A Model of Temporal Connective Acquisition

Mark A. Gorenstein[1,2,*], Zhang Cedegao[1,*], Steven T. Piantadosi[1]
{gorenstein, cedzhang, stp}@berkeley.edu
[1]Department of Psychology, [2]Helen Wills Neuroscience Institute
University of California, Berkeley, CA 94720, USA
*These authors contributed equally to this work.

Abstract

Temporal connectives are function words that relate events in time. Despite their ubiquity and utility, children acquire the meanings of temporal connective words late in development. Experimental work has uncovered patterns in the acquisition of temporal connectives that clarify the learning challenge that these words pose to children. In particular, developmental studies have identified differing acquisition trajectories across connective types, asymmetries in learning within pairs of related connectives, and monotonic increases in comprehension with age. Expanding on prior theoretical accounts, we formalize temporal connective acquisition in a computational word learning framework. We demonstrate that each of the empirically determined acquisition patterns emerges in the learning behavior of our computational model. Finally, we discuss our findings in relation to earlier theories and to general learnability concerns in language acquisition.

Keywords: computational modeling; language acquisition; learnability; semantics; time

Introduction

Did you brew a cup of tea before settling into your desk chair? Were you listening to a Charles Mingus record while the kettle boiled? Temporal connective words like ‘before’ and ‘while’ enable speakers to express temporal relations between events and may facilitate sophisticated forms of causal and event reasoning. Temporal connectives are present in many languages and some researchers have proposed that they function as semantic universals (von Fintel & Matthewson, 2008). Despite the ubiquity and utility of temporal connectives in natural language, children do not fully acquire their meanings until after age 7 (Feagans, 1980). Moreover, children exhibit distinct learning trajectories for each of the temporal connectives, complicating efforts to explain temporal connective acquisition within a unified framework. In the present study, we develop a new model of temporal connective acquisition that synthesizes formal semantic analyses of the meanings of temporal connectives with statistical learning methods capable of inferring these meanings from developmentally plausible amounts of data. In what remains of the introduction, we first review the empirical findings that pertain to theories of temporal connective acquisition, then present two prominent theoretical accounts, and lastly introduce our word learning framework.

Empirical Findings

Developmental studies have identified unique patterns of acquisition that distinguish temporal connective types. According to Feagans (1980), children first understand temporal connectives expressing sequence (‘before’ and ‘after’), then temporal connectives expressing simultaneity (‘while’), and, lastly, temporal connectives expressing both sequence and duration (‘since’ and ‘until’). Investigating children’s comprehension of all three types of connectives, Feagans (1980) found that children understand sentences containing ‘before’ and ‘after’ at age 3 and those containing ‘while’ at age 7, but do not exhibit above-chance understanding of sentences containing ‘since’ and ‘until’ at any age in that range. A cross-linguistic study confirmed the general tendency of this learning trajectory in English, Thai, and Lisu (Winskel, 2003).

Beyond the developmental differences found across temporal connective types, several studies have identified unique acquisition trajectories among the words within a connective type. Beginning with Clark (1971), experimenters have routinely observed that children understand the connective ‘before’ prior to ‘after,’ in spite of their superficial similarity (Feagans, 1980; Winskel, 2003; Blything, Davies, & Cain, 2015). Additionally, both Feagans (1980) and (Winskel, 2003) report that ‘until’ is better comprehended than ‘since’ among cohorts of English-speaking children, although this comparison has attracted less theoretical attention.

Finally, several studies provide evidence for monotonic increases of temporal connective comprehension with age. Feagans (1980), Winskel (2003), and Blything et al. (2015) all report main effects of age in their linear model analyses, and Clark (1971) also reports a strong negative correlation between a subject’s error rate and age. Crucially, these improvements persist beyond the specific age group in which children were judged to have acquired the meaning of a connective, which is typically determined by a performance threshold in the experimental task.

