# Meaningful questions: The acquisition of auxiliary inversion in a connectionist model of sentence production

Hartmut Fitz
Max Planck Institute for Psycholinguistics Nijmegen

Franklin Chang
School of Psychology, Liverpool University

## Author Note

Hartmut Fitz, Max Planck Institute for Psycholinguistics Nijmegen, the Netherlands
Franklin Chang, School of Psychology, University of Liverpool, United Kingdom

Correspondence should be addressed to:
    Hartmut Fitz, Max Planck Institute for Psycholinguistics Nijmegen, the Netherlands
    P.O. Box 310, 6500 AH Nijmegen, the Netherlands
    Email: hartmut.fitz@mpi.nl
    Phone: +31-(0)24-352 1524

Abstract

Nativist theories have argued that language involves syntactic principles which are unlearnable from the input children receive. A paradigm case of these innate principles is the structure dependence of auxiliary inversion in complex polar questions (Chomsky, 1968, 1975, 1980). Computational approaches have focused on the properties of the input in explaining how children acquire these questions. In contrast, we argue that messages are structured in a way that supports structure dependence in syntax. We demonstrate this approach within a connectionist model of sentence production (Chang, 2009) which learned to generate a range of complex polar questions from a structured message without positive exemplars in the input. The model also generated different types of error in development that were similar in magnitude to those in children (e.g., auxiliary doubling, Crain & Nakayama, 1987; Ambridge, Rowland, & Pine, 2008). Through model comparisons we trace how meaning constraints and linguistic experience interact during the acquisition of auxiliary inversion. Our results suggest that auxiliary inversion rules in English can be acquired without innate syntactic principles, as long as it is assumed that speakers who ask complex questions express messages that are structured into multiple propositions.

*Keywords:* Language acquisition, sentence production, subject-auxiliary inversion, polar questions, conceptual structure, connectionist modeling

# Meaningful questions: The acquisition of auxiliary inversion in a connectionist model of sentence production

## 1   Introduction

A central debate in language acquisition concerns the question which aspects of our knowledge of language are learned from experience and which are part of our biological endowment for language. Nativist theories have argued that there are syntactic principles that are impossible to learn given the input that children receive (Chomsky, 1968, 1980). A prominent example of such an unlearnable principle is the structure dependence of linguistic operations which seems ubiquitous in language. It is most commonly illustrated in terms of auxiliary inversion in English yes-no questions (also called *polar* questions). In transformational grammars (Chomsky, 1981), polar questions are derived from declarative sentences by auxiliary movement. For instance, declaratives with progressive verbs place the auxiliary *is* before the verb *jumping* as in (1) and this auxiliary is moved to sentence initial position in polar questions (2).

(1)     The boy is jumping.

(2)     Is the boy _ jumping?

Put this way, the auxiliary inversion rule is simple and should be learnable but it becomes more challenging when there are multiple auxiliaries in complex declaratives like (3) with two clauses.

(3)     The boy that is jumping is happy.

When the polar question version of (3) is created as in (4), the main clause auxiliary is moved to sentence initial position, rather than the embedded clause auxiliary as in (5).

(4)     Is the boy that is jumping _ happy?

(5)     *Is the boy that _ jumping is happy?

Chomsky (1980) argued that the syntactic knowledge that supports the correct structure (4) is not obviously explained by the input to children. If children only hear single-clause questions with one auxiliary (2) or right-branching questions with two auxiliaries like

(6)     Is the boy _ chasing the cat that was lazy?

then the input would be consistent with an incorrect rule that creates ungrammatical questions like (5) where the linearly first auxiliary is placed in sentence initial position.

In order to learn the correct rule, children would need to hear center-embedded polar questions with two auxiliaries (4), but these questions appear to be rare in child-directed speech. MacWhinney (2004) found only one occurrence in approximately 3 million utterances (up to age 5;0) in the CHILDES database (MacWhinney, 2000; see also Legate & Yang, 2002). The absence of positive exemplars has been argued to demonstrate the *poverty of the stimulus* with respect to auxiliary inversion in complex questions. On this view, the input to children does not contain enough information to acquire the linguistic rules that adult speakers appear to use (Chomsky, 1980; Crain, 1991; Laurence & Margolis, 2001; Berwick, Pietroski, Yankama, & Chomsky, 2011). The *poverty of the stimulus* and the compatibility of the input with an incorrect linear order rule have been taken as evidence in support of innate syntactic knowledge. This argument has made auxiliary inversion in

complex questions the prototypical defense of linguistic nativism and it is presented in virtually all introductions to syntactic theory (e.g., Carnie, 2012; Radford, 2004) and first language acquisition (e.g., Karmiloff & Karmiloff-Smith, 2002; E. V. Clark, 2002; Ambridge & Lieven, 2011).

An alternative to nativist accounts of auxiliary inversion are constructivist, or usage-based, theories of language (Tomasello, 2003; Goldberg, 2006; Bybee, 2010). These theories argue that the input from which children learn is richer than has been assumed by nativists, and that children use powerful statistical learning mechanisms to gather information from this data (Saffran, Aslin, & Newport, 1996; Gómez, 2007). For example, Reali & Christiansen (2005) have suggested that word co-occurrence statistics suffice to render grammatical questions (4) more probable than their ungrammatical counterparts (5) and Pullum & Scholz (2002) have argued that complex questions other than (4) can provide evidence for the correct auxiliary inversion rule (see also Sampson, 1989). Thus, usage-based theories have shifted the explanatory burden away from innate syntactic knowledge and onto the learning mechanism and its input.

Although these approaches differ in where they seek relevant constraints on complex question formation, they both locate these constraints within the domain of language. English auxiliary inversion rules are either learned from linguistic input or preset by innate syntactic constraints. In the present work, we propose an alternative account where auxiliary inversion arises from constraints that are outside of language. In particular, we focus on the role of the message that is conveyed when complex questions are being produced. Our claim is that the message is structured into nonlinguistic propositions and this structure constrains the way grammars can develop in the language system. Critically, we argue that the message structure is available before children acquire the ability to produce complex questions. We provide a computational model that is able to use the structure of meaning when learning to produce grammatical complex polar questions from impoverished input.

### 1.1   Framing the learning problem

Nativist and constructivist theories generally agree that some aspects of question formation are learned from the input. This is because auxiliary inversion in interrogatives is typologically rare and limited to a handful of predominantly West-Germanic languages (Siemund, 2001). Most languages mark polar questions in other ways and therefore different realizations of interrogative form must be acquired through experience. Speakers of English learn to front auxiliaries in interrogatives and in nativist theories, learning is modulated by a constraint on the grammar space that blocks the incorrect rule from being adopted (Crain & Pietroski, 2001; Crain & Thornton, 2012). This constraint has been called *structure dependence* because transformations are sensitive to the hierarchical phrase structure of sentences. For example, sentence (3) would have a syntactic structure where each clause is dominated by its own $\mathcal{S}$ node (matrix $\mathcal{S}1$, embedded $\mathcal{S}2$) as in (7).

(7)     [The boy [that is jumping]$_{\mathcal{S}2}$ is happy.]$_{\mathcal{S}1}$

Structure dependence constrains the language system to consider only rules where auxiliaries do not move out of their $\mathcal{S}$ domain, and this would guarantee the acquisition of the correct rule where only the matrix auxiliary can be displaced. The constraint is domain-specific because it is formulated in terms of linguistic categories.

Constructivist theories, on the other hand, have questioned whether an innate syntactic constraint is required to learn auxiliary inversion (Pullum & Scholz, 2002; A. Clark & Lappin, 2011; Ambridge, Pine, & Lieven, 2014) and the nature of errors that children make throughout development has been an important source of evidence in this debate. In an elicited production study, children between 3;2 and 4;7 never omitted the embedded clause auxiliary as in (5), and this seemed to support the innateness of a structure-dependent constraint (Crain & Nakayama, 1987). Using the same paradigm, however, Ambridge, Rowland, & Pine (2008) found that such errors could be elicited when target items contained modal auxiliaries or plural copulas, undermining the nativist line of reasoning. Moreover, 62% of the responses in Crain & Nakayama's study contained other kinds of errors and older children [6;3–7;9] tested in Ambridge et al. still produced 52.6% incorrect questions. This suggests that the acquisition of complex questions is a slow, gradual process that requires accumulation of relevant evidence over an extended period of time. A frequent error type in both studies was *auxiliary doubling* where the main clause auxiliary was repeated after the relative clause (e.g., *Is the boy who is jumping is happy?*) which is difficult to reconcile with the idea that question formation involves auxiliary movement as a displaced constituent should not persist in its canonical location. If questions are not derived from declaratives by movement (Rowland & Pine, 2000; Dąbrowska & Lieven, 2005; Ambridge, Rowland, Theakston, & Tomasello, 2006; Bouchard, 2012), there would be no need for a structure-dependent constraint restricting auxiliary inversion. Others have argued that these errors are due to performance limitations (Crain & Nakayama, 1987), but then a theory is needed which explains why these limitations change over development.
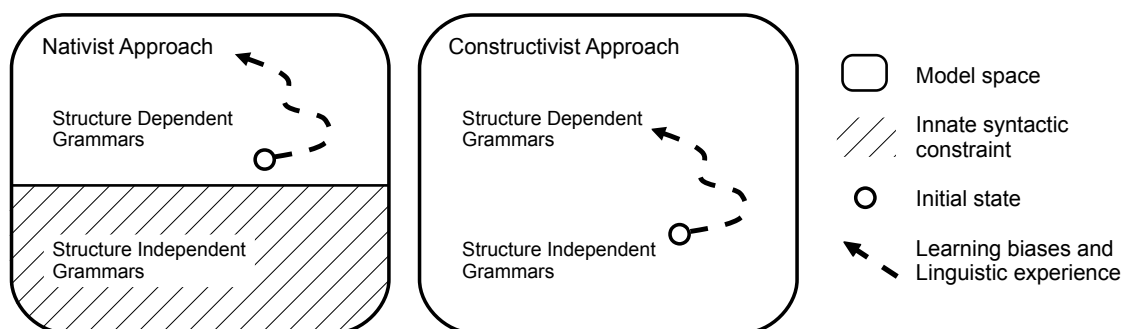


*Figure 1.* Schematic of nativist and constructivist accounts of auxiliary inversion.

Nativist and constructivist approaches are often treated as polar opposites and it is helpful to delineate their assumptions from a more abstract perspective. Language is a neurobiological system whose configuration is determined by the values of its internal parameters (e.g., the strength of $\sim 10^{14}$ synaptic connections). The set of realizable values is the language network's model space. Each point in the space implicitly represents the grammar currently adopted by the learning child (Figure 1). During acquisition, input interacts with learning biases and maturational constraints to force the system into a region of model space which represents adult knowledge of language. Linguistic nativism assumes that the learner never entertains grammars where structure dependence is violated because an innate syntactic constraint removes such grammars from the model space (shaded region in left panel of Figure 1). This constraint could be expressed in terms of the system's initial state, for instance the connectivity structure of the language-ready brain. Within this approach, the role of experience is to learn the particular realization of structure dependence

in a given language. On the constructivist view, structure-independent grammar regions exist prior to experience and while the language system is in a state of immaturity, various production errors are observed. Constructivists also assume that there are 'acquisition paths' (Estigarribia, 2009) for structure dependence through the model space (dashed line in right panel of Figure 1). Linguistic experience and domain-general learning biases ensure that the language system eventually ends up in the correct region. These learning biases can take many different forms (see Elman et al., 1996; Mitchell, 1997; Pearl & Lidz, 2009; Perfors, Tenenbaum, & Regier, 2011) and it is likely that multiple biases are at work during language acquisition. For example, a bias can be present in the way the language system attends to and categorizes linguistic input, in the perceived similarity of input with knowledge already acquired (A. Clark & Lappin, 2011), and in the dynamics of the learning mechanism itself (Petersson & Hagoort, 2012). The key issue in this debate, then, is to elucidate the nature of these biases and how they interact with experience of language.

*1.2   Previous computational models of auxiliary inversion*

       To understand how input and biases might interact, it is useful to look at computational models that have attempted to explain the acquisition of complex questions without innate syntactic constraints on auxiliary movement. In this brief review, we make a broad distinction between approaches that only use word sequences as input and those that assume the input is coded in terms of some preordained set of syntactic categories. One example of the word-based approach is the Simple Recurrent Network (SRN for short; Elman, 1990) which is a layered neural network that predicts the next word in a sentence based on previous words. It is able to acquire and represent syntactic knowledge in its connection weights (Elman, 1991). Lewis & Elman (2001) showed that this model tended towards structure-dependent processing in the absence of positive exemplars like (4), when trained on an artificial English-like language. At the critical position inside the relative clause (e.g., *Is the boy who is smoking crazy?*) it activated the auxiliary *is* more strongly than participles like *smoking.* This was taken as evidence that the correct generalization could be based on substructure frequency because the bigram *who is* occurred in complex declaratives within the training set, whereas the bigram *who smoking* did not, and SRNs have an architectural bias to encode such local regularities (Elman et al., 1996; Rohde & Plaut, 1999). Unfortunately, however, the model was only tested on a *single* complex question which severely limits the generality of this approach. It has been noted, for example, that other types of questions do not violate the co-occurrence statistics of English when the embedded auxiliary is omitted (e.g., *Can the boys who (can) run jump?*; Ambridge et al., 2008) and therefore bigram statistics might not favor the grammatical form in all complex questions.
       The idea of Lewis & Elman (2001) was explored more rigorously in the reduced model of Reali & Christiansen (2005) which counted bigrams in a small corpus of child-directed speech (10,705 sentences; Bernstein-Ratner, 1984). Bigram counts were then used to compute sentence probabilities for 100 complex polar questions like (4) and their ungrammatical version (5), and the grammatical form was preferred 96% of the time. Based on these results, the authors argued that the input that children receive might provide distributional information relevant to auxiliary inversion even if direct evidence (4) is absent. These findings were subsequently scrutinized by Kam, Stoyneshka, Tornyova, Fodor, & Sakas (2008) who only obtained 87% accuracy on a slightly different version of the corpus. The items tested by Reali and Christiansen almost exclusively contained present tense, subject-relative clauses with the singular embedded auxiliary *is.* When tested on questions with object-relative clauses (e.g., *Is the wagon that your sister is pushing red?*) or those that

required *do*-support (e.g., *Does the boy who plays the drum want a cookie?*), the model's accuracy dropped below chance. Subject-relativized questions were supported by the critical bigram *that/who is* because English relative pronouns are homographs of other unrelated, more frequent forms (e.g., *that* as a deictic pronoun in *that is nice*) and similar bigram evidence did not exist for different question types. The model also failed to identify the grammatical form in Dutch where polar questions can be formed by subject-verb inversion. Thus, the bigram-based approach only worked for a narrow subclass of complex polar questions, a wider range of other auxiliary inversion constructions was not supported by the input. The occasional congruence between bigram statistics and grammaticality might reflect an accidental fact of English rather than point towards a causal mechanism for the acquisition of complex yes-no questions.

The model of Reali & Christiansen (2005) acted as a classifier based on sentence probabilities and it has not been demonstrated that n-gram related biases for choosing between whole sentence alternatives are sufficient for language production or comprehension. A model capable of generating utterances was the alignment learner of A. Clark & Eyraud (2007). It had a bias to substitute word strings for one another if they had been experienced in the same context within a text. For example, if the learner encountered *the boy is happy* and *the boy who is jumping is happy* it could replace *boy* in a simple question like *is the boy tall?* and thus generate the complex question *is the boy who is jumping tall?* while avoiding the ungrammatical form *\*is the boy who jumping is tall?*. Clearly, this model could not generalize to novel questions for which supportive contexts had not occurred in the input and it also overgenerated in undesirable ways because English is not a substitution language (Berwick et al., 2011).

In contrast to the word-based approaches, a wide range of computational models convert words in the input into syntactic categories and learn regularities over these categories (Niyogi & Berwick, 1996; Sakas & Fodor, 2001; Yang, 2002; Hsu & Chater, 2010). An example of this approach applied to auxiliary inversion is the data-oriented parsing model of Bod (2009). It was trained on 10,280 adult utterances from the Eve corpus (R. Brown, 1973) that were annotated for syntactic dependencies. By recombining treelets extracted from the input it could parse novel sentences. In this framework, the probability of the most probable derivation was used to measure grammaticality. For example, grammatical complex questions had a probability of 4.4e-17 whereas ungrammatical ones obtained a probability of 2.1e-18. These differences are exceedingly small and it was not tested if they held up for corpora other than Eve. Furthermore, in sentence generation only 40.5% of the generated questions fell under the syntactic pattern *Is P who is Q R?* and most of these items were semantically unacceptable (e.g., *Is that who is that some noodles?*). Complex questions with different relative clause types and variation in auxiliaries were not tested. Despite the strong assumption of syntactically labelled input, the model did not reach a human adult state.

Finally, a Bayesian learner was proposed to model grammar selection for a corpus of child-directed speech (Perfors, Tenenbaum, & Regier, 2011). The corpus (Adam, R. Brown, 1973) was tagged with parts-of-speech and three grammars were handcrafted to parse the resulting sequences of word categories. Due to a stipulated simplicity bias, the model preferred a hierarchical grammar that could parse complex polar questions which had not occurred in the training corpus. One concern with this approach is that it does not explain how grammars are acquired in the first place. Another challenge is to explain children's errors. A grammar that generates errors would have to be included in the hypothesis space and Bayesian inference would first have to select this deficient grammar before recovering

to a grammar that approximates the adult knowledge of language. One would then have to explain why the optimal grammar favored by the simplicity bias is not desirable for long periods of development (up to 7 years of age).

Overall, the reviewed computational approaches have shown that probabilistic language models can develop some preference for grammatical complex questions. Word-based models have to assume that the appropriate input regularities for each combination of structure, tense, aspect, and modality are available to children. Category-based approaches assume that different auxiliaries are part of the same category from the beginning of development and this assumption is at odds with many current theories of language acquisition (e.g., Tomasello, 2003; O'Grady, 2005; Ambridge & Lieven, 2011). One reason why the auxiliary inversion debate has been difficult to resolve is that there is very little agreement about the nature of the input. Each of the above models is based on a different input corpus and there is no evidence that these corpora are truly representative of the input that children receive to learn auxiliary inversion. Different adult corpora display substantial variability in lexical and structural probabilities (Kilgarriff, 2001; Roland, Dick, & Elman, 2007) and these differences are magnified greatly when looking at smaller corpora of child-directed speech, for children at different points in language development (Chang, Lieven, & Tomasello, 2008). Moreover, if a large corpus was created that provided learning data to a particular child for each combination of structure, tense, aspect, and modality, we would still not be able to tell whether some children might learn the abstract rule without some combinations in their input. A theory of language acquisition should ensure that the correct constraints are being learned, even when there are gaps in the input.

