Because is why: Children's acquisition of topoi through why questions

Christine Howes, Ellen Breitholtz and Vladislav Maraev
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
christine.howes@gu.se, ellen.breitholtz@ling.gu.se, vladislav.maraev@gu.se

Abstract

In this paper we look at how children learn the underlying principles of commonsense reasoning, sometimes referred to as *topoi*, which are prevalent in everyday dialogue. By examining the utterances of two children in the CHILDES corpus for whom there is extensive longitudinal data, we show how children can elicit topoi from their parents by asking why-questions. This strategy for the rapid acquisition of topoi peaks at around age three, suggesting that it is a critical step in becoming a fully competent language user.

1 Introduction

Children pick up language with remarkable ease. From not being able to speak at all they learn in a few years to be fully competent language users. This does not just mean being able to communicate meaning coded in words and phrases – it also involves inference and association from the linguistic expressions and non-linguistic actions used to a meaning in use. Breitholtz (2020) discusses how such inferences draw on globally accepted facts ("the sun sets in the west"), norms ("one loves one's family"), and other principles of reasoning ("If you can do a and a is more difficult than b you can also do b"). Principles like these often implicitly underpin conversational moves, episodes in conversation and entire discourses, and have been claimed to be essential to capturing linguistic meaning in use (Ducrot, 1988; Anscombre, 1995). In their account Ducrot and Anscombre draw on Aristotelian dialectic and rhetoric, and use the term *topoi* (sg. *topos*) for such principles. Familiarity with the topoi that are acceptable in a community is also important for being proficient in a new language, as well as interpreting the behaviour of others. We see evidence of this in (1), discussed in Breitholtz and Howes (2020) where a father and son engage in a discussion about whether or not Lee, the son, could still play football even though he is not going to school because of illness. Both Dave and Lee are reasoning in a pragmatically competent way, despite evoking different topoi such as "one should rest when one is ill", "disease spreads less outdoors" (and possibly "fresh air is healthy") and "if one is well enough to do something less important and more exerting, one is also well enough to do something more important and less exerting".

(1) Dave: . . . you're gonna be home from football until four, you gonna have your dinner, want a bath.
 Lee: Yeah, but I might not go to school tomorrow.
 Dave: Why?
 Lee: Cos of my cough.
 Dave: How can you play football and not go to school then?
 Lee: Cos I was going out in the fresh air, I'm alright when I'm out in the fresh air.
 Dave: So why aren't you going to school then?
 Lee: I'm in the class room all day dad.
 [BNC KBE 10554-10561]

As this dialogue illustrates, a pre-teen child is capable of sophisticated argumentation drawing on principles which are also recognised by his adult discussion partner. However, the topoi Lee draws on in (1) have been learnt by him by explicit instruction, but also by via inference and induction. Breitholtz and Howes (2020) discuss how younger children, around four years of age, can be shown to have adopted topoi which they then generalise in non conventional ways. They also point out that one way for children to acquire topoi is through an extensive use of why-questions, and show that these peak at around 3 years (consistent with extensive evidence about children's stages of acquisition of wh-questions Bloom et al., 1982; Rowland

48

et al., 2003; Valian and Casey, 2003). In this paper we probe this finding by looking at the longitudinal use of why-questions by two children in the CHILDES corpus for whom extensive longitudinal data is available.

