COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2017) 1–24

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12567

Bayesian Word Learning in Multiple Language Environments

Benjamin D. Zinszer,[a] Sebi V. Rolotti,[b] Fan Li,[c] Ping Li[d]

[a]*Department of Communication Sciences and Disorders, University of Texas at Austin*
[b]*Department of Neuroscience, Columbia University*
[c]*Department of Biostatistics and Bioinformatics, Duke University*
[d]*Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University*

Received 20 January 2017; received in revised form 28 August 2017; accepted 9 October 2017

Abstract

Infant language learners are faced with the difficult inductive problem of determining how new words map to novel or known objects in their environment. Bayesian inference models have been successful at using the sparse information available in natural child-directed speech to build candidate lexicons and infer speakers' referential intentions. We begin by asking how a Bayesian model optimized for monolingual input (the Intentional Model; Frank et al., 2009) generalizes to new monolingual or bilingual corpora and find that, especially in the case of the bilingual input, the model shows a significant decrease in performance. In the next experiment, we propose the ME Model, a modified Bayesian model, which approximates infants' mutual exclusivity bias to support the differential demands of monolingual and bilingual learning situations. The extended model is assessed using the same corpora of real child-directed speech, showing that its performance is more robust against varying input and less dependent than the Intentional Model on optimization of its parsimony parameter. We argue that both monolingual and bilingual demands on word learning are important considerations for a computational model, as they can yield significantly different results than when only one such context is considered.

Keywords: Language acquisition; Word learning; Bilingualism; Bayesian modeling

1. Introduction

Language learners face the difficult inferential problem of mapping words they hear to the referents in their environment intended by a speaker. Developmental research in the last two decades has identified a number of potential constraints and biases that children may bring to bear on language learning to mitigate this problem, and subsequently many

Correspondence should be sent to Benjamin D. Zinszer, Department of Communication Sciences and Disorders, University of Texas at Austin, Austin, TX 78712. E-mail: bzinszer@gmail.com

computational simulations have been proposed to describe word learning. However, many early models were limited to learning from highly constrained input or, conversely, allowing many-to-many mappings in models' learned lexicons (i.e., the set of mappings between word forms and referents inferred from the training data). Parsimony and ecologically relevant learning have therefore gained increasing importance for word learning models, and several models are now making progress toward producing highly plausible predictions based on naturalistic training materials (Fazly et al., 2010; Frank, Goodman, & Tenenbaum, 2009; McMurray et al., 2012).

One such example, the Intentional Model of Frank et al. (2009) offered a major improvement in the simulation of infant word learning through Bayesian inference (see Xu & Tenenbaum, 2007b, for a review of Bayesian models in word learning). The Intentional Model uses a conservative assumption for assigning prior probabilities to lexicons based on parsimony; that is, learning fewer mappings is generally better than learning more. This assumption may reflect the child learner's own bias toward parsimony in word learning. The Intentional Model's success over alternate models of word learning suggests that parsimony is indeed a key constraint. However, the Intentional Model's specific constraint on the *total* size of a lexicon could be problematic for learning from bilingual input wherein approximately twice as many words exist for the same set of objects. To date, the Intentional Model has been tested only with monolingual corpora of child-directed speech, leaving open the question of whether it would provide reasonable predictions in a bilingual language environment. In this paper, we examine the performance of the Intentional Model when presented with bilingual input and propose a modestly revised approach to parsimony that, we argue, better accommodates new learning contexts, such as in simultaneous bilingual language acquisition.

1.1. Parsimony in word learning

Confronted with a multiword utterance in a natural environment with many objects, how does a learner locate the intended referent and map it to the correct word? This has been named the "word-to-world" mapping problem in the literature (the classical "indeterminacy problem" as per Quine, 1960; but also see Markman, 1990, 1994; Smith & Yu, 2008; among others). Developmental researchers have proposed a number of heuristics that a child may be using to reduce the dimensionality of the word-by-object mapping problem, thereby making word learning a psychologically tractable task that both children (Liitschwager & Markman, 1994) and adults (Yurovsky & Yu, 2008) can solve quickly and robustly in evidence-poor natural learning scenarios.