Although the reviewed models differ considerably in mechanistic terms, they all select the most probable word, sentence or grammar based on statistical regularities in the input using different biases. In sentence production, however, speakers do not merely generate the most probable sequence of words but sentences that convey an intended meaning or *message*, regardless of how (un)likely these sentences are vis-á-vis linguistic experience. Hence, there is more to explain about auxiliary inversion than just statistical preferences. Since production is the primary source of evidence with respect to the acquisition of complex questions (Nakayama, 1987; Crain & Nakayama, 1987; Ambridge et al., 2008) it is unclear how existing approaches relate to these data, in particular the patterns of error observed in children over development. Even with large, representative input corpora, these models would not explain how meaning is used to produce complex questions. A complete account of auxiliary inversion needs to address production data in acquisition.

In our view, it is unlikely that the input conspires for each and every child to ensure the correct generalizations are achieved across a range of relative clause types and auxiliaries differing in tense, number and modality. Our approach therefore argues that the *structure of meaning* can help to smooth over gaps in the input allowing for broader generalizations. By embedding auxiliary inversion within sentence production, we can potentially explain a wider range of behaviors, including errors in development. In the following section we will characterize the assumptions about message structure that are needed to support the acquisition of complex questions in production.

## 1.3   Putting meaning into questions

Meaning has been argued to be important in grammatical development in both usage-based (Tomasello, 2003; Goldberg, 2006) and generative approaches (Culicover & Jackendoff, 2005) but the role of meaning in the acquisition of complex polar questions has

not been spelled out. Existing computational models (Lewis & Elman, 2001, Reali & Christiansen, 2005; Bod, 2009; A. Clark & Eyraud, 2007) have focused on what can be extracted from sentence form alone and meaning plays no role in these models. In contrast to these approaches, we propose a computational model that learns to produce sentences from meaning and we argue that the structure of meaning can bias the system towards structure-dependent rules.

When producing a sentence, speakers convey a message and the message needs to distinguish between *The boy kisses the girl* and *The girl kisses the boy* (Fodor & Pylyshyn, 1988; Levelt, 1989; Chang, 2002). As in many theories of meaning (Gruber, 1976; Jackendoff, 1990), this distinction can be captured by thematic roles such as Agent (the cause or initiator of an action) and Patient (the event participant that is affected by the action). For instance, the message for *The girl eats the cake* can be represented by thematic roles that are bound to conceptual content as in Agent–GIRL, Action–EAT, and Patient–CAKE which we sometimes abbreviate as the proposition EAT(GIRL, CAKE). Language-specific processing strategies can then be used to link the thematic structure in the message to linear order in sentence forms (Jackendoff & Wittenberg, 2014). For example, in English agents tend to precede actions, and actions tend to precede patients. Although some theories view thematic roles and propositions as linguistic elements, we assume that they are based in nonlinguistic event/action representations (Zacks & Tversky, 2001). Monkeys and apes do not acquire human-like languages and hence they should not have human-like semantic roles. Nonetheless, they understand thematic relations of *who does what to whom* in social contexts and situations that are relevant for survival (Cheney & Seyfarth, 2007). Thus, conceptual structure occurs in nonlinguistic animals, independent of language (Jackendoff, 2014). Furthermore, there is evidence that these abilities appear early in human development where infants understand some aspects of events such as causality and goal-directedness before they acquire language (Leslie & Keeble, 1987; Luo, 2011).

Declarative sentences assert a proposition whereas yes-no questions request information about a proposition and this difference in pragmatic function can be represented in the message:

(8)     The boy jumps.          Meaning: JUMP(BOY)

(9)     Does the boy jump?      Meaning: JUMP(BOY) + YSNO

Augmenting the declarative message with a YSNO feature is a parsimonious way to distinguish speech acts and it is motivated by the fact that sentence (8) would be an appropriate answer to question (9). Hence, declarative and interrogative forms are systematically linked by shared conceptual content in the underlying messages. The YSNO feature itself does not prescribe auxiliary inversion and allows the surface form of polar questions to differ across languages. For instance, in Japanese the message of (9) would be expressed as the declarative (8) plus a question marker (e.g., declarative *shonen ga jyanpu*, question *shonen ga jyanpu ka?*). Thus, a different mapping between meaning and form would have to be learned for languages without auxiliary inversion.

In order to use tense and aspect correctly, adult speakers must know whether they are referring to completed or ongoing events in the past or present. For example, sentence (10) is the past progressive version of sentence (9), and we assume that features in the message encode these distinctions (e.g., PAST, PROG).

(10)     Was the boy jumping?      Meaning: JUMP(BOY) + YSNO + PAST + PROG

Tense and aspect interact with the speech act feature to constrain the production of verb phrases in questions and declaratives.

Auxiliary inversion in complex questions can now be linked to these features. Questions like (4) can be broken up into two clauses and each clause has its corresponding nonlinguistic meaning:

(11)     The boy is jumping          Meaning: JUMP(BOY) + PROG

(12)     Is the boy running?         Meaning: RUN(BOY) + PROG + YSNO

Critically, it is necessary to distinguish which predicate is being questioned (the *running* predicate is marked with a YSNO question feature), otherwise speakers would be unable to distinguish *Is the boy running?* from *Is the boy jumping?*. The division of the message into two propositions with one question marker is motivated by discourse constraints on production. Speakers produce embedded clauses to identify referents when there are multiple possible ones (Fox & Thompson, 1990; Diessel, 2004; Haywood, Pickering, & Branigan, 2005). It is pragmatically felicitous to say *the boy that was jumping* if there is another boy who was not jumping and both boys are salient in the mind of the speaker and the listener (they are given and in the common ground; H. H. Clark, 1996). Since the jumping boy is shared information between the speaker and the listener, there is no reason for the speaker to ask whether the boy jumped. Information is typically requested about a proposition that is private knowledge possessed by the listener, but not the speaker (Austin, 1962; Grice, 1975; Van Valin Jr. & LaPolla, 1997; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). If meaning was not structured into propositions, and instead encoded as a unordered list (e.g., JUMP, RUN, BOY, BOY, YSNO, PROG, PROG), speakers could not know which boy they are requesting information about. Therefore, adult production of complex polar questions presupposes that there are two propositions where one is shared (11) and another is questioned (12).

Although children may not always infer adult meaning from the start of language acquisition, they must eventually infer enough to acquire the ability to use language in an adult-like way and convey appropriate messages. In our approach, we assume that the production of complex questions is preceded by the development of basic conceptual and pragmatic distinctions. They include the ability to represent tense/aspect, different speech acts, and multiple propositions. Data on the development of tense and aspect comes from several studies. A basic understanding of tense semantics starts to develop around 2 years of age (Weist, 1991; Wagner, 2001), 3 year-olds can reliably distinguish past, present and future (Harner, 1981). Progressive aspect develops before past tense (R. Brown, 1973) and is initially restricted to activity verbs (Shirai & Anderson, 1995). Although progressive forms increase dramatically with age (Smith, 1980), 3 year olds can already use progressive verb morphology productively (Akhtar & Tomasello, 1997). Thus, children mostly distinguish between various tenses and aspects by around 3 years of age and it is reasonable to assume that they activate PROG and PAST features to represent these differences.

Questions are some of the most frequent elements in child-directed speech (Legate & Yang, 2002; Cameron-Faulkner, Lieven, & Tomasello, 2003) comprising at least one third of all utterances. Evidence for speech act identification in children comes from the spontaneous production of yes-no questions, including non-canonical forms, which has been found as early as 1;6 years of age (Dabrowska, 2000; Estigarribia, 2009). Around age 2;6 inverted yes-no questions begin to outnumber non-inverted ones in elicited production (Erreich, 1984). In verb learning studies, children understand that questions request

information and respond accordingly before the age of 3 (e.g., Brooks & Tomasello, 1999). Hence we can assume that children distinguish between questions and declarative speech acts (e.g., in terms of a YSNO feature) quite early in development.

Finally, the production of complex sentences with multiple clauses requires the ability to represent different actions with separate thematic roles. Wynn (1996) provides evidence that six month old infants can individuate events in multiple event sequences. Further evidence that children can activate two propositions simultaneously comes from theory-of-mind tasks where children have to maintain a reality proposition of where an object is actually hidden and keep it distinct from a belief proposition about where one of the characters *thinks* that the object is hidden. These abilities appear early in development (Onishi & Baillargeon, 2005) and become robust in verbal tasks by around 4 years (Baron-Cohen, Leslie, & Frith, 1985; Wellman, Cross, & Watson, 2001). Thus, before children begin to spontaneously produce complex polar questions, they are developing the ability to maintain two propositions in memory.

Based on these findings, we assume that messages are structured into multiple propositions that can distinguish speech acts and tense/aspect before children extend auxiliary inversion to complex structures. During acquisition, children learn to link this pre-linguistic conceptual structure in the message to overheard utterances. In Table 1, we outline how children would learn particular meaning-form mappings attested in the primary linguistic data (structures (i)–(iv)) that can then be combined productively to generate novel utterances (v). Simple declaratives (i) support the mapping of thematic roles into appropriate

Table 1
*Meaning-based generalization*

| Utterance Type | Meaning | Example Sentence |
|---|---|---|
| (i)　　　　Simple declarative | JUMP(BOY) + PROG | *The boy is jumping.* |
| (ii)　　Simple yes-no question | HAPPY(BOY) + YSNO | *Is the boy happy?* |
| (iii)　　Complex declarative | HAPPY(BOY) | *The boy that is jumping is happy.* |
| | JUMP(BOY) + PROG | |
| (iv)　Right-branching yes-no question | CHASE(BOY,CAT) + PROG + YSNO　LAZY(CAT) | *Is the boy chasing the cat that is lazy?* |
| (v) Center-embedded yes-no question | HAPPY(BOY) + YSNO　JUMP(BOY) + PROG | *Is the boy that is jumping happy?* |

*Note.* Declaratives and questions are distinguished by a YSNO feature that encodes the type of speech act. Two propositions in the sentence message enable the production of relative clauses (individual propositions are shown on separate lines in the table). Novel complex yes-no questions inherit semantic features from simple questions and complex declaratives. Learned partial meaning-form mappings combine to generalize to novel constructions.

sentence positions (e.g., agent $\rightarrow$ subject) and from exposure to simple questions (ii), children can learn word order differences between speech acts and place auxiliaries sentence-initially in English polar interrogatives. Complex declaratives in the input (iii) show that distinct propositions map to clausal units where the content of the main clause proposition is disrupted by the center-embedded relative clause proposition in sequential left-to-right production. When children encounter right-branching yes-no questions (iv), they can link the speech act feature to the proposition expressed in the main clause which then comes to signal the location of the questioned proposition in novel complex questions (v). Although the primary linguistic data contains only few (if any) center-embedded yes-no questions (MacWhinney, 2000; Legate & Yang, 2002), children might be able to combine

knowledge from the frequent constructions (i)–(iv) to generate appropriate complex polar questions (v) from the underlying messages (Goldberg, 2006). On the proposed account, messages derive from nonlinguistic propositions which are the same across differences in tense, aspect, modality and speech act types. Thus, the meaning-form mappings learned from utterance types (i)–(iv) in Table 1 are not dependent on subtle distributional regularities in the input and generalize more widely, and this might explain why most children eventually converge on the correct auxiliary inversion pattern in production.

In the present work, we investigate whether a computational model of syntactic development can simulate this process of meaning-based generalization. We first describe the model architecture, its learning environment and methods for training and evaluation (section 2). We then show that the model generalizes in a structure-dependent way (section 3) and explain how this ability depends on the message structure (section 5). The approach has broader coverage than previous computational accounts in that our model can produce complex questions with copulas, different tensed and modal auxiliaries, *do*-support, and a variety of relative clauses. In addition, we show that the model's learning curves are consistent with fine-grained developmental differences regarding auxiliary and relative clause types, and relative clause length (section 3). In section 4, we replicate the quantity and time-course of the distinct types of error children produced in behavioral studies. We identify input structures that are necessary for children's recovery from auxiliary doubling errors in the course of development (section 6). Finally, we motivate assumptions about the model's input by means of a corpus of child-directed speech which is an *order of magnitude* larger than the corpora that have been used in the computational approaches discussed above. Our approach goes beyond previous work which did not include meaning and hence could not model the production of complex questions and the errors observed in children.

## 2    A structured message account of auxiliary inversion

To instantiate a message-based account of auxiliary inversion, we used a connectionist network called the Dual-path model that was designed to explain how people learn to produce sentences from meaning (Chang, 2002). The model addresses important production issues that are relevant for complex questions such as learning to activate thematic roles in language-specific ways and linking tense/aspect features to verbs. It was able to use messages with multiple propositions to generate complex sentences with embedded clauses (Chang, 2009; Fitz, 2009) and could account for relative clause development in English (Fitz, Chang, & Christiansen, 2011). It has also been shown to learn abstract syntactic structures (Chang, Dell, & Bock, 2006), which is critical in explaining the breadth of generalizations in complex question production. During syntactic development, the Dual-path model generated errors that mirrored the errors that are produced by children in learning verb-structure biases (Twomey, Chang, & Ambridge, 2014). Unlike some of the previous models of auxiliary inversion, that only included English syntactic knowledge, it was able to learn typologically-different languages like German (Chang, Baumann, Pappert, & Fitz, 2015) and Japanese (Chang, 2009), which suggests that it is a more general theory of language acquisition. The present model version retains the basic design features of the Dual-path architecture (Chang, 2002; Chang et al., 2006) and hence structure dependence in complex questions will be argued to result from assumptions that were motivated to explain these other production and acquisition phenomena.

## 2.1  Dual-path architecture

The Dual-path model had two pathways, a sequencing pathway and a meaning pathway (Figure 2). The sequencing pathway was an SRN (Elman, 1990) that learned
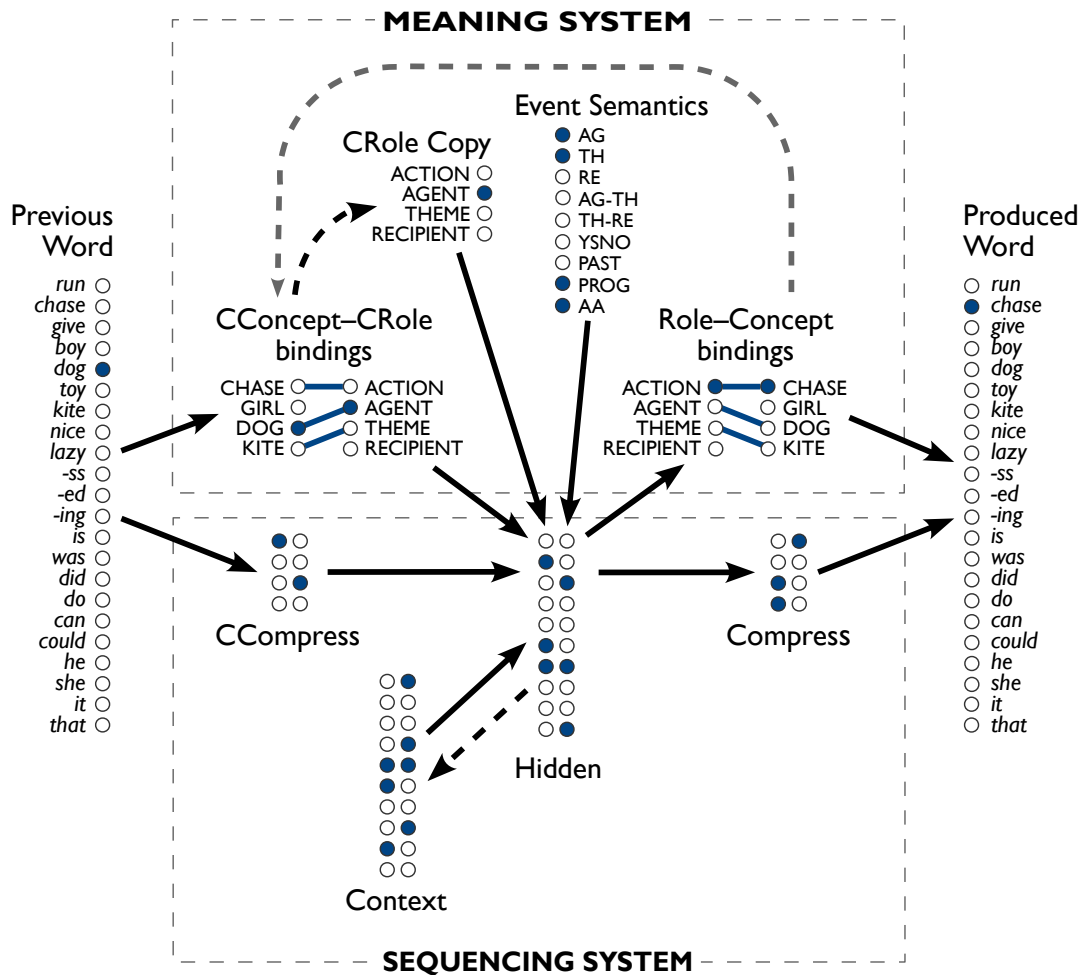


*Figure 2.* Architecture of the Dual-path model (Chang, Dell, & Bock, 2006). Word input arrived at the PREVIOUS WORD layer and activation spread through the sequencing and meaning pathways to the PRODUCED WORD output layer. The sequencing system represented word categories at the CCOMPRESS and COMPRESS layers and learned abstract syntactic frames over these categories. The meaning system represented a sentence-specific message that the model learned to activate incrementally at the right point in time. The two pathways intersected at the HIDDEN layer and competed/cooperated at the output layer. Solid arrows indicate connection weights that were adapted during learning, fixed copy connections are represented by dashed arrows. Solid links between roles and concepts in the respective layers were set during production and enabled binding. The conceptual structure of the target utterance was represented by features in the EVENT SEMANTICS layer which also signaled the intended speech act.

syntactic representations by predicting the next word in sentences, one word at a time. The difference between the model's predictions and the actual next word (error) was used by the learning algorithm (Rumelhart, Hinton, & Williams, 1986) to adjust the weights in the

network and this process gradually made future predictions more accurate. The sequencing pathway mapped from the Previous Word layer through a set of compression layers (Compress, CCompress) and a Hidden layer that eventually projected to the Produced Word output layer. The compression layers only allowed a small number of lexical distinctions to be represented and therefore the model had to learn syntactic categories which were suitable for next word prediction. The Hidden layer received activation from a Context layer that held a copy of the previous Hidden layer activation state. This provided the network with memory, allowing it to process temporally extended sequences of word categories and learn abstract syntactic frames. By itself, the SRN biased the system towards grammatical sequences, but it could not select a specific sequence to convey a particular message.

