

A model of Frame and Verb Compliance in language acquisition

Rutvik Desai*

Indiana University, Bloomington, IN, USA

Received 8 February 2005; received in revised form 1 February 2006; accepted 8 February 2006

Available online 2 January 2007

Abstract

Researchers studying word learning have discovered that the syntactic frame in which a word appears plays an important role in the interpretation of the word, and this importance diminishes gradually with increasing age. Interpretation of sentences based on the frame and the verb is known as Frame and Verb Compliance, respectively. Here, a connectionist model is presented that explains this shift from Frame of Verb Compliance in terms of competition between two cues—the frame and the verb—that predict causality. The model learns a miniature language by associating sentences with the corresponding “scenes”. The most frequent cues, the frames, are learned first, resulting in frame compliance in the initial phase of learning. As the learning progresses, the less frequent but more powerful cues, the verbs, are learned and prevail over the frames resulting in verb compliance. It is argued that these phenomena can be attributed to the interaction between properties of the input and the learning mechanism, and it is not necessary to invoke specialized principles.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Language acquisition; Verb learning; Verb compliance; Syntactic frames; Connectionism

1. Introduction

Children learn new words rapidly. A common-sense explanation for vocabulary acquisition is that word meanings are learned by observing real-world contingencies of their use. The meaning of *jump* is learnt from noticing that it occurs in the presence of jumping events. However, this simple explanation has several difficulties when attempting to account for acquisition of meaning of all words. Many of these problems are listed by Gleitman [17]: (a) This theory fails to account for the fact that children with radically different exposure conditions (e.g., the blind and the sighted) acquire similar meanings [21], (b) many verbs are used for the same events and only provide a perspective on an event (e.g., *chase* and *flee*), and (c) many verbs only differ in the level of specificity at which they describe single events (e.g., *see*, *look*, *orient*).

In light of these problems, it has been suggested that children use another rich source of information, namely, the syntactic context in which the words occur. This

proposal is known as *syntactic bootstrapping* [17,16,21]. According to this hypothesis, children can use the knowledge of syntax to predict meanings of words. The learner observes the real world situations and also observes the language structures in which various words appear. If there is a correlation between meanings and a range of syntactic structures, the meaning (or some components of the meaning) of an unknown word can be predicted when it appears in a familiar structure.

1.1. Verb Compliance and Frame Compliance

One way to study the effect of syntax on the acquisition of word meaning is to use familiar words in a different or incorrect syntactic context, and examine the effect on the interpretation of the word. For example, we can insert a transitive (T) verb in an intransitive (I) frame and examine how children interpret the sentence. If children are still learning about a verb, then they may more readily accept its occurrence in an incorrect frame. Casual meaning is associated with T frames, and noncasual meaning with I frames. They are more likely to reject an incorrect frame when they have fully acquired the verb. If a child interprets the sentence in accord with the frame, she are said to be

*Corresponding author. Current address: Department of Neurology, Medical College of Wisconsin, 8701 Watertown Plank Rd, MEB 4550, Milwaukee, WI 53226, USA. Tel.: +1 414 456 4483; fax: +1 414 456 6562.
E-mail address: rhdesai@mcw.edu.

Frame Compliant. If the interpretation fits more with the verb, she is deemed *Verb Compliant*.

Frame and Verb Compliance are interesting for another theoretical reason. While children's verb use is overwhelmingly correct, a major exception to this appears somewhere around the age of 3. As reported by Bowerman [3,4], children sometimes use verbs in incorrect sentence frames, as in **Don't fall that on me* (to protest the impending dropping of an object by someone). Thus, children overgeneralize, e.g., they use a verb transitively when only intransitive use is allowed, or vice versa. Children must learn eventually which uses are "licensed" for which verbs. For example, they must learn that *sink* can be used either transitively or intransitively, but *fall* and *go* allow only noncausal interpretation. How children recover from these overgeneralizations is a major question in language acquisition. It is essentially the same as asking why children become Verb Compliant at some stage, assuming that comprehension and production share some of the same mechanisms. When children show verb compliant behavior, they have sufficient confidence in their knowledge of verb meaning that they reject contradictory cues, which is exactly the requirement for eliminating overgeneralizations.

Now we look at some empirical evidence for Compliance effects.

1.2. The data

Naigles and colleagues [31,33] conducted experiments involving the approach described above. They asked 120 children, from 2.5 to 12 years of age, as well as adults, to enact grammatical and ungrammatical sentences using "Noah's Ark" and wooden toy animals as props. Ungrammatical sentences were constructed by placing T verbs (*bring, take, push, put*) in I frames (e.g., **The lion puts in the ark, *The zebra brings*). Similarly, I verbs (*come, go, fall, stay*) were inserted in T frames (e.g., **The elephant comes the giraffe*). The children's enactment was deemed to be Frame Compliant if they modified the meaning of the verb to conform to the frame in which it was encountered (e.g., the elephant pushing or carrying the giraffe). It was considered Verb Compliant if they followed the restrictions of the verb (e.g., the elephant moving independently of the giraffe).

Their overall results indicated that younger children, especially the 2-year-olds, were more Frame Compliant, enacting the ungrammatical sentences according to the demands of the frame and altering the meaning of the verb. They allowed novel frames to influence the interpretation of the familiar verbs. Older children, and especially the adults, were more Verb Compliant, following the restrictions of the verb and repairing the sentence. Children at the intermediate ages were en route to the adult state, showing intermediate levels of Frame and Verb Compliance.

Similar experiments have been conducted with children with Down Syndrome (DS) [32]. The linguistic skills of

children with DS are split in an interesting way. Relative to their syntactic knowledge (often measured by MLU or auxiliary use) their vocabulary growth is advanced. It was reported [32] that children with DS who had a "vocabulary age" of 6 years were syntactically like 3-year-olds. While children with DS were more Frame Compliant than their chronological-age mates, they also exhibit the move from Frame to Verb Compliance. Adolescents with DS show more Verb Compliance than gradeschoolers with DS. Thus, with the advance in syntactic knowledge, DS children also move toward Verb Compliance.

In this paper I present a connectionist model that attempts to explain the mechanisms by which this shift occurs. First, various theories of Compliance are briefly discussed, and a new proposal is presented. Then, a network implementation of the proposal is described. The network's behavior with respect to the compliance effects is then examined. We end with a discussion of the results and their implications.

2. Theories of compliance

Frame and verb compliant behavior is closely related to the well-known overgeneralization errors made by children, and their recovery from those errors. We now consider some of the theories related to overgeneralization as well as Frame and Verb Compliance, and ask what the implications of the model are in the light of these results.

2.1. Maturation

A maturation-based account is offered by Pinker [35]. Very briefly, verbs become organized into semantic subclasses known as *narrow range subclasses* as their representations are refined. The semantics of the verb class determines whether the verbs allow alternations (e.g., causal and noncausal use in T and I frames) or not. When the representation of a verb matches that of another verb in the same subclass that is known to alternate between causal and noncausal, the former is allowed to alternate as well. For example, motion verbs that encode path (e.g., *bring, take, go*) can be used either causally or noncausally, but not in both ways. On the other hand, motion verbs that encode manner, like *roll* and *bounce* can be used both transitively and intransitively. Overgeneralization occurs because a verb is used in the same manner as other semantically similar verbs in a subclass.

The shift to Verb Compliance may occur because the verb representations are elaborated to the extent that they have formed grammatically relevant narrow range subclasses. Some verbs no longer allow causal interpretation because they do not fit the semantic specification of the subclasses that are causal. At the time of puberty, those subclasses of verbs for which there has been no evidence of alternation become fixed or "closed". After that, no new information about the verb is accepted, resulting in Verb Compliant behavior. For example, because *come* and *go*

do not encode manner of motion, they do not match the specification of the alternating subclass of motion verbs (that includes *roll* and *bounce*). This subclass is closed at maturation, so *come* and *go* no longer allow causal interpretations.

As pointed out by Naigles et al. [31,32], there are factors other than age that appear to affect Compliance and present a serious problem for this account. If the maturation-based account was correct, one would expect to observe an across-the-board shift from Frame to Verb Compliance, for all verbs and frames. But this is not the case. Some verbs elicit more Frame Compliance than others. For example, in the Naigles et al. [31] study, in the NVN frame, *come* and *go* elicited significantly more verb compliance than *stay* and *fall*. *Stay* and *fall* also differed from each other significantly. In the NV frame, *bring*, *take*, and *put* showed significantly more Frame Compliance than *push*. Secondly, some frames induce frame compliance to a later stage than others. For example, the shift to Verb Compliance for the NV frame is effectively complete at age 5. On the other hand, even 12-year-olds and adults continue to exhibit Frame Compliance for their NVNPN frame (e.g., **Elmo jumps the ball on the ground*). I frames shifted earlier than T frames. Furthermore, the move towards Verb Compliance can occur at different ages, anywhere between 2 and 7 years of age.

In the study of children with DS [32], it was found that although these children were more Frame Compliant than their chronological age mates, they exhibit this shift also. Because their maturational progress is dissociated, one would expect that prepubescent children with DS will be no less Verb Compliant than their adolescent counterparts if the maturational account is correct. This appears not to be the case and children with DS also move to become more Verb Compliant.