2 Background

2.1 Reasoning in Dialogue

Reasoning is essential in communication since interacting with others frequently involves making non-logical common-sense inferences linking context, background knowledge and beliefs to utterances in the dialogue in order to understand one another. These underpinning principles of reasoning – referred to as topoi – have been discussed at length in the literature on rhetoric and argumentation (e.g. Toulmin, 2003, a.o.). However, the idea of rules of thumb available to language users, which justify statements, suggestions or other types of utterances goes back to dialectic and rhetoric. In modern times, the concept of topos was introduced in linguistics as a theory of linguistic meaning where parts of discourse are perceived as connected by topoi (Ducrot, 1988). On this view the topoi accessible to an individual do not constitute a monolithic logical system, but represents a set of resources at the disposal of a dialogue participant for producing and interpreting utterances and discourse contributions. Breitholtz (2020) shows how a theory of topoi relates to semantic-pragmatic theories such as Gricean implicature theory and Relevance theory (Grice, 1975; Sperber and Wilson, 1995), and how it can explain puzzles such as bridging inferences and certain types of discourse coherence (Clark, 1975; Asher and Lascarides, 2003).[1] Consider for example the exchange in (1), where Lee is trying to persuade his father Dave that he is well enough to play football but not well enough to go to school: In (1), both Lee and Dave base their argument on a generally accepted topos that being ill restricts certain activities, with Dave drawing on topoi about exertion, like "if you can do x and x is more exerting than y, you can also do y" – in fact a version of the "more and the less"-topos mentioned in the introduction – and Lee on other topoi having to do with the spread of disease and the health benefits of fresh air. In this dialogue sequence we see that

---

[1]We should also note that, as pointed out by one of our reviewers, our approach theoretically and methodologically resonates with The Geneva Model of discourse analysis (see e.g. Filliettaz and Roulet, 2002).

an everyday conversation involves reasoning which cannot be accounted for using only traditional pragmatic theories where implicatures (Grice, 1975; Sperber and Wilson, 1995) are reached via assumptions of rationality and relevance. It also requires familiarity with a variety of topoi – principles about how it is acceptable to reason in different situations. Breitholtz and Howes (2020) suggest that topoi are learned through interaction with other agents and the world and show examples of where children draw on non-conventional topoi that they have learned by overextending inferences made in other instances of discourse. One such example is (2), where Greta, at 4 years and 3 months old in March 2020 demonstrates awareness of a topos related to the corona pandemic, namely that old people who contract the disease are more likely to die:

(2)   Greta: What would happen if you drank the sea water?
   Mother: It would make you poorly.
   Greta: Really poorly?
   Mother: Yes.
   Greta: Old people would die. I don't know about us though.
   [from Breitholtz and Howes (2020)]

In this example Greta overextends the topos that the elderly are more likely to die, if they contract coronavirus, to another situation where a young person would get ill.

2.2 Acquisition through interaction

Although traditional linguistics and developmental psychology started with the premise that there must be an innate language learning facility due to the presumed 'poverty of the stimulus' of a child's linguistic input (Berwick et al., 2011), there is a large body of evidence that refutes this position, from both a computational (Clark and Lappin, 2010) and a social perspective. This work (e.g. Halliday, 1975; Tomasello, 1992) emphasises the nature of language as action, and makes explicit the role of interaction in language acquisition. Specifically, research on child language acquisition underscores the importance of the social environment for the language learning child (Stephens and Matthews, 2014). Children are active in interactions with their caregivers long before they produce language and evidence suggests that it is this learning to interact (e.g. through gaze, Gredebäck et al., 2010; and turn-taking, Hilbrink et al., 2015; Casillas, 2014) which

bootstraps language acquisition (Levinson, 2006; Rączaszek-Leonardi et al., 2019). In a longitudinal study of the CHILDES-corpus (MacWhinney, 2000), Hiller and Fernández (2016) show that the type and amount of corrective feedback received by children affects their acquisition of particular grammatical phenomena. We hypothesise that adults' responses to why-questions and corrective feedback directed at topoi evoked by children will affect children's ability to seamlessly draw on topoi in conversation.

3 Why why?

Previous research shows that in adult conversation, topoi can be elicited by using why-questions (Schlöder et al., 2016). In many instances, what constitutes a good answer to a why-question constitutes an acceptable enthymeme when combined with the queried utterance, as is the case in (1), where "because of my cough" is an acceptable enthymematic reason for not going to school because of the acceptability of, for example, an underlying topos that when you are ill you should not mix with other people because you are contagious.

Where a dialogue participant cannot access or accommodate an appropriate topos, the asking of a why-question should be a particularly useful strategy to get one's interlocutor to make the topos more explicit. Indeed, when asked (fake) why-questions in a text-based dialogue experiment, people do provide the "missing" premises (Axelsson-Nord et al., 2021).