To generate a sentence for an intended message, the model had a meaning pathway which encoded the target message in fast-changing links between a Role and a Concept layer. There were four thematic role variables (Action, Agent, Theme, Recipient) in the Role layer which could be bound temporarily to conceptual content. For instance, the message CHASE(DOG, KITE) for *the dog is chasing the kite* was instantiated as a fixed connection between a Role layer unit for Agent and a Concept layer unit for DOG (top right of Figure 2). Likewise, there was a fixed connection between the Theme unit in the Role layer and the KITE unit in the Concept layer. These links were set by message planning and were not changed by learning like the other links in the model. The Concept layer was connected to the Produced Word output layer, and this allowed the model to learn which words were associated with which concepts. Since the Hidden layer in the sequencing system was connected to the Role layer, the model could learn to incrementally activate the Role layer units which then activated message-specific concepts and appropriate words at particular points in sentences. For example, at the beginning of the above sentence, the model must decide whether to produce *dog* or *kite* and it learned that it could activate the Agent role unit to produce the correct word.

A reverse message in the CConcept and CRole layers was set at the same time as the production message and the CConcept layer received input from the Previous Word layer (top left of Figure 2). This system allowed the model to determine the thematic role of the previously produced word. The Hidden layer could then use this information to activate the next role in the production message which was dependent on earlier choices. For example, once the word *dog* had been produced this would activate the Agent role in the CRole layer on the subsequent time-step and signal to the Hidden layer that the Action role should be activated next in order to produce a verb phrase. To enhance memory for the roles that had already been produced, there was a CRole Copy layer which held a running average of the CRole layer activations.

The final part of the message was the Event Semantics which represented the pre-linguistic conceptual structure of the target utterance (top middle of Figure 2). It had features that encoded the number of arguments in the message. For example, active features AG and TH in the Event Semantics would signal that there were two event participants in the transitive sentence *the dog is chasing the kite*, and hence there were arguments for the Agent and Theme roles in the intended message. The Event Semantics also encoded the tense, aspect and speech act type of the target utterance. This information was provided to the Hidden layer and guided the model to select appropriate structures in production. Simple, present tense declaratives were treated as the default construction with no additional active Event Semantics features. Past tense and progressive aspect were marked by features PAST and PROG relative to this default, and polar questions were signalled by the

YSNO feature. When the YSNO feature was turned on, a question was the target utterance, when it was switched off, a declarative had to be produced. These defaults were motivated by the fact that present tense auxiliaries were more than three times as frequent in child-directed speech than past tense auxiliaries (Thomas corpus; Lieven, Salomo, & Tomasello, 2009). Similarly, Cameron-Faulkner et al. (2003) found declaratives to be more frequent (39%, copular, subject-predicate and complex) than questions (32%, *wh*-questions and polars), and there were 56% declaratives (44% questions) in the Nina corpus (Legate & Yang, 2002). Defaults kept the number of features in the EVENT SEMANTICS small and improved generalization.

Auxiliaries could occur in copular constructions (e.g., *She is nice*), progressive verb forms (e.g., *The boy runs*) or in modal verb phrases (*She can run*). Copulas were the default verb phrase type, an action feature AA and a MODAL feature distinguished target non-modal and modal verb phrases in the EVENT SEMANTICS, respectively. Any auxiliary could be fronted in polar questions (e.g., *Is she nice*; *Can she run?*), other cases were expressed with *do*-support (e.g., *Does the dog chase the kite?*). All distinctions made in the EVENT SEMANTICS are illustrated by the example message-sentence pairs shown in Tables 4 and 5 of the Appendix.

## 2.2 Encoding complex utterances

In order to represent messages with two propositions, there were two sets of features in the EVENT SEMANTICS and duplicate role units in the ROLE and CROLE layers. For example, the dative relative clause sentence *The boy that the girl gives the book to jumped* was associated with the message ACTION1–JUMP, THEME1–BOY, ACTION2–GIVE, AGENT2–GIRL, THEME2–BOOK, RECIPIENT2–BOY and the EVENT SEMANTICS features TH1, AG2, TH2, RE2, PAST1, TH1-RE2 (the roles in Figure 2 correspond to the roles marked as 1 here). The first proposition JUMP(BOY) + PAST was instantiated as a temporary connection between the ROLE layer unit for THEME1 and the CONCEPT layer unit for BOY. The second proposition GIVE(GIRL, BOOK, BOY) involved links between the ROLE unit for AGENT2 and the CONCEPT unit for GIRL, the ROLE unit for THEME2 and the CONCEPT unit for BOOK, and the ROLE unit for RECIPIENT2 and the CONCEPT unit for BOY. Thus, it was assumed that the main idea that was being expressed was represented by the first proposition. The co-reference feature TH1-RE2 helped to identify cases where the same referent occurred in two different roles and there was such a feature for each combination of roles in the EVENT SEMANTICS. Here, it signalled that the boy THEME1 of the first proposition was the same as the boy RECIPIENT2 of the second proposition and this licensed the use of a relative clause. Without this binding feature, we would assume that two different boys were being referred to in each proposition. Since the same concept BOY was also linked to both the THEME1 and RECIPIENT2 units in the CROLE layer, the model learned that it should initiate an embedded clause by sequencing the word *that* after having produced *the boy*. When the segment *the boy that the girl gave the book to* had been produced, the co-reference feature TH1-RE2 informed the HIDDEN layer that the RECIPIENT2 role had to be omitted from the relative clause surface form. Activation that had accumulated in the CROLE and CROLE COPY layers signalled that all of the roles in the target utterance had been produced except the main clause verb. The model's learned representations would cause the HIDDEN layer to activate the ACTION1 role to produce the verb *jump* next, whose simple past tense was encoded by the active PAST1 and the silent PROG1 features. Like the ROLE units, tense and aspect features were distinguished for each proposition. Messages for complex yes/no-questions were encoded in the same way as those for complex declaratives, except for the single,

additional speech act feature YSNO in the EVENT SEMANTICS.

The sequencing and meaning pathways were cooperating (or competing) at the PRODUCED WORD output layer. Content words were activated by the meaning system, auxiliaries and other function words were produced by the sequencing system. Since the model was learning to use both pathways concurrently, errors could arise when these systems acquired mismatching representations in early development. The sequencing pathway was an SRN and hence it was similar to other statistical learning models in its sensitivity to frequency and n-gram regularities in the input. The acquisition of different constructions was therefore modulated by linguistic experience. However, the fast-changing links in the message enabled variable-like behavior and endowed the model with the syntactic productivity that SRNs lack (Fodor & Pylyshyn, 1988; Marcus, 2003; Chang, 2002). Moreover, the message was not an unstructured bag of concepts ('semantic soup'), but instead was structured into propositions with thematic role slots and some aspects of syntactic behavior which had to be learned from the input in other models can be explained by properties of the message. For example, if the model learned how to use the YSNO and PROG1 features to front auxiliaries in simple main clauses, then the same acquired representations could support auxiliary inversion in complex polar questions because the second proposition did not change this partial meaning-form mapping. In this respect the Dual-path model differed from some of the previously mentioned learners that could only use surface sequences of words where there is ambiguity about whether the main or embedded clause auxiliary should be placed in front.

## 2.3   Input distribution

With these provisions, the model was trained on a language containing copular (e.g., *The boy was lazy*), intransitive (e.g., *The dog jumps*), transitive (e.g., *The cat is carrying the toy*), and dative structures (e.g., *The boy gives the toy to the dog*). Sentences with relative clauses were assembled from these basic structures and all participant roles of the underlying construction could be relativized. For instance, the matrix head noun could be the agent, theme or recipient of a dative relative clause. The input grammar also generated canonical questions with subject-auxiliary inversion, including questions with relative clauses that could be attached to any lexical noun phrase. Verb tense (present, past) and aspect (simple, progressive) were represented by inflectional morphemes that were treated as separate words (e.g., *push -ed*, *give -ing*). Noun phrases were either lexical nouns preceded by a determiner, or pronouns (*he, she, it, him, her*) which created positional variation in the language. There were construction-specific restrictions on animacy. For example, agents in transitives were living things whereas themes in dative structures were inanimate objects (see the Appendix for more details on the input language). The lexicon contained 78 words in 12 categories which allowed the generation of approximately $5.7 \times 10^9$ distinct utterances.

Across model subjects, 95.1% of all training items were single-clause structures, half of them simple polar questions, and 4.9% of the input items contained a relative clause. Right-branching and center-embedded relative clauses were approximately equally frequent to ensure that both constructions were learned to an adult degree. 1.6% of the input items were questions with a relative clause and 0.3% had a center-embedded relative clause. Both assumptions were based on corpus data that will be discussed in more detail in section 6 below. Crucially, the learning environment did not contain complex questions with a center-embedded relative clause and an auxiliary in both clauses as these are the structures that were used to test the model's ability to generalize. When auxiliaries occurred in two clauses, these sentences were either assertions or questions with a right-branching relative

clause (e.g., *Could he touch the cup that was small?*). With these properties, the input was consistent with the assumptions made by Chomsky in formulating the problem of auxiliary inversion, and with previous corpus evidence that children rarely, if ever, experience questions of type (4) during acquisition. Thus, the model's input should support the incorrect structure-independent generalization where the linearly first auxiliary from the embedded clause is placed in sentence-initial position.

*2.4   Training and evaluation*

In training, sentences were paired with messages that encoded their meaning. Twenty-five different sets of 50,000 message-sentence pairs were randomly created to train 25 model subjects for a total of 500,000 input patterns. Thus, the training set covered less than 0.001% of the distinct sentences that the input grammar could potentially generate. Before each target utterance, the sentence message was set in the meaning system and it did not change during production. After exposure to each message-sentence pair the weights in the network were updated. Model subjects were tested after every 20,000 training sentences (which we call one *epoch*) on a set of novel items made up of simple declaratives, simple polar questions, complex declaratives with a right-branching or center-embedded relative clause, and their corresponding polar questions (50 items each). Center-embedded polar questions in the test set contained auxiliaries in both clauses (copula, progressive aspect, or modal) and this type of question had not occurred in the learning environment. Each model subject was tested on 150 distinct questions of this type; items in the test set had a mean utterance length (MLU) of 10.3 words (MLU 6.4 in training). Performance was evaluated in terms of *sentence accuracy* where the model had to produce the target utterance exactly, word-by-word. A production error occurred if the most active word in the model's output layer mismatched the target word in one or more sentence positions. Errors and their types were automatically extracted from the model's output using a syntactic coder script. No errors of any kind were permitted for an utterance to be counted as correct. Thus, the performance measure was maximally strict and not based on comparing word distributions as in previous neural network studies of auxiliary inversion (e.g., Lewis & Elman, 2001; Reali & Christiansen, 2005).

## 3   Meaning-based generalization to novel complex yes-no questions

When the model was trained and tested in the described way, it learned to produce the target language with 99.3% accuracy after around half a million episodes. Figure 3 shows the mean performance across subjects over time on the various structural types in the language input (single-clause questions and assertions, multi-clause assertions, right-branching questions), and on the novel center-embedded complex questions with an auxiliary in each clause. Simple declaratives and questions were both learned by the first test point and this mirrors studies where children showed similar accuracy for single-clause declaratives and subject-auxiliary inversion questions over development (Santelmann, Berk, Austin, Somashekar, & Lust, 2002). Single-clause structures were followed by constructions with a relative clause. To examine the developmental patterns in these complex structures, a mixed effects model was applied to the percentage of correctly produced utterances for each model subject, with speech act (declarative, question) and sentence type (right-branching, center-embedded) crossed. Unless stated otherwise, linear mixed models reported throughout the paper used logit-transformed percentages or proportions (Jaeger, 2008). Model subject was a random intercept with speech act and sentence type as random
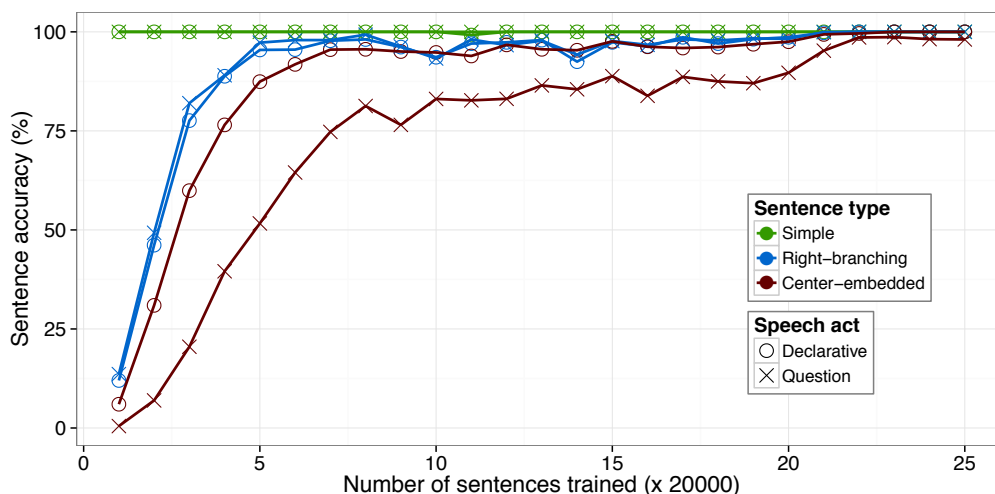
*Figure 3.* Model performance for all structural types in the input language, which were either single-clause, right-branching or center-embedded questions and declaratives. Center-embedded questions were novel polar questions with two auxiliaries which were not present in the model's learning environment.

main-effect slopes (Barr, Levy, Scheepers, & Tily, 2013), $p$-values were obtained by model comparisons (likelihood-ratio tests). There was no difference between right-branching declaratives and questions and this is consistent with the lack of differences in children's elicited production of these structures (Diessel & Tomasello, 2005). Center-embedded structures were learned more slowly by the model than right-branching structures ($\beta = -0.32$, SE $= 0.05$, $\chi^2(1) = 70.5$, $p < 0.001$) and children also find center-embedding more difficult than right-branching, both in comprehension (Clancy, Lee, & Zoh, 1986; Côrrea, 1995; Kidd & Bavin, 2002) and production (Diessel & Tomasello, 2000). Finally, there was an interaction of speech act and sentence type due to the slower acquisition of center-embedded questions with two auxiliaries, and there is evidence that these structures are difficult even for older children (52.6% errors in children aged 6.3–7;9; Ambridge et al., 2008).

The ability to produce complex questions is preceded by the acquisition of relative clauses in complex declaratives. Several studies found that English speaking children have more difficulty comprehending and producing transitive object-relative clauses (P-relatives such as *The boy that the cat chases...*) than transitive subject-relative clauses (A-relatives such as *The boy that chases the cat...*) (see O'Grady (1997) and Diessel (2004) for overviews), although processing difficulty in P-relatives is modulated by many factors, e.g., head noun animacy and the presence of pronoun subjects (Kidd, Brandt, Lieven, & Tomasello, 2007; Reali & Christiansen, 2007). Moreover, children find intransitive subject-relative clauses (S-relatives such as *The boy that runs...*) easier than A-relatives, and indirect object-relative clauses (IO-relatives such as *The cat that the boy gave the toy to...*) harder than P-relatives (Diessel & Tomasello, 2005). The model's language input contained all these relative clause types and they occurred with equal frequency in the learning environment. A mixed effects model was applied to the percentage of correctly produced complex declaratives with relative clause type (S, A, P, and IO) as a fixed factor which was coded numerically, centered around 0. Model subject was a random intercept with by-subject slopes for relative clause type. There was a significant main effect ($\beta = 0.34$, SE $= 0.04$, $\chi^2(1) = 39.1$, $p < 0.001$), indicating that the development of relative clause structures in the model corresponded to

the order of acquisition in the human data (S > A > P > IO, where X > Y means that X was acquired before Y). Similar results with the Dual-path model were previously obtained in Fitz, Chang, & Christiansen (2011) where these fine-grained developmental differences were explained in terms of substructure regularities. For example, the subsequences THAT AUX and THAT VERB were shared across several of the input structures (e.g., copular relative clauses, S-relatives and A-relatives) and A-relatives were aided by exposure to similar constructions (see also Abbot-Smith & Behrens, 2006) because the SRN component of the Dual-path model was sensitive to distributional patterns of word categories. Subsequences like THAT DET NOUN and THAT PRONOUN were less common in the input and hence P-relatives were more difficult to learn than A-relatives despite occurring with the same construction-level frequency in the learning environment. Taken together, these results on the production of complex declaratives suggest that the Dual-path model can capture some aspects of relative clause acquisition in children.

Last to develop were the novel center-embedded yes-no questions that were absent in training. Even though the learning environment was consistent with a structure-independent rule where the linearly first auxiliary is displaced, the model eventually generalized to the correct rule with high production accuracy (mean 98.1% at the end of training, SD = 2.6%). Nearly half of the model subjects achieved perfect generalization (100% accuracy) and each of the 25 model subjects reached at least 90% accuracy. Remaining errors occurred almost exclusively in the post-relative clause region (99% of all errors), indicating sources of production difficulty which were unrelated to structure-dependent generalization. Tested questions could have copular, intransitive, transitive or dative relative clauses, which could relativize subjects, direct objects or indirect objects (Table 2). Auxiliaries could function as a copula, modal auxiliary, or as an auxiliary

Table 2

*Examples of correctly produced complex questions*

| Example question | Modality | Auxiliary | Relative clause | Gap | Tense |
|---|---|---|---|---|---|
| *Was the brother that is lazy touching it?* | Non-modal | Different | Copular | Subject | Mixed |
| *Could the boy that could carry the apple run?* | Modal | Same | Transitive | Subject | Past |
| *Can the girl that was sitting jump?* | Mixed | Different | Intransitive | Subject | Mixed |
| *Is the mother that the aunt is holding waiting?* | Non-modal | Same | Transitive | Object | Present |
| *Can the man that could throw it to him eat the toy?* | Modal | Different | Dative | Subject | Mixed |
| *Is the pencil that she can show to the boy nice?* | Mixed | Different | Dative | Object | Present |

*Note.* Novel center-embedded questions with two auxiliaries that were produced correctly at the end of training include a wide variety of structures with differences in tense, modality, and relative clause type.

in a progressive aspect verb phrase, and different auxiliaries could occur across clauses. Verb tense could be present or past and aspect could be simple (copulas, modal auxiliaries) or progressive. Hence, there was far more diversity in the tested items than in previous computational studies of auxiliary inversion (Lewis & Elman, 2001; Reali & Christiansen, 2005; A. Clark & Eyraud, 2007; Bod, 2009; Perfors et al., 2011) and the fact that the Dual-path model could produce all of these question types correctly demonstrates the robustness of meaning-based generalization.