2.2. Mutual Exclusivity

Another proposal about recovery from overgeneralizations is the Mutual Exclusivity (ME) Principle (also called Contrast, Uniqueness, or Pre-emption) [4,9,10,26]. In brief, this principle is that children will allow only one lexical entry to occupy a semantic niche. When two words are determined to have similar meanings, one of them is pre-empted and removed from the lexicon. For example, causal *come* is basically equivalent to *bring*. Using Bowerman's [4] example, during the period in which overgeneralized (causal) *come* is frequent in production, *bring* is practically nonexistent. When *bring* becomes more frequent eventually, the causal *come* declines. This can explain why some verbs elicit verb compliance. For example, transitive *bring* and *take* pre-empt causal uses of *come* and *go*, respectively. We do not thoroughly discuss ME here (see [27,2,11]) for a critical evaluation of ME and other constraint-based accounts), but note that while ME may have some role to play in recovering from overgeneralizations, it does not account for all the effects found in the data. For example,

it does not explain why intransitive *push* causes Verb Compliance earlier than intransitive *bring* or *take*. This principle also does not work for all the verbs, since for some verbs, it is difficult to find a similar meaning verb that can pre-empt its use in the right way.

2.3. Lexical Knowledge

A different account, based on lexical knowledge, is offered by Naigles et al. [33]. This account relies on children's knowledge of individual verbs. Children's conjectures about verb meanings are refined by ongoing events as well as the structures in which they appear. At early stages of vocabulary acquisition, open-minded children assume that not all structures have as yet been heard and therefore certain properties of verbs (such as whether they encode causality) may be unknown to them. In this case they make use of the structural information provided by the frames. At some point, however, older children and adults feel warranted to believe that all the relevant information about the meaning has been obtained. Then they would perceive a novel structure as simply ill-formed, causing Verb Compliant behavior.

This theory can explain various effects in the data well. For example, the shift towards Verb Compliance is a function of individual verbs and frames because different amounts of knowledge is accrued for them due to their differing frequencies in the input. While this account is supported by data, some important details remain unclear. One might ask where the so-called "open-mindedness" in the initial stages and the confidence about the meaning at later stages come from. After hearing a verb in a certain number of contexts, exactly what makes a child more or less open-minded to accept new meanings? An answer is not offered in Naigles et al. [33], but one possibility is to invoke some type of innate parameter or threshold that allows children to determine whether a certain amount of experience with a verb is enough to warrant confidence in the meaning of that verb.

2.4. Lexical knowledge and innate principles

More recently, Lidz et al. [23] offer an explanation that involves both lexical knowledge and innate principles. It is best summarized by the following quote:

"The deduction of verb meaning based on an analysis of the surface structure is a learning heuristic. The learning device is asking itself, in effect: Assuming Principles [the *Theta Criterion*¹ and the *Projection Principle*²], what could be the meaning of the verb now heard, such that

¹The Theta Criterion states that there is a one-to-one correspondence between theta-roles (e.g., "agent" and "patient") and arguments (e.g., "subject", "object").

²The Projection Principle states that the lexical properties of each lexical item must be preserved at every level of representation (e.g., Deep Structure, Surface Structure, Logical Form).

these principles projected this observed (surface) structure for it? Such a deductive procedure will be invoked only when the learner does not have secure knowledge of the verb in question”. (p. 37)

Thus, when children do not have secure knowledge of the verb, they invoke innate principles that state that participants in an event will line up one-to-one with noun-phrases in the clause [8], resulting in Frame Compliance. The shift to Verb Compliance occurs due to increasing lexical knowledge, which presumably overrides the need to invoke the Principles.

This theory has a problem similar to that of the Lexical Knowledge account: Exactly how do children determine whether they have “secure knowledge” of the verb? Furthermore, subjects even show different compliance effects for the same verb in different frames [31,32]. If the decision to invoke principles is based only on the knowledge of the verb, the frame in which the verb is embedded should have no effect.

2.5. Competing Cues

I propose that Frame and Verb Compliance are emergent consequences of the statistical properties of the input, and the learning of those properties. There are two cues available to the learner to predict the causality of an event from an utterance: the “frame” or “construction” that a verb appears in, and the verb itself. Frames can be thought of as frequently occurring linguistic patterns that associate form with meaning, or certain types of meaning [15,17,18,40]. Very frequent patterns, and patterns in which some aspect of the form or function is not predictable from their component parts or from other frames, are considered frames. Here, we consider two simple frames used in the experiments with children: the T frame “N is V-ing a N” and the I frame “N is V-ing”. We start with the assumption that T frames are associated with causal events more than with noncausal events, and the reverse is true for I frames. For the present purposes, following the behavioral studies, a “causal event” is defined as an agent causing an overt change in the state of a patient. Certainly, there are T verbs that are associated with verbs that can be considered noncausal (e.g., *hold*, *watch*). In spite of such verbs, it is assumed that T frames are associated with causal events because causal events are hypothesized to be more salient and hence receive a higher weight in the input. Note that the verbs that provide a strong cue to causality are those that do not allow a causative alternation,³ or those alternating verbs that are used chiefly in either causal or noncausal form.

Why are young children Frame Compliant? A critical factor is that frames are far more frequent than any particular verb. For the language learner, this means that

³Causative alternation occurs when the transitive use of a verb V can be described roughly as ‘cause to V-intransitive’. For example: (1) The dog is turning. (2) The boy is turning the dog.

there are many more opportunities to learn the frame-causality correspondence than there are to learn the verb-causality correspondence. To take a simple example, if a child hears n T/causal verbs m times each, then there are $m \times n$ T frames but only m occurrences of any given verb. The child can (probabilistically) predict the causality of an event by learning this frequent cue early.

Given that frames are more frequent than any individual verb at any age, why should there ever be a shift to Verb Compliance? The reason for the shift is that while frames are more frequent, verbs provide a more powerful and efficient cue for predicting causality. Verbs are more powerful in two ways. First, they allow prediction of causality *earlier* than frames do. Because a (nonalternating) verb can predict causality by itself, the resources of the listener can be directed elsewhere. As soon as the listener hears “John is bringing...”, she can predict that an object is being brought, without processing the entire frame. This allows the listener to be more efficient and direct their resources towards other processing needs. Secondly, nonalternating verbs are more powerful cues than frames because they provide a more *consistent* relationship with causality. As noted above, T frames provide a probabilistic cue to causality, but (nonalternating) T verbs provide a deterministic cue. The verb-causality mapping is learned well relatively late because of the lower frequency of individual verbs, but once learned, it provides a stronger prediction than the frame-causality mapping.

I now present a preliminary connectionist network simulation that implements this proposal.

3. The network

The task of the connectionist network was to take utterances generated by a simple grammar as input, and attempt to predict the “scene” described by the utterance. The architecture of the network is shown in Fig. 1. It contained recurrent connections in the hidden layer as in a simple recurrent network (SRN) [14] to handle temporal sequences of words. The main modification here to the commonly used SRN was the recurrent layer on the output. The motivation for this addition was that the task of the network (see below) was to gradually build the correct representation of the output scene. Recurrent connections on the output layer to make it easier for the network to remember what has been already learned from the portion of the sentence processed thus far. The first hidden layer, with few units, causes the network to convert localist representations used in the input to distributed, overlapping representations.⁴

⁴The space of possible network architectures and learning parameters is very large. It is impractical to explore this space extensively and compare results at many different points. Only a few variations in the network architecture were tested during early stages of this project, and they did not appear to make a qualitative difference in the network behavior. For example, increasing the number of units in the two hidden layers by 10, or attaching a recurrent layer to the first hidden layer instead of the second

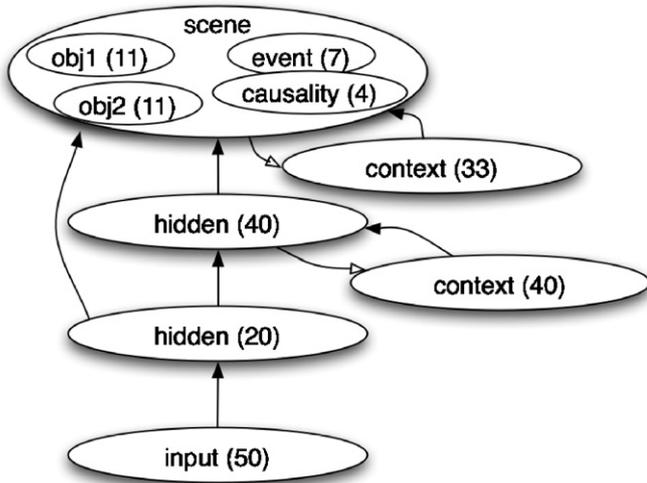


Fig. 1. The network architecture. The numbers in brackets indicate the number of units in that layer. Solid arrows indicate full trainable connections between layers, while hollow arrows indicate fixed one-to-one connections.

Then input to the network consisted of sentences or noun phrases (called “utterances”) describing one or two objects and optionally an action, generated by the grammar shown below:

```

S →      NP | NP1 | NP is IV | NP1 are IV | NP is
         TV NP | NP is ALTV | NP is ALTV NP
NP →     DET N | DET SIZE N
NP1 →    NP and NP
N →      boy | girl | dog | mouse | ...
IV →     sleeping | dancing | falling |
         staying | ...
SIZE →   large | small
TV →     pushing | feeding | bringing |
         holding | ...
ALTV →   breaking | opening | moving | ...
DET →    a
    
```

One can divide the utterances generated by this grammar into five basic types: (a) N, (b) NN, (c) NV, (d) NNV, and (e) NVN. With optional adjectives describing the size, utterances such as *a girl and a big dog are dancing* or *a small dog and a big mouse are* are obtained. The utterances of types (c) and (d) are intransitive, while those of type (e) are transitive. These utterances were presented to the network sequentially, one word at each time step. Words were represented in a localist manner by turning on a single bit in the input layer. Also, *ing* was treated as a separate word, with the assumption that it can be discerned from the word stem as a separate unit. An end-of-utterance marker, *stop*, was presented after the last word of each utterance, at which point all context units were reset.