Mutual exclusivity is one such constraining assumption that both provides important clues for word–object mapping but poses a serious difficulty for children in many natural language environments. The mutual exclusivity constraint stipulates that if the name of an object is already known, a new name should be applied to a novel object instead of the already named object. Frank et al.'s (2009) Intentional Model exhibits a mutual exclusivity preference resulting from its far more general parsimony constraint (discussed in detail below). Thus, this constraint for mutual exclusivity appears to be important in word learning, and it has plausible analog in a successful word learning architecture. However, this success might be partly attributed to an implicit assumption of many word learning studies and models: monolingualism.

Monolingual children confront violations of the mutual exclusivity constraint to a limited degree with synonymy, but bilingual children are forced to violate mutual exclusivity for every referent they learn to name in both languages (Au & Glusman, 1990; Davidson & Tell, 2005). Empirical studies of child bilingualism have indicated that bilingual children relax the mutual exclusivity criterion (Byers-Heinlein & Werker, 2009; Houston-Price, Caloghiris, & Raviglione, 2010) once they reach a metalinguistic awareness that multiple names (translation equivalents) must be assigned to the same object (Byers-Heinlein & Werker, 2013).

Bilingualism is a norm rather than exception among the world's language users (Grosjean, 2010), but computational models of word learning have made limited progress in ecologically plausible bilingual word learning. Analogous hierarchical learning problems, wherein the very assumptions that support learning must be periodically reevaluated for their fitness to the data, are being addressed through the Bayesian approach (Qian, Jaeger, & Aslin, 2012). The next crucial step for word learning models, therefore, is to understand how the state of the art in monolingual learning generalizes to bilingualism.

1.2. The Intentional Model

Bayesian models have emerged in the past decade as an important class of word learning models that incorporate some basic constraints in the model's learning procedure. Bayesian models have significantly improved the psychological plausibility of early language acquisition and replicated several of the benchmarks set out by Xu and Tenenbaum (2007a) as requisite features for simulations of word learning. Indeed, there is accumulating behavioral evidence that inductive problem solving of the sort faced by children in word learning is Bayesian in character (Bonawitz & Griffiths, 2010; Qian et al., 2012; Xu & Tenenbaum, 2007a,b). Bayesian word learning models rely on relatively simple statistical biases to evaluate lexicons prior to observing data (the prior probability) and in comparison to the observed data (the likelihood). Each possible lexicon is treated as a competing hypothesis, compared for its fitness to the observed data, and assigned posterior probability or a quantification of the model's belief in that particular hypothesis (lexicon) given the new observations (Perfors, Tenenbaum, Griffiths, & Xu, 2011). By this method, no hypothesis is technically eliminated, but decreasing probabilities are assigned to the more unlikely hypotheses.

This relationship can be expressed mathematically using Bayes' theorem:

$$p(H_i|D) = \frac{p(D|H_i)p(H_i)}{\Sigma_j p(D|H_j)p(H_j)} \tag{1}$$

Here $p(D|H_i)$, or the likelihood probability (described in greater detail in Experiment 1), represents the probability of observing these data given the hypothesis H_i while $p(H_i)$ is the prior probability of hypothesis i. The better the fit to the observed data *and* the more likely (a priori) the hypothesis, the higher the posterior probability $p(H_i|D)$ that hypothesis i is the correct one. The denominator in the above expression is a normalization factor, ensuring that the posterior probabilities for all hypotheses sum to one. While this expression represents an evaluation of the hypotheses, the manner in which the likelihood and prior are calculated, as well as how a specific lexicon is generated and chosen from the hypothesis space in the first place, is model-specific.

Given the considerable power and success of the Bayesian Intentional Model (Frank et al., 2009) compared to other computational models of word learning, we focus in this study on the Intentional Model as our initial basis for simulating children's word learning from natural language corpora. Although the Intentional Model has not directly incorporated mutual exclusivity as a learning bias or mechanism, the model successfully reproduces a strong preference for mutual exclusivity when confronted with the decision to associate a new word with either an object whose name is known or a novel object (Frank et al., 2009). Parsimony in the total number of mappings is implemented in this model by biasing the prior probability estimates against larger lexicons. In effect, this bias discourages adding new word–object pairs to the lexicon unless they significantly improve the lexicon's fitness to the data (see Frank et al., 2009, for more details).