### 3.1    Auxiliary type

Behavioral studies have found that the acquisition of complex questions with different auxiliaries varies over development. Children aged 3;2–5;11 in Crain & Nakayama (1987) produced 59% correct questions with either copulas or progressive verb forms in each clause, whereas children from the same age group only produced 21% correct questions with a modal auxiliary in one of the two clauses. In another study, Ambridge et al. (2008) found that children aged 6.3–7;9 produced 47.4% correct questions with modal auxiliaries in both clauses. Since these children were older than the ones in Crain & Nakayama (1987) and yet produced more errors, these results suggest that modal complex questions were more difficult to learn than copular or progressive forms.

To test for these differences in the model, we split the production data by auxiliary type. Modal questions had modal auxiliaries in both clauses, non-modal questions had a copula or progressive auxiliary in both clauses, and mixed-modal questions had a modal auxiliary in one of the clauses and a copula or progressive auxiliary in the other. Figure 4 (left panel) depicts the intermediate performance on these three types of questions for epochs 5–7 (one epoch corresponds to 20,000 training items). For statistical analysis all data
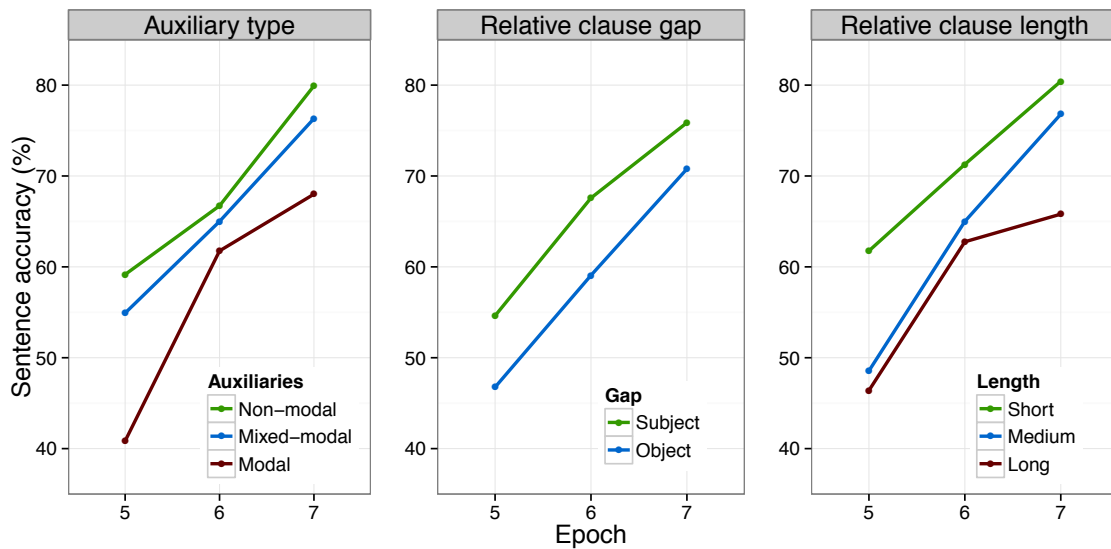


*Figure 4.* Cross-section of production accuracy on complex questions distinguished by auxiliary type in the main and embedded clause (modal, non-modal, mixed-modal), relative clause type (subject, object gap), and relative clause length (short, medium, long).

from epochs 2–21 was used, where model performance was neither zero, nor at ceiling (see Figure 3). A mixed model with auxiliary type (non-modal, mixed-modal, modal) as fixed factor and random intercepts for subjects, with auxiliary type as random main-effect slopes, revealed a main effect for auxiliary type ($\chi^2(2) = 30.2$, $p < 0.001$). Orthogonal contrasts showed that questions with non-modal auxiliaries were learned earlier than the other two question types ($\beta = 0.88$, SE $= 0.16$, $\chi^2(1) = 23.6$, $p < 0.001$) and questions with mixed auxiliaries developed before questions with exclusively modal ones ($\beta = 0.36$, SE $= 0.14$ , $\chi^2(1) = 6.6$, $p < 0.05$). Thus, the model reproduced the order of acquisition for complex questions with different auxiliaries from the developmental literature (Crain & Nakayama, 1987; Ambridge et al., 2008). The bias for non-modal auxiliaries was due to the occurrence of auxiliary forms like *is/was* in a range of different constructions (e.g., as copulas and

progressives). Since the rules for question formation were the same across these constructions, non-modal auxiliaries were being trained more than modal ones in interrogative speech acts, leading to earlier acquisition and better performance. This account is supported by the observation that the non-modal auxiliaries occurring in the model's input language were nearly twice as frequent in the Thomas corpus (Lieven et al., 2009) than the modal ones.

### 3.2 Relative clause type

A second study by Nakayama (1987) investigated 3–5 year old children's elicited production of center-embedded questions with different relative clause types and found that they performed better for items with a subject-relative clause than items with an object-relative clause (66% versus 44% correct). To test whether the model showed similar behavior, a mixed model was fitted to the percentage of correctly produced center-embedded polar questions, with relative clause type as fixed factor. Model subject was a random intercept with by-subject slopes for relative clause type. There was a main effect for type ($\beta$ = 0.37, SE = 0.13, $\chi^2(1)$ = 7.7, $p < 0.01$), indicating that complex questions with subject-relative clauses were learned before questions that contained an object-relative clause, and this is in line with the behavioral results in children (Nakayama, 1987). The explanation for this effect was similar to the account for the A/P-relative clause difference in complex declaratives, in that word category subsequences in subject-relatives (e.g., THAT AUX) were overall more common in the learning environment than subsequences supporting object-relatives (e.g., THAT DET) and the Dual-path model's sequencing system was sensitive to these substructure frequencies during syntactic development.

### 3.3 Relative clause length

Nakayama (1987) also found that children had more difficulty with complex questions containing long relative clauses than with short ones. We split center-embedded questions in the model's test set into three groups, those with a short relative clause (copular + intransitive), those with a medium-length relative clause (transitives) and those with a long relative clause (datives). As in Nakayama's study, we only included subject-relative clauses in the comparison. A mixed model was applied to the percentage of correct productions with relative clause length as fixed factor, a random intercept for subjects, and by-subject random slopes for relative clause length. There was a main effect of length ($\chi^2(2)$ = 8.6, $p < 0.05$), suggesting that longer relative clauses reduced production accuracy in complex questions. Orthogonal contrasts revealed that short relative clauses were learned better than the longer ones ($\beta$ = 0.38, SE = 0.16, $\chi^2(1)$ = 3.9, $p < 0.05$) and medium-length relative clauses were acquired before long ones ($\beta$ = 0.36, SE = 0.17, $\chi^2(1)$ = 4.7, $p < 0.05$). These results are consistent with the developmental data of Nakayama (1987) and the model showed this behavior because of substructure similarity across the different sentence types. Short subject-relative clauses like *the boy that is sleeping...* shared abstract substructures (THAT AUX VERB) with longer ones in transitive and dative structures. Thus, learning to produce the longer relative clauses helped the model to also produce the subsequences in the shorter subject-relatives, which contributed to this length effect.

To summarize these findings, the model generalized auxiliary inversion in a structure-dependent way with 98% accuracy, despite the fact that center-embedded questions with multiple auxiliaries had not occurred in the learning environment. It correctly produced a diverse set of tested questions (Table 2), with differences in tense,

modality and relative clause type, including object-relatives containing distinct auxiliaries across clauses. High accuracy and broad coverage demonstrate that generalization based on message structure is a powerful mechanism for language acquisition. Developmental differences in acquisition were found to be consistent with what is currently known about the production of complex questions in children with respect to different auxiliaries, relative clause structure, and length. Input distributional patterns such as auxiliary frequency and local substructure regularities *modulated* learning and processing but these factors were distinct from the causal mechanism that enabled generalization in our account of auxiliary inversion. From input structures that are widely available to children the model learned how the structure of thought is connected to language-specific forms, and these meaning-form mappings extended to complex questions when novel messages were expressed.

## 4   Errors in development

Having modeled differential learning and the adult state, we now turn to examining more specific production errors observed in development. Errors and their distribution at different stages offer a direct window into the intermediate representations that children develop, providing clues towards a characterization of the acquired linguistic knowledge and the underlying learning mechanism. Two elicited production studies have recorded errors in the development of complex polar questions (Crain & Nakayama, 1987; Ambridge, Rowland, & Pine, 2008). These studies used a similar method in that an experimenter requested children to pose a question to a puppet (*"Ask the puppet if X"*, where *X* contained a relative clause, e.g., *"...the boy that is watching TV is happy"*). Crain and Nakayama tested children around 4;7 years of age and found no production errors that violated structure dependence where the embedded clause auxiliary instead of the main clause auxiliary was placed in sentence-initial position (e.g., *\*Was the boy who _ holding the plate is crying?*). They argued that this provided support for the view that children are endowed with an innate structure-dependent constraint that rules out these errors. Ambridge et al. (2008), however, questioned the design of this study. They argued that structure-dependent errors were unlikely to be observed with the materials used by Crain and Nakayama because auxiliary extraction from the embedded clause would create bigrams that strongly violate the co-occurrence statistics in English (e.g., *who holding* in the above example). In their replication of the study, they instead used complex questions with modal auxiliaries where structure-dependent errors would engender legal bigrams (e.g., *\*Can the boy who run fast can jump high?*). They found that children in the age range 4;7–5;7 (mean 5;1) produced around 5% of such errors in modal questions, and children between 6.3–7;9 years of age (mean 6;9) also committed around 5% of these errors in questions with singular/plural copulas and progressive forms (e.g., *\*Is the lion that eating the man is hungry?*). These errors resembled violations of structure dependence in that the relative clause auxiliary was missing. However, due to the use of identical auxiliaries across clauses in their items, it is not possible to tell whether the initial auxiliary was a copy of the main clause auxiliary, or a displaced embedded clause auxiliary. We therefore label these errors descriptively as *embedded auxiliary* errors. Taken together, these studies have shown that children are unlikely to make errors related to the relative clause auxiliary, suggesting that some constraint influences complex question formation, but this constraint does not seem to be absolute.

A more common error type in both studies was *auxiliary doubling* (34% of all errors across the four experiments) where the embedded clause auxiliary was present and the main clause auxiliary was repeated in its canonical post-relative clause position (e.g., *\*Is the boy who is watching Mickey Mouse is happy?*). Crain & Nakayama (1987) explained these errors

as a result of memory deficiencies; children copy an auxiliary into sentence-initial position and forget to delete it later. Hence, they view this type of error as being caused in performance, outside the core operation of a structure-dependent constraint. Ambridge et al. (2008), on the other hand, argued for a statistical learning account of auxiliary inversion where the main clause auxiliary is reactivated when it forms a high-frequency word category chunk with relative clause-final words (e.g., NOUN AUX in the example above). Both are *prima facie* plausible accounts of why auxiliary doubling might occur, but they beg the question of how production deficits and local attraction are overcome when children retreat from these errors over time. Moreover, these accounts do not explain how children eventually come to know that the main clause auxiliary should be deleted.

Complex questions developed gradually in the Dual-path model and we wanted to determine if errors produced along the way were comparable in kind and magnitude to the error patterns found in children. Replicating errors provides strong, additional constraints for computational approaches, and a complete account of auxiliary inversion should align with these patterns. To this end, model output was coded as an *embedded auxiliary* error if the utterance started with an auxiliary, had a relative pronoun and the embedded verb in the target positions, but the embedded verb phrase auxiliary was omitted (e.g., *Is [...] that _ running [...]*). If the embedded clause auxiliary was present and, in addition, the initial auxiliary was repeated in the main clause verb position, an *auxiliary doubling* error had occurred (e.g., *Is [...] that is running <u>is</u> [...]*). Finally, we tested whether the model committed genuine *structure-dependent* errors in mixed-modal items similar to the ones used in Crain and Nakayama's (1987) Experiment 2. These errors were coded like *embedded auxiliary* errors with the additional requirement that the sentence-initial auxiliary in the model's output had to be the embedded clause auxiliary from the target question, not the main clause auxiliary. This distinction was possible because the mixed-modal items by design contained different auxiliaries in each clause. The three error types are exemplified in Table 3. As before, a correct polar question was produced when the utterance matched the target

Table 3
*Example error types*

| Declarative | Question | Error type |
|---|---|---|
| The boy who is holding the plate is crying. | Is the boy who Ø holding the plate is crying? | Embedded auxiliary |
| The boy who is jumping was happy. | Was the boy who is jumping **was** happy? | Auxiliary doubling |
| The boy who can run is jumping. | **Can** the boy who Ø run is jumping? | Structure dependence |

*Note.* Bold face and null constituents mark the error location.

question word-by-word. In order to compare errors produced by the model against the child production data, we used the weighted average (by number of subjects) from the two experiments in age range 3;2–5;11 (Crain & Nakayama, 1987) and the data in the age groups 4;7–5;7 and 6.3–7;9 from Ambridge et al. (2008) since these groups represented three different mean ages. For comparison, the human ages were mapped onto model epochs that matched the overall pattern across the three studies. Figure 5 shows that the model generated the error patterns that were found in the experimental studies when production accuracy was comparable between the model and children. *Embedded auxiliary* errors were frequent in the model initially, but quickly diminished over development. *Auxiliary doubling*
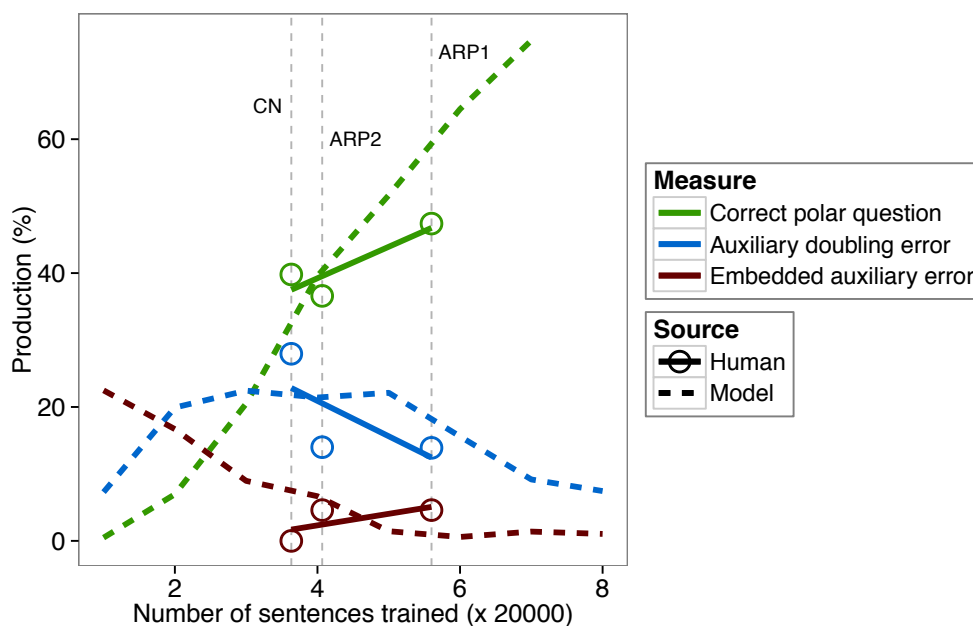
*Figure 5.* Error comparison between the model and child production data from three age groups over development. CN denotes the weighted average of Experiments 1 and 2 in Crain & Nakayama (1987), ARP1 and ARP2 refer to Experiments 1 and 2, respectively, in Ambridge et al. (2008).

errors followed an inversely U-shaped curve, starting off low, but rising to approximately 20% by epoch 2, and this level was maintained until epoch 5 where they began to decline towards 0.4% at the end of learning. Furthermore, the model produced *no* genuine *structure-dependent* errors in the age range that matched the age of children in the study of Crain and Nakayama (epochs 3–4). In other words, the embedded clause auxiliary was never placed sentence-initially which is consistent with the available child production data. Thus, the very same model that reached adult production accuracy (Figure 3) and reproduced developmental differences in complex questions (Figure 4), also replicated the types and magnitudes of errors found in children—the absence of *structure-dependent* errors in mixed-modal items (Crain & Nakayama, 1987), the moderate occurrence of *embedded auxiliary* errors (Ambridge et al., 2008), and the large number of *auxiliary doubling* errors found in both these studies. This confluence suggests that the Dual-path model captures critical properties of the mechanism underlying the acquisition of complex questions in children.

To understand the nature of these errors and why the model was able to replicate them it is useful to reassess the process of auxiliary inversion. In generative accounts, an auxiliary is moved or copied into sentence-initial position and this right-to-left operation creates a choice in questions with multiple auxiliaries. The prominence of *auxiliary doubling* errors across studies, however, is incompatible with movement and the *embedded auxiliary* errors observed in Ambridge et al. (2008) argue against a copy account. Complex question production, in humans and our model, is an incremental left-to-right process that starts with a conceptualization of the intended message, and the communicative intention to elicit information precedes formulation. The first production step is to generate an auxiliary which belongs to the questioned proposition. Input evidence for this association is

unequivocal since it holds in all single-clause and complex questions in the learning environment. Thus, children can learn from abundant experience that question production starts with an auxiliary that is part of the questioned proposition and adopt this default when producing other simple and complex questions. Put the other way round, input evidence to initiate a question with an auxiliary other than the one linked to the questioned proposition is completely absent. From this perspective, it is no longer puzzling how children can come to know which auxiliary has to be placed in front. The sentence-initial auxiliary always belongs to the questioned proposition and therefore does not mark a choice point between main and embedded clause auxiliary. As a consequence, errors where the auxiliary from an unquestioned proposition initiates production are not to be expected, and this explains the absence of *structure-dependent* errors in child production data (Crain & Nakayama, 1987; Ambridge et al., 2008).

When the relative clause with a modal auxiliary or progressive verb is being produced, the first genuine choice point is reached which is a potential source of error. The production system needs to link the initial auxiliary to some verb phrase and the embedded verb is the first option. This link is supported by all simple questions in the input, and when main and embedded auxiliaries are identical (as in the materials of Ambridge et al. (2008)) *embedded auxiliary* errors can occur. At the same time, complex declaratives provide strong input evidence *against* omitting the relative clause auxiliary and this might explain why these errors were infrequent, in particular within the elicited production paradigm where children could simply repeat the target relative clause from the experimenter's instructions.

The second choice point occurs at the transition back into the main clause where the auxiliary needs to be suppressed. Here, again, complex declaratives provide strong evidence for retaining the auxiliary and local attraction might also support this choice (Ambridge et al., 2008). Therefore, a large number of *auxiliary doubling* errors should be expected and this is in line with the experimental evidence. Thus, incremental left-to-right production in the model predicts the pattern of errors observed in children's elicited production. Error types and their distribution over development in turn suggest that complex question formation is the product of several, independent processing steps that have to be acquired separately (see also Mayer, Erreich, & Valian, 1978; Rowland, 2007):

(1)    FRONT-AUX        Start with auxiliary from questioned proposition.
(2)    EMBEDDED-AUX    Produce relative clause with full verb phrase.
(3)    SUPPRESS-AUX    Omit auxiliary in main clause canonical position.