One reason that a dialogue participant may not have access to appropriate topoi is that there may be more than one applicable topoi available. However, in the case of young children it is often the case that a child lacks any topos that would make an argument coherent altogether.

We hypothesise that asking why-questions to increase the acceptable topoi one has access to is also a learning strategy for children, in line with evidence that children's why-questions are used for explanations and arguments (Bova and Arcidiacono, 2013)[2] once they have acquired a sufficient grounding in areas such as syntax (Cooper et al., 2023). Such a strategy – extrapolating and applying general principles of reasoning from minimal input (even when these go awry as in (2)) shows

---

[2]It should be noted that why-questions can be used to express frustration, and not seek reasons in any real sense, but we leave this distinction to one side for future work. We thank one of our anonymous reveiwers for this point.

how children are capable of utilising informative learning signals to learn from limited data.

4 Method

For this exploratory study, we used two longitudinal cases from CHILDES. The specific sources and their characteristics are described in Table 1 (Henry, 1995; Rowland, 2007; Lieven et al., 2009). These were chosen based on the data collection being sufficiently fine-grained, and covering the proposed critical period for why-question acquisition at around age 3 (as shown in Figure 1 taken from Breitholtz and Howes, 2020).

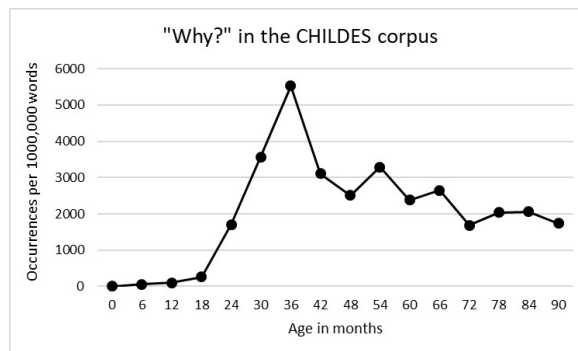| Source	Description
Lara	Eng-UK/Lara; 120 recordings between age 1;9.13 – 3;3.25 (at home)
Thomas	Eng-UK/Thomas; 379 recordings between age 2;0.12 – 4;11.20 (mostly at home)

Table 1: Sources of data used

Figure 1: Frequency of 'why' in child language by age

We used PyLangAcq (Lee et al., 2016) to process the data and extracted all uses of 'why' split between those produced by the child and those produced by any other dialogue participant. While this will inevitably also pick up instances of 'why' which do not result in the giving of reasons (e.g. "I don't know why she did it") we believe it is a reasonable starting point for analysis with more fine-grained study left aside for future work. For comparison, we also extracted instances of 'because', once again split by child/any other dialogue participant. Because is often use to provide explanations (Eaton et al., 1999), and can thus also be analysed as making enthymemes more explicit in dialogue (as seen in example 1, where Lee respond
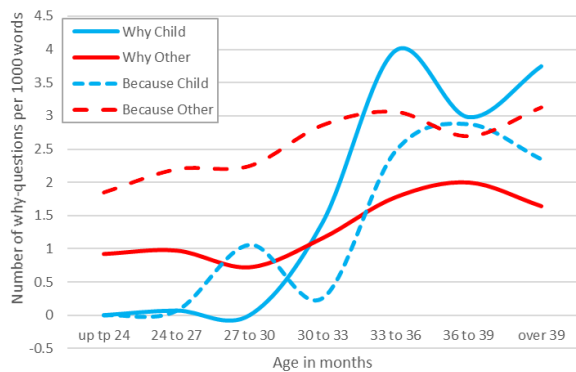
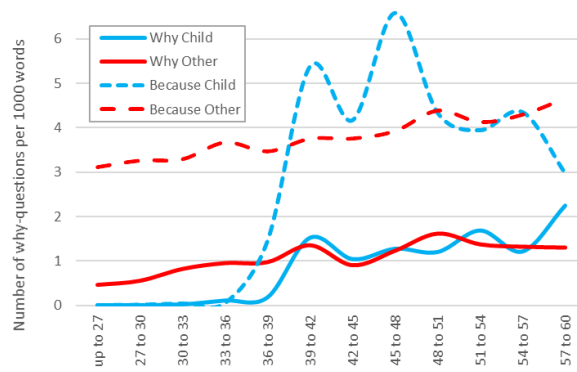Figure 2: Lara: Frequency of 'why' and 'because' per 1000 words by age in months

Figure 3: Thomas: Frequency of 'why' and 'because' per 1000 words by age in months

to his father's questions with because clauses that serve to illuminate the topoi that Lee is relying on in the dialogue).