Theoretical Accounts

Theoretical treatments of temporal connective acquisition have attempted to account for the empirical findings just outlined in terms of broader developmental principles. A promising direction of research has related semantic analyses of the temporal connectives to their relative ease of acquisition. An early theory offered by Clark (1971) posits that distinct combinations of binary-valued and hierarchically-organized features comprise the meanings of the temporal connectives. In order of generality to specificity, the features Clark proposes are: time (indicating whether the word refers to a temporal relation), simultaneous (indicating whether events in relation overlap), and prior (indicating that one event precedes another). The theory predicts that children acquire features in general-to-specific order, and that within each stage of feature acquisition, the positive value of the feature is learned first. Accordingly, children will first learn the correct meaning of 'when' (+time, +simultaneous), then the meaning of 'before' (+time, -simultaneous, +prior), and lastly the meaning of 'after' (+time, -simultaneous, -prior).

An alternative theory from Feagans (1980) proposes that the logical complexity of temporal connectives determines their order of acquisition. Feagans (1980) surveys formal semantic analyses of the temporal connectives and notes that their relative complexity matches the order of acquisition present in empirical data. That is, children first acquire the temporal connectives that possess the fewest logical elements ('before' and 'after') and then proceed to acquire connectives consisting of more logical elements ('while' then 'since' and 'until'). Rather than positing the existence of several distinct temporal features, Feagans contends that children construct temporal connective meanings from just a single primitive relation, the temporal ordering of two events.

Word Learning Framework

Our study develops a word learning model that can formally assess the predictions of Feagan's theory. We hypothesize that learners build semantic expressions for the temporal connectives from simpler representational primitives. We take inspiration from *language of thought* (LOT) models of concept and word learning, which provide a formalism for uniting compositional semantic representations with statistical learning mechanisms that can infer target meanings from observed data (Fodor, 1975; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Piantadosi, 2011; Goodman, Tenenbaum, & Gerstenberg, 2014; Piantadosi, Tenenbaum, & Goodman, 2016; Piantadosi & Jacobs, 2016). Prior research has shown that LOT models can learn the meanings of quantifiers, another important class of function words, under assumptions and with learning targets comparable to those of the current study (Piantadosi, Tenenbaum, & Goodman, 2012). In the next section, we introduce the target temporal connective meanings that our model will attempt to learn.

Semantic Analysis of Temporal Connectives

A rich and long-standing tradition in formal semantics has analyzed the usage of temporal connectives and produced precise logical expressions of their meanings. In these analyses, temporal connectives are treated as functions that map event contexts to binary truth values (*true* or *false*). We first present the basic intuitions behind influential formal semantic accounts of the five temporal connectives and then describe the interval-based event representation used in this work.

'Before' and 'after': Perhaps surprisingly, leading analyses of the semantics for 'before' and 'after' suggest that they are not exact converses of one another (Anscombe, 1964; Beaver & Condoravdi, 2013). That is, it is not the case that for all true utterances of the form '*A* after *B*', '*B* before *A*' is also true. For example, consider that Alice was in California from 2015 to 2019 and Bob was in California from 2016 to 2018. On the basis of these facts, we can state that Alice was in California after Bob was. However, we cannot say that Bob was in California before Alice was (presumably because Alice arrived there first). Our target expressions for 'before' and 'after' follow the analysis of Anscombe (1964) and preserve their asymmetries.

'Since' and 'until': These two words can be viewed as special cases of 'after' and 'before,' respectively. That is, whenever '*A* since *B*' is true, '*A* after *B*' must be true, and the same relationship holds for the 'until-before' pair. In a sentence containing 'since' or 'until,' the main clause expresses an event whose duration extends into a reference time point determined by the context. Our target expressions for these two words follow the analysis in temporal logic from Kamp (1968), which is widely accepted in the field of formal semantics (Zwarts, 1995; Condoravdi, 2010).[1] Kamp's analysis focuses only on the retrospective cases of 'since' and the prospective cases of 'until.' In these cases, the contextual reference points can be naturally modelled as the moments of utterance. To illustrate this, consider the following two utterances of retrospective 'since' and prospective 'until': 'He has been playing piano since he attended grade school' and 'I will be working in the office until she comes back.' For the 'playing piano' event, we assume an ongoing process that persists into the moment of utterance. For the 'working' event, we assume that it started before the moment of utterance and will end at some point during the 'coming back' event.