While achieving plausible lexicons and demonstrating a mutual exclusivity bias under a rather small set of constraints is an important step for computational models of child language acquisition, the success of the Intentional Model in this regard may partly hinge on the use of monolingual training data, which has fewer violations of parsimony assumptions than analogous bilingual or multilingual input. In other words, can the model learn a lexicon containing two words for every object? This is the challenge that every bilingual child faces. Frank et al. (2009) demonstrate that in some circumstances, the Intentional Model succeeds in learning synonymous relationships (e.g., *dog* and *doggie*), but such examples are few and are systematically disadvantaged by their model. In addition, their model is optimized to learn from a particular corpus of monolingual child-directed speech (videos me06 and di03 of the Rollins data in CHILDES; MacWhinney, 2000), which allows the model to be optimized for a single parsimony parameter controlling the level of bias against large lexicons. It remains unclear how this optimization process may impact the Intentional Model's application to a different monolingual corpus or whether the procedure can be generalized to bilingual input.

1.3. The present study

Bayesian word learning models have so far remained confined to the domain of monolingual acquisition. Despite the limited success of such models, including the Intentional Model, in accounting for mutual exclusivity, Bayesian word learning models have thus far not considered how these constraints might be applied in bilingual language environments. How does the Bayesian learning framework such as that of Frank et al. (2009)

generalize to these language environments? What should a model look like in order to simulate the child's behavior in both monolingual and bilingual learning contexts?

In the simulations reported in this study, we first test the utility of the Intentional Model of Frank et al. (2009) for new training data: (a) a new monolingual corpus, similar to Frank et al.'s original training corpus, and (b) two bilingual input corpora, based on translations of the original data and the new monolingual corpus, respectively, in which the number of word types per object increases as a result of the new language. We predict that a Bayesian inference model designed with the assumption that adding new words to the lexicon is less parsimonious than retaining an existing lexicon will perform more poorly in the bilingual conditions than in the monolingual condition because the target lexicon in the bilingual input condition is highly incompatible with this prior assumption. We further investigate how bilingual learning compares with the learning of each of the counterpart monolingual corpora, and to what extent the optimization of the parsimony parameter (the model's prior term) accommodates these new training sets.

Next, we propose a revised algorithm for the computation of the prior probability through a more direct implementation of a mutual exclusivity constraint by penalizing the number of words that are already mapped to given objects rather than penalizing the lexicon size as a whole (the latter of which was adopted in Frank et al., 2009). We investigate how optimization of the two models' parsimony parameters accommodates these new training sets. We ask how the demands of word learning are best met across diverse circumstances; specifically, whether a single unique optimization may satisfy all of the above conditions, or whether different input environments require different parameter adaptations in the learning algorithm for each condition. Finally, we evaluate these possibilities and each model's performance under different conditions against known empirical evidence.

2. Experiment 1

2.1. Model and training

The Intentional Model learns word–object mappings based on corpora of child-directed speech and object presentations transcribed from parent–infant interactions in a laboratory. The specifics of the model can be found in its original presentation in Frank et al. (2009), but for clarity and completeness, we summarize the basics of the model here. The model repeatedly generates hypotheses for lexicons by proposing incremental changes (i.e., adding, removing, or swapping word–object mappings), which are subsequently scored according to their posterior probability. The posterior probability for each lexicon is computed according to Bayes' rule by multiplying the prior and likelihood probabilities: $p(L|C) \propto p(\mathrm{L}) \times p(C|L)$. The lexicon space is searched using a simulated tempering strategy whereby a number of searches with differing degrees of greediness are run in parallel. The model's search and scoring process proceeds for 50,000 moves, Frank and colleagues' upper estimate for the number of moves required for the model to converge on a solution (2009, supplemental materials).

