**L1-acquisition of wh-functions: A case study of Czech**

*Radek Šimík and Klára Matiasovitsová* (Charles University)

**Background on wh-functions** Wh-words are multi-functional expressions used in interrogatives (matrix/MI and embedded/EI), (un)conditionals/(Un)C, different kinds of relatives (correlatives/CoR, free relatives/FR, light-headed relatives/LHR, headed relatives/HR), and indefinites. These functions are not on a par. Various asymmetries have been observed, which are summarized in what we call the wh-function hierarchy in (1) (indefinites are left aside). For example, Lehmann (1984) observed that if a language uses wh-words in HRs, then it can also do so in FRs or LHRs, but not conversely; later research (Authors) has shown that analogous implications tend to hold for other cut-off points of the hierarchy. Schwartz (1971) observed that wh in-situ is allowed in Is and CoRs, but not in (L)HRs; later research has shown that while wh in-situ is possible also in UnCs, it is extremely rare if possible at all in FRs (Demirok 2017). As a final example, the morphology of wh-words gets more complex for functions farther right on the hierarchy (e.g. addition of *a-* to the interrogative wh-base in Hungarian around the CoR level or the switch from *w-* to *d-*morphology in German around the LHR level); Authors.

(1) *Wh-function hierarchy:* MI > EI > (Un)C > CoR > FR > LHR > HR

**Background on wh-acquistion** The hierarchy in (1) is valid for typological, syntactic, morphological, and diachronic generalizations. The existing evidence from L1-acquisition is limited but consistent with (1): the findings reported in Flynn & Lust (1980), Labelle (1990), or Guasti & Shlonsky (1995) suggest that wh-words in Is are acquired much earlier than HRs (in English and French); Clauss (2017) furthermore finds that FRs are acquired in between the two (in English), suggesting the hierarchy I > FR > HR – a proper part of (1). To our knowledge, a study considering more than two functions at a time is missing.

**Present study** Czech is particularly well-suited for the study of wh-function acquisition, as it uses wh-words in all the functions in (1). Our hypothesis is that children acquire wh-functions in line with (1), i.e., they should start with using wh-words for Is and acquire the (L)HR function of wh-words as last, with functions such as C or FR in between. The two independent variables are thus WH-FUNCTION (with levels corresponding to (1)) and AGE (two subsequent and equally long age periods); the dependent variable is the FREQUENCY of individual wh-functions in the two periods. We look at how the frequency changes with age.

**Corpus and method** Our data come from Czech CHILDES (Chromá et al. 2020) and particularly the Viktor subcorpus. Viktor was recorded from 2;6.23 until 3;8.06 [Y;M.D] during 32 sessions. There is a total of 29853 child tokens and 39267 parent tokens. Using CLAN we extracted all wh-tokens, including sufficient context, to allow for annotation. The data were annotated for WH-FUNCTION (and a number of auxiliary categories, to be detailed in presentation) by the second author and then controlled by the first author; the few disagreements were resolved by consensus.

**Results** Tab. 1 shows the observed and expected (bracketed) frequencies of all attested functions in child speech in two age periods (the few CoR have been added to C). Viktor produced all the listed wh-functions from the onset of recording. Therefore, we will draw inferences from the observed rises/dips in wh-function frequency from one age period to the next.

| | MI | EI | C | FR | LHR | HR | Total |
|---|---|---|---|---|---|---|---|
| 2;06–3;01 | 286(253) | 19(33) | 20(27) | 21(19) | 3(9) | 28(36) | 377 |
| 3;02–3;08 | 294(327) | 56(42) | 42(35) | 23(25) | 18(12) | 54(46) | 487 |
| Total[%] | 580[67] | 75[9] | 62[7] | 44[5] | 21[2] | 82[9] | 864 |

Table 1: Observed (expected) frequencies of wh-functions in two age periods (child)