'While': For '*A* while *B*' to be true, *A* and *B* just need to overlap with one another in time (Monens, 1971; Bennett & Partee, 2004). Suppose that Charlie slept from 2 PM to 4 PM and that it rained from 3 PM to 5 PM; in this case, one can validly assert both that Charlie was sleeping while it was raining and that it was raining while Charlie was sleeping. A common application of '*A* while *B*' establishes that *A* is a part of *B*, as in 'She took a psychology course while she was a freshman in college.' Since the sub-event relation is a special case of the overlapping relation, our target expressions for 'while' conform to such usages.

Interval-Based Event Representation

The formal semantic analyses reviewed above are often expressed in first-order logic. For the present study, we translate first-order logical formulas into semantically equivalent expressions in a simpler representational language that assumes an interval-based representations of events. In our representation, time is discrete and linear, so time points are represented by integers. An event E is represented by an interval defined by two integers e_1 and e_2, which are the start and end points of E, respectively. Our formulation diverges from the set-based representations typically employed by semanticists, but confers two main advantages: firstly, intervals are

[1] Kamp (1968) and Zwarts (1995) are not readily available online. For an overview of Kamp's semantics, see Section 4 of Goranko and Rumberg (2020) or Chapter II of van Benthem (1991).

Table 1: Target hypotheses of the semantics

Word	Target Hypothesis
A before *B*	$a_1 < b_1$
A after *B*	$b_1 < a_2$
A since *B*	$(a_1 < t) \wedge (t \leq a_2) \wedge (a_1 \leq b_2) \wedge (b_1 < t)$
A until *B*	$(a_1 \leq t) \wedge (t < a_2) \wedge (b_1 \leq a_2) \wedge (t < b_2)$
A while *B*	$(b_1 < a_2) \wedge (a_1 < b_2)$

cognitively plausible representations of events in time (Ivry & Hazeltine, 1995), and secondly, interval representations are more memory-efficient and computationally tractable than set representations, and so they are commonly used in computational systems that perform temporal reasoning (Allen, 1983).

Importantly, for all of the five temporal connectives, we obtain quantifier-free translations of the quantified first-order semantics using just the integer comparators $=$, $<$, $\leq$, and the Boolean connectives. For example, the expression $(\exists a \in A)\ (\forall b \in B)\ a < b$ is a first-order logic formula that expresses our target semantics for '*A* before *B*.' It says that there exists a time point a in A such that for any time point b in B, a precedes b. Given our assumptions about the structure of time (discrete and linear) and events (the start point preceding the end point), this proposition is equivalent to $a_1 < b_1$ in our notation, which says that the start point of A, a_1, precedes the start point of B, b_1. To see why, note that b_1 precedes all other points in B. Therefore, a_1 precedes *all* points in B. a_1 is in A, so a point in A that precedes all points in B exists.[2] The final target hypotheses for our model are shown in Table 1. In the next section, we proceed to describe the key components of our learning model.

The Learning Model

We define $w = \langle$before, after, since, until, while$\rangle$ as the words whose hypothetical meanings $m = \langle m_1, \ldots, m_5 \rangle$ the learner is representing such that m_i is the meaning of w_i. For example, the target m_5 for $w_5 =$ while is

$$\lambda\, A\, B\, t\ .\ (b_1 < a_2) \wedge (a_1 < b_2)\,.$$

A given data point for learning consists of a tuple $\langle u_i, c_i \rangle$, where $u_i \in w$ is the utterance attached to a context c_i. A context, c_i, consists of two events, A and B, and a moment of utterance, t. The components of a context are the input arguments to the Boolean functions that learners construct. An illustration of our learning setup is shown in Figure 1.

We are interested in computing $P(m \mid u, c)$, the probability of a set of meanings m, given observed contexts c paired with utterances u. By Bayes' rule,

$$P(m \mid u, c) \propto P(u \mid m, c) \cdot P(m)\,. \quad (1)$$

[2] Although the five word meanings considered here have equivalent quantifier-free expressions in the interval representation, not *all* first-order logical expressions admit of such translations.