The prior probability for a lexicon is calculated according to a parsimony assumption, awarding each lexicon L_i a score inversely proportional to its size (number of mappings):

$$p(L_i) \propto e^{-\alpha|L_i|} \tag{2}$$

The parsimony parameter α controls the degree to which the model biases lexicons toward smaller total sizes. The value for this parameter ($\alpha = 7$) was selected by Frank et al. to produce optimal performance on their training corpus. We retain that value in Experiment 1, applying it to both the original corpus and a new corpus.

The likelihood function, which calculates the probability $p(C|L_i)$ of observing the corpus C of situations given a lexicon, is based on a number of interdependencies and assumptions. These assumptions include the following: (a) For the objects O_s, intentions I_s, and words W_s in each situation S, I_s is a subset of O_s, and every subset is equally likely to be intentionally referred to, that is, $p(I_s|O_s) \propto 1$; (b) given I_s, a speaker's utterance W_s depends upon both I_s and the lexicon L; (c) speakers have a certain probability γ of using a word referentially in any given context. In addition to these assumptions, we consider two distinct probabilities: first, the probability $p_R(w|o, L_i)$ of choosing a word $w \in W_s$ uniformly at random from the set of valid labels to refer to a given object $o \in O_s$ with lexicon L_i, and second, the probability $p_{NR}(w|L_i)$ of choosing a word to be used non-referentially. A parameter κ dictates how likely words in the lexicon are to be used non-referentially relative to words outside the lexicon (i.e., because we choose $\kappa < 1$, words in the lexicon are less likely to be used non-referentially). The Intentional Model's parameters γ and κ are set to the maximum a posteriori values estimated by Frank et al. (2009; that is, 0.1 and 0.05, respectively), which we do not expect to greatly differ across languages or new (but similarly structured) corpora. The final likelihood probability is thus defined:

$$p(C|L_i) = \Pi_{S\epsilon C}\Sigma_{I_S\subseteq O_S}\Pi_{W\epsilon W_S}\left[\gamma \cdot \Sigma_{o\epsilon I_S}\frac{1}{|I_S|}p_R(W|O,L_i) + (1-\gamma)\cdot p_{NR}(W|Li)\right] \tag{3}$$

After training, the model is scored both on the accuracy of its lexicon and on the accuracy of the inferences it makes about speakers' referential intentions given this lexicon. These scores are measured relative to a gold standard lexicon and intention set generated by a human coder. The gold standard lexicon includes every noun (including plurals and baby talk, excluding pronouns) used to refer to an object at least once in the data. The gold standard intents were based on the speakers' referents in Fernald and Morikawa's (1993) videos of mother–child interaction (see Frank, Tenenbaum, & Fernald, 2013, for further details about coding). The measures of accuracy used were *precision* (proportion of mappings made that were correct), *recall* (proportion of the total gold standard mappings that were found), and *F* score (the harmonic mean of precision and recall, commonly used as a standard measure of a model's degree of accuracy). *F* scores were calculated for both the lexicon (mappings of words to objects) and the intents (selection of referent objects for each situation).

Finally, we also compared the Intentional Model to the IBM Machine Translation Model I (Brown, Della Pietra, Della Pietra, & Mercer, 1993). This model provided the best performance among the associative probability-based approaches used for comparison in Frank et al.'s (2009) analysis. The Translation Model computes association probabilities both for objects given words and for words given objects. After calculating a word-by-object matrix of association values, the model compares a number of lexicons created at different probability threshold values, retaining only word–object pairs with an association higher than the threshold. The lexicon resulting from the threshold value that yielded the highest posterior score was kept for each model.

2.2. Material

In the present experiment, the Intentional Model was tested on four different datasets, two monolingual corpora and two bilingual translations of the same corpora. The first training corpus, drawn directly from Frank et al.'s (2009) study, was entirely in English. To create the corresponding bilingual corpus, a native speaker of Spanish translated approximately 50% of the situations in the monolingual corpus into Spanish, recreating the child-directed style of speech that characterized the English utterances and providing a roughly balanced input of English and Spanish situations to simulate the bilingual environment. Transparency (i.e., the translation's Spanish native-likeness) was prioritized over fidelity (i.e., the extent to which the translation exactly renders the meaning of the English) whenever possible. Besides language differences, the resulting bilingual corpus is similar to the monolingual corpus, as illustrated in Table 1. The most important difference between the corpora is the larger number of word types in the bilingual input, a result that is expected from the use of two languages and by extension the regular use of at least two different word types to indicate the same object.