(footnote continued)

hidden layer yielded very similar results. Thus, the results reported here appear to be robust with respect to such changes in the network architecture.

On the output or the semantic end, the descriptions of scenes corresponding to the input utterances were presented as a 33-bit fixed-width vector. There were two slots for objects, and one for the action or the event taking place. Each object slot was divided into two slots of 4 and 7 units each, which represented the attribute large (1100) or small (0011) and type of object respectively. In the 11-bit event slot, the first 4 bits indicated whether the action was causal or noncausal (with activations 1100 and 0011, respectively), and the remaining 7 bits described other features of the action. The causality units were set to 0 for utterances without a verb. A distributed representation for each individual object and event was generated by setting 3 randomly chosen bits in its slot to 1. If each bit is viewed as representing a feature, this created representations with partially overlapping features. The slots for an attribute, object, or action not described in the utterance were set to 0.

The task of the network was to predict the entire output scene after each word. The complete target was held constant for the duration of the entire utterance. The task of predicting the entire scene from the beginning of the sentence is more difficult than the task used in some simulations where the semantic representation of a word is presented only when the word is presented to the network, because here the network is forced to predict the entire output from incomplete information. However, it more realistic in the sense that the learner is presented with a scene and a corresponding utterance, but does not know in advance which parts of the utterance correspond to which parts of the scene. This correspondence is part of what the network has to learn, in addition to specific bit patterns that may correspond to words. The present task also encourages the network to process the words as soon as they arrive, and attempt to predict the meaning of the entire utterance as soon as possible.

Ten nouns were used in the simulations, without any semantic constraints (i.e., all nouns were allowed to appear with all verbs). There were three types of verbs in the input to the network: I, T, and alternating (A). There were 7 I verbs that described noncausal events (e.g., *sleep, dance, fall*), and 7 T verbs that described causal events (e.g., *lift, push, pull*). In addition, there were 3 T but noncausal verbs (e.g., *hold, hug*). Finally, there were 7 A verbs (verbs that allow causative alternation, e.g., *open, move, turn*). These verbs were coded as causal when presented in a T frame, and as noncausal when presented in an I frame.⁵ The 7 output bits representing the arbitrary part of the meaning were identical in both cases. These type frequencies

⁵For simplicity, there were no verbs in the input that can be used both transitively and intransitively, but are noncausal in both cases (e.g., *watch, see*). If such verbs are included, it would result in weakening the syntactic cues provided by the A verbs (similar to the way in which noncausal T verbs weaken the link between T verbs and causality). A higher proportion of A verbs in the input would be necessary to achieve quantitatively similar results. The omission of such verbs does not represent an inherent limitation of the model, although this is not demonstrated here.

(10 nouns and 24 verbs) were held constant for all simulations; only the token frequency was varied as described below.

3.1. Comprehension

The network was tested first for its capability of performing the basic task of producing the correct scene corresponding to an input utterance. A training set of 400 utterances was generated with an equal sampling probability for all sentence types. The network was trained using backpropagation, using sum-squared error, on the utterances in the training set. The initial weights were sampled from a uniform random distribution between -0.5 and 0.5 . A learning rate of 0.01 and momentum of 0.1 were used. The weights were updated after presentation of every 50 words. Training was continued till there was no significant improvement in the error.

To assess the performance of the network, if the activation of an output unit was less than 0.5 , it was considered OFF, and it was taken as being ON otherwise. An utterance was declared to be processed correctly if, at the end of the utterance, all output units had the desired ON or OFF activations. With this criterion, in a average of 10 runs, 99.1% (SD 0.8) accuracy was achieved on a training set of 400 utterances and 96.2% (SD 5.3) of utterances were processed correctly in a separate set of 500 utterances of the testing set, which the network had not seen during training. The network was, then, largely successful in this task of producing semantics given an utterance, or comprehension.

3.2. Frame and Verb Compliance in the network

3.2.1. Training

We now turn to Frame and Verb Compliance effects. To examine the effects of differing frequencies, nouns and verbs were sampled from a normal distribution with a standard deviation of 2, as shown in Fig. 2. Words are represented by vertical lines, and the height of the line is the sampling probability for that word. Words that are sampled from near the peak of the distribution are more frequent in the input, while those sampled from near the tails are less frequent. The group of 10 nouns and each group of verbs (I/noncausal, T/causal, T/noncausal, A) were sampled separately in this manner. Forty such networks were created with initial weights randomly sampled between -0.5 and 0.5 , and a training set of 400 utterances chosen randomly according to this probability distribution of nouns and verbs. The probability for different utterance types was as follows: 25% I/noncausal (type (c) and (d)), 25% T/causal (type (e)), 3% T/noncausal (type (e)), 25% A (type (c) and (e)). The remaining set contained utterances of types (a) and (b) in equal proportion. Adjectives were not used. The other training parameters were same as in testing comprehension.

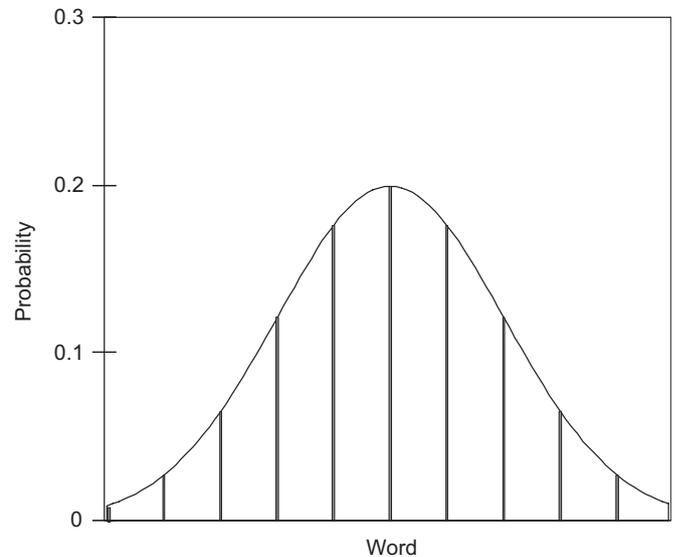


Fig. 2. The sampling probability distribution of nouns and verbs. The distribution is normal with a standard deviation of 2.

3.2.2. Testing

Two types of ungrammatical utterances were generated to test the network. The first was an NVN sentence with a known I verb (e.g., *a boy is sleeping a dog*) while the second was an NV sentence with a known T verb (e.g., *a boy is bringing*). T verbs were inserted in I sentences, and vice versa. As in the experiments with children, familiar nouns were used in the frames. Two such incompatible frames were prepared using each verb. The 7 I and 7 T/causal verbs used in training were divided in 3 groups: 2 low frequency (from near the tails of the distribution in Fig. 2), 2 medium frequency (between the peak and the tail), and 3 high frequency (near the peak). The mean frequency of the low frequency verbs was $7.5/100$, that of the medium frequency verbs was $12.5/100$, and that of the high-frequency verbs was $20/100$. The categories of “high”, “low” and “medium” frequencies are only meant to be understood relative to each other, and only qualitatively represent words that may be considered high, low, or medium frequency in child-directed speech (CDS).

We are interested in examining whether the network interprets these verbs in incompatible frames as depicting causal events or noncausal ones. The networks were trained for 250 epochs⁶ and probed periodically during the training with the utterances with incompatible verbs.⁷ Recall that there are four units in the network’s output indicating causality, where the pattern 1100 stands for a causal meaning while 0011 stands for a noncausal interpretation. To assess the network’s response, a variable δ was defined as the mean activation of the first two units

⁶One epoch represents one presentation of each word in the training set.

⁷The networks were trained for 100 additional epochs to examine any changes in the behavior. No qualitative changes were observed, hence the results are reported for the first 250 epochs.

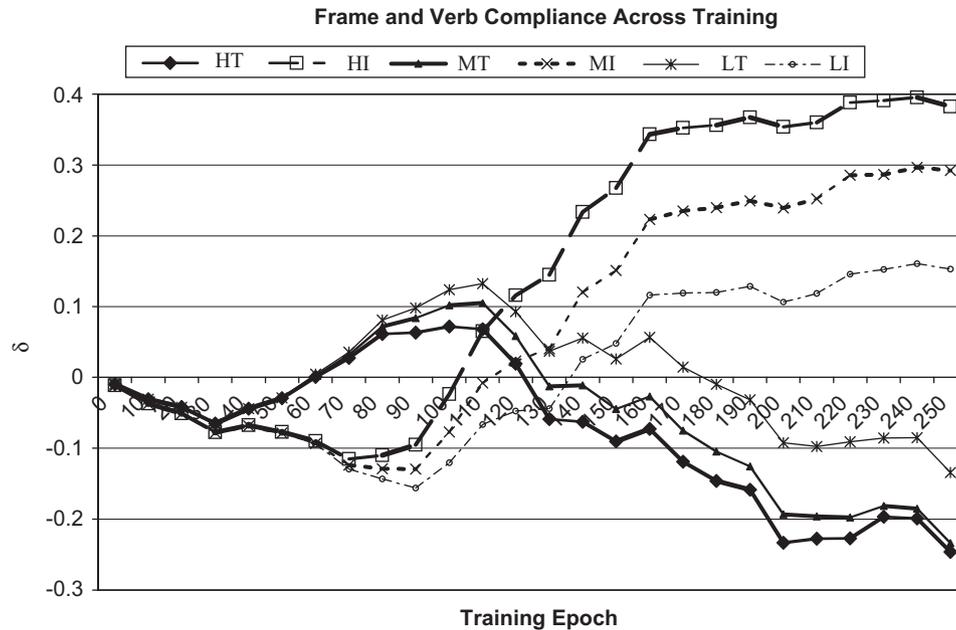


Fig. 3. The mean value (for 40 networks) of the parameter δ across training. δ is calculated at the end of the utterances, for ungrammatical utterances that use verbs in incompatible frames. A positive δ implies a causal interpretation of the verb while a negative δ indicates noncausal interpretation. H = high frequency verbs, M = medium frequency verbs, L = low frequency verbs, T = transitive frame, I = intransitive frame.

minus the mean activation of the last two units. A positive δ indicates a causal response, while a negative δ suggests a noncausal interpretation of the verb. The mean value of δ , at 10-epoch intervals, calculated at the end of transitive and intransitive utterances with incompatible verbs, is shown in Fig. 3.