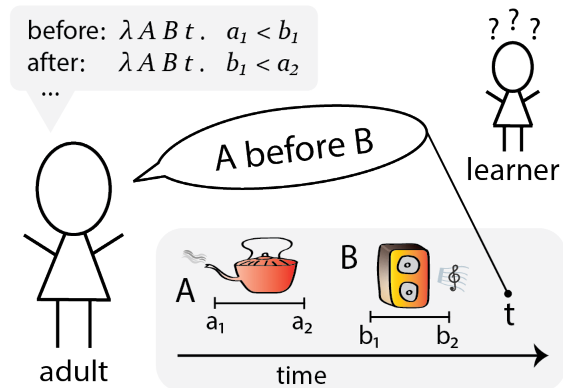

Figure 1: Word learning setup. Adults produce utterances according to the target temporal connective meanings, and the learner's goal is to infer these meanings from observed utterances and contexts. Suppose that an adult utters 'Jack set the kettle to boil before he turned on the radio' at time t. Here the event 'Jack sets the kettle to boil' (A), the event 'Jack turns on the radio' (B), and time t form a context. In this case, the learning target depends only on A and B, since the moment of utterance, t, is not part of the semantic expression for 'before.'

In the following subsections we describe the grammar that generates expressions m, a probabilistic extension of the grammar that defines prior probabilities $P(m)$, and a function that computes the likelihood of utterances given a set of meanings and observed contexts, $P(u \mid m, c)$.

Grammar

The target temporal connective meanings we previously described can be generated by a context-free grammar (CFG). Such grammars consist of non-terminal variables that expand, through the successive application of production rules, into strings consisting exclusively of terminal symbols. Our CFG, which is specified in Table 2, always returns a Boolean function of the input. It supports the standard Boolean connectives, *not*, *and*, and *or*, which can define any other binary Boolean operation (e.g., *implies*).[3] The grammar also includes integer comparators $=$, $<$, and $\leq$. The int arguments of the comparators can be any of the input integers: a_1, a_2, b_1, b_2, or t.

One should note that this is a very expressive grammar that can generate many logical expressions that are irrelevant to the temporal relations under consideration. For example, the grammar can generate hypotheses like $(a_2 < t \wedge b_2 < t)$, which means that both A and B ended before the time of utterance. The target expressions for 'since' and 'until' are functions that consist of three Boolean connectives, of which there are 2^{16} possibilities in total. For each of those possibilities, one can obtain still different expressions by changing the integer-comparison arguments ($a_1 = t$, $b_2 \leq a_2$, etc.) for the Boolean

[3] The technical notion here is that the three functions form a *functionally complete* set of Boolean operations (Enderton, 2001).

Table 2: The grammar used to generate meanings of temporal connectives.

Nonterminal		Expansion	Description
START	→	λ A B t . BOOL	A function of A, B, and t that returns a truth value
BOOL	→	*true*	Always true
	→	*false*	Always false
	→	(*and* BOOL BOOL)	True if and only if (iff) both arguments are true
	→	(*or* BOOL BOOL)	True iff at least one argument is true
	→	(*not* BOOL)	True iff the argument is false
	→	(= int int)	True iff the two arguments are equal
	→	(< int int)	True iff the first argument is less than the second
	→	(≤ int int)	True iff the first argument is less than or equal to the second
int	→	a_1	The first element of argument A (the start point of A)
	→	a_2	The second element of argument A (the end point of A)
	→	b_1	The first element of argument B (the start point of B)
	→	b_2	The second element of argument B (the start point of B)
	→	t	Argument t (the time point of the utterance)

connectives. So the entire hypothesis space of expressions containing at most three Boolean connectives is, by combinatorial approximation, on the order of 10^8. Therefore, learning specific temporal relations in our grammar is highly nontrivial as it requires identifying plausible relations in a large space of possible hypotheses.

Prior

We define prior probabilities, $P(m)$, over hypothetical meanings using an extension of our grammar. Having specified the CFG in Table 2, we transform it into a probabilistic context-free grammar (PCFG) by assigning probabilities to the expansion rules. For every non-terminal, we assign uniform probabilities to each of its expansion rules so that the prior does not prefer any particular expansions.

This PCFG induces a prior distribution over all possible hypotheses. The prior probability of a hypothesis is proportional to the product of the probabilities of its constituent grammar expansions. Our prior specification penalizes grammar expansions and therefore biases learning towards shorter hypotheses. Thus, we build a simplicity bias into our model, which has been shown to capture the behavioral tendencies of experimental subjects (Feldman, 2000; Chater & Vitányi, 2003).