Since the frequencies are relatively small, we analyze the sample based on Tab. 2, where we have collapsed two adjacent function pairs – C+FR and LHR+HR. The former collapse is motivated by the fact that the absolute majority of FR were string-identical to temporal/conditional clauses (introduced by *když,* differing only in the attachment – left-peripheral for C and right-peripheral for FR). The latter collapse is motivated by the very low frequency of LHR but also by the fact that both LHR and HR exhibit a comparable frequency rise. We refrain from collapsing MI and EI because their frequency change profile is very different. In result, Tab. 2 shows a rise (↑) in (L)HR and EI frequencies and a dip (↓) in MI frequency (all $p < 0.006$; sign. level Bonferroni-corrected to $p = 0.0125$). The frequency of C+FR remains constant ($\sim$; $p = 0.3$).

| | MI | EI | C+FR | LHR+HR | Total |
|---|---|---|---|---|---|
| 2;06–3;01 | 286(253) | 19(33) | 41(46) | 31(45) | 377 |
| 3;02–3;08 | 294(327) | 56(42) | 65(60) | 72(58) | 487 |
| Total[%] | 580[67] | 75[9] | 106[12] | 103[12] | 864 |
| rise \| dip \| level $\chi^2(1)$ | ↓ 7.60 | ↑ 10.21 | $\sim$ 1.06 | ↑ 7.67 | |

Table 2: Observed (expected) frequencies of wh-functions in two age periods (child; function collapse)

Tab. 3 shows adult frequencies in the two periods. The overall distribution of wh-words to functions is similar to the one in Tab. 2 and so are some frequency shifts, particularly the drop of MI and rise of EI; there is also a steep rise in C+FR frequency (all $p < 0.01$). The frequency of (L)HR remains constant in this sample.

| | MI | EI | C+FR | LHR+HR | Total |
|---|---|---|---|---|---|
| 2;06–3;01 | 1079(1017) | 128(148) | 68(105) | 78(82) | 1353 |
| 3;02–3;08 | 609(671) | 118(98) | 107(70) | 58(54) | 892 |
| Total[%] | 1688[75] | 246[11] | 175[8] | 136[6] | 2245 |
| rise \| dip \| level $\chi^2(1)$ | ↓ 9.41 | ↑ 6.97 | ↑ 33.5 | $\sim$ 0.48 | |

Table 3: Observed (expected) frequencies of wh-functions in two age periods (adults; function collapse)

**Discussion** The results are partly consistent with our hypothesis and with existing observations. The MI wh-function is clearly dominant at an early stage of acquisition (here, until 3;01), accounting for 76% of all wh-tokens. Just about half a year later the MI frequency drops to 60% and other wh-functions gain momentum, particularly (L)HR (from 8% to 15%). The frequency of C and FR, which are in between MI and (L)HR on the wh-hierarchy in (1), remains stable in the observed period (approx. 12%). The function that stands out is EI, which behaves more on a par with (L)HR than MI or C. We hypothesize that an additional factor – on top of (1) – is embedding; a comparison with embedded declaratives would be needed to test that hypothesis. A comparison to the adult data reveals a surprisingly adult-like performance of Viktor. The adult data show that the initial high frequency of MI in children can be attributed to input: the overall frequency of MI in the adult sample is very high, esp. in the earlier period (80%). Finally, this pilot research demonstrates (i) that more data is needed to draw inferences about individual wh-functions due to their overall low frequency, (ii) that a child can acquire all core wh-functions at a surprisingly young age; this must be taken into account if we wish to observe the onset of wh-acquisition. Both these limitations will be addressed in future research.

**Selected references • Chromá A, F Smolík & K Matiasovitsová 2020.** Chromá Czech corpus. CHILDES Talkbank. • **Clauss, M 2017.** The form and acquisition of free relatives. UMass PhD dissertation. • **Demirok, Ö 2017.** Free relatives and correlatives in wh-in-situ. *Proceedings of NELS 47.* GLSA Publications. • **Flynn, S & B Lust 1980.** Acquisition of relative clauses. *Cornell Working Papers in Linguistics 1,* 33-45. • **Lehmann, C 1984.** *Der Relativsatz.* Gunter Narr. • **Schwartz, A 1971.** General aspects of relative clause formation. *Working Papers on Language Universals 6,* 139-171. Stanford University.