4. Results

As one might expect, in the very early phases of training, the interpretation of two types of ungrammatical utterances does not differ significantly with regard to causality. As the network begins to learn regularities in the input related to causality, at 60–70 epochs, the value of δ moves towards positive values for the T frames (see the statistical results below), and towards negative values for the I frames. This indicates a Frame Compliant response. The network has learned the frame-causality correspondence more so than it has learned verb-causality association, and hence when there is conflict, the frame-based or syntactic cue tends to prevail over the verb-based or lexical cue. As the learning progresses, this trend begins to reverse without any change in the learning parameters or the input. Now the verb-specific cues are learned with greater accuracy, and by 200 epochs, both utterances show a tendency of Verb Compliance. The high-frequency verbs induce the strongest Verb Compliance, because they are learned with greater accuracy and exert stronger influence on the frame. The reverse is true for low-frequency verbs.⁸

⁸There are some asymmetries in the curves for T and I frames. These are due to a number of factors, one being that there are TFs that are noncausal but there are no IFs that are causal. There are also two types of

The data were analyzed with repeated measures ANOVA with three factors: frame (T, I), frequency (high, medium, low) and training (26 levels for training epochs in steps of 10). Main effects of all factors and interactions between all factors were obtained. There was a main effect of frame ($F(1, 39) = 46.89, p < 10^{-6}$), main effect of frequency ($F(2, 78) = 9.25, p < 0.0002$), and main effect of training ($F(25, 975) = 5.46, p < 10^{-6}$). There was also interaction between frame and frequency ($F(2, 78) = 89.24, p < 10^{-6}$), between frame and training ($F(25, 975) = 184.01, p < 10^{-6}$), between frequency and training ($F(50, 1950) = 7.07, p < 10^{-6}$), and between frame, frequency and training ($F(50, 1950) = 68.52, p < 10^{-6}$). These interactions can be seen in Fig. 3.

Post-hoc Newman–Keuls tests were conducted to assess differences in δ at particular stages in training. For simplicity, results at 0, 90 and 200 epochs of training are

(footnote continued)

IFs, with different lengths, one being shorter than the TFs and one being longer, and there is only type of T frame. This affects the total number of words for which causal or noncausal output is required. The number of words in a frame after the verb also affects how well the verbs are learned. By learning the verb-causality relationship, the network can predict causality correctly for all subsequent words. If there are many subsequent words, there is more impetus to learn this relationship because it results in a greater reduction in total error. This is why T verbs have a stronger influence on I frames than I verbs have on T frames. This is consistent with the behavioral results of Naigles et al. [31], where I frames showed Verb Compliance earlier than T frames. It is also possible to adjust the frequencies of different types of frames and alter the exact shape of these curves. The basic tendency of the network to learn the frame-based cues first and then move to verb-based cues is observed in such variations also, although the quantitative details of the result clearly depend on the values of various parameters, including the frequency of different utterance types.

reported. Similar results were obtained for 60–90 epochs and 200–250 epochs (individual epochs or collapsing across that range) respectively. At 90 epochs, high-frequency T frames (i.e., T frames with a high frequency I verb) and high-frequency I frames were different ($p < 0.00003$). There was an increase in δ for these T frames from epoch 0 ($p < 0.00001$), and a decrease in δ for I frames ($p < 0.00006$). At 200 epochs, the high-frequency T and I frames were again different ($p < 0.00006$). There was a change in δ from epoch 0 for both T ($p < 0.00004$) and I ($p < 0.00003$) frames. Similar results were obtained for medium and low-frequency frames for individual epochs, and also after collapsing across epochs 60–90 and across epochs 200–250. There was a significant difference between all three frequency levels within each frame type (3 pairwise comparisons) at 90 as well as at 200 epochs (all $p < 0.0001$).

For comparison, as one might expect, if grammatical sentences are used for testing, δ increases with training for T frames/T verbs, and is > 0.9 at 200 epochs. Similarly, it is < -0.9 for I frames/I verbs at 200 epochs.

4.1. Timecourse of activation

One way to gain further insight into the network's behavior is to examine the change in output activation as an utterance is presented. Fig. 4 shows the value of δ for an I frame and a T frame with incompatible verbs, averaged over all 7 verbs, after 90 and 200 epochs of training. Repeated measures ANOVA was performed separately for I and T utterances to evaluate the results. For the I frame (Fig. 4a), factors of words (6 levels) and training (3 levels at 0, 90, and 200 epochs) were used. There were main effects of words ($F(5, 195) = 109.07, p < 10^{-6}$) and training ($F(2, 78) = 5.11, p < 0.01$), and an interaction between words and training ($F(10, 390) = 100.09, p < 10^{-6}$). Post-hoc Newman–Keuls tests revealed that at 90 epochs, all words other than the first three words caused a change in δ (all $p < 0.02$) compared to its value at 0 epochs. The early part of the utterance (*a boy is*) is common to both frames and does not provide any information regarding causality. At 200 epochs, the last three words caused a change in δ compared to epoch 0 (all $p < 0.00001$), and also compared to epoch 90 ($p < 0.00003$). At epoch 0, the presentation of words did not cause a change in δ over the previous value of δ (all $p > 0.9$). At 90 epochs, presentation of the verb caused a decrease in δ ($p < 0.00001$) which was further changed by *-ing* ($p < 0.01$). The last word *stop* provided the syntactic cue, increasing the δ ($p < 0.00002$). Similar pattern in the change of δ was observed at 200 epochs. The change caused by the verb was stronger than at 90 epochs ($p < 0.00002$). The final word *stop*, however, did not change δ significantly compared to its value at the presentation of the verb ($p > 0.12$). Thus, change in the relative strengths of lexical and syntactic cues can be seen through training.

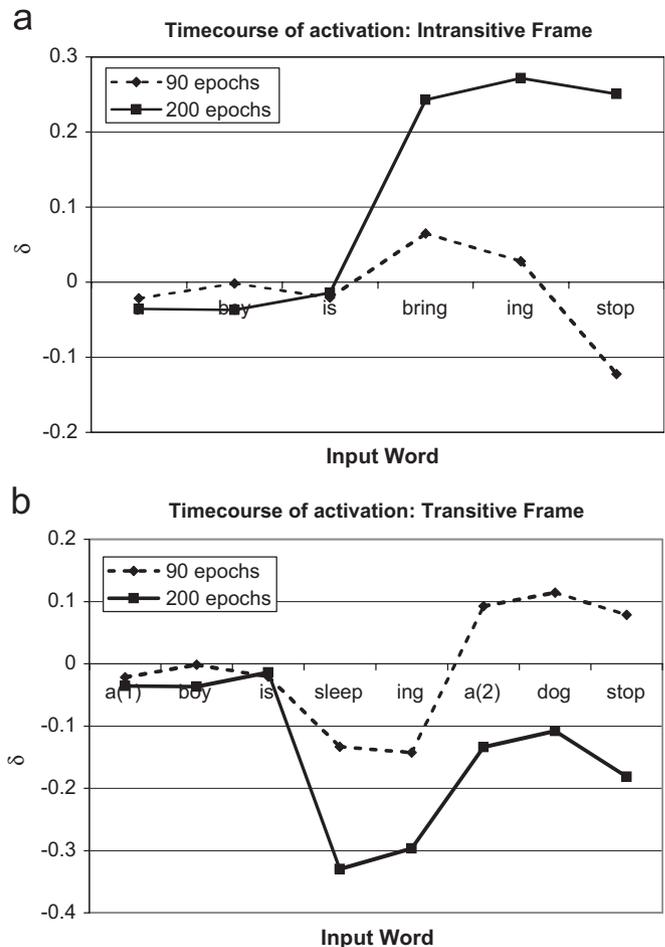


Fig. 4. The timecourse of the mean value of δ for two ungrammatical utterances with incompatible verbs (averaged across all incompatible verbs) at 90 and 200 training epochs. (a) Intransitive frame and (b) transitive frame (the two 'a's in the utterance are distinguished as 'a(1)' (first occurrence) and 'a(2)' (second occurrence)).

Analogous observations can be made for the T frame (Fig. 4b). The same statistical results hold for the T frame (where 8 levels were used for the words factor). Presentation of the verb caused a significant change in δ , and more so at 200 epochs than at 90 epochs. The syntactic cue, starting with *a(2)*, caused a significant increase in δ at both 90 and 200 epochs, but at 200 epochs it was not sufficient to make it positive.