If children apply FRONT-AUX, *structure-dependent* errors are ruled out. They are ruled out not by a hard-wired constraint on a movement operation, but by knowledge acquired from copious input evidence that initial auxiliaries in polar questions group with the questioned proposition. If children apply FRONT-AUX but fail to apply EMBEDDED-AUX, this results in the *embedded auxiliary* errors observed in Ambridge et al. (2008). If they apply FRONT-AUX and EMBEDDED-AUX but fail to apply the SUPPRESS-AUX step, a complex polar question with an *auxiliary doubling* error will be produced. When the learning task is re-conceptualized in terms of incremental left-to-right production, rather than right-to-left auxiliary fronting, errors are naturally expected to occur at sentence-internal choice points while representations for language are still developing, and this is consistent with the experimental data.

To reach an adult state, the Dual-path model needed to acquire these three components—FRONT-AUX, EMBEDDED-AUX and SUPPRESS-AUX—which involved both pathways. For instance, speech act distinctions were represented in the meaning system,

whereas the frequency of different word sequences was encoded in the sequencing system, and both pathways contributed to production at the output layer. Furthermore, the Front-Aux, Embedded-Aux and Suppress-Aux components could depend on message structure and input evidence in different ways and were therefore learned at different rates. When combinations of these components were still to be learned, the different error types were produced. Thus, the model provides a unified account of the acquisition of the correct rule as well as the generation of errors at different points in development. Rather than attributing these errors to deficits extraneous to language, like memory limitations, the model explains errors as a result of the same learning mechanism that allowed it to acquire the language in the first place.

In the remainder of this paper we will investigate what the precise role of the message and the influence of linguistic input was in learning the Front-Aux, Embedded-Aux and Suppress-Aux steps of complex question production. In section 5, we contrast different model versions where features in the message were manipulated to trace their contribution to the Front-Aux and Embedded-Aux components. In section 6, we identify input structures required to learn the Suppress-Aux step and validate model-based discoveries by a corpus analysis. These comparisons will shed light on how the model could gradually recover from the error types associated with these three production steps, until an adult performance level was achieved.

## 5   Role of the message

The unique character of the approach taken here is that question formation is modeled under the assumption that language is used to convey meaning. Syntactic rules develop in the process of learning to link sentence meaning to sentence forms in the input and the message is made up of different propositions that support the acquisition of structure-sensitive rules. Incremental left-to-right production creates choice points in complex questions that could be inferred from the observed error patterns in children's elicited production. The model had to learn how these choice points were associated with different parts of the message.

### 5.1   Speech act encoding

The first step was to produce the auxiliary belonging to the questioned proposition at the start of the sentence (Front-Aux). The target auxiliary type (copula, modal, or progressive) was determined by a combination of tense, aspect and action features in the Event Semantics. For instance, the combination of PROG + AA features mapped to *is + verb*, the MODAL feature mapped to *can + verb*, and the silence of all features (PROG, AA, and MODAL) mapped to the copula *is + adjective*. The Event Semantics also distinguished speech acts in terms of a YSNO feature which signalled for both simple and complex messages that a question instead of a (default) declarative sentence had to be produced. Single-clause sentences were represented by a message with one proposition. For example, the sentence *The boy is jumping* was encoded by the message Action1–JUMP, Theme1–BOY and Event Semantics features PROG1, AA1, TH1. When the YSNO feature was present, the target utterance changed to *Is the boy jumping?*. Through experience of these simple questions in the learning environment, the model could learn to associate the YSNO feature with the roles coded as Action1 and Theme1, and the Event Semantics features indexed by 1. This association linked the YSNO feature with the questioned proposition and the link was gradually strengthened by every single-clause question in the language input. Complex

questions, on the other hand, were supported by a message with two propositions. For example, *Is the boy that was jumping running?* was represented by the Action1–RUN, Theme1–BOY, Action2–JUMP, Theme2–BOY and Event Semantics features PROG1, AA1, TH1, PROG2, PAST2, AA2, TH2, TH1-TH2, together with the neutral YSNO feature. Due to the learned association between the speech act marker YSNO and the proposition indexed by 1, over development the message came to encode knowledge of which of the two propositions was questioned. Since the questioned proposition was always represented in the same spatial location of the Event Semantics, this knowledge transferred to multi-clause questions. Thus, the model could activate the correct auxiliary *is* from the questioned proposition (rather than *was*) in the Front-Aux production step. With this knowledge in place, no errors occurred where a question was initiated with an auxiliary from the proposition expressed in the embedded clause and the absence of these errors is consistent with the existing child production studies on complex questions (Nakayama, 1987; Crain & Nakayama, 1987; Ambridge, Rowland, & Pine, 2008).

To demonstrate that the YSNO feature itself carried no clause-specific information, the model was compared to a version where questions were the default speech act and declaratives were marked instead by a feature DECL in the Event Semantics. Representing questions by the absence of a feature entails that there was no overt marker in the message that signalled the questioned proposition. The model had to learn from experience which proposition the sentence-initial auxiliary was linked to. All other parameters were kept constant and the model was trained and tested as before. Results from this comparison indicated that the two model versions performed identical in every relevant aspect. At the end of training, accuracy on novel complex questions with two auxiliaries reached 99.1% compared to 98.9% for the model with declaratives as default structures (see Figure 3). Thus, the representation of questions by means of an active YSNO feature was not necessary for structure-dependent generalization. The model version with an interrogative default also produced similar auxiliary doubling patterns (ranging between 25.3% initially and 0.1% at the end of training) and matched children's errors in development equally well. This model comparison shows that the specifics of semantically encoding the speech act distinction were negligible. What was crucial for learning the Front-Aux step was that different propositions were systematically distinguished in the conceptual structure of the message such that interrogative forms could reliably be linked to the questioned proposition through linguistic experience.

### 5.2   *Reduced influence of the message*

After the initial auxiliary has been produced in the Front-Aux step, it needs to be linked to one of two non-adjacent verbs. In the critical complex questions, the embedded clause verb is the linearly first, but incorrect, option. To generalize in a structure-dependent way, the model had to retain the embedded clause auxiliary and produce a full verb phrase inside the relative clause. Thus, the second component in complex question formation was to produce the embedded clause auxiliary (Embedded-Aux). Surface distributional evidence for the auxiliary-verb link was conflicting. In simple yes-no questions, the sentence-initial auxiliary and its corresponding verb form were separated by the subject noun phrase (*Is the boy jumping?*), in complex declaratives (*The boy that is jumping is running*) the first verb was always immediately preceded by its auxiliary. Since simple questions were substantially more frequent than complex declaratives, it could be expected that there was a tendency for the model to associate the initial auxiliary with the first verb and therefore fail on the Embedded-Aux step early in development. To overcome this input bias the model had to

learn that a relative clause was necessary to express multiple propositions in the same sentence and that one of the propositions was linked to the embedded clause verb phrase. Acquiring a notion of clause structure was a prerequisite for generalizing auxiliary inversion to complex questions.

In order to test these hypotheses about the EMBEDDED-AUX component, the model from section 3 was compared with a version that received a reduced message where activation values in the EVENT SEMANTICS were scaled down from 1.0 to 0.05. This allowed the model to still produce complex questions but it diminished the influence of the message on learning and generalization in two ways. First, active message features were less distinct from silent ones which made different speech acts similar and more confusable. And secondly, it increased the burden on the sequencing system in that learned representations were influenced more by input distributional properties over sentence forms than by their underlying messages. Apart from the EVENT SEMANTICS, all other parameters were kept constant, and 25 model subjects were trained on the same 500,000 input items and tested as before. The results of this message manipulation are depicted in Figure 6. Solid lines represent generalization in the model of Figure 3, dashed lines show behavior when the activation strength of the message was reduced. Correct generalization here means that the
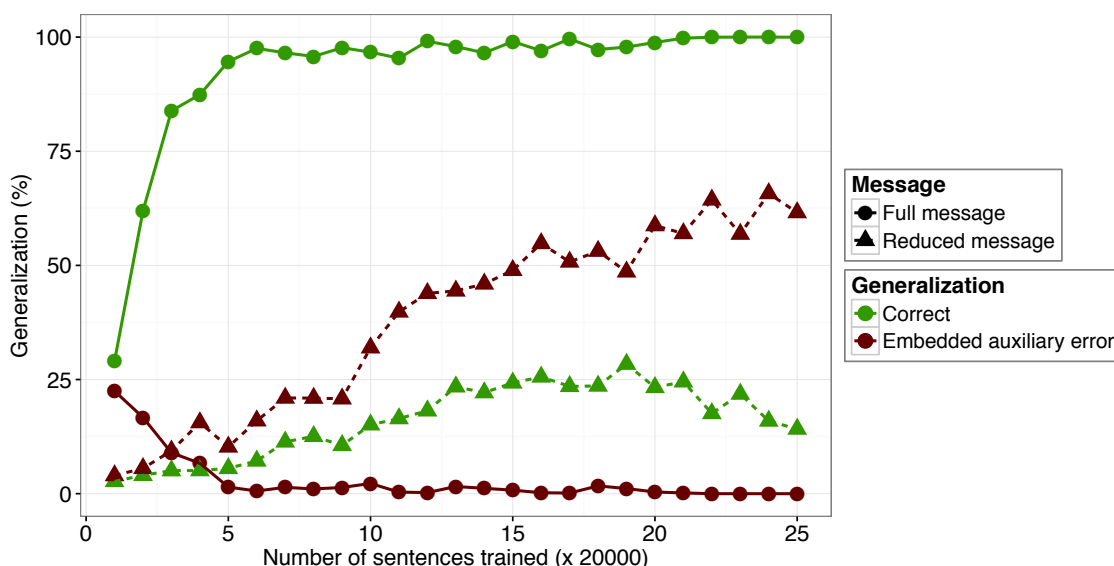


*Figure 6.* Different generalization abilities in the model when the influence of sentence meaning was reduced in the EVENT SEMANTICS, compared with a full message.

model produced the target sentence-initial auxiliary (FRONT-AUX), a relative pronoun in the target position and an embedded clause verb phrase with the target auxiliary (EMBEDDED-AUX). Incorrect generalization was defined by the same pattern except that the embedded clause auxiliary was missing (EMBEDDED-AUX error). With the full message, performance converged on the correct generalization after around 5 epochs. With the reduced message, the model considered two different auxiliary inversion rules early in development, one where the initial auxiliary was associated with the relative clause verb phrase (and thus the auxiliary was omitted in this position), and another, correct rule where the initial auxiliary belonged to the main clause verb (and thus the embedded auxiliary was produced). After 3 epochs the erroneous generalization started to become dominant. Eventually, the model lacking the full message arrived at the incorrect rule and produced

more complex questions with Embedded-Aux omitted (14.2% correct and 61.6% incorrect productions at the end of training). Correct generalization was higher with the full message than the reduced message ($\beta = 7.1$, SE $= 0.09$, $\chi^2(1) = 245$, $p < 0.001$), indicating that meaning was critical for structure-dependent auxiliary inversion.

Learning from identical input, the same architecture could display diverging behavioral patterns. Thus, surface distributional regularities, such as lexical or word category substructure frequencies, did not explain the model's capacity to generalize to the correct rule (see also Kam et al., 2008). At first glance this observation would seem incompatible with results obtained by distributional learning accounts such as Reali & Christiansen (2005). This study, however, modeled sentence classification rather than incremental production and the tested items lacked the structural diversity of questions in our test corpus. Both task and test items differed from the Dual-path model approach. To show consistency, we implemented the bigram-based learner of Reali and Christiansen and trained it on the input to our model. It was then used to classify the complex questions with subject-relative clauses from our test set. The grammatical version was preferred 100% of the time, for each of the 25 distinct training sets our model subjects were exposed to. Thus, in terms of bigram support, the input to our model was indistinguishable from the corpus of child-directed speech used by Reali and Christiansen. Nevertheless, in sentence production without meaning our model converged on the incorrect rule. This underlines our remark in the introduction that biases for classification may be insufficient for production and highlights the need to model auxiliary inversion in a language processing task (rather than whole sentence classification). It also indicates that the addition of meaning to models of language acquisition *qualitatively* changes what is being learned. Bigram frequencies at the lexical or word category level (Lewis & Elman, 2001; Reali & Christiansen, 2005; Ambridge et al., 2008) may only have an explanatory role within assumptions about the meaning that is associated with these bigrams.

Although meaning constraints proved crucial for generalization, the model with full message did not immediately arrive at the correct rule either but generated 22.5% Embedded-Aux errors early in training (first test point, Figure 6). While the system had not yet learned how the structure of meaning was linked to sentence forms, it had to rely mainly on sequencing regularities in the input. In the majority of language input, a sentence-initial auxiliary belonged to the sequentially first verb form (e.g., in all single-clause and right-branching questions). Thus, early in development the initial auxiliary was matched to the first verb in complex questions, resulting in Embedded-Aux errors. Over time, the model learned that the message could represent distinct propositions that were systematically expressed in different clauses and in this manner, a notion of clause structure was gradually acquired from experience. The model also learned that there were message components that controlled embedded verb phrases and that relative clause forms never had an auxiliary extracted. Message representations for relative clauses were identical across speech acts and once the model started learning these partial mappings, Embedded-Aux errors were decreasing.

The two model variants contrasted here instantiate distinct hypotheses about how the learning problem for auxiliary inversion constraints can be conceptualized. According to Chomsky's formulation (Chomsky, 1968, 1975, 1980), children should adopt a linear, structure-independent rule given the evidence they receive from the ambient speech. The reduced message version can be viewed as a model that tests this prediction. When critical items with auxiliaries in both clauses were absent from the learning environment, and the model had to induce auxiliary inversion rules from surface distributional properties alone, it

converged on behavior where the sentence-initial auxiliary was linked to the embedded verb phrase, and this supports Chomsky's claim. On the alternative view that we propose here, the induction of syntactic rules from surface forms is not the primary target in language acquisition. Instead, our model learned to link sequences of words to an intended message, and these mappings combined into larger meaningful utterances. The message reflected semantic and pragmatic constraints on question formation, and partial meaning-form mapping regularities generalized to novel constructions. Syntactic rules for auxiliary inversion developed as a byproduct of learning the meaning and communicative function of questions and relative clauses. Despite the fact that critical items were not encountered in training, the model was driven towards behavior that resembled structure-dependent auxiliary fronting.

Initial overgeneralization shows that the model did not have an innate syntactic constraint built-in that would categorically prohibit the omission or displacement of an embedded clause auxiliary. The message was not forcing the learning system to adopt the correct rule from the outset. It was only through experience of meaning-form relations for simple questions and complex declaratives that the correct pattern was being acquired. With reduced message input, linguistic experience alone was insufficient and led to the wrong generalization. Thus, neither message structure nor input properties in isolation explained how auxiliary inversion was acquired. It was the complex interplay between these factors from which structure-dependent generalization emerged.

## 5.3 Development of meaning

At this point it might be useful to clarify the status of the message in the proposed account. As the previous simulation has shown, a structured message distinguishing speech acts and multiple propositions was necessary for generalizing auxiliary inversion to complex questions. Here we show that it is not required that this semantic/pragmatic infrastructure be in place from the start of language acquisition. Our prior focus was on modeling the development of auxiliary inversion from meaning, rather than the development of meaning itself. Providing the full message from the beginning, we have merely taken a computational shortcut which should not be understood as a commitment to meaning innateness. To illustrate that generalization could also happen when the message became available incrementally in the course of acquisition, we simulated the gradual development of meaning where message components were partially blanked out. These components could be individual ROLE-CONCEPT bindings or the EVENT SEMANTICS (see Figure 2). Thus, the message could be missing information about who was the agent, theme or recipient in a target utterance, or information about the entire event structure, including tense/aspect, modality, and speech act distinctions. In this training protocol, early in acquisition none of the message components were present and every 50,000 episodes another ~10% of the features were randomly added in. The full message was only available to the model during the final 50,000 training sentences. In all other ways the model was identical to the previous version. Figure 7 shows that structure-dependent generalization was delayed but eventually an adult state was reached at the end of training (95.2% accuracy). The comparison demonstrates that the assumption of full message input was not essential for learning the FRONT-AUX and EMBEDDED-AUX steps in complex question production. This suggests that as children gradually acquire the meaning-form relations of basic constructions such as simple interrogatives and relativization, correct generalization to more complex, composite constructions becomes increasingly likely. The development of structured meaning itself, which the present work does not address, might be grounded in elemental, presumably
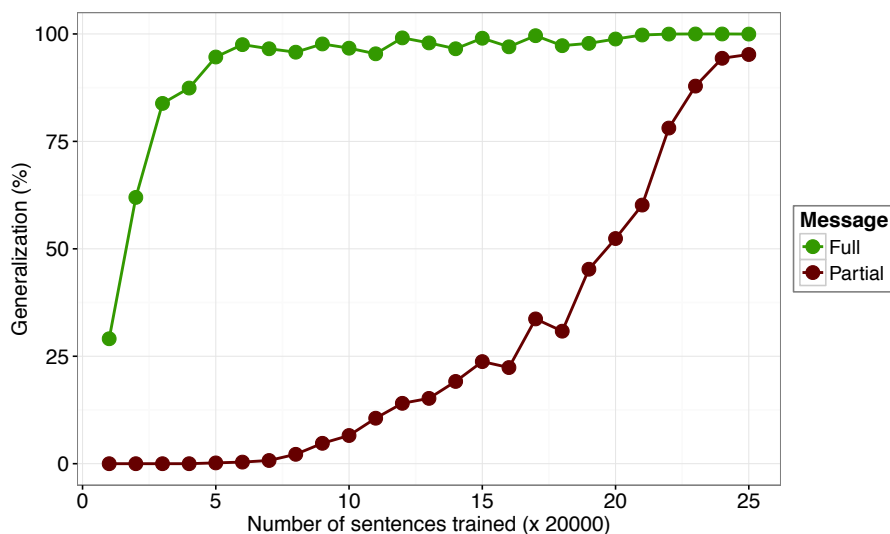
*Figure 7*. Structure-dependent generalization when messages were developing over time, compared with a model where full messages were available from the beginning of acquisition.

innate, cognitive primitives for categorization, word learning, and the conceptual structuring of perceptual experience (Palermo, 1982; A. L. Brown, 1990; Carey, 2009; Jackendoff, 2002).