Likelihood

We formulate a likelihood that represents language production from the learner's perspective. We assume that the learner determines how likely a speaker would have been to utter a meaning in a hypothetical set m given observed event contexts c. That is, the likelihood measures the probability that utterances $u_i \in u$ would be produced in contexts $c_i \in c$ given that the speaker holds m to be the meaning of the temporal connectives.

We assume that each utterance u_i depends only on the hypothetical meanings m and the context c_i, but not any of the other utterances or their corresponding contexts, and thus can rewrite the likelihood as:

$$P(u \mid m, c) = \prod_{i=1}^{n} P(u_i \mid m, c_i)\,. \tag{2}$$

Importantly, multiple utterances can be true in a given context under some hypothetical meanings m. So, to compute $P(u_i \mid m, c_i)$, we first partition w into two sets by evaluating each w_x on the function represented in its corresponding m_x. w_{true} contains those words that are true under the current context c_i and meanings m,

$$w_{true}(m, c_i) = \{w_x \in w \mid m_x(c_i) = true\}, \tag{3}$$

while w_{false} consists of the remaining words:

$$w_{false}(m, c_i) = w - w_{true}(c_i). \tag{4}$$

We assume that speakers generate utterances that are true of a context with probability $\alpha = 0.95$, which characterizes the amount of noise in the data. In the likelihood, we model this production noise as a speaker's choosing u_i randomly from the set w, either ignoring or misapplying the functions present in m.

If $u_i \in w_{true}$, the utterance's inferred meaning applies to the context and:

$$P(u_i \mid m, c_i) = \frac{\alpha}{|w_{true}(c_i)|} + \frac{1-\alpha}{|w|}. \tag{5}$$

The two terms correspond to the two ways in which a true utterance could have been generated: either from sampling with uniform probability from the elements of w_{true} or by choosing u_i randomly from the set of all temporal connective words w. If, instead, $u_i \in w_{false}$, then only this second possibility applies and thus:

$$P(u_i \mid m, c_i) = \frac{1-\alpha}{|w|}. \tag{6}$$

The likelihood for a true utterance given in Equation 5 follows the size principle (Tenenbaum, 1999), a crucial feature of our model that biases the learner to prefer meanings that are true in fewer contexts.

Inference

Given the expressiveness of our grammar, many of the generated hypotheses will have low probability, either because of their long length (small prior) or inability to explain the data (low likelihood). To approximate the posterior distribution, we employ Markov chain Monte Carlo (MCMC) methods. As a practical computational constraint, we set the maximum number of grammar expansions in a hypothesis to 15. We run an incremental sampling procedure beginning at 25 data points (consisting of five examples of each word) and increment in steps of 25 up to a maximum data amount of 600. The training data for all words are generated according to the likelihood given by Equation 5 assuming the target semantics given in Table 1. When generating data for a given word, we randomly select event intervals A and B, a time point t, and accept that context with its likelihood probability. At each data increment, we run the Metropolis-Hastings algorithm for 50,000 steps to sample the posterior, $P(u|m,c)$, for many possible settings of m, and store the ten sets of meanings with the greatest posterior probability. We repeat the entire sampling procedure 30 times.

Results

Following machine learning best-practices, we evaluate the top hypotheses our model learns on a testing dataset. This testing dataset is generated as follows: for each word, we produce five-thousand positive and five-thousand negative examples according to the target hypotheses in Table 1. Chance performance (e.g., a hypothesis that is trivially true for all contexts) on this task is 50%. Evaluating our model on a testing dataset allows us to identify partially-learned hypotheses that exhibit above-chance performance but may lack components of the target semantics.

We retrieve the top ten hypothetical meanings m for each interval step. Next we partition the hypotheses in the ten high-probability settings of m by word and count the proportion of the 3,000,000 test examples (10 hypotheses · 30 repeats · 10,000 test examples per word) that are correctly labeled at that data amount. We plot these accuracy values for each data amount in Figure 2, which summarizes our results.

Our model learns the target meanings for all of the temporal connectives in w. For each connective, the model learns hypotheses that perform significantly above chance, even at the smallest data amount (five training examples per word). Moreover, the learning curve for each temporal connective exhibits monotonic improvement up to a plateau. Assuming a 95% accuracy criterion for successful acquisition, the order of acquisition our model attains is: 'before' $\prec$ 'after'$\prec$ 'while' $\prec$ 'since' $\approx$ 'until.'