Fig. 5 illustrates the timecourse of activation in another way. The value of δ throughout training is shown after each word is presented to the network. In both T and I utterances, the cue given by the verb is learned gradually and gains in strength throughout training. The syntactic cues, however, cause the network to show frame compliance early and the network is only "pulled" by the verb towards Verb Compliance later. Note that syntactic cues are not simply individual or "local" word cues. Words such as *a*, *-ing*, and *stop* affect the interpretation of causality differently depending on the context in which they are encountered.

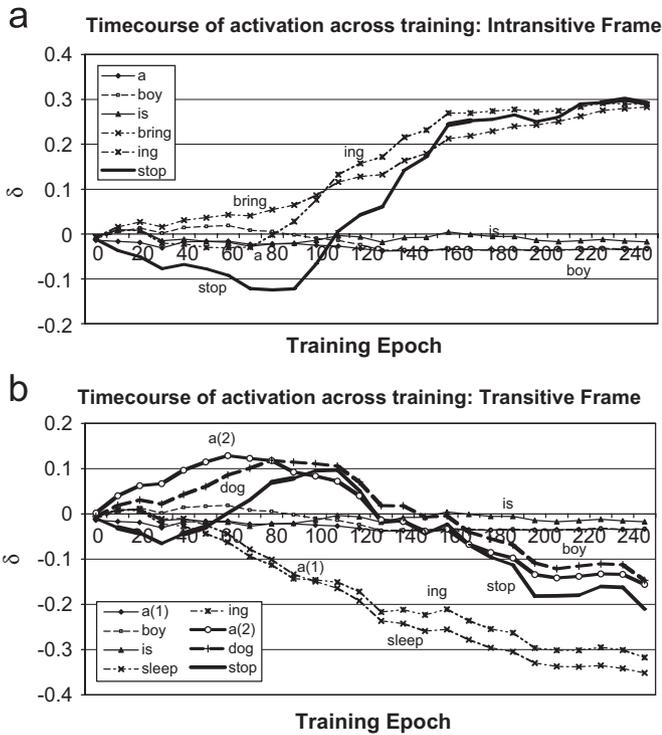


Fig. 5. The timecourse of the mean value of δ after presentation of each word across training, for ungrammatical utterances with incompatible verbs (averaged across all incompatible verbs). (a) Transitive frame (the two ‘a’s in the utterance are distinguished as ‘a(1)’ (first occurrence) and ‘a(2)’ (second occurrence)) and (b) intransitive frame.

4.2. The role of alternating verbs

The account of Compliance presented here relies on competition between lexical and syntactic cues to causality. Frame Compliance is due to syntactic cues being learned first, and the move to Verb Compliance is due to lexical cues becoming stronger than syntactic cues. It follows that if one of the cues is made stronger, the corresponding effects on Compliance should be observed. The frequency of verbs is closely related to the strength of the lexical cue. With higher frequency, the verb meaning is learned more accurately affecting the strength of lexical cue, resulting in relatively weaker Frame Compliance and stronger Verb Compliance, which can be observed in Fig. 3. We can examine analogous effects of the syntactic cues by manipulating the amount of alternating (A) verbs in the input. Alternating verbs necessitate the use of syntactic cues to correctly predict causality, because causality of an utterance depends only on whether the verb is used in a T or I frame. If Frame Compliance is a result of syntactic cues being learned early, then boosting these cues by increasing the number of utterances using A verb should result in stronger Frame Compliance and weaken Verb Compliance. Reducing or removing utterances with A verb should have the opposite effect.

Fig. 6a shows the values of δ through training, for utterances with incompatible verbs, when A verbs were

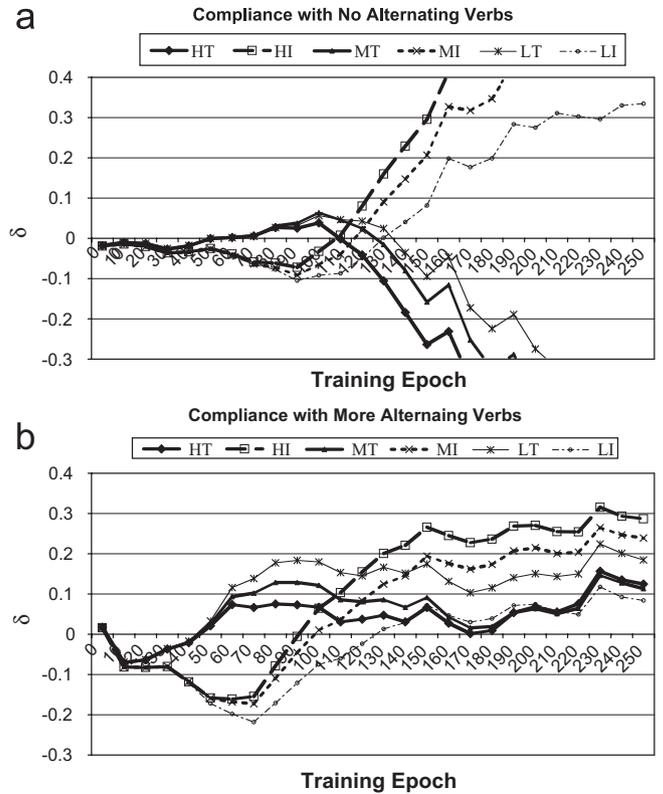


Fig. 6. The mean value of δ across training when the number of utterances with alternating verbs is changed in the input. The number of all other utterances is identical to the first experiment (Fig. 3). (a) No alternating verbs in the input. (b) When the number of alternating verbs is doubled from the first analysis. “More alternating verbs” refers to an increase in the token frequency, not the type frequency. H = high frequency verbs, M = medium frequency verbs, L = low frequency verbs, T = transitive frame, I = intransitive frame.

eliminated from the training set. The initial frame compliance is significantly weaker, and verb compliance in later stages is stronger because there is little competition for the lexical cues. Fig. 6b shows δ when the number of utterances with A verbs was doubled. Now Frame Compliance is stronger as predicted. Verb Compliance for the T frame is not achieved within 250 epochs and Verb Compliance for the I frame is also weaker. Further training for 100 epochs did result in verb compliance for T frames. In both these simulations, the number of all other types of utterances was identical to the main simulation (Fig. 3), and the network received identical amount of training on them. Hence, this change is not due to any modifications of the input other than the amount of utterances with A verbs.

Repeated measures ANOVA with a between-subjects factor of alternation (3 levels) and within subjects factors of frame (2 levels) and training (26 levels) was preformed. The 3 levels of frequency was collapsed for this analysis. Main effects of all factors as well as interactions between all factors were obtained. There were main effects of alternation ($F(2,117) = 14.07, p < 0.000003$), frame ($F(1, 117) = 241.58, p < 10^{-6}$), and training ($F(25, 2925) = 15.91,$

$p < 10^{-6}$). There were also interactions between alternation and frame ($F(2, 117) = 101.92, p < 10^{-6}$), alternation and training ($F(50, 2925) = 4.66, p < 10^{-6}$), frame and training ($F(25, 2925) = 599.91, p < 10^{-6}$), and alternation, frame and training ($F(50, 2925) = 91.53, p < 10^{-6}$).

Planned comparisons were used to compare the three levels of alternation within each frame type. The strength of Frame Compliance was assessed by collapsing δ between epochs 50–100 at each of the three levels of alternation, and that of Verb Compliance was assessed by collapsing across epochs 200–250. Each of the three pairwise comparisons between the three levels of alternation within both T and I frames yielded significant differences (all $p < 0.0001$), such that Frame Compliance was stronger with higher proportions of A verbs in the input, and Verb Compliance was stronger with lower proportions of A verbs.

We can also compare the speed with which the networks reach frame compliance and verb compliance with changes in the amount of utterances with A verbs. The networks were considered to have reached frame compliance when the δ for T frames was significantly positive and that for I frames was significantly negative. There were considered to have reached verb compliance using the same criterion for the two frames in the opposite direction. With higher proportion of A verbs in the input, the networks reached frame compliance faster and verb compliance later ($p < 0.0001$ for all pairwise comparisons).

It is interesting to observe the independent effect of syntactic cues on causality with differing amounts of A verbs in the input. We can assess the independent effect of the frame by inserting a novel verb into it, because a novel verb does not exert any (systematic) influence on causality. Fig. 7 shows δ when a novel verb is inserted in the two frames. The strength of the syntactic cue is tied to the amount of A verbs in the training set. Because the only way to correctly classify the sentences with A verbs is to rely on the frame, higher proportion of A verbs in the input results in better learning of syntactic cues contained in the frame.

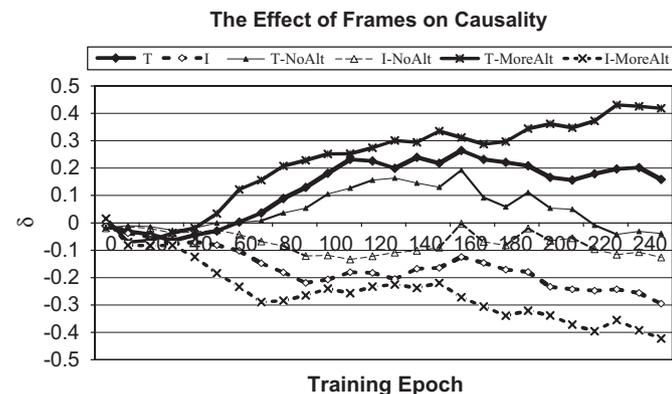


Fig. 7. The mean value of δ for frames with novel verbs, for networks with different proportions of alternating verbs in the input. T = transitive frame, I = intransitive frame (the number of alternating verbs corresponding to the first experiment, Fig. 3). NoAlt = no alternating verbs in the input, MoreAlt = more (double) alternating verbs in the input.