5 Results & Discussion

As can be seen in Figures 2 and 3, both children have a peak of why-questions. For Lara, this occurs between 27 and 36 months whilst for Thomas this occurs between 36 and 42 months. Interestingly, while both show a distinct peak in why-questions (from asking none prior to this peak), this is relatively lower for Thomas, who peaks at approximately 1.5 why-questions per 1000 words, compared to Lara's 4. These differences in individual children are not apparent in the data shown in Figure 1. In this regard it is also informative to consider the input each child received in terms of why-questions, since acquisition of wh-questions in general has been linked to the input from the caregiver(s) (Rowland et al., 2003). While we cannot, of course, extrapolate from the available data to the total exposure of each child to why-questions, it is notable that in the available data, Thomas is also exposed to fewer why-questions than Lara, though the general patterns of why-questions they encounter is similar in both cases, rising steadily as the child asks more why-questions themselves. More fine-grained analysis is necessary to see how and whether these apparent contingencies have a direct impact on the child's interactive behaviour.

In terms of the use of because, once again we see that both children have a peak at around the same age range. Interestingly though, Thomas' peak in the use of 'because' coincides with his peak in the

use of 'why', but is greater (in the order of 5-6 words per 1000). This corresponds to the relatively greater input of 'because' from other speakers that we see in the Thomas data, as compared to Lara.

Interestingly, these children seem to have potentially different strategies for acquiring topoi, with the differences not fully explainable by broad differences in input that we have looked at here.

5.1 Qualitative results

We now turn to some examples from the Lara corpus to illustrate how why-questions can elicit topoi. Example (3) is an early example of a why-question from Lara, which does elicit a topos regarding what types of behaviour are naughty. Note that in this case a similar question to an adult might have instead been answered by providing some motivation for Peter Rabbit's naughtiness, rather than the topos supplied here. We hypothesise that this is because of the expectation that competent adult users of the language will already have access to a topos which licences "stealing (lettuces from Mr McGregor's garden) is naughty", so the question would in fact be interpreted differently if the asker were an adult. This suggests that people interacting with small children who lack some rhetorical resources, are sensitive to this fact (even if this is not a conscious awareness), but this is an empirical question which future work should try to investigate.

(3) CHI: is that rabby all by himself ?
    MOT: yes . he's in Mr McGregor's garden .
         he's naughty . [. . .]
    CHI: is he naughty ?
    MOT: yes .

51

CHI: why ?
MOT: why what ?
CHI: why he naughty ?
MOT: because he's gone into Mr McGregor's garden and he's stealing all his carrots and that's naughty , isn't it ?

[Lara 2;08;02]

Examples (4) and (5) are examples of the child asking more than one why-question in a row. This behaviour is very familiar to parents of children of around three, and we suggest that the initial answer may not satisfy the child's desire to access or accommodate an appropriate topos. Further work is needed to see how common such chains of why-questions are and whether these also occur at a critical age point or around the acquisition of particularly complex topoi.

(4) ELS: watch you don't break them now , Lara darling .
ELS: cause
CHI: why ?
ELS: cause Auntie Linda bought me them .
ELS: because I hafta look after them .
ELS: be very careful with them .
ELS: that's a good girl .
CHI: why ?
ELS: because I like them .
ELS: they're my special things .
CHI: them your special things ?
ELS: yes . *[Lara 2;10;14]*

Example (5) is also interesting as the topos it conveys is a normative one that in this particular family there is a 'rule' that one does not open a new treat if you already have one open. This may be a common rule in families, but there may also be differences between how children acquire such normative principles as opposed to, for example, globally accepted facts (for example, children learn that if you drop something it falls to the ground in the preverbal period). It is also not clear that children make this distinction at around the age they are producing a lot of why-questions. In example 6, Lara produces an enthymematic utterance which is underpinned by a normative topos which her mother rejects (that you have to have gloves on if you're gardening). In this case, Lara's why-questions seem to be targeting finding out what it is about this situation which means the normative topos that she has previously acquired (when you do gardening you wear gloves) does not apply.