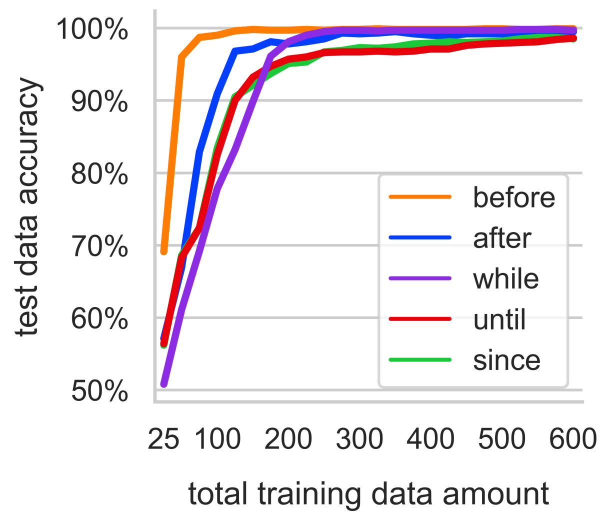

Figure 2: Modeling results. The top ten hypotheses learned at each training data amount and inference chain are evaluated on 10,000 test examples, split equally between those that are true and false of each word. Chance performance is thus 50%.

Discussion

Our learning model recapitulates each of the developmental findings outlined in the introduction. Firstly, the model acquires accurate representations of the sequential connectives ('before' and 'after'), followed by the simultaneous connective ('while'), and lastly the connectives expressing both sequence and duration ('since' and 'until'). Secondly, the model acquires 'before' prior to acquiring 'after.' Thirdly, our model learns increasingly more precise meanings for the connectives, resulting in monotonically increasing accuracy.

The learning framework we employ captures aspects of the theoretical proposals offered by both Feagans (1980) and Clark (1971). The semantic feature theory of Clark (1971) emphasizes the incremental character of temporal connective acquisition, which is argued to proceed in discrete stages as children acquire the specific temporal features necessary for understanding each of the connectives. Our model also exhibits incremental learning because the prior introduces a preference for simple meanings, so that complex expressions are assigned high probability only when simpler expressions fail to account for observed utterances. Our account diverges from Clark's because the features that constitute our target meanings are available to the model from the outset as primitives, and so the learner's task is to infer combinations of primitive elements that best account for observed utterances.

Our theoretical approach is most consistent with that of Feagans (1980), which proposes that logical complexity determines the order of temporal connective acquisition. Indeed, we found that the logical complexity of our target connective meanings determined the amount of training data required for successful acquisition. Our study extends Feagans' work by demonstrating that a statistical learner can acquire accurate meanings for the temporal connectives from developmentally-plausible amounts of data. Feagans (1980) observes that the set of temporal connective meanings can be constructed from a single operation that compares the relative ordering of two events. In our framework, this operation is achieved by applying the integer comparators $=$, $<$, and $\leq$ to pairs of event time points. Beyond its representational simplicity, Feagans' proposal is compatible with theories positing that a small set of primitive cognitive operations, which may include the operators relevant to this study, can support learning across a range of domains (Fodor, 1975; Piantadosi, 2011; Piantadosi & Jacobs, 2016).

A critical feature of our model is its ability to overcome the subset problem in language acquisition (Wexler & Manzini, 1987; Crain, Ni, & Conway, 1994). Our learning framework is potentially susceptible to the subset problem because several of our target expressions overlap in meaning. Namely, the contexts in which 'until' is true is a small subset of those in which 'before' is true, and the contexts in which 'since' and 'while' are true are small subsets of those in which 'after' is true. The latter connectives are thus logically stronger than the former connectives. Because our model learns exclusively from positive data, the subset problem introduces the worry that our learner will fail to distinguish the meanings of 'until' from 'before' and 'since' and 'while' from 'after.' However, the setup of our likelihood, which implements the size principle (Tenenbaum, 1999), solves this problem by preferring meanings that are true in fewer contexts. Our model is biased to learn the correct and logically stronger meanings for 'until,' 'since,' and 'while' because these meanings assign higher likelihood to observed utterances.