Planned comparisons were used to evaluate the differences in δ for three levels of A verbs, within each frame type. For both T and I frames, there were differences between each of the three pairwise comparisons between three levels of alternation (all $p < 0.0001$).

Note that when no A verbs are included in the training set, the strength of the syntactic cues increases initially, separating T and I frames. It then reduces and both T and I frames become similar. Using a planned comparison and only considering the last epoch, the difference between T and I frames is significant at 250 epochs ($p < 0.01$) but becomes nonsignificant at 300 epochs (not shown in the figure). This is due to the fact that when the network is Verb Compliant, causality can be predicted from the verb alone, and the syntactic cues are no longer necessary. The resources used for learning the syntactic cues (network weights in case of the model) can be employed for other learning or processing. When utterances with A verbs are present in the input, it is necessary to retain this learning in order to correctly classify them.

4.3. Verb meaning

The δ parameter is the main indicator of the causal interpretation of an utterance in the network. The verb “meaning”, however, is represented by 7 units (randomly set to 1 or 0), in addition to the causality units. We can examine this arbitrary portion of verb meaning for any effects of Frame or Verb Compliance. We can imagine that children, when acting out incompatible sentences in a Frame Compliant manner, modify the “core” meaning of the verb to some extent. The interpretation of the changed meanings in the network, however, is not straightforward because there is no semantic structure among different classes of verbs. The ideal representation of a T verb may be closest to the ideal representation of another T verb, an I verb, or an A verb. Hence, modifying one T verb to be similar to an I verb may be easy (requiring a small change in the bit pattern) and may be difficult (requiring a change in almost all bits) for another T verb. Here, the distances between the ideal (teacher) and actual (output) representations of T/causal, I, and A verb classes were calculated. The distances were averaged for all 7 verbs in a class. The results are shown in Fig. 8.

The results were analyzed with a repeated measures ANOVA with the factors of training (0, 90, and 200 epochs), output verb type (T, I) and the teacher verb type (T, I, and A). The “output verb type” factor represents the actual outputs, at the end of the utterance, for T and I verbs when inserted into incompatible frames. The “teacher verb type” factor levels are the desired representations given to the network for training. There were main effects of training ($F(2, 78) = 126.80, p < 10^{-6}$), output type ($F(1, 39) = 85.70, p < 10^{-6}$), and teacher type ($F(2, 78) = 3.71, p < 0.03$). Interactions between training and output type ($F(2, 78) = p < 10^{-6}$), between output type and teacher type ($F(2, 78) = 34.85$), and between training,

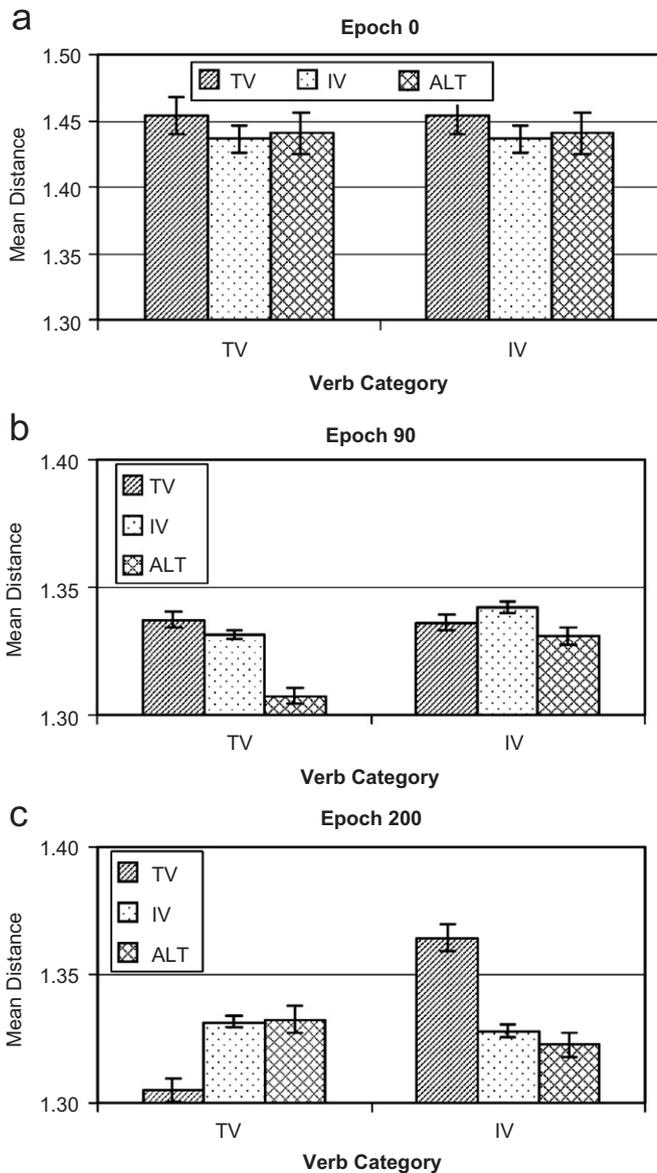


Fig. 8. Changes in the (arbitrary part of) verb meaning at three stages in training. For transitive (causal) and intransitive verbs, the mean distance between the actual and the ideal outputs of transitive (causal), intransitive, and alternating verbs is shown. X-axis is the output verb type, and Y-axis shows the distance to the teacher representations of the three verb classes. (a) Before training, (b) at 90 epochs, and (c) at 200 epochs. TV = transitive verbs, IV = intransitive verbs, ALT = alternating verbs. Error bars show the Standard Error of Mean.

output type, and teacher type ($F(4, 156) = 111.70, p < 10^{-6}$) were obtained. There was no interaction between training and teacher type ($F(4, 156) = 0.90, p > 0.46$).

Post-hoc Newman–Keuls tests showed that before training (epoch 0, Fig. 8a) both output T and I verbs were farthest from teacher T verbs ($p < 0.00002$) and their distance from the other two types was not significantly different. Because of the random assignment of output training patterns (identical for all networks), the teacher representations of T, I, and A verbs were not equidistant, giving rise to these differences before training. It is

interesting to see whether the networks modified the representations when Frame Compliant. Indeed, at epoch 90 (Fig. 8b), both output T and I verbs were closer to teacher A verbs than to the two other types ($p < 0.0002$). The difference between their distances from teacher T and I verbs was not significant for both T and I verbs. The network modified the representations of T and I verbs in incompatible frames to be similar to those verbs that occur in both T and I frames.

At 200 epochs, the pattern changed as the networks became more Verb Compliant. Both output T and I verbs were closer to their own teacher representations than to the opposite category (I and T, respectively) (both $p < 0.0002$). The output T verbs were closest to their own teacher category than to I or A verbs ($p < 0.0002$). The difference between their distance to I or A verbs was nonsignificant. The output I verbs, on the other hand, were closest to teacher I and A verbs (difference n.s.), followed by teacher T verbs. The asymmetry in the results is most likely due to the initial bias in the teacher representations (I and A verbs were more similar to each other than to T verbs).

In summary, when the networks were Frame Compliant in the initial phases of training, even the arbitrary parts of verb those were modified so as to make them similar to those of A verbs. In the later stages, when the networks were Verb Compliant, these modifications were much weaker and outputs for T and I verbs were more similar to their own teacher representations than to those of the opposite class. This behavior is similar to children when acting out sentences with incompatible verbs.

5. Discussion

An account of Frame and Verb Compliance was presented that is based on competition between syntactic and lexical cues to causality. Frame Compliance is result of the fact that there are always more frames than individual verbs in linguistic input, which makes frames easier to learn for a novice learner, and frames can (probabilistically) predict causality. Lexical cues are more powerful in that they are more reliable than syntactic cues, and allow prediction of causality earlier, making input processing more efficient. Hence, when the learner is sufficiently advanced to have learned the lexical cues well, they often trump syntactic cues when there is a conflict between the two. Both cues have graded strength, and hence different outcomes can result depending on their relative strength for a given frame/verb pair. When the strength of the lexical cue is increased (through verb frequency), stronger Verb Compliance (weaker Frame Compliance) is observed. Similarly, when the strength of the syntactic cue is varied (through the proportion of A verbs in the input), corresponding effects on compliance are observed. This account can be viewed as an extension of the Lexical Knowledge theory, in that increasing lexical knowledge of verbs causes a move towards Verb Compliance in an

input-driven manner. However, it does not rely on unspecified mechanisms for determining whether the child has sufficient knowledge of the verb to be Verb Compliant. The Competing Cues account also explains why the same verb may elicit differing compliance effects in different frames, a fact that is problematic for other accounts. The strength of the lexical cue is identical in different utterances, but because the competing syntactic cue can have differing strengths for various frames, a different outcome may result.

The competition between availability vs. specificity of cues is not unique to frames and verbs discussed here. This tension is also central to Rosch et al.'s [36] well-known ideas about basic-level categories, and to the global nature of infants' first categories [24]. Frequently available cues with useful predictive power are often learned first, followed by less frequent but more diagnostic cues.