A second monolingual dataset was generated by drawing from an additional set of annotated transcriptions of English-speaking mothers interacting with their infants from

Table 1
Statistics describing each corpus

	Corpus 1 Monolingual	Corpus 1 Bilingual	Corpus 2 Monolingual	Corpus 2 Bilingual
Corpus information				
Object types	22	22	22	22
Word types	419	629	321	486
Mean objects/situation	2.04	2.04	2.93	2.93
Total situations	619	619	571	571
Gold standard lexicon				
No. of mappings	34	50	34	49
Mean words per object	1.94	2.88	1.57	2.29

Note. Corpus information includes type counts for words and objects in the corpora. Each corpus also has a gold standard lexicon, against which lexicons learned by the model are compared.

Fernald and Morikawa (1993); these transcriptions are similar to the transcriptions used to generate the first monolingual corpus, providing a second set of training data comparable in size although having a slight increase in the number of objects for each situation (M = 2.93 objects/situation). The bilingual version of this corpus was generated by the same method as the previous corpus, which involved a random selection and translation of approximately 50% of the monolingual input (see Table 1 for details).

2.3. Results

The Intentional Model was run for 50,000 iterations in each condition to yield five estimated lexicons from the simulated tempering process. Per the procedure of Frank et al. (2009), the best lexicon was selected among these five to represent the outcome in each condition. For each model, we generated F scores for Lexicon (word–object mappings) and Intents (inference about the referent, if any, in each situation) based on both the precision and recall for each measure. Additionally, we report the model's complete lexicon (set of mappings) in each condition.

Comparing across the two monolingual corpora, the model performed better in Corpus 1, for which its learning parameters were previously optimized. Despite producing similarly sized lexicons for both corpora (Corpus 1: 24 mappings, Corpus 2: 25 mappings), the Lexicon and Intents were much less accurate for Corpus 2 (see Table 2). The Intentional Model's performance sharply declined with bilingual input, with an F score decrease of 0.17 relative to monolingual input in Corpus 1 and a 0.13 decrease in Corpus 2. Correspondingly, the intentional inference (Intents) also decreased slightly (Corpus 1: 0.13, Corpus 2: 0.10) between the monolingual and bilingual versions of the corpus.

The Intentional Model outperformed the Translation Model in Lexicon accuracy for all corpora. In the two bilingual corpora, the Translation Model yielded higher scores for its

Table 2
F scores and lexicon size achieved by Intentional Model and Translation Model in each condition.

Corpus	Language	Model	Lexicon (F)	Intents (F)	Lexicon Size
Corpus 1	Monolingual	Bayes	**0.48**	**0.55**	24
		Translation (w|o)	0.11	0.46	167
		Translation (o|w)	0.10	0.45	103
	Bilingual	Bayes	**0.31**	0.42	20
		Translation (w|o)	0.10	0.39	107
		Translation (o|w)	0.09	**0.53**	584
Corpus 2	Monolingual	Bayes	**0.34**	**0.31**	25
		Translation (w|o)	0.17	0.30	14
		Translation (o|w)	0.07	0.25	492
	Bilingual	Bayes	**0.21**	0.21	27
		Translation (w|o)	0.14	**0.36**	65
		Translation (o|w)	0.07	0.18	285

Note. F scores are the harmonic mean of precision and recall. Max F scores in each condition are marked in bold.

inferences about speakers' referential intentions (Intents). The translation Model's lexicons were much larger than the Intentional Model's lexicons in almost all circumstances. In one exception, for the monolingual version of Corpus 2, the Translation ($w|o$) lexicon was considerably smaller (14 mappings vs. 25); however, this small lexicon also produced far lower performance in both Lexicon and Intents (see Table 2).

In the Appendix, we report the Intentional Model's best lexicon from each condition to illustrate its ability to associate names provided in the input with the presented objects. The lexicons include correct mappings (such as "bunnies" and the object *bunny*), situationally related (but rated as incorrect) mappings (e.g., "red" and *truck*), and spurious mappings (e.g., "ruff" and *pig*). In general, models in the monolingual corpora produced many more correct mappings than models in the bilingual corpora. With bilingual input, the model learned words in each language, although it tended to learn either an English or a Spanish name for an object but rarely both.

