argument for the strong continuity hypothesis. In Barbara Lust, Margarita Suñer & John Whitman (eds.), *Syntactic theory and first language acquisition: Cross-linguistic perspectives, Volume 1: Heads, projections, and learnability*, 85–118. New Jersey, NJ: Lawrence Erlbaum.
- Lust, Barbara C., Claire Foley & Cristina D. Dye. 2009. The first language acquisition of complex sentences. In Edith L. Bavin (ed.), *The Cambridge handbook of child language*, 237–258. Cambridge: Cambridge University Press.
- Lyutikova, Ekaterina & Andrei Sideltsev. to appear. Relative construction in Hittite: A corpus-based case study in syntax–prosody interface. *Journal of Historical Linguistics*.
- MacWhinney, Brian. 1995. *The CHILDES project: Tools for analyzing talk, Second edition*. Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, Brian. 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31. 883–914.
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- Meyer, Roland. 2017. The C system of relatives and complement clauses in the history of Slavic languages. *Language* 93(2). e97–e113. <https://doi.org/10.1353/lan.2017.0032>.
- Mitrenina, Olga V. 2012. The syntax of pseudo-correlative constructions with the pronoun *kotoryj* (‘which’) in Middle Russian. *Slověne* 1(1). 61–73. <https://doi.org/10.31168/2305-6754.2012.1.1.4>.
- Nirmala, C. 1980. Relativization in children’s speech. *Osmania Papers in Linguistics* 6. 53–70.
- Packard, Jerome L. 1988. The first language acquisition of prenominal modification with *de* in Mandarin. *Journal of Chinese Linguistics* 16(1).
- Partee, Barbara. 1987. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in discourse representation and the theory of generalized quantifiers, GRASS 8*, 115–143. Dordrecht: Foris.
- Polinsky, Maria. 2015. Tsez syntax: A description. Manuscript, University of Maryland. <https://ling.auf.net/lingbuzz/002315>.

- Rawlins, Kyle. 2013. (Un)conditionals. *Natural Language Semantics* 40(2). 111–178. <https://doi.org/10.1007/s11050-012-9087-0>.
- Reilly, Judy Snitzer. 1986. The acquisition of temporals and conditionals. In Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson (eds.), *On conditionals*, 309–331. Cambridge: Cambridge University Press.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94. <http://logica.ugent.be/philosophica/fulltexts/27-4.pdf>.
- van Riemsdijk, Henk. 2017. Free relatives. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax: Second edition*, chap. 116. Wiley Blackwell. <https://doi.org/10.1002/9781118358733.wbsyncom116>.
- Rizzi, Luigi. 1994. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3(4). 371–393. https://doi.org/10.1207/s15327817la0304_2.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: A handbook of generative syntax*, 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2001. On the position of “Int(errogative)” in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 267–296. Oxford: Elsevier North-Holland.
- Roelofsen, Floris. 2019. Semantic theories of questions. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*, Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.504>.
- Rooth, Mats. 1985. *Association with focus*. Amherst, MA: University of Massachusetts dissertation. <https://scholarworks.umass.edu/dissertations/AAI8509599>.
- Rowland, Caroline F., Julian M. Pine, Elena V. M. Lieven & Anna L. Theakston. 2003. Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language* 30(3). 609–635. <https://doi.org/10.1017/S0305000903005695>.
- Schwartz, Arthur. 1971. General aspects of relative clause formation. In *Working Papers on Language Universals 6*, 139–171. Stanford, CA: Stanford University. <https://eric.ed.gov/?id=ED094567>.
- Sharma, V. 1974. *A linguistic study of speech development in early childhood*. Agra: Agra University dissertation.
- Siddiqi, Daniel. 2018. Distributed morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.15>.
- Šimík, Radek. 2010. Free relatives at the interface. In Jan-Wouter Zwart & Mark de Vries (eds.), *Structure preserved: Studies in syntax for Jan Koster*, 321–328. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.164.34sim>.