## 6   The role of the input

The FRONT-AUX and EMBEDDED-AUX components in left-to-right question production, and therefore structure-dependent generalization, were directly related to the structure of meaning. The third production step was to omit the main clause auxiliary after exiting from the embedded clause (SUPPRESS-AUX) and here we investigate how the composition of the input could modulate the learning of this component. Auxiliaries at this position were normally associated with the PROG feature when paired with complex declarative utterances. Cameron-Faulkner et al. (2003) estimated that complex utterances with two lexical verbs make up around 6% of child-directed speech and the number was comparable in the model's input ($\sim$5%). Hence, there was positive evidence that could support the mapping of PROG to an auxiliary before the main clause verb. In the target complex yes-no questions with progressive aspect (and similarly for modal auxiliaries and copulas), however, the YSNO feature signalled that there should be no auxiliary immediately preceding the verb form as in all simple polar questions. Because of these conflicting features, the model produced many auxiliary doubling errors early in development and only gradually acquired SUPPRESS-AUX towards the end of training.

High rates of auxiliary doubling have also been found in two production studies (Crain & Nakayama, 1987; Ambridge et al., 2008) and this is difficult to explain for accounts where auxiliary inversion is governed by innate movement rules. These findings rather indicate that omission of the main clause auxiliary may be acquired through relevant experience over an extended period of time. The Dual-path model approach was assuming that direct evidence might be required to learn SUPPRESS-AUX in that 0.3% of the model's input were center-embedded polar questions with an auxiliary in the main clause, e.g.

(13)     Is the boy that eats happy?

In this section we examine whether this input assumption influenced the rate of auxiliary doubling errors (Suppress-Aux) and whether it was critical for structure-dependent generalization (Front-Aux and Embedded-Aux).

### 6.1   Manipulating the input distribution

To test the role of examples like (13), the same model was trained and tested as before but without these structures in the input. The comparison shows that the exclusion of these items from the learning environment had little effect on structure-dependent generalization (Figure 8, right panel). In both input conditions the model adopted the correct auxiliary
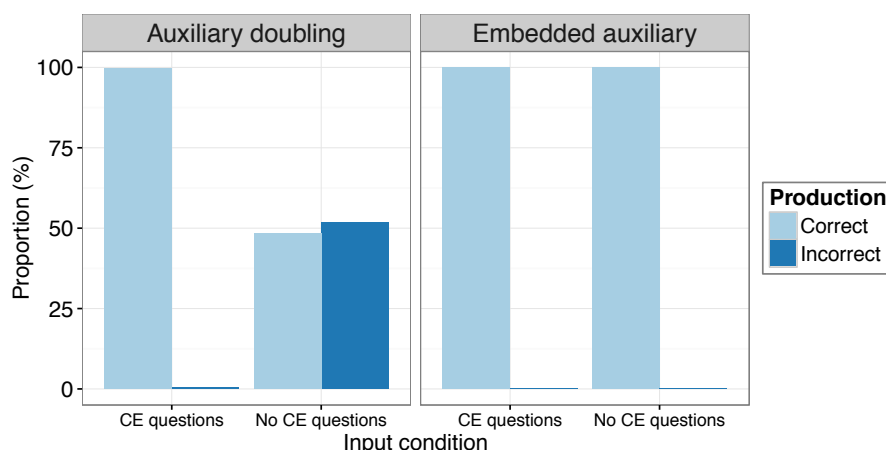


*Figure 8.* Model performance with and without center-embedded yes-no questions with a main clause auxiliary in the input, averaged over the final three test points. Neither input condition contained center-embedded questions with two auxiliaries.

inversion pattern at the end of training ($\chi^2(1) = 2.4$, $p = 0.12$). Thus, the inclusion of a small number of questions like (13) in the model's input was not necessary to achieve structure-dependent generalization. Simple polar questions and complex declaratives in the learning environment were sufficient for the model to converge on the correct behavior (Front-Aux and Embedded-Aux steps). The amount of auxiliary doubling errors, on the other hand, was strongly influenced by the presence of items (13) in the model's input (Figure 8, left panel). Without these questions, the model more frequently reactivated the main clause auxiliary after sequencing the relative clause (around 53% of the time) and the difference between input conditions was significant ($\chi^2(1) = 48.7$, $p < 0.001$). In other words, the model was not reliably learning the Suppress-Aux step in the absence of direct evidence.

Based on these results, we hypothesize that children require exposure to these, or similar kinds of questions, in order to retreat from auxiliary doubling, and the high rate of such errors found late in development supports this possibility in that even older children aged 6;3–7;9 produced 24% auxiliary doubling errors out of all valid attempts (Ambridge et al., 2008). In addition, the input manipulation shows that auxiliary doubling and structure-dependent processing were largely independent in the model. Convergence on the correct inversion rule was similar in both input conditions but auxiliary doubling differed significantly. Together with the results from section 5, this suggests that Front-Aux, Embedded-Aux, and Suppress-Aux are three separate aspects of learning complex questions which are governed by distinct, local meaning-form mapping regularities, and the model needed to learn each of these regularities from relevant experience.

## 6.2 Auxiliary doubling and bigram frequency

It has been argued that auxiliary doubling errors might be related to bigram probabilities at the transition from relative clauses into the main clause (Ambridge et al., 2008). On this view, if auxiliary repetition creates high frequency bigrams of word categories (e.g., relative clause-final NP + auxiliary), this should cause more doubling errors than lower frequency bigrams (e.g., relative clause-final VP + auxiliary). However, while bigram frequency might correlate with error rates in children, it is not clear what the role of bigram frequency is for learning to *avoid* auxiliary doubling, as adults typically do. The Dual-path model produced large numbers of these errors early in development and eventually learned to omit the auxiliary in the post-relative clause region, but it failed to do so when relevant evidence was not provided (Figure 8, left panel). To determine the role of bigrams in explaining this difference, we calculated the frequency of bigrams occurring in test items at the transition from the relative clause into the main clause. For example, if the test item was *Is the boy that is happy nice?* the number of occurrences of *happy is* and *happy nice* in the training set was counted, both at the lexical and word category level. The former bigram would support auxiliary doubling, whereas the latter would support omission. In the input condition without direct evidence, bigrams that supported doubling were approximately 23 times more frequent, lexically, and 4.7 times more frequent at the category level than bigrams that supported omission. This would seem to suggest that auxiliary doubling can be explained in terms of bigram frequency at the main clause transition and similar arguments have been made to explain doubling in children (Ambridge et al., 2008). However, in the input condition that did provide evidence against doubling, only 0.3% of the model's training set differed and the effect on bigram frequencies was therefore marginal (the above factors changed to 22 and 4.6, respectively). Bigram probabilities still supported auxiliary doubling, and yet the model's behavior changed dramatically from high to low error rates. This suggests that surface distributional properties of the input, such as the frequency of lexical and word category subsequences, may not be critical, unless one considers how they are associated with meaning. It also indicates that frequency-based accounts, where child-language data is correlated with input statistics need to be tested in concrete learning models that are sensitive to these properties. Causal explanations can then be obtained through manipulations of the input environment from which the models learn. The Dual-path model could retreat from auxiliary doubling based on sparse evidence because it was learning to link word order regularities at the main clause transition in a few positive examples to an underlying message representation. In this manner, it could overcome surface distributional biases and acquire the correct syntax for complex questions where main clause auxiliaries are not duplicated. As with structure dependence, adding meaning into the equation qualitatively changed what was being learned even if differences in the input were very small. Consequently, postulating absolute frequency thresholds for the learnability of constructions (Legate & Yang, 2002) may be premature.

## 6.3 Support from child-directed speech

In order to learn SUPPRESS-AUX, the model required direct evidence against auxiliary doubling in the form of complex questions with a main clause copula or auxiliary (e.g., *Is the boy that eats happy?*). To determine if this learning account was supported by child-directed speech, we searched the Thomas corpus (Lieven, Salomo, & Tomasello, 2009) from the CHILDES database (MacWhinney, 2000) for similar utterances. Our version of the corpus was not morpho-syntactically annotated, so PERL regular expressions were used to extract

items which were then manually inspected. Since there is not guarantee that all relevant items were picked up, the reported estimates are conservative. The corpus contained a total of 336,030 maternal utterances recorded over a period of three years (age 2;0–4;11). Mother and child were recorded for 5 hours a week between age 2;0 and 3;2 (5 hours × 56 weeks = 280 hours) and for 5 hours within one week for every month until 5 years old (22 months × 5 hours = 110 hours). There were 98,131 questions in the corpus (29.2% of all utterances), 52,628 of which were marked by intonation or question tags. The remaining 45,503 questions (46%) were canonical in form, i.e., polar questions with subject-auxiliary inversion or *wh*-questions. Short form questions without predicate were not counted as canonical (e.g., *Are you?*). 1,456 canonical questions had a relative clause with or without overt relative pronoun, including fused relative clauses (e.g., *Is that what you're saying?*) but excluding reduced relative clauses (e.g., *Can you see the hands moving on the clock?*). Thus, complex questions appeared to be quite frequent (3.2% of all canonical questions) and relative clauses in a question context could help children to recognize that their surface form and semantic-pragmatic function does not differ across speech acts.

Out of the canonical complex questions, 1,194 items were polar questions (77%) and 262 were *wh*-questions (23%). 1,422 (98%) items had an auxiliary in the main clause (copula, progressive aspect, modal, or *do*-support), and 651 (45%) had an auxiliary in the relative clause. Auxiliaries in both clauses occurred in 630 questions (43%). Except for the two items listed below, complex questions were right-branching and the relative clause did not interrupt the main clause verb phrase.

(14)    How can yoghurt that comes out of the fridge be hot? [2;11]

(15)    Was that cheese you had from the fruit lorry not nice? [4;01]

Since neither one of these questions had an auxiliary in the relative clause, these findings are consistent with the claim that direct evidence for structure dependence is rare in typical child-directed speech (Chomsky, 1980; Legate & Yang, 2002; MacWhinney, 2004). Nevertheless, these utterances are important because they provide positive evidence for auxiliary omission when returning to the main clause from an embedding. Although two items are few (density 2/390 hours = 0.005128), the corpus is just a small subset of the input Thomas received. If we arbitrarily assume that the child heard language for four hours a day, every day for 14 years (365 × 4 × 14 = 20,440 hours), then Thomas should have heard at least 104 examples of type (14) and (15) which provide evidence against auxiliary doubling. Further direct evidence for omission comes from complex *wh*-questions like

(16)    Where is the little lorry that came with this? [4;09]

which can be viewed as the interrogative counterpart to the declarative

(17)    The little lorry that came with this is [in the box/on the shelf/etc.]

where the main clause auxiliary does not recur after the relative clause. The Thomas corpus contained 174 such complex *wh*-questions with a main clause auxiliary (0.38% of all canonical questions) and based on their distribution in the corpus we estimated that their frequency was increasing by age (b = 0.0026, $p < 0.05$). In addition to items like (14) and (15), these questions can provide children with evidence for auxiliary omission in the post-relative clause region. By the same density argument as above, Thomas should have heard at least 9,198 complex *wh*-questions with a main clause auxiliary by the age of 14, and these numbers will be higher if older children get more complex input than a five-year old

and hear language for longer than four hours a day. Thus, there is clearly a substantial amount of input that children can use to learn about suppressing the main clause auxiliary following a relative clause.

Learning auxiliary omission in the Dual-path model (Suppress-Aux) required positive evidence and corpus data suggests that such evidence might be available to children in various forms. The presence of complex questions with a main clause auxiliary was not necessary for achieving structure-dependent generalization, but it helped to reduce auxiliary doubling errors over time. A small amount of these items was sufficient because the model could link substructures at the main clause transition to meaning representations and thus overcome input distributional biases towards auxiliary doubling. Manipulations of the model's message and input environment demonstrate the complex interplay between meaning-based constraints and language experience in the acquisition of auxiliary inversion.

## 7   Discussion

The present work investigated whether a computational model of sentence production could simulate the acquisition of auxiliary inversion from meaning and generalize to novel polar questions in a structure-dependent way. Structure dependence is often discussed in terms of *learned* versus *innate* constraints, which has created an unfortunate divide (see also Yang, 2004; Pearl & Lidz, 2009; A. Clark & Lappin, 2011) between seemingly incompatible views and one of the aims of this work was to find some common ground. Auxiliary inversion is a rare phenomenon in languages of the world (Siemund, 2001; Goldberg, 2006) and all theories assume that it requires learning about English auxiliaries, progressive verbs, and embedded clause structures. Learning and appropriate input are therefore essential in all accounts of this phenomenon. In nativist accounts, structure dependence is part of Universal Grammar and encoded into the initial state of the language system (Crain, 1991; Berwick et al., 2011; Crain & Thornton, 2012). Together with linguistic experience it becomes a particular learned grammar (Chomsky, 1983). Although this approach is conceptually simple, a computational implementation that instantiates its solution is still lacking. It is unclear how prior syntactic knowledge could be represented by the initial state, how it would interact with experience, and how observed production errors (Nakayama, 1987; Crain & Nakayama, 1987; Ambridge et al., 2008) could be explained in this framework. Constructivist approaches have argued that the input itself provides constraints that learning and experience can turn into a structure-dependent grammar (Reali & Christiansen, 2005; Ambridge et al., 2008; Ambridge et al., 2014). The explanatory burden, however, cannot be shifted entirely onto the input and its distributional make up. Generalization to novel instances requires inductive biases that cannot themselves be learned (Mitchell, 1997; Pinker, 2004). Thus, even if the input contains relevant direct (Pullum & Scholz, 2002; Scholz & Pullum, 2002) or indirect evidence (Reali & Christiansen, 2005), it would still have to be explained which biases are needed within a formal learning account to make use of this information and generalize beyond experience. Input is the oil that lubes the acquisition machinery, but it is not the machinery itself.

Existing computational models have employed a range of biases to demonstrate that auxiliary inversion can be learned in the absence of positive exemplars, without assuming innate syntactic, or other constraints specific to language. These domain-general biases, however, appear to be too weak (see also Berwick et al., 2011). The models of Reali & Christiansen (2005) and Bod (2009) recognized ungrammatical forms which is distinct from the ability to produce grammatical sentences. The SRN of Lewis & Elman (2001) activated a probability distribution over words at each sentence position but did not generate

well-formed questions. By design, the learner of A. Clark & Eyraud (2007) could not generalize to novel questions which were unsupported by relevant contexts, and the Bayesian model of Perfors et al. (2011) did not explain how grammars in the hypothesis space were learned in the first place. Thus, approaches that have focussed on distributional properties of the input have thus far not yielded working models that can learn to generate auxiliary inversion questions to an adult degree. The approach proposed here modeled the production of complex questions from meaning and acquired adult knowledge of auxiliary inversion where less than 1% of produced utterances were incorrect. It has wider coverage than previous models in terms of the distinct complex question types that were produced correctly. These included questions with copulas, different tensed and modal auxiliaries, *do*-support, and a variety of relative clauses. Moreover, the model's learning trajectories were found to be consistent with fine-grained developmental differences regarding different auxiliaries, relative clause type, and relative clause length.

In our approach, we have attempted to make explicit both the nature of the biases as well as the role of the input that is required to learn complex question formation. Our results suggest that constraints on auxiliary inversion can arise not only from statistical regularities in the language input, or from preset syntactic knowledge, but also from the very meaning that is conveyed in sentence production. If language acquisition involves learning to map between meaning and form, the structure of meaning can constrain the way the language system interacts with experience and restrict the space of learnable grammars. This allowed the model to produce complex polar questions in a structure-dependent way without positive exemplars in the input. When language acquisition was rendered 'meaningless', however, the model converged on the wrong behavior. When the sentence message was made available in piecemeal fashion over development, the model generalized to the novel questions equally well. The assumption that the full message was present from the beginning was therefore not critical in our approach, it was sufficient when structured meaning developed gradually due to other cognitive abilities and their own set of biases. Furthermore, embedded auxiliary omission occurred early in development even in the presence of a structured message. Hence, structure-independent grammars were accessible during acquisition, and not removed from the model space by prior knowledge. Structured meaning was acting as a soft bias on the distributional learning mechanism, rather than as a categorical constraint on the system's initial state. These findings suggest that the syntax of auxiliary inversion was not reducible to the conceptual structure in the message, it was only through the interaction of nonlinguistic meaning and linguistic experience that structure dependence could emerge.

One assumption of our approach is that messages for sentence production were divided up into individual propositions (e.g., JUMP(BOY) and RUN(BOY)) that mapped onto clause-like units. We also assumed that speech act distinctions were represented in the message. This approach differs from non-structured theories of meaning such as distributional semantics (Landauer & Dumais, 1997; Schütze, 1998). If the meaning of words, phrases and sentences like *Is the boy that is jumping running?* would be a point in a multi-dimensional state space, without internal structure, it would be non-trivial to extract the dimensions that encode each proposition. Given the difficulty in separating propositions, it would be hard to mark the fact that *the boy is jumping* is given information and shared between the speaker and the listener, while the fact that *the boy is running* is perhaps known by the listener, but not the speaker. Without the ability to entertain each of these two propositions in isolation, it becomes difficult to understand why a complex question would be asked in the first place. In contrast, the structured approach to meaning argues that

humans perceive events in terms of proposition-like units. Structured meaning is assumed in most theories of adult language (Bock & Levelt, 1994; St. John & McClelland, 1990; Mayberry et al., 2009; Kintsch, 1988; Anderson et al., 2004; Chang et al., 2006; Jackendoff, 2002, Gärdenfors, 2014), and the only question, in our view, is whether these abilities become available early enough to support auxiliary inversion. There is substantial evidence for the early development of sophisticated social cognition abilities, which require multiple propositions (e.g., theory of mind tasks; Onishi & Baillargeon, 2005), and which involve marking some propositions as shared and some as private (Tomasello et al., 2005). There is also evidence for some of these abilities in other species (e.g., chimpanzee A understands the proposition that chimpanzee B has seen the food hidden under a particular bucket; Tomasello et al., 2003), which suggests that structured meaning may have evolved in human evolution. Many of these early social cognition results would be difficult to explain within a non-structured theory of meaning, where the basic unit of an event or proposition is not distinctly represented.