2.4. Discussion

Experiment 1 demonstrated that the Intentional Model, as proposed by Frank et al. (2009), performed better overall than a competing associative model on three new corpora. This advantage supports the overall usefulness of the Bayesian inference approach to understanding lexical learning in young children. However, the large performance deficits for the new corpora relative to the original corpus (Corpus 1 Monolingual) suggest room for improvement in simulating real child learners.

The high performance achieved by the Intentional Model on its original training corpus was considerably reduced when faced with another monolingual (and very similar) corpus. Corpus 2 mapped a similar number of words and objects as Corpus 1, learned across a similar number of situations. Critically, the model's high performance in Corpus 1 was, in part, due to the maximum a posteriori optimization of its training parameters, which we expected to readily generalize to a new, similar corpus. The parameters γ and κ describe more general properties of the language input (likelihood of null references and likelihood of referential expressions, respectively) and thus seem unlikely causes of the differences between conditions. However, the parsimony parameter α is more directly linked to the differences between the training corpora, which differed in the number of words assigned per object. We explore this issue further below.

Two differences between these monolingual corpora may have contributed to a change in performance. First, in each situation, Corpus 2 presented more objects on average (Corpus 1: 2.04 objects/situation, Corpus 2: 2.93), marginally increasing the difficulty of the intentional inference and thus the word–object mapping. Second, the gold standard lexicon for Corpus 2 Monolingual mapped fewer words to each unique object (1.57 words/object) than Corpus 1 Monolingual (1.94 words/object). In this case, the Intentional Model's parsimony bias may not have been strict enough to optimize learning of Corpus 2. This difference highlights a potential weakness of the Intentional Model. Specifically, the model's performance might rely on optimizing the value of the parsimony parameter, even between two very similar corpora of child-directed speech. We explore this issue in the next experiment.

Comparing the bilingual to monolingual versions of each corpus, the Intentional Model did not appear to overcome its parsimony constraints to learn the larger bilingual lexicons. The bilingual and monolingual lexicons learned by the model were similar in size, yielding far worse performance for bilingual lexical learning. This result is consistent with our prediction that bilingual input would systematically violate the model's preference for smaller lexicons and severely curb the addition of new words from each language. One possible explanation is that the bilingual corpora specifically challenged the model by presenting increased number of word types overall, thereby reducing the co-occurrence frequency between any given word–object pair. However, the model also did not distribute the mapped words (in either language) across objects as widely as it did for the monolingual corpora. Of the 18 objects in Corpus 2 Bilingual, five had multiple mappings, but none representing translation equivalents. Thus, on the one hand, the model was restricted by parsimony in its learning of two languages, but on the other hand, this parsimony did not prevent the model from making multiple spurious mappings to the same object instead of distributing words across the objects.

These results bring into focus the Intentional Model's specific mechanism for measuring parsimony in its estimate of prior probability. The formula for computing the prior probability of any given lexicon (see Eq. 2 above) relies on that lexicon's total size in mappings, not the number of names any given object is assigned. By this method, a lexicon that adds three new names for the same object has an equal prior probability to a lexicon that adds one new name for each of three objects. This constraint, as it is implemented in the current Intentional Model, differs considerably from the mutual exclusivity constraint as proposed by Markman (1990, 1994) or its Bayesian implementation proposed by Tenenbaum and Xu (2000). However, while the current prior does not explicitly bias proposed lexicons toward one-to-one word–object mappings, the model strongly prefers one-to-one mappings and displays mutual exclusivity-like behavior when the model is applied to naturalistic input (Frank et al., 2009). Thus, the Intentional Model's approach has the advantage of simplicity, and when it is applied to a corpus that favors one-to-one word–object mappings it indeed succeeds in simulating the experimental results of a mutual exclusivity task. However, given bilingual input, learning new names for a given object and learning a name for a new object is an important distinction, and the Intentional Model offers no apparent means of adapting to this type of input. It is an empirical question whether the model can ever accommodate these violations, which the next experiment explores.

