

## Caregiver modelling and feedback in the context of early questions

**Authors: Johannes Heim, Maryam Bala & Arabella Sinclair**

We aim to shed new light on the question of what children can learn from caregiver input in language development, specifically through identifying the pragmatic properties of questions. Following accounts that situate language learning in collaborative interaction (e.g. Clark 2019), we depart from much previous literature (e.g. Marcus 1993) by expanding the distinction of positive vs. negative evidence, i.e. providing adultlike data vs. corrective feedback, into a fourfold distinction of caregiver intervention. Support for such a distinction comes from a largescale analysis of question-answer sequences between children and their caregivers on the English section of the CHILDES database (McWhinney 2000) over the course of 12 to 48 months of development. To facilitate our data-driven analysis, we automate part of the data processing, leading to more efficient human annotation, which in turn we plan to use to develop classifiers to automatically detect different feedback types, investigating whether properties extracted from large language models (LLMs) can serve as additional features.

Previous explorations of caregiver feedback have focused on corrective or communicative feedback to non-adult syntax in mostly small-scale investigations. What has been overlooked is the richness of conscious modelling and exemplifying of question-answer behaviour that gives rise to a novel typology of intervention. We propose that caregiver intervention includes both proactive and reactive elements that transcend the traditional divide between positive and negative evidence. We submit that many of the child-directed questions are not information-seeking. Instead, caregivers and children cooperate in rehearsing adult-like question-answer behaviour. The resulting typology of modelling, exemplifying, correcting, and affirming such behaviour therefore grounds in a distinction of the context of this behaviour, of who displays this behaviour, and of how the caregiver responds to the context or behaviour.

	<i>Type of Caregiver Intervention</i>			
	<b>Modelling (1)</b>	<b>Exemplifying (2)</b>	<b>Correcting (3)</b>	<b>Affirming (4)</b>
<b>Context</b>	event	child question	child question	event
<b>Sequence</b>	caregiver	caregiver	caregiver	child
<b>Feedback</b>	none	positive	negative	positive

We use this typology to explore how advancement in language learning is affecting caregiver behaviour. Natural language processing (NLP) provides us with the means to trace distribution and development of caregiver intervention across many children and conversations, thereby helping us understand the proactive role of caregivers in shaping the syntactic and pragmatic development of their children. Question-answer pairs serve particularly well for this task as they usually coincide with turn-takes (Casillas & Frank 2014), a reliable sequence of punctuation marks, and a constrained pragmatic phenomenon with large lexical overlap, lending credit to the notion of question-answer congruence (Reich 2002). This makes identifying adultlike question-answer behaviour sufficiently predictable for NLP. We follow Hiller & Fernandez (2016) in using lexical overlap as primary means of identifying repetitions across 2,026 transcripts from 46 monolingual, English-speaking children from age 12 to 48 months (mean  $MLU_{word}$  1-7.353) on the CHILDES database. Punctuation and turn-taking distribution between children and caregiver in adjacent utterances helped distinguishing the different types of intervention. This filter resulted in identifying 15,709 raw question-answer sequences, of which 800 were manually annotated by the first author. Our analysis (see Figs. 1 & 2) shows clear decline of modelling and correcting while affirming and exemplifying show inverse skewedness in parallel to what the child needs in development.

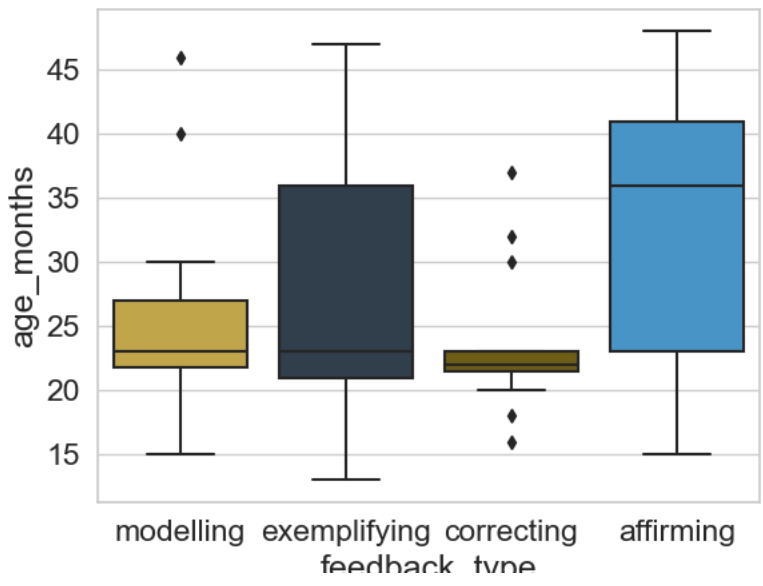


Fig. 1: Age range per intervention type

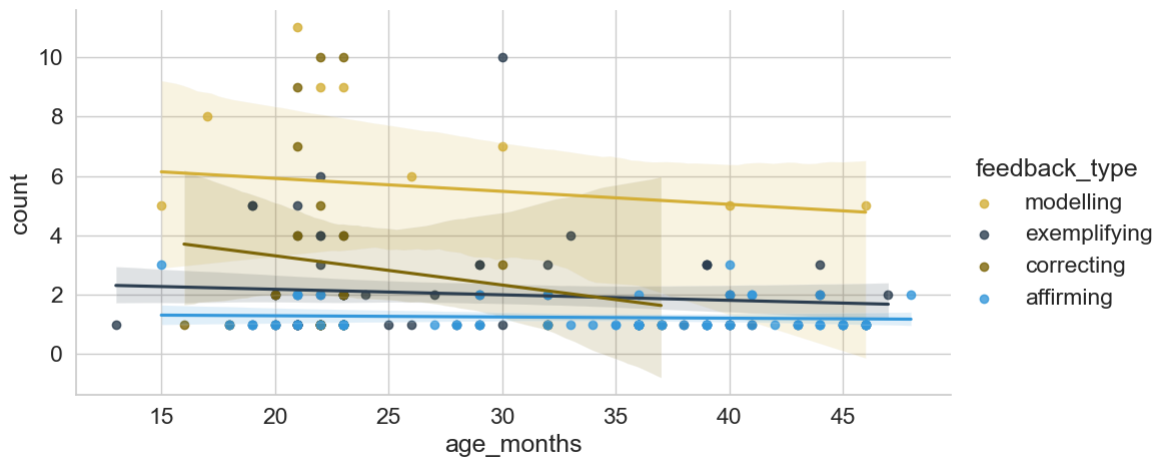


Fig. 2: Count of caregiver interventions per type by age in months