While the importance of increasing lexical knowledge of verbs in compliance is acknowledged by other theorists, the Competing Cues account also emphasizes the role of frames in the early interpretation of verbs. This emphasis is in agreement with considerable evidence that frequent linguistic patterns, variously called “frames”, “constructions”, or “items”, play an important part in language acquisition, and even in adult language. To take a few examples, Mintz [28] analyzed six corpora of CDS and found that words surrounded by the same “frequent frames” were accurately classified together, suggesting that children (and adults) are sensitive to frame-like units. In another analysis of CDS, Cameron-Faulkner and Tomasello [5] found that a large part of maternal utterances consisted of repetitive phrases or items, and children used many of the same phrases, in some cases at a rate that was highly correlated with the maternal use. A large collection of cross-linguistic evidence that children's early language is based on specific linguistic items and expressions they comprehend and produce, is provided by Tomasello [38,39,40].

5.1. Key features of the model

What are the essential properties of the model presented here that cause early Frame Compliance and a shift to Verb Compliance? The first critical feature for Frame Compliance is the input where one cue to causality is very frequent and hence easy to learn. The second feature is the “greedy” nature of learning in the model. The network attempts to minimize the error as quickly as possible (using sum-squared error function and gradient-descent learning) using any cues available. As a result, it learns cues that may be sub-optimal (with respect to reducing the overall error) but frequently available throughout the training set and cause *some* immediate reduction in the error. I suggest that children, in some respects, are greedy learners as well. In many situations, they attempt to learn any regularities in the available input at a given time that allow them to make some useful predictions about the environment at that

time, rather than waiting for cues that may be more optimal.

The task of predicting the output is an important feature of the model that results in Verb Compliance. Verbs are better cues to causality than frames partly because they allow prediction of causality early in many cases. There is ample evidence that listeners attempt syntactic and semantic prediction and form expectations in the act of comprehension (e.g., some recent evidence from ERP and eye-tracking studies includes [1,20,42,41,12]). If the task of the network was to produce the correct output pattern corresponding only to the current input word, there would be no benefit in prediction, and the advantage of frames over verbs would be reduced considerably. Lexical cues would be only advantages in cases where frames are not very reliable predictors of causality, e.g., in the case of T frames.

The model learns by receiving feedback about the correctness of its response. Children do not frequently receive explicit feedback about syntax or word meanings, and hence this aspect of the model can be considered unrealistic. The argument that negative evidence is necessary to learn language but is not provided to children has been put forth by many theorists (e.g., [25]). While children may not receive negative *feedback*, they may receive negative *evidence* from their own internal mechanisms. Grammatical and lexical learning may be thought of as hypothesis testing, where children make predictions and form expectations from limited input. These hypotheses are subsequently confirmed, constituting positive evidence, or violated, constituting negative evidence. The feedback given to the model can be conceptualized as a simplified implementation of this internal mechanism.

5.2. Predictions

The Competing Cues theory makes some unique predictions about Compliance effects. Because the Compliance effects are due to the competition between lexical and syntactic cues, if the relative strengths of these cues are modified, it should have a direct effect on Compliance. If there are few A verbs in a language, the strength of the syntactic cues should be less and stronger Verb Compliance should be achieved. On the other hand, in languages where many verbs are allowed to alternate and where such usage is frequent, stronger frame compliance should be observed. For example, in Gujarati (an Indo-European language spoken in Western India by about 46 million people; [13]), morphological transformations can be used to convert most verbs into causal forms, but there are a few exceptions.⁹ The account presented here predicts that

⁹For example, *fenk-vu* (to throw) → *fenkaav-vu* (to have thrown or to cause to be thrown), *tod-vu* (to break) → *todaav-vu* (to have broken), *su-vu* (to sleep) → *suvaav-vu* (to put to sleep), *khench-vu* (to pull) → *khenchaav-vu* (to have pulled), but not **javu* (to go) → *javdaav-vu* (to cause to go) or **aav-vu* (to come) → *aavdaav-vu* (to cause to come).

stronger Frame Compliance should be obtained in such cases, with possibly even adults being Frame Compliant for some verbs, because the morphological cue to causality is common and consistent. Accounts based on maturation, innate principles, or lexical knowledge do not make this prediction. The maturation-based account predicts that once the narrow-range subclasses are closed at maturation, new and conflicting information will not be accepted and verb compliance would result. Accounts that rely on only lexical knowledge for a move towards verb compliance also suggest that for older children and adults, there should be sufficient lexical knowledge of the verb in question to induce Verb Compliance. The account presented here is based on competition between both syntactic and lexical cues, and hence predicts that when syntactic or morphological cues are strong, they can provide substantial competition to lexical cues even at a later age.

One of the reasons that the network switches to Verb Compliance in the later stages of training is that verbs allow prediction of causality earlier than frames do. This gives an advantage to verbs over frames even if both are equally consistent in predicting causality. Hence, the strength of the lexical cue offered by the verb is related to its position in the sentence. The Competing Cues account predicts that when a verb tends to be encountered early in the sentence, it will induce Verb Compliance earlier, because in this case the verb has more predictive power with respect to causality and hence is associated more strongly with it. It follows that children learning a verb-initial language such as Welsh or Arabic [6] will show greater Verb Compliance than English-learning children of the same age. By the same token, children learning a verb-final language, such as Hindi or German, should exhibit greater Frame Compliance than English-learning children. In English, T verbs should induce greater verb compliance than I verbs. This prediction is born out both in behavioral data [31] and the model. These predictions are not made by any other account of compliance, because the position of the verb in a sentence is irrelevant in these accounts.

5.3. The nature of input

A fixed set of utterances with verbs and nouns of varying frequency was used here for simulations. However, there is some evidence that the linguistic environment of children is not constant but changes with age. CDS to younger children is syntactically and semantically simplified, is less diverse, and contains more high-frequency words [7,19,34]. Caretakers restrict their vocabularies when talking to young children [7,19], and the type-to-token ratio increases with age in CDS [7]. What effect is this change likely to have on Compliance behavior? I suggest that such a change is likely to aid Verb Compliance, because there is evidence that syntactic diversity of verb use in CDS, along with frequency, is an important factor in verb learning [29,30]. With higher syntactic diversity (and associated semantic diversity), lexical knowledge about verb should be acquired

more rapidly and can hence aid Verb Compliance. Here, verbs were varied only in frequency and not in syntactic diversity. If syntactic diversity of verbs is also increased as the training progresses, it is plausible that better learning and possibly stronger Verb Compliance for syntactically diverse verbs can be modeled. Increased syntactic/semantic diversity is likely to aid the model in learning the verb meaning, because the verb form to verb meaning correspondence would be the only constant in varied uses of the verb and hence easier to isolate. Indeed, in his story-processing network, St. John [37] found good generalization only when the input was highly combinatorial, so that nouns and verbs occurred with many other nouns and verbs.

5.4. Semantic structure

Random bit patterns were used in the simulations here to represent the portion of verb meaning that does not directly indicate causality. It has been suggested, however, that the behavior of a verb, especially with regard to the interpretation and expression of its arguments, is guided to a large extent by its meaning. Levin [22] provides an extensive listing of semantic classes of English verbs whose syntactic behavior is similar. Semantic structure can be included in the model so that, for example, various I/noncausal verbs are similar to each other and more distant from various causal verbs. There will tend to be greater overlap in the internal representations of verbs in the same class, and less overlap between classes. This overlap can aid transfer of learning between verbs, such that each verb is not learned as a completely separate unit. For example, a low-frequency verb can show verb compliance early because although the verb itself is not well learned, it has overlapping high-frequency neighbors causing it to adopt their behavior.

5.5. Accounting for empirical data

The model presented here attempts to account for the basic Frame and Verb Compliance behavior shown by children and adults. Compared to the complexity of CDS and children's environment, syntactic and semantic inputs to the network used here were clearly impoverished, and hence quantitative comparisons between children's performance on individual verbs and the model's performance are not appropriate. However, there is at least one way in which the model's performance does not mirror empirical data. Strong Frame Compliance (approximately 70–80% causal act-outs) is achieved for I verbs *come* and *go*, which is higher than some lower frequency verbs. The model, as well as the Lexical Knowledge theory, however, suggest that high-frequency verbs elicit weaker Frame Compliance because they are learned more accurately and have a higher amount of lexical knowledge associated with them.

One way to explain this contradiction is to consider the actual semantic features of these verbs. Note that strong Frame Compliance is observed only for specific

high-frequency verbs, and not across-the-board for all high-frequency verbs. Some I verbs are similar, in their core meaning, to some T verbs, though the T verbs require an agent and a patient. When there is contextual pressure to act out an I verb in a causal manner, it is plausible that it is easier to do so for verbs that have causal semantic neighbors. T versions of *come* and *go* can be interpreted as some forms of *push*, *take*, or *pull*, which are also relatively high-frequency verbs. When enacting “zebra goes the giraffe”, zebra making the giraffe “go” can be thought of as the zebra taking or pushing the giraffe to a new location. Even if there were no such verbs in English as *push* or *take*, the familiarity of an event where an agent is taking a patient somewhere is likely to be relevant. Furthermore, one animal making another animal “go” is relatively straightforward to enact using simple toys with no moving parts. By contrast, “zebra stays the giraffe” can elicit less frame compliance although *stay* is a lower frequency verb than *go*. An agent physically forcing a patient to stay in one place is a relatively rare occurrence and does not cleanly correspond to a verb that children are likely to be familiar with. In addition, the enactment of one toy animal forcing another to stay in one place is relatively complex, especially in the absence of any other props such as corners, walls, doors, or ropes.