One limitation of our model is its reliance on a uniform sample from true meanings in the likelihood specified by Equation 5. The model of quantifier learning in Piantadosi et al. (2012) integrated a production probability based on utterance informativeness into its likelihood. These production probabilities were computed over a large simulated dataset in support of a Gricean pragmatic framework. By weighing utterances according to their informativeness, the learning model will be biased towards more specific hypotheses, reflecting the presuppositional assumption that speakers will tend to utter the most specific word that applies to a context.

Our model learns in an idealized setup that diverges in important respects from the situation children encounter. Our learning framework was intentionally designed to assess whether previously reported patterns of acquisition would emerge in a statistical learner under ideal conditions; having determined that these patterns do emerge from the formal learning problem, we now discuss several factors beyond the current scope of the model that affect children's learning. Firstly, unlike our model, real-world learners must contextually disambiguate the temporal meanings of the connective words from non-temporal alternatives (e.g., the spatial meaning of 'before'). Secondly, our model maintains perfect memory of previously encountered training examples, whereas children are likely to focus their learning on recently experienced episodes. Thirdly, our model is not subject to psycholinguistic processing constraints that have previously been suggested to influence temporal connective acquisition, like working memory limits during sentence parsing (Blything et al., 2015; Clark, 1971). Future work could expand upon our study by building each of these limitations into computational models and evaluating their impact on word learning.

Acknowledgements

We thank Terry Regier for helpful comments on this work.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843.

Anscombe, G. E. M. (1964). Before and after. *Philosophical Review*, *73*(1), 3–24.

Beaver, D., & Condoravdi, C. (2013). A uniform analysis of 'before' and 'after'. In *Proceedings of the 13th semantics and linguistic theory conference* (pp. 37–54).

Bennett, M., & Partee, B. H. (2004). Toward the logic of tense and aspect in English. In B. H. Partee (Ed.), *Compositionality in formal semantics* (pp. 59–109). Malden, MA, USA: Blackwell Publishing Ltd.

van Benthem, J. (1991). *The logic of time : A model-theoretic investigation into the varieties of temporal ontology and temporal discourse* (2nd ed.). Dordrecht: Kluwer Academic Publishers.

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922–1934.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266–275.

Condoravdi, C. (2010). NPI licensing in temporal clauses. *Natural Language & Linguistic Theory*, *28*(4), 877–910.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing, and modularity. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 443–467). Hillsdale, NJ: Lawrence Erlbaum Associates.

Enderton, H. (2001). *A mathematical introduction to logic* (2nd ed.). Boston: Academic Press.

Feagans, L. (1980). Children's understanding of some temporal terms denoting order, duration, and simultaneity. *Journal of Psycholinguistic Research*, *9*(1), 41–57.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

von Fintel, K., & Matthewson, L. (2008). Universals in semantics. *The Linguistic Review*, *25*(1-2), 139–201.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, *32*(1), 108–154.

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2014). Concepts in a probabilistic language of thought. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 59–109). Cambridge, MA: MIT Press.

Goranko, V., & Rumberg, A. (2020). Temporal logic. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2020 ed.). Metaphysics Research Lab, Stanford University.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 3.

Kamp, J. A. W. (1968). *Tense logic and the theory of linear order*. Doctoral dissertation, UCLA.

Monens, M. (1971). *Tense, aspect and temporal reference*. Doctoral dissertation, University of Edinburgh.

Piantadosi, S. T. (2011). *Learning and the language of thought*. Doctoral dissertation, MIT.

Piantadosi, S. T., & Jacobs, R. A. (2016). Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science*, *25*(1), 54–59.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). *Modeling the acquisition of quantifier semantics: A case study in function word learnability.* (Unpublished manuscript)

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, *123*(4), 392–424.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In *Advances in neural information processing systems* (pp. 59–68).

Wexler, K., & Manzini, M. R. (1987). Parameters and learnability in binding theory. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 41–76). Dordrecht: Springer Netherlands.

Winskel, H. (2003). The acquisition of temporal event sequencing: A cross-linguistic study using an elicited imitation task. *First Language*, *23*(1), 65–95.

Zwarts, F. (1995). Nonveridical contexts. *Linguistic Analysis*, *25*(3-4), 286 - 312.