(5) DAD: you're not opening that one until you've eaten all that one .
DAD: that's the
CHI: why ?
DAD: that's the rules , isn't it ?
CHI: why ?
DAD: er if you go in there and open it in there you're gonna be in big trouble .
MOT: we'll take it away from you .
DAD: you won't eat it .
CHI: pardon me ?
DAD: did you hear what I said ?
MOT: if you open it mummy will eat it .
DAD: do you hear what I'm saying , sugar ?
CHI: yes . *[Lara 2;10;14]*

(6) MOT: I was looking at what else we could we could plant in the garden .
CHI: you've gotta have gloves on .
CHI: but I haven't got any
MOT: you don't hafta have gloves on .
CHI: why ?
MOT: well .
MOT: you don't hafta .
CHI: why ?
MOT: well .
MOT: it's only (be)cause your hand get dirty .

[Lara 3;2;11]

6   Conclusions

As we have shown, children tend to have a peak of why-questions at around 3 years of age, which we speculate is due to their rapid acquisition of topoi at around this age. The development of the two children we have looked at in this paper is consistent with this. They also exhibit a peak in the use of 'because', although even in our small sample, the ways in which they use the available resources differs between the children suggesting there may be different pathways to acquiring topoi. One hypothesis is that if children are exposed to more explicit topoi (in the form of 'because' explanations) they may not have such a necessity to ask explicit why-questions. Further exploration of children's use of linguistic and pragmatic markers and their relationship to the interactive input is necessary to further elucidate these issues.

One of our plans for future research is to look at if and how 'why' and 'because' are complementary (and to what degree). These are intuitively codependent strategies (if you ask me why, I might expect

a because), but how productive these strategies are has, to the best of our knowledge, not previously been investigated – particularly in child language data. More fine-grained analysis is necessary to investigate whether there are other aspects of the children's pragmatic acquisition strategy that co-varies with these two linguistic markers, but we leave such analysis to future work.

Although our data does not conclusively say that children whose parents use more 'because' explanations do the same, they suggest a connection between child behaviour and the behaviour of care givers in this respect. However, further work is needed to look at the relations between the frequency of why-questions and because-clauses in the language produced by caregivers and children.

Example (6) also suggests avenues for future work, since it indicates that the child has already acquired a topos and is now concerned with how far this topos can be generalised (in this case, the child has learnt that one usually wears gloves when gardening to keep one's hands clean, but that this is not necessary). Learning the scope and range of topoi is a critical –and non-trivial– task for the language learning child, as demonstrated by Greta's overextension of the topos in (2). It is noteworthy in this regard that young children are able to pick up and modify the topoi they have access to from very little input – something that is still beyond the capabilities of conversational AI.

Acknowledgements

References

Jean-Claude Anscombre. 1995. La théorie des topoi: Sémantique ou rhétorique? *Hermés*, 15.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Ebba Axelsson-Nord, Vladislav Maraev, Ellen Breitholtz, and Christine Howes. 2021. What do you mean? Eliciting enthymemes in text-based dialogue. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*. SEMDIAL.

Robert C. Berwick, Paul Pietroski, Beracah Yankama, and Noam Chomsky. 2011. Poverty of the stimulus revisited. *Cognitive Science*, 35(7):1207–1242.

Lois Bloom, Susan Merkin, and Janet Wootten. 1982. "wh"-questions: Linguistic factors that contribute to the sequence of acquisition. *Child development*, pages 1084–1092.

Antonio Bova and Francesco Arcidiacono. 2013. Investigating children's why-questions: A study comparing argumentative and explanatory function. *Discourse Studies*, 15(6):713–734.

Ellen Breitholtz. 2020. *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Brill, Leiden.

Ellen Breitholtz and Christine Howes. 2020. Communicable reasons: How children learn topoi through dialogue. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue (WatchDial)*.