Additionally, common expressions such as “Let’s go!” and “Come on!”, though not strictly causative, have a pragmatic causative sense associated with them. Other expressions such as “Make it go!” (with reference to a remote-controlled toy) also imply causation. Such implied social causation may also affect the act-outs of particular verbs.¹⁰

Thus, frequency of frames and verbs are not the only factors that can affect children’s act-outs. Semantic features, semantic neighborhood, experiment-specific factors, and pragmatic usage of verbs may also be relevant. As noted before, the model presented here does not contain internal semantic structure for verbs, but in principle, it is possible to account for such effects as well. Note that other theories of Compliance based on lexical knowledge also do not account for Frame Compliance for such high frequency verbs, because there should be enough lexical knowledge of *come* and *go* to induce Verb Compliance after hearing them possibly thousands of times.

6. Conclusions

An account of Frame and Verb Compliance was presented that is based on competing lexical and syntactic cues to causality. A preliminary implementation of the theory was presented in the form of a connectionist network that learns to comprehend utterances of a miniature language by associating them with the corresponding scenes. During the course of training, the network exhibits Frame Compliance in the early stages

and Verb Compliance in the later stages, without any change in the input or training parameters. Frequent syntactic cues are learned first, resulting in Frame Compliance. More powerful but less frequent lexical cues are learned later, causing a shift to Verb Compliance. It also makes unique predictions about Compliance effects in other languages. Unlike other proposals, this account does not entail innate principles and explicit rules or reasoning mechanisms. It also makes unique predictions about compliance effects in other languages. There is nothing in the network designed specifically to produce these effects; they emerge as a result of the network attempting to accomplish the task of associating utterances with scenes. The network’s behavior is an emergent consequence of the interaction between characteristics of the input, task, and the learning algorithm of the network. Children’s compliance behavior, similarly, may change automatically when the strength of lexical and syntactic cues change without an explicit decision-making processes, and may indeed be different if the characteristics of the input are different. This work supports the view that specific mechanisms or behaviors can arise as a result of the nature of the input, task, and the general characteristics of the tools employed to perform the task, without the presence of dedicated mechanisms.

Acknowledgments

I thank Letitia Naigles, Gedeon Deák, and an anonymous reviewer for their detailed and very helpful comments and suggestions.

References

- [1] G. Altmann, Y. Kamide, Incremental interpretation at verbs: restricting the domain of subsequent reference, *Cognition* 73 (1999) 237–264.
- [2] P. Bloom, *How children learn the meanings of words*, MIT Press, Cambridge, MA, 2000.
- [3] Bowerman, M., 1977. The acquisition of rules governing “possible lexical items”: evidence from spontaneous speech errors, in: *Proceedings of Research in Child Language Development*, vol 13, pp. 148–156.
- [4] M. Bowerman, Starting to talk worse: clues to language acquisition from children’s late speech errors, in: S. Strauss (Ed.), *U-shaped Behavioral Growth*, Academic, New York, 1982.
- [5] T. Cameron-Faulkner, M. Tomasello, A construction based analysis of child directed speech, *Cogn. Sci.* 27 (2003) 843–873.
- [6] A. Carnie, E. Guilfoyle, *The Syntax of Verb Initial Languages*, Oxford University Press, New York, 2000.
- [7] R.S. Chapman, Mother-child interaction in the second year of life: its role in language development, in: R.L. Schiefelbusch, D. Bricker (Eds.), *Early Language: Acquisition and Intervention*, University Park Press, Baltimore, 1981.
- [8] N. Chomsky, *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht, Foris, 1981.
- [9] E.V. Clark, The principle of contrast: a constraint on language acquisition, in: B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*, Erlbaum, Hillsdale, NJ, 1987.
- [10] E.V. Clark, Acquisitional principles in lexical development, in: S. Gelman, J. Byrnes (Eds.), *Perspectives on Thought and Language*:

¹⁰I thank Gedeon Deák for this suggestion and examples.

- Interrelations in Development, Cambridge University Press, Cambridge, 1991.
- [11] G. Deák, Chasing the fox of word meaning: why “constraints” fail to capture it, *Dev. Rev.* 20 (2000) 29–80.
- [12] K. DeLong, T. Urbach, M. Kutas, Probabilistic word pre-activation during language comprehension inferred from electrical brain activity, *Nature Neurosci.* 8 (8) (2005) 1117–1121.
- [13] R. Doctor, *A Grammar of Gujarati*, Jain Publishing, Fremont, CA, 2001.
- [14] J.L. Elman, Finding structure in time, *Cogn. Sci.* 14 (1990) 179–211.
- [15] C.J. Fillmore, The mechanisms of ‘Construction Grammar’, *Berkeley Linguistics Society Meeting 14* (1988) 35–55.
- [16] L. Gleitman, H. Gleitman, A picture is worth a thousand words, but that’s the problem: The role of syntax in vocabulary acquisition, *Curr. Dir. Psychol. Sci.* 1 (1992) 31–35.
- [17] L. Gleitman, The structural sources of verb meanings, *Language Acquisition* 1 (1990) 3–55.
- [18] A. Goldberg, *Constructions: A Construction Grammar Approach to Argument Structure*, University of Chicago Press, Chicago, 1995.
- [19] Q. Hu, A study of some common features of mothers’ vocabularies, in: J.L. Sokolov, C.E. Snow (Eds.), *Handbook of Research in Language Development Using CHILDES*, Erlbaum, Hillsdale, NJ, 1994.
- [20] Y. Kamide, C. Scheepers, G. Altmann, Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English, *J. Psycholinguistic Res.* 32 (2003) 37–55.
- [21] B. Landau, L. Gleitman, *Language and Experience: Evidence From the Blind Child*, Harvard University Press, Cambridge, MA, 1985.
- [22] B. Levin, *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, Chicago, 1993.
- [23] J. Lidz, H. Gleitman, L. Gleitman, *Kidz in the ‘Hood: Syntactic Bootstrapping and the Mental Lexicon*. (Tech. Rep. IRCS-01-01), University of Pennsylvania, Institute of Research in Cognitive Science, Philadelphia, PA, 2001.
- [24] J.M. Mandler, Perceptual and conceptual processes in infancy, *J. Cogn. Dev.* 1 (2000) 3–36.
- [25] G. Marcus, Negative evidence in language acquisition, *Cognition* 46 (1993) 53–95.
- [26] E. Markman, How children constrain the possible meanings of words, in: U. Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge University Press, Cambridge, 1987.
- [27] W.E. Merriman, J.A. Evey-Burkey, J.M. Marazita, L.H. Jarvis, Young two-year-olds’ tendency to map novel verbs onto novel actions, *J. Exp. Child Psychol.* 63 (1996) 466–498.
- [28] T. Mintz, Frequent frames as a cue for grammatical categories in child directed speech, *Cognition* 90 (2003) 91–117.
- [29] L. Naigles, E. Hoff-Ginsberg, Input to verb learning: evidence for the plausibility of syntactic bootstrapping, *Dev. Psychol.* 31 (1995) 827–837.
- [30] L. Naigles, E. Hoff-Ginsberg, Why are some verbs learned before other verbs? Effects of input frequency and structure on children’s early verb use, *J. Child Lang.* 25 (1998) 95–120.
- [31] L. Naigles, A. Fowler, A. Helm, Developmental shifts in the construction of verb meanings, *Cogn. Dev.* 7 (1992) 403–428.
- [32] L. Naigles, A. Fowler, A. Helm, Syntactic bootstrapping from start to finish with special reference to Down Syndrome, in: M. Tomasello, W. Merriman (Eds.), *Beyond Names of Things: Young Children’s Acquisition of Verbs*, Erlbaum, Hillsdale, NJ, 1995.
- [33] L. Naigles, H. Gleitman, L. Gleitman, Children acquire word meaning components from syntactic evidence, in: E. Dromi (Ed.), *Language and Cognition: A Developmental Perspective*, Ablex, Norwood, NJ, 1993.
- [34] J. Pine, The language of primary caregivers, in: C. Gallaway, B.J. Richards (Eds.), *Input and Interaction in Language Acquisition*, Cambridge University Press, Cambridge, UK, 1994.
- [35] S. Pinker, *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press, Cambridge, MA, 1989.
- [36] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, P. Boyes-Braem, Basic objects in natural categories, *Cogn. Psychol.* 8 (1976) 382–439.
- [37] M.F. St. John, The story gestalt: a model of knowledge-intensive processes in text comprehension, *Cogn. Sci.* 16 (1992) 271–306.
- [38] M. Tomasello, The item-based nature of children’s early syntactic development, *Trends Cogn. Sci.* 4 (4) (2000) 156–163.
- [39] M. Tomasello, Do young children have adult syntactic competence?, *Cognition* 74 (2000) 209–253.
- [40] M. Tomasello, *Constructing a Language*, Harvard University Press, Cambridge, MA (London), 2003.
- [41] J. Van Berkum, C. Brown, P. Zwitserlood, V. Kooijman, P. Hagoort, Anticipating upcoming words in discourse: evidence from ERPs and reading times, *J. Exp. Psychol.: Learn., Mem., Cogn.* 31 (2005) 443–467.
- [42] N. Wicha, E. Moreno, M. Kutas, Anticipating words and their gender: an event related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading, *J. Cogn. Neurosci.* 16 (2004) 1272–1288.



Rutvik H. Desai obtained his Bachelor’s degree in Computer Engineering from Gujarat University, India, his Master’s degree in Computer Science from Utah State University, and his Ph.D. in Computer Science and Cognitive Science from Indiana University. He received postdoctoral training in cognitive neuroscience and functional neuroimaging from the Medical College of Wisconsin, where he is currently an Assistant Professor of Neurology. His interests include language acquisition, neuroimaging of language, and artificial intelligence.