In Experiment 2, we address two questions: How does the Intentional Model's lexical learning improve when the parsimony parameter (α) is freed to optimize for a given corpus, and how could a *different* approach to parsimony better accommodate both monolingual and bilingual input. To this end, we propose and test a direct implementation of the mutual exclusivity constraint in a revised model. Simulations in Experiment 2 search the parsimony parameter space of the Intentional Model and this revised model side-by-side. We ask whether a mutual exclusivity term (hereafter, the *ME* parameter) proves more robust to variation across training corpora than the Intentional Model's analogous parameter (α).

3. Experiment 2

Given the limitations of the Intentional Model's lexical learning in Experiment 1, we hypothesized that adjusting the parsimony parameter for the Intentional Model may improve its performance for another monolingual lexicon. Can the Intentional Model be optimized for a bilingual corpus? Learning translation-equivalent words for objects might require drastically reducing the model's parsimony, or this task could be outside the model's capability at any degree of parsimony. If the latter case proves true, a revised model that changes the calculation of the prior probability might better accommodate input from either monolingual or bilingual environments.

The revised model described below attempts one such approach: We shift the focus of parsimony from minimizing the size of the lexicon to minimizing the number of words that are mapped to the same object. This new method computes the prior probability of a lexicon as inversely proportional to the average number of words mapped to each object. In this way, the model does not penalize the lexicon for mapping a new word to any new, unnamed object. When adding names to objects that have already been mapped to another name in the lexicon, mapping the word to an object with the fewest number of names already maximizes the prior probability. While this constraint still disadvantages learning of translation-equivalent words, it may be a more reasonable representation of the prior assumption children must overcome to learn two languages. We expect this approach to more equitably distribute words over objects rather than learning several spurious names for relatively few objects, particularly when the model is making inferences from noisier data (i.e., fewer co-occurrences per mapping). This reduction in spurious mappings indirectly supports learning alternate correct names for objects, which were previously rejected due to an over-conservative preference for minimizing lexicon size. While we search for the values of the free-parameter α that maximize the Intentional Model's performance for all four lexicons, we also search the values of the new mutual exclusivity parameter in the revised model to explore its effect on learning.

3.1. Model and training

In the Intentional Model, prior probabilities are inversely proportional to the size of the lexicon, such that lexicons with a large number of word–object pairings are penalized with lower prior probabilities, regardless of the distribution of words per object. As illustrated by the results of Experiment 1, this was found to be problematic for learning corpora other than the original corpus for which the model was optimized: with only a modest change between the monolingual corpora, the model's performance scores dropped significantly, and the best lexicons for bilingual corpora were constrained to the same size as those of the monolingual lexicons, despite the bilingual lexicon's need to map a greater number of words to each object.

In the revised model, hereafter the ME Model, we compute the prior probability as inversely proportional to the ratio of the number of words in the lexicon and the number of objects named in the lexicon, or the average words per object in the lexicon:

$$p(L_i) \propto e^{-(\alpha \times L) \times ME \times (|words|/|objects|)} \tag{4}$$

Under this revised definition, L is a constant defined by the size of the gold standard lexicon for Corpus 1 Monolingual ($L = 34$), hence the absence of iteration (formerly L_i) in the new exponential term. This constant makes no assumption about the learner's metalinguistic knowledge about lexicon size, but instead it is set to allow comparison between the Intentional Model described in Experiment 1 and the present ME Model. Thus, when α is held at 7 (as in Experiment 1) and ME is set to 1, the prior for a lexicon with only one-to-one mappings will be weighted exactly the same as a target (gold standard) lexicon for Corpus 1 Monolingual. However, adding and subtracting new mappings will affect the prior differently than the Intentional Model, giving the ME Model a prior bias toward lower word-to-object ratios.