We are, of course, not the first to argue that children can use meaning, or other rich sources of information, in the acquisition of grammar (e.g., Wexler & Culicover, 1980; Pinker, 1984; Steedman, 1996; see the discussion in A. Clark & Lappin, 2011). On these approaches, meaning is inferred from a shared visual environment, situational context, and speaker intentions, and then gets mapped onto syntactic categories in acquisition. This idea is considered problematic because meaning-form relations can be opaque in any given language, they vary across languages, and conceptual structure itself is shaped by language and culture (Levinson, 2003). In the Dual-path model, the acquisition relation between meaning and form was more complicated. Abstract syntactic frames developed through distributional learning in the sequencing system. In situated comprehension, inferred meaning and developing syntax jointly generated internal predictions about upcoming words which in turn drove syntactic development through error-based learning. In production, conceptual structure constrained how syntactic knowledge was used to express an intended message and learned syntax, in turn, could inform the meaning system when to activate conceptual content in novel messages. The link between meaning and form was bidirectional, and language was acquired through interacting pathways for syntax and meaning. The learned grammar was more than a simple one-one correspondence between conceptual and syntactic categories. It got encoded into thousands of synaptic connections which were fine-tuned in the acquisition process to satisfy multiple constraints—including structure dependence—on how meaning relates to form. The addition of meaning into language acquisition models has also been studied in formal learning theory, within the framework of 'identifiability in the limit'. When language acquisition is conceptualized as grammar induction from surface strings, plausible classes of candidate grammars for natural language are unlearnable (Gold, 1967, but see Shinohara, 1994). However, if grammar denotes a mapping between phonetic and meaning representations, rather than a device for generating strings, language acquisition becomes the process of identifying a function. Assuming that pairs of overheard utterances and their inferred meanings are available to children, every class of computable mappings turns out to learnable in this paradigm (Jain, 1999). In a formally precise sense, procedures for language acquisition that have access to meaning are more powerful than those for extracting natural language grammars from sentence form alone.

Errors in development provide important information about the acquisition paths children take in learning auxiliary inversion. Previous accounts were not able to replicate the errors that children typically make, e.g., the highly frequent auxiliary doubling errors.

When the model of Bod (2009) was used to generate (rather than parse) complex questions, it produced around 7% of sentence forms that resembled auxiliary doubling which is well below the error rates found in children. It also generated 46% of error types that were not observed in children. Models that attribute errors to high frequency n-grams (Lewis & Elman, 2001; Reali & Christiansen, 2005) cannot explain how children recover from these errors, unless they assume that the input distribution or the learning mechanism changes over development. In our approach, we replicated the distinct types of children's errors in elicited production and their distribution over development within the same model that had reached adult production accuracy. Informed by these error patterns, we proposed an account of complex questions where incremental processing was factored into three independent components (Front-Aux, Embedded-Aux, and Suppress-Aux). A step-wise learning account has also been suggested for auxiliary inversion in simple questions (Estigarribia, 2009). We then identified input structures that were necessary for the model to recover from auxiliary doubling errors and found evidence for these structures in a large corpus of child-directed speech. Again, the role of meaning was critical here since a small amount of positive evidence allowed the model to overcome strong distributional biases and recover from auxiliary doubling. Thus, meaning can decorrelate frequency of occurrence and successful acquisition, and our findings suggest that it might play a particularly important role in the acquisition of syntactic principles when input evidence is sparse.

Reproducing errors required an architecture that was more complicated than previous approaches which modeled language acquisition as the probabilistic selection of appropriate utterances, structures or grammars from a set of alternatives (Reali & Christiansen, 2005; Bod, 2009; Perfors et al., 2011). One could imagine a meaning-based variant of this task, where meaning is used to reduce the set of alternatives that are being considered (see e.g., Fitz, 2010). This would not change the fact, however, that the task assumes utterance planning to involve the generation of multiple whole sentence alternatives from which one is selected, and this is inconsistent with a large body of evidence supporting the view that language processing is incremental (e.g., Tanenhaus et al., 1995; Altmann & Kamide, 1999; Sedivy et al., 1999; Ferreira & Swets, 2002). To implement an incremental account of language production, the Dual-path model augmented the sequence processor of Lewis & Elman (2001) with a meaning system to guide structural choices. It generated sentences one word at a time and it had to learn language-specific constraints from input structures and meaning that informed this word-by-word generation process.

One consequence of incremental production was the model's the ability to generate errors which resulted when one of the components of question production failed. For example, auxiliary doubling errors occurred when the main clause auxiliary was not yet deleted from its position next to the verb, even when it had been produced at the start of the sentence. The model was able to explain how these errors changed over development because it learned each component of the task separately and they developed at different rates. For instance, Suppress-Aux took more time to learn and hence doubling errors were created for a longer period in development. Thus, incremental production enabled us to explain a wider range of data about the acquisition of complex questions. Another aspect was that the Dual-path model captured the unique semantic relationship between question forms and their declarative counterparts (Berwick et al., 2011). Questions were systematically related to declaratives in that they shared a common message distinguished only by a feature that determined the speech act type. This allowed the model to express the same proposition in declarative or interrogative form. Complex questions were unambiguously tied to their correct interpretation, not via a transformational rule over

forms, but via shared meaning. This link between speech acts was causal in that message similarity enabled the model to generalize partial meaning-form mappings from one construction to another. As the model was learning to incrementally produce one form, the acquired representations directly influenced the acquisition of the complementary form.

In line with constructivist approaches, previous computational models have focused on the role of the input in explaining auxiliary inversion in complex questions. Each of these models used a different corpus from a different child and there is little evidence that they would work if trained on other child language data. Most of these corpora come from play sessions in which the parent and child are playing in a joint space where there is often shared information about many objects. Auxiliary inversion questions are used to request information about the past history of objects that is only known by one participant. Thus, existing child-directed spoken corpora may not be the most representative corpora for the utterances that could support complex question acquisition. Our meaning-based approach argues that the composition of the input might be less important than previous accounts have claimed. Instead, the input is linked to different parts of the message for each proposition, and the message smoothes over variation in the input. The Dual-path model learns to activate semantic roles at particular sentence positions and this helps to address gaps in linguistic experience. For example, if the model hears an object-relativized declarative, it can learn to activate the agent role of the embedded proposition and then the auxiliary. If the agent changes in a novel message, the same processing steps can yield an object-relativized complex question with different event participants. This form of generalization is related to systematicity in language (Fodor & Pylyshyn, 1988) and the Dual-path model has been shown to behave strongly systematic in Chang (2002) and Fitz (2009). The message allows the system to correctly produce utterances that have not appeared in the learning environment.

Although input is required in all different accounts of auxiliary inversion, there is a cline in the amount that is needed. Word-based learning accounts like Lewis & Elman (2001) and Reali & Christiansen (2005) require that every child hear every bigram that is necessary to select the appropriate structure. Syntactic category-based learners (Bod, 2009; Perfors et al., 2011) require less input as they need just one example of a target sequence of categories for each child. Generative accounts have the smallest input requirements as they only need enough to learn English words and any structures that trigger auxiliary inversion rules. The Dual-path model occupies the middle ground as its message and semantic categories help it to generalize over words. Meaning representations in the present account were much the same as in previous work with the Dual-path model (Chang, 2002; Chang, Dell, & Bock, 2006; Fitz, 2009; Twomey, Chang, & Ambridge, 2014; Chang, Baumann, Pappert, & Fitz, 2015). Both Chang (2009) and Fitz, Chang, & Christiansen (2011) used a message with multiple propositions to produce complex declaratives with simple and progressive aspect. Only the inclusion of the YSNO feature associated with interrogative forms was novel in the current model. Due to the similarity of the present representations with those in previous model versions, the message format is motivated by its capacity to explain a wide range of psycholinguistic phenomena like structural priming, heavy NP shift and relative clause acquisition.

In this work, we have emphasized the role of a structured message in explaining the acquisition of structure-dependent representations. It is important to keep in mind, however, that structured meaning needs to be embedded in a theory of syntactic development and incremental production for which the Dual-path model provides a template. We are able to explain the developmental time-course of auxiliary inversion because FRONT-AUX,

Embedded-Aux, and Suppress-Aux regularities are learned gradually at different rates and are activated incrementally in production. Future work needs to address whether structure-dependent syntactic principles beyond auxiliary inversion might be rooted in the structure of meaning as well.

## Appendix

### A.1 Model specifications

Simulations were run in the LENS connectionist software package (version 2.63; Rohde, 1999). Unless otherwise stated, default parameters of the simulator were used. To allow for exact replication, the complete model on which simulations were based is available at: `https://sites.google.com/site/sentenceproductionmodel/home/auxinversion`

The model's sequencing system mapped from the Previous Word layer (83 units) to the CCompress layer (15 units) which connected to a Hidden layer (70 units), through the Compress layer (15 units) and to the Produced Word layer (83 units). A Context layer (70 units) held a copy of the Hidden layer activation at the previous time step and was fully connected to the Hidden layer. At the start of each utterance, all Context units were reset to 0.5. The Produced Word layer used the softmax activation function

$$o_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{1}$$

to create a continuous winner-take-all bias for that layer, where $x_i$ is the net input to the $i$th output unit and $o_i$ is its output. The Previous Word layer received one-to-one inputs from all of the Produced Word layer units and from the previous target outputs, and a winner-take-all filter was applied to increase the output signal-to-noise ratio. During learning from the speech of others, the Previous Word input was set to the sum of the overheard target word and the model's own internal word predictions.

Messages were stored in the weights between the Role-Concept bindings (Figure 2, top panel, right), which consisted of the Role layer (8 units) and the Concept layer (64 units). The Hidden layer connected to the Role layer, which connected to the Concept layer. The Concept layer, in turn, connected to the Produced Word layer. The model assumed that this "production message" was set by the shared visual scene when learning from others, or by the speaker's own message planning during production. The weights between the Role and Concept layers were initially cleared, then for a particular message the Role-Concept bindings between appropriate units (e.g., Agent1=DOG) were set to a weight of 6 and these weights did not change with learning. Similarly, to allow the model to recognize the role of previously produced words, a "comprehension message" was employed (Figure 2, top panel, left). It was identical to the production message, except that the direction was reversed, mapping from concepts to roles, via weights between the CConcept (64 units) and CRole (8 units) layers. The CRole-CConcept bindings were set simultaneously with the Role-Concept bindings, with a weight of 6.

The Previous Word layer connected to the CConcept layer, which connected to the CRole layer, which in turn connected to the Hidden layer (Figure 2, top panel, left). To ensure the model could avoid producing roles that had already been produced, there was also a CRole Copy layer (8 units; Figure 2, top panel, center) which averaged a copy of its own activation with the previous activation of the CRole layer. To help the model learn the links between the previous word and its appropriate concept (i.e., the weights between the Previous Word and CConcept layers, Fig. 1, top panel, left), the previous activation of the Concept layer was used as a training signal for the CConcept layer (light grey line, Figure 2). Finally, the meaning system also included an Event Semantics layer (22 units) connected to the Hidden layer. The Role-Concept links in the production message, the CConcept-CRole links in the comprehension message, and Event Semantics activations were all set before a training or test sentence was processed.

Unless specified otherwise, units in all layers used the logistic activation function, with activation values running between 0 and 1. Weights were initially set to values uniformly sampled between -1 and 1. Units were unbiased in order to make layers more dependent on their inputs for their behavior. However, CONCEPT and CCONCEPT units were biased to -3 to ensure that they had a low default activation level.

## A.2   Training

A version of backpropagation was used to train the model where derivatives were clipped at 1.0 (Doug's momentum; Rohde, 1999). Momentum was set to 0.9 throughout training. Weights were updated after each message-sentence pair had been trained; the term *episode* refers to the time taken to train one message-sentence pair, an *epoch* comprises 20,000 episodes. To simulate the gradual reduction in plasticity over development (Johnson & Newport, 1989), the learning rate started at 0.17 for the first 400,000 training pairs and was gradually lowered to 0.001 over the remaining 100,000 pairs. the learning rate for connections from PREVIOUS WORD to CCOMPRESS was set to a constant of 0.25 to counteract a vanishing gradient in lower layers. To prevent overfitting and improve generalization, connection weights were set to decay by a fraction of $5 \times 10^{-7}$ following each episode. Training ended after 500,000 sentences had been processed.

Error during training was measured in terms of divergence between target and produced words:

$$\text{Divergence error} = \sum_{i=1}^{n} t_i \log \frac{t_i}{o_i} \tag{2}$$

where $t_i$ is the target activation of the $i$th output unit and $o_i$ its actual output value on the current word input. Divergence is the natural choice of loss function for softmax output units (Bishop, 1996).

Training began by randomizing all weights and the same seed was used for all runs. Model subjects differed in terms of the set of training items they were exposed to. At the start of each utterance, the message was set and did not change throughout production. After the sentence was generated, the sequence of PRODUCED WORD activations was processed by a decoder program that yielded the produced sentence. Sentences were then processed by a syntactic coder program that added the syntactic and message tags. The model's output was compared with the target sentence and an utterance was considered accurate if all the words and inflectional morphemes were correctly produced.

## A.3   Input grammar

Input to the model consisted of message-sentence pairs generated from a grammar. The grammar included various concepts that were organized into several categories, for example LIVING (MAN, WOMAN, CAT, DOG, BOY, GIRL, etc.), and OBJECT (BALL, STICK, TOY, KITE, APPLE, CAKE, etc.). There were three action categories, each containing eight action concepts. One of these action categories corresponded to the intransitive verb class (ACTION-INTR, e.g., JUMP, DANCE), one to the simple transitive class (ACTION-TRANS, e.g., PUSH, HIT), and one to the dative class (ACTION-DAT, e.g., GIVE, THROW). Message-sentence pairs were generated from paired message/sentence templates. The message template specified the set of roles that should be filled and the type of category that should fill these roles. For example, the dative template would have role slots for AGENT1, THEME1 and

RECIPIENT1, with the AGENT1 and RECIPIENT1 role filled by the LIVING category. To create a specific message, each role was filled by randomly selecting a concept from the appropriate category. For example, when a message stipulated AGENT1=LIVING, the AGENT1 role might be filled with the concept MAN. The sentence template that was paired with the message template specified how the message mapped onto a word order. For example, the transitive sentence template specified that AGENT1 came first, followed by ACTION1, then THEME1.

The input grammar also included a copular construction where the predicate was instantiated from a class ADJ of eight adjective concepts (NICE, PRETTY, BIG, SMALL, etc.). Copular verbs were drawn from the AUX class which contained the auxiliaries *is, was, did, does, can, and could*. To distinguish the copular construction from action constructions, the latter used a feature AA in the EVENT SEMANTICS. Messages also contained information about tense (past and present) and aspect (simple and progressive) which was represented by TENSE and ASPECT features in the EVENT SEMANTICS. Simple present tense was the default (tense and aspect features were silent) and past tense and progressive aspect were marked by active PAST and PROG features, respectively. This default was motivated by the fact that present tense auxiliaries were roughly three times as frequent as past tense auxiliaries in child-directed speech (Thomas corpus; Lieven, Salomo, & Tomasello, 2009). Inflectional suffixes for tense and aspect were represented as separate words (e.g., *run -ing*; *push -ed*) to keep the number of word input and output units small (PRODUCED WORD and PREVIOUS WORD layers).

Constructions could be instantiated in declarative or interrogative form and a speech act feature represented this distinction in the EVENT SEMANTICS. Since in child-directed speech declaratives were slightly more frequent than questions (Cameron-Faulkner, Lieven, & Tomasello, 2003), they were made the default speech act (silent feature) and questions were marked with an active YSNO feature. Each noun phrase in the language was either a lexical noun preceded by a definite determiner or a case-marked pronoun (*he, she, it, him, her*). Pronouns were represented by a link from the ROLE layer to a concept PRN in the CONCEPT layer and the activation of the lexical content of the noun phrase was half as strong (weight = 3) than usual. Although the input grammar generator contained information about syntactic categories, verb classes, structural types, and word order differences between questions and declaratives, the model was not given this information. Rather, it had to develop appropriate representations from the variability and distributional properties of the input.

Action constructions could also occur with modal auxiliaries (e.g, *The dog can carry the toy*; *Could the boy run?*) and modality was encoded in the EVENT SEMANTICS by replacing the action feature AA with a MODAL feature. When questions occurred with simple aspect, *do*-support was enabled (e.g., *Does the dog carry the toy?*; *Did the boy run?*). Thus, the combination of tense, aspect, action and speech act features determined the verb phrase that had to be produced and the model had to acquire these distinctions and how they were expressed in English from experience. Table 4 shows all basic constructions permitted by the grammar.

From these basic constructions, complex templates for utterances with relative clauses were created. To limit the complexity of the model's language input, these constructions could have copular, intransitive, transitive and dative relative clauses but no dative main clauses. In total, there were 27 distinct complex construction templates in the grammar. They generated right-branching and center-embedded relative clauses with subject (e.g., *The boy was holding the orange that was big*; *The boy that was pretty was holding the toy*), direct object (e.g., *The apple that the boy was holding was nice*) and indirect object gaps (e.g., *The cat*

*that the boy gave the cake to was cute*). All constructions could occur in declarative and interrogative form, including complex *do*-support questions (e.g., *Does the boy that eats the cake run?*). Moreover, tense, aspect and modality permitted in the basic constructions could occur in any combination within relative clause constructions. Complex center-embedded questions with two auxiliaries, however, never occurred in the training set (e.g., *Is the boy that can jump lazy?*). They were used only in testing to assess the model's ability to generalize auxiliary inversion to these novel utterances.

Table 4
*Basic constructions in the input grammar*

| **Copular** | | |
|---|---|---|
| | *Assertion* | *Question* |
| Event semantics | TH1, PAST1 | TH1, YSNO |
| Message template | THEME1 = LIVING, OBJECT | THEME1 = LIVING, OBJECT |
| | ACTION1 = ADJ | ACTION1 = ADJ |
| Sentence template | THEME1 AUX ACTION1 | THEME1 AUX ACTION1 |
| Sentence examples | *She was pretty.* | *Is the apple big?* |

| **Intransitive** | | |
|---|---|---|
| | *Assertion* | *Question* |
| Event semantics | TH1, AA1, PAST1 | TH1, AA1, PROG1, YSNO |
| Message template | THEME1 = LIVING | THEME1 = LIVING |
| | ACTION1 = ACTION-INTR | ACTION1 = ACTION-INTR |
| Sentence template | THEME1 ACTION1 | AUX THEME1 ACTION1 |
| Sentence examples | *The boy jump -ed.* | *Is the boy run -ing?* |

| **Transitive** | | |
|---|---|---|
| | *Assertion* | *Question* |
| Event semantics | AG1, TH1, AA1, PROG1 | AG1, TH1, AA1, PAST1, YSNO |
| Message template | AGENT1 = LIVING | AGENT1 = LIVING |
| | ACTION1 = ACTION-TRAN | ACTION1 = ACTION-TRAN |
| | THEME1 = OBJECT | THEME1 = OBJECT |
| Sentence template | AGENT1 ACTION1 THEME1 | AUX AGENT1 ACTION1 THEME1 |
| Sentence examples | *The woman is push -ing the ball.* | *Was the girl hit -ing the toy?* |

| **Dative** | | |
|---|---|---|
| | *Assertion* | *Question* |
| Event semantics | AG1, TH1, RE1, AA1, PAST1 | AG1, TH1, RE1, AA1, PROG1, YSNO |
| Message template | AGENT1 = LIVING | AGENT1 = LIVING |
| | ACTION1 = ACTION-DAT | ACTION1 = ACTION-DAT |
| | THEME1 = OBJECT | THEME1 = OBJECT |
| | RECIPIENT1 = LIVING | RECIPIENT1 = LIVING |
| Sentence template | AGENT1 ACTION1 THEME1 TO RECIPIENT1 | AUX AGENT1 ACTION1 THEME1 TO RECIPIENT1 |
| Sentence examples | *The cat give -ed the toy to the dog.* | *Is the man throw -ing the stick to the dog?* |

*Note.* Features in the EVENT SEMANTICS signaled the thematic nature of event participants (AG1 = AGENT1, TH1 = THEME1, RE1 = RECIPIENT1), and the tense, aspect, action and speech act type of the target utterance.