Marisa Casillas. 2014. Turn-taking. In Danielle Matthews, editor, *Pragmatic development in first language acquisition*, pages 53–70. John Benjamins Amsterdam, The Netherlands.

Alexander Clark and Shalom Lappin. 2010. *Linguistic Nativism and the Poverty of the Stimulus*. John Wiley & Sons.

Herbert H. Clark. 1975. Bridging. In *TINLAP '75 Proceedings of the 1975 workshop on Theoretical issues in natural language processing*, pages 169–174. Association for Computational Linguistics.

Robin Cooper, Jonathan Ginzburg, and Staffan Larsson. 2023. Learning language games probabilistically: From crying to compositionality. In J-P Bernardy, R Blanck, S Chatzikyriakidis, S Lappin, and A Maskharashvili, editors, *Probabilistic Approaches to Linguistic Theory*, pages 361–422. CSLI Publications.

Oswald Ducrot. 1988. Topoï et formes topiques. *Bulletin d'Etudes de Linguistique Francaise*, 22:1–14.

Judy H Eaton, Glyn M Collis, and Vicky A Lewis. 1999. Evaluative explanations in children's narratives of a video sequence without dialogue. *Journal of Child Language*, 26(3):699–720.

Laurent Filliettaz and Eddy Roulet. 2002. The geneva model of discourse analysis: an interactionist and modular approach to discourse organization. *Discourse Studies*, 4(3):369–393.

Gustaf Gredebäck, Linn Fikke, and Annika Melinder. 2010. The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental science*, 13(6):839–848.

H.P. Grice. 1975. Logic and Conversation. *Syntax and Semantics*, 3(S 41):58.

Michael Alexander Kirkwood Halliday. 1975. Learning how to mean. In *Foundations of language development*, pages 239–265. Elsevier, Amsterdam.

Alison Henry. 1995. *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford University Press on Demand.

Elma E. Hilbrink, Merideth Gattis, and Stephen C. Levinson. 2015. Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in psychology*, 6:1492.

Sarah Hiller and Raquel Fernández. 2016. A data-driven investigation of corrective feedback on subject omission errors in first language acquisition. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 105–114.

Jackson L. Lee, Ross Burkholder, Gallagher B. Flinn, and Emily R. Coppess. 2016. Working with chat transcripts in python. Technical Report TR-2016-02, Department of Computer Science, University of Chicago.

Stephen C. Levinson. 2006. On the human "interaction engine'. In N. J. Enfield and Levinson S. C., editors, *Roots of Human Sociality: Culture, Cognition and Interaction*. The Wenner-Gren Foundation, Oxford.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. 2009. Two-year-old children's production of multi-word utterances: A usage-based analysis. *Cognitive Linguistics*.

Brian MacWhinney. 2000. *The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database*. MIT Press, Cambridge, MA.

Joanna Rączaszek-Leonardi, Nicole Rossmanith, Iris Nomikou, and Katharina J Rohlfing. 2019. Levels of coordination in early semantic development. *Psychology of Language and Communication*, 23(1):212–237.

Caroline F Rowland. 2007. Explaining errors in children's questions. *Cognition*, 104(1):106–134.

Caroline F Rowland, Julian M Pine, Elena VM Lieven, and 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.

Julian Schlöder, Ellen Breitholtz, and Raquel Fernández. 2016. Why? In *Proceedings of JerSem*, pages 5–14.

Dan Sperber and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*, second edition. Blackwell, Oxford.

Gemma Stephens and Danielle Matthews. 2014. The communicative infant from 0-18 months: The social-cognitive foundations of pragmatic development. In Danielle Matthews, editor, *Pragmatic Development in First Language Acquisition*, pages 13–36. John Benjamins, Amsterdam.

Michael Tomasello. 1992. The social bases of language acquisition. *Social development*, 1(1):67–87.

Stephen E Toulmin. 2003. *The Uses of Argument*. Cambridge University Press, Cambridge.

Virginia Valian and Lyman Casey. 2003. Young children's acquisition of wh-questions: The role of structured input. *Journal of child language*, 30(1):117–143.