- Šimík, Radek. 2011. *Modal existential wh-constructions*. Groningen: University of Groningen dissertation. http://www.lotpublications.nl/Documents/269_fulltext.pdf.
- Šimík, Radek, Klára Matiasovitsová & Filip Smolík. 2023. From correlativization to relativization: A view from Czech L1-acquisition. In Paris Gappmayr & Jackson Kellogg (eds.), *BUCLD 47: Proceedings of the 47th annual Boston University Conference on Language Development*, vol. 2, 702–714. Somerville, MA: Cascadilla Press. <https://www.lingref.com/buclD/47/BUCLD47-56.pdf>.
- Šimík, Radek. 2018. Ever free relatives crosslinguistically. In Uli Sauerland & Stephanie Solt (eds.), *Proceedings of Sinn und Bedeutung 22*, vol. 2, 375–392. Berlin: Leibniz-Center General Linguistics. <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/112>.
- Šimík, Radek. 2021. Free relatives. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas E. Zimmermann (eds.), *The Wiley Blackwell companion to semantics*, vol. 2, 1033–1070. John Wiley & Sons. <https://doi.org/10.1002/9781118788516.sem093>.
- Šimík, Radek. in press. From interrogatives to relatives: A comprehensive account of wh-constructions. In Sjeff Barbiers, Norbert Corver & Maria Polinsky (eds.), *The Cambridge Handbook of Comparative Syntax*, Cambridge University Press. <https://ling.auf.net/lingbuzz/007142>.
- Šimík, Radek, Petr Biskup, Kateřina Bartasová, Markéta Dančová, Eliška Dostálková, Kateřina Hrdinková, Gabriela Kosková, Jaromír Kozák, Klára Lupoměská, Albert Maršík, Edita Schejbalová & Illia Yekimov. 2022. Extraction from clausal adjuncts in Czech: A rating study. In Breanna Pratley, Özge Bakay, Eva Neu & Peyton Deal (eds.), *NELS 52: Proceedings of the Fifty-Second Annual Meeting of the North East Linguistic Society, Volume Three*, 109–123. Amherst, MA: GLSA Publications. <https://ling.auf.net/lingbuzz/006632>.
- Somashekar, Shamitha. 1999. *Developmental trends in the acquisition of relative clauses: Cross-linguistic experimental study of Tulu*. Ithaca, NY: Cornell University dissertation.
- Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Knut Tarald Taraldsen (eds.), *Tromsø Working Papers on Language and Linguistics: Nordlyd 36.1 [Special issue on nanosyntax]*, 1–6. Tromsø: CASTL. <http://www.ub.uit.no/baser/nordlyd>.
- Steedman, Mark. 2017. The emergence of language. *Mind & Language* 32(5). 579–590. doi:<https://doi.org/10.1111/mila.12160>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12160>.
- Steinkrauss, Rasmus. 2009. *Frequency and function in wh question acquisition: A usage-based case study of German L1 acquisition*. Groningen: University of Groningen dissertation.
- Tanaka, Nozomi, Elaine Lau & Alan L. F. Lee. online first. On the universality of the subject preference in the acquisition of relative clauses across languages. *First Language* <https://doi.org/10.1177/01427237241226734>.

- Thornton, Rosalind. 2016. Acquisition of questions. In Jeffrey L. Lidz, William Snyder & Joe Pater (eds.), *The Oxford handbook of developmental linguistics*, 310–340. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199601264.013.14>.
- Truswell, Robert & Nikolas Gisborne. 2015. Quantificational variability and the genesis of English headed wh-relatives. In Eva Csipak & Hedde Zeijlstra (eds.), *Proceedings of Sinn und Bedeutung 19*, 638–655. University of Göttingen: Göttingen. <https://doi.org/10.18148/sub/2015.v19i0.255>.
- Veselovská, Ludmila. 2021. *Wh-questions: A case study in Czech*. Olomouc: Palacký University Press. <https://doi.org/10.5507/ff.21.24459653>.
- de Vries, Mark. 2005. The fall and rise of universals on relativization. *Journal of Universal Language* 6(1). 125–157. <https://doi.org/10.22425/jul.2005.6.1.125>.
- Williamson, Gregor. 2019. Conditional antecedents as polar free relatives. In Katherine Blake, Forrest Davis, Kaelyn Lamp & Joseph Rhyne (eds.), *SALT 29: Proceedings of the 29th Semantics and Linguistic Theory Conference*, 496–508. Washington, DC: Linguistic Society of America. <https://doi.org/10.3765/salt.v29i0.4615>.
- Zaefferer, Dietmar. 1991. Conditionals and unconditionals: Cross-linguistic and logical aspects. In Dietmar Zaefferer (ed.), *Semantic universals and universal semantics*, Dordrecht: Foris.
- Zimmermann, Malte. 2018. Embedded questions and concealed relative questions in Hausa and Akan. In Elizabeth Bogal-Allbritten & Elizabeth Coppock (eds.), *Proceedings of TripleA 4*, 1–16. Tübingen: University of Tübingen. <https://doi.org/10.15496/publikation-24505>.