In order to encode the meaning of complex utterances with multiple propositions, two sets of roles were used in the ROLE and CROLE layers (e.g., AGENT1 and RECIPIENT2). Likewise, there were duplicate sets of features in the EVENT SEMANTICS to encode tense and aspect (e.g., PAST1, PROG2) and the number of event participants in each proposition (e.g.,

TH1 and AG2). There was only one speech act feature YSNO which was neutral between propositions. Thus, the model had to learn from the structures in the input that the questioned proposition was always in the main clause. To express two propositions using a relative clause, speakers must know that one participant is co-referential across propositions; it is the same boy who is big who is also running in *The boy that is running is big*. To encode co-reference in the EVENT SEMANTICS, binding features were used for each combination of thematic roles. For example, if the AGENT1 of the main clause event was the THEME2 of the embedded clause event, a feature AG1-TH2 represented the fact that the event participants linked to these roles were co-referential. To illustrate the structure of the message underlying complex utterances, 3 of the 27 constructions were selected which cover the critical distinctions (Table 5). Here, we focus on complex questions, messages for the corresponding assertions had the YSNO feature omitted and were identical otherwise.

Table 5
*Example complex question constructions in the input grammar*

| **Progressive aspect and modal auxiliary, right-branching direct object-relative clause** | |
|---|---|
| Event semantics | AG1, TH1, AA1, PROG1; AG2, TH2, MODAL2; TH1-TH2; YSNO |
| Message template | AGENT1 = LIVING, ACTION1 = ACTION-TRAN, THEME1 = OBJECT |
| | AGENT2 = LIVING, ACTION2 = ACTION-TRAN, THEME2 = OBJECT |
| Sentence template | AUX AGENT1 ACTION1 THEME1 that AGENT2 AUX ACTION2 |
| Sentence examples | *Is the boy hit -ing the toy that he can push?* |
| **Do-support, one auxiliary, center-embedded subject-relative clause** | |
| Event semantics | TH1, AA1; AG2, TH2, AA2, PAST2; TH1-AG2; YSNO |
| Message template | THEME1 = LIVING, ACTION1 = ACTION-INTR |
| | AGENT2 = LIVING, ACTION2 = ACTION-TRAN, THEME2 = OBJECT |
| Sentence template | AUX THEME1 that ACTION2 THEME2 ACTION1 |
| Sentence examples | *Does the dog that push -ed the toy jump?* |
| **Progressive aspect, two auxiliaries, center-embedded indirect object-relative clause** | |
| Event semantics | TH1, PAST1; AG2, TH2, RE2, PROG2; TH1-RE2; YSNO |
| Message template | THEME1 = LIVING, OBJECT, ACTION1 = ADJ |
| | AGENT2 = LIVING, ACTION2 = ACTION-DAT, THEME2 = OBJECT, |
| | RECIPIENT2 = LIVING |
| Sentence template | AUX THEME1 that AGENT2 AUX ACTION2 THEME2 to ACTION1 |
| Sentence examples | *Was the woman that she is give -ing the cake to nice?* |

*Note.* Two sets of roles and features in the EVENT SEMANTICS encoded the two different propositions. Binding features encoded the co-reference of event participants for the production of complex utterances, e.g., TH1-AG2 signaled that the main clause THEME1 was the embedded clause AGENT2. The first two example complex questions could occur in the training set (see section 6), the third example could occur in the test set only.

## A.4   Input and test distribution

To train the model, 50,000 message-sentence pairs were randomly generated from this grammar, subject to the following distributional restrictions. 95.1% of the training items were single-clause structures, half of which were assertions and half simple polar questions. This ratio reflects corpus evidence that both types of speech acts are similar in frequency within

the speech to children, with assertions being slightly more frequent (Cameron-Faulkner et al., 2003; Legate & Yang, 2002). Basic constructions in the grammar (Table 4) occurred with the same frequency in the training set for both assertions and questions. 4.9% of the input items contained a relative clause which is consistent with corpus-based findings concerning complex construction frequency (Cameron-Faulkner et al., 2003). Assertions with a right-branching or center-embedded relative clause were approximately equally frequent to ensure that both constructions were learned to an adult degree at the end of training. 1.6% of items in the input were questions with a relative clause and 0.3% had a center-embedded relative clause. Both assumptions were motivated by child language data from the Thomas corpus (Lieven, Salomo, & Tomasello, 2009, see section 6.3). The model's input did not contain yes-no questions with a center-embedded relative clause and an auxiliary in both clauses as these structures are extremely rare in child-directed speech (Legate & Yang, 2002; MacWhinney, 2004; Lieven et al., 2009). Complex sentences with auxiliaries in each clause were either assertions or questions with a right-branching relative clause. Across speech acts, subject-relatives were more frequent (58%) than object-relative clauses (42%). Present and past tense were equally frequent in the input. Around 29% of all auxiliaries were modal auxiliaries (*can, could*), 71% were non-modal (*is, was, does, did*), and this ratio was similar in the Thomas corpus for the same auxiliaries (27% modal).

The input grammar generated approximately $5.7 \times 10^9$ distinct message-sentence pairs from which 50,000 items were randomly selected for training. Hence, the training set represented less than 0.001% of the utterances licensed by the grammar. The model was trained on these items ten times in succession for a total of 500,000 episodes. Because different children hear different input during language development, 25 model subjects were created and each subject was exposed to a different randomly generated training set. Test sets contained 400 randomly generated message-sentence pairs composed of single-clause declaratives, single-clause yes-no questions, complex right-branching and center-embedded declaratives, and right-branching yes-no questions (50 items each). In addition, there were 150 complex center-embedded questions in the test set, 50 with two modal auxiliaries, 50 with two non-modal auxiliaries and 50 items with one modal and one non-modal auxiliary (in any order). Questions of this kind had not occurred in the learning environment. The model was tested on these items periodically after every 20,000 training episodes (= 1 epoch). Generalization to novel complex yes-no questions shown in Figure 3 represents the mean performance over 25 model subjects, tested on 150 questions each.

## A.5   Error coding

Production errors in complex questions were automatically extracted from the model's Produced Word output. They were coded as embedded auxiliary errors, auxiliary doubling, or structure dependence errors in the following way. An *embedded auxiliary error* occurred if questions started with an auxiliary, had a relative pronoun and the embedded verb in the target positions, but the auxiliary preceding the embedded verb was omitted (e.g., *Is [...] that _ running [...]*). Errors of structure dependence were only examined in questions where main and embedded clause auxiliaries differed, otherwise it would not be possible to determine whether the sentence-initial auxiliary belonged to the main or embedded clause. For a *structure-dependent error* to occur, the embedded clause auxiliary had to be produced sentence-initially, it had to be omitted within the embedded clause, and the main clause auxiliary had to occur in its canonical location (e.g., *Can [...] that _ jump is [...]* when the target question was *Is the boy that can jump lazy?*). An *auxiliary doubling error* occurred when questions started with an auxiliary, had a relative pronoun and the target embedded

verb phrase in the correct position, and the main clause auxiliary was repeated following the relative clause (e.g., *Is [...] that was run -ing is [...]*). Only if the produced utterance matched the target question word-by-word it was labelled as a correct polar question.

# References

Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science*, *30*(6), 995–1026.

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, *33*(6), 952–965.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2014). Child language acquisition: Why universal grammar doesn't help. *Language*, *90*(3), e53–e90.

Ambridge, B., Rowland, C. E., & Pine, J. M. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science*, *32*(1), 222–255.

Ambridge, B., Rowland, C. F., Theakston, A. L., & Tomasello, M. (2006). Comparing different accounts of inversion errors in children's nonsubject wh-questions: 'what experimental data can tell us?'. *Journal of Child Language*, *33*(3), 519–557.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language*, *11*(3), 557–578.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, *35*(7), 1207–1242.

Bishop, C. M. (1996). *Neural networks for pattern recognition*. Clarendon Press.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London: Academic Press.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*, *33*(5), 752–793.

Bouchard, D. (2012). Solving the UG problem. *Biolinguistics*, *6*(1), 1–31.

Brooks, P., & Tomasello, M. (1999). Young children learn to produce passives with nonce verbs. *Developmental Psychology*, *35*(1), 29–44.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science*, *14*(1), 107–133.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish

shared from private information when interpreting questions during interactive conversation. *Cognition, 107*(3), 1122–1134.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge: Cambridge University Press.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*(6), 843–873.

Carey, S. (2009). *The origin of concepts.* New York: Oxford University Press.

Carnie, A. (2012). *Syntax: A generative introduction* (3rd ed.). Wiley-Blackwell.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26*(5), 609–651.

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language, 61*(3), 374–397.

Chang, F., Baumann, H., Pappert, S., & Fitz, H. (2015). Do lemmas speak German? A verb position effect in German structural priming. *Cognitive Science, 39*(5), 1113–1130.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research, 9*(3), 198–213.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind.* Chicago: University of Chicago Press.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace, Jovanovich.

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon.

Chomsky, N. (1980). On cognitive structures and their development. In M. Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 35–54). London: Routledge and Kegan.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures.* Berlin: Mouton de Gruyter.

Chomsky, N. (1983). Some conceptual shifts in the study of language. In L. S. Cauman, I. Levi, C. D. Parsons, & R. Schwartz (Eds.), *How many questions? Essays in honor of Sidney Morgenbesser* (pp. 154–169). Indianapolis, IN: Hackett Publishing.

Clancy, P., Lee, H., & Zoh, M.-H. (1986). Processing strategies in the acquisition of relative clauses: Universal principles and language-specific realizations. *Cognition, 24*(3), 225–262.

Clark, A., & Eyraud, R. (2007). Polynomial time identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research, 8*, 1725–1745.

Clark, A., & Lappin, S. (2011). *Linguistic nativism and the poverty of the stimulus.* Oxford: Wiley-Blackwell.

Clark, E. V. (2002). *First language acquisition.* Cambridge: Cambridge University Press.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Côrrea, L. M. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research, 24*(3), 183–203.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14*(4), 597–650.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63*(3), 522–543.

Crain, S., & Pietroski, P. M. (2001). Nature, nurture, and universal grammar. *Linguistics and Philosophy*, *24*(2), 139–186.

Crain, S., & Thornton, R. (2012). Syntax acquisition. *WIREs Cognitive Science*, *3*, 185–203.

Culicover, P., & Jackendoff, R. (2005). *Simpler syntax*. UK: Oxford University Press.

Dabrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics*, *11*(1/2), 83–102.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, *11*(1/2), 131–151.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, *81*(4), 882–906.

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*(2), 195–225.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Erreich, A. (1984). Learning how to ask: Patterns of inversion in yes-no and wh-questions. *Journal of Child Language*, *11*(2), 579–592.

Estigarribia, B. (2009). Facilitation by variation: Right-to-left learning of English yes/no questions. *Cognitive Science*, *34*(1), 68–93.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84.

Fitz, H. (2009). *Neural syntax*. ILLC dissertation series, University of Amsterdam.

Fitz, H. (2010). Statistical learning of complex questions. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2692–2698). Austin, TX: Cognitive Science Society.

Fitz, H., Chang, F., & Christiansen, M. H. (2011). A connectionist account of the acquisition and processing of relative clauses. In E. Kidd (Ed.), *The acquisition of relative clauses—processing, typology and function* (Vol. 8, pp. 39–60). Amsterdam: John Benjamins.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3–71.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, *66*(2), 297–316.

Gärdenfors, P. (2014). *The geometry of meaning. semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gold, M. E. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447–474.

Goldberg, A. (2006). *Constructions at work: The nature of generalisations in language*. Oxford: Oxford University Press.

Gómez, R. L. (2007). Statistical learning in infant language development. In M. G. Gaskell (Ed.), *The oxford handbook of psycholinguistics* (pp. 601–616). Oxford: Oxford University Press.

Grice, H. P. (1975). Speech acts. In P. Cole & J. L. Morgan (Eds.), *The Oxford handbook of psycholinguistics* (pp. 41–58). New York: Academic Press.

Gruber, J. (1976). *Lexical structures in syntax and semantics.* Amsterdam: Elsevier/North Holland.

Harner, L. (1981). Children talk about the time and aspect of actions. *Child Development*, *52*(2), 498–506.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, *16*(5), 362–366.

Hsu, A. S., & Chater, N. (2010). The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science*, *34*(6), 972–1016.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Jackendoff, R. (2014). *Genesis of a theory of language: From thematic roles (source) to the Parallel Architecture (goal).* Retrieved from `http://ase.tufts.edu/cogstud/jackendoff/papers/GenesisofPA.pdf` (Ms., Tufts University)

Jackendoff, R., & Wittenberg, E. (2014). What you can say without syntax: A hierarchy of grammatical complexity. In F. Newmeyer & L. Preston (Eds.), *Measuring linguistic complexity* (pp. 65–82). UK: Oxford University Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jain, S. (1999). *Systems that learn: An introduction to learning theory.* Cambridge, MA: MIT Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Kam, X., Stoyneshka, I., Tornyova, L., Fodor, J., & Sakas, W. (2008). Bigrams and the richness of the stimulus. *Cognitive Science*, *32*(4), 771–787.

Karmiloff, K., & Karmiloff-Smith, A. (2002). *Pathways to language: From fetus to adolescent.* Cambridge, MA: Harvard University Press.

Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*(6), 599–617.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A crosslinguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860–897.

Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, *6*(1), 97–133.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for the Philosophy of Science*, *52*(2), 217–276.

Legate, J. A., & Yang, C. D. (2002). Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review*, *18*(1/2), 151–162.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288.

Levelt, J. M. W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levinson, S. C. (2003). Language and mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 25–46). Cambridge, MA: MIT Press.

Lewis, J., & Elman, J. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In B. Skarabela, S. Fish, & A. H. J. Do (Eds.), *Proceedings of the 26th annual boston university conference on language development* (Vol. 2, pp. 359–370). Somerville, MA: Cascadilla Press.

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3), 481–508.

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453–460.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Mahwah, NJ: Erlbaum.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, *31*(4), 883–914.

Marcus, G. F. (2003). *The algebraic mind: Integrating connectionism and cognitive science.* Cambridge, MA: MIT Press.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*(3), 449–496.

Mayer, J. M., Erreich, A., & Valian, V. (1978). Transformations, basic operations and language acquisition. *Cognition*, *6*(1), 1–13.

Mitchell, T. M. (1997). *Machine learning.* New York, NY: McGraw-Hill, Inc.

Nakayama, M. (1987). Performance factors in subject-auxiliary inversion by children. *Journal of Child Language*, *14*(1), 113–125.

Niyogi, P., & Berwick, R. C. (1996). A language learning model for finite parameter spaces. *Cognition*, *61*(1/2), 161–193.

O'Grady, W. (1997). *Syntactic development.* Chicago, IL: The University of Chicago Press.

O'Grady, W. (2005). *How children learn language.* Cambridge University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Palermo, D. S. (1982). Theoretical issues in semantic development. In S. A. Kuczaj (Ed.), *Language development: Syntax and semantics* (pp. 335–364). Psychology Press.

Pearl, L., & Lidz, J. (2009). When domain-general learning fails and when it succeeds: Identifying the contribution of domain-specificity. *Language Learning and Development*,

*5*(4), 235–265.

Perfors, A., Tenenbaum, J., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338.

Petersson, K. M., & Hagoort, P. (2012). The neurobiology of syntax: Beyond string-sets. *Philosophical Transactions of the Royal Society, B*, *367*, 1971–1883.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pinker, S. (2004). Clarifying the logical problem of language acquisition. *Journal of Child Language*, *31*(4), 949–953.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *18*(1/2), 9–50.

Radford, A. (2004). *English syntax: An introduction.* Cambridge: Cambridge University Press.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*(6), 1007–1028.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

Rohde, D. L. T. (1999). *LENS: The light, efficient network simulator* (Tech. Rep. No. CMUCS-99-164). Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*(1), 67–109.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379.

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

Rowland, C. F., & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: What children do know? *Journal of Child Language*, *27*(1), 157–181.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 318–362). Cambridge, MA: MIT Press.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sakas, W. G., & Fodor, J. D. (2001). The structural triggers learner. In S. Bertolo (Ed.), *Language acquisition and learnability* (pp. 172–233). Cambridge: Cambridge University Press.

Sampson, G. (1989). Language acquisition: Growth or learning? *Philosophical Papers*, *18*, 203–240.

Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language*, *29*(4), 813–842.

Scholz, B. C., & Pullum, G. K. (2002). Searching for arguments to support linguistic nativism. *The Linguistic Review*, *18*(1/2), 185–223.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*, 97–123.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147.

Shinohara, T. (1994). Rich classes inferable from positive data: Length-bounded elementary formal systems. *Information and Computation, 108*, 175–186.

Shirai, Y., & Anderson, R. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language, 71*(4), 743–762.

Siemund, P. (2001). Interrogative constructions. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 1010–1028). Berlin: de Gruyter.

Smith, C. (1980). The acquisition of time talk: Relations between child and adult grammars. *Journal of Child Language, 7*(2), 263–278.

Steedman, M. (1996). Phrasal intonation and the acquisition of syntax. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 331–342). Hillsdale, NJ: Lawrence Erlbaum Associates.

St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence, 46*(1/2), 217–257.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states–the question is which ones and to what extent. *Trends in Cognitive Sciences, 7*(4), 153–156.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–735.

Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology, 73*, 41–71.

Van Valin Jr., R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning and function.* Cambridge: Cambridge University Press.

Wagner, L. (2001). Aspectual influences on early tense comprehension. *Journal of Child Language, 28*(3), 661–681.

Weist, R. (1991). Spatial and temporal location in child language. *First Language, 11*, 253–267.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.

Wexler, K., & Culicover, P. W. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*(3), 164–169.

Yang, C. D. (2002). *Knowledge and learning in natural language.* Oxford: Oxford University Press.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451–456.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.