When the number of words in the lexicon matches or exceeds the number of objects, the ME Model does not penalize the lexicon for adding any new, unnamed object but rewards it by decreasing the mean labels per object overall. This behavior contrasts with the Intentional Model in which the addition of new mappings, even correct ones, was penalized. As a consequence of minimizing the word-to-object ratio, in situations where the number of objects in a proposed lexicon exceeds the number of words, the ME Model penalizes the addition of new one-to-one mappings (which would increase the word-to-object ratio); however, this tendency is mitigated by the $p\langle C|L_i\rangle$ term, which weighs the fitness of the proposed lexicon to the corpus (i.e., the probability of observing the corpus, given the lexicon). Decreasing the value of ME allows the word–object ratio to increase. That is, a greater number of words may be mapped to each object with the same prior probability as the one-to-one lexicon would be assigned under $ME = 1$. In this way, the model can appropriately shift its prior weighting to bias lexicons in different language input environments.

Setting the reference of $\alpha = 7$ for comparison to $ME = 1$ allows us to compare the analogous parameter spaces for α and ME by looking at the effects from relative changes in each parameter. In this experiment, we search multiples up to two times ($\alpha = 14$, $ME = 2$) each parameter. An iteration of each model was run for multiples of 0.05 from 0.05 to 2. For the Intentional Model, this yielded values of α from 0.35 to 14. For simplicity, Lexicon F scores for each model are plotted over the space of 0 to 2, allowing direct comparison for analogous levels of parsimony (e.g., $\alpha = 6.30$, $ME = 0.90$).

We predicted that the ME Model would produce improvements over the Intentional Model when generalizing to new corpora by specifically penalizing multiple words per object mappings instead of new mappings added to the lexicon. However, we did not know whether learning in the bilingual corpora would improve in the ME Model, given the ME Model's preference for one-to-one mappings. The search of α and *ME* parameter spaces allowed us to explore the relationship between each model's parsimony bias and learning of various types of input.

3.2. Material

The same corpora from Experiment 1 were used in Experiment 2, and performance was again assessed using the *F* score (harmonic mean between precision and recall) for Lexicon and Intents.

3.3. Results

Table 3 describes the Intentional and ME Models' respective performance at levels of parsimony analogous to Experiment 1. Overall the two models were quite similar when parsimony is held constant (α = 7, *ME* = 1). In Corpus 1 Monolingual, the Intentional Model's lexicon scored considerably better than the ME Model's lexicon (Intentional Model: *F* = 0.48, ME Model: *F* = 0.43), but intentional inference was equally accurate in the two models (*F* = 0.55). In Corpus 1 Bilingual and both version of Corpus 2, the two models performed almost identically for this level of parsimony, producing similarly sized and similarly accurate lexicons.

3.3.1. Optimized performance

In Experiment 1, we speculated that the performance deficits of the Intentional Model for new input might be attributable to non-optimal parsimony, which had been set according to Corpus 1 Monolingual. In this experiment, we searched a wide range of values for the parsimony parameter in each model, ranging from 5% (0.05) to 200% (2.00) of the Experiment 1 baseline value. Table 4 describes each model's best Lexicon *F* scores over the entire search space for the α and *ME* parameters and the multiple of the baseline

Table 3
Mean *F* scores and lexicon size achieved by Intentional Model (Int.) and ME Model when respective parsimony parameters are held at analogous Experiment 1 level (α = 7, *ME* = 1)

Corpus	Language	Model	Lexicon (*F*)	Intents (*F*)	Lexicon Size
Corpus 1	Monolingual	Int.	0.48	0.55	24
		ME	0.43	0.55	26
	Bilingual	Int.	0.31	0.42	20
		ME	0.29	0.42	25
Corpus 2	Monolingual	Int.	0.34	0.31	25
		ME	0.33	0.30	26
	Bilingual	Int.	0.21	0.21	27
		ME	0.24	0.22	26

Table 4
Mean F scores and lexicon size achieved by Intentional Model (Int.) and ME Model at their optimized levels of parsimony within the space of 0.05–2 times baseline ($\alpha = 7$, $ME = 1$)

Corpus	Language	Model	Parameter	Lexicon (F)	Intents (F)	Lexicon Size
Corpus 1	Monolingual	Int.	$\alpha \times 0.90$	0.47	0.58	32
		ME	$ME \times 0.80$	0.44	0.55	29
	Bilingual	Int.	$\alpha \times 1.20$	0.33	0.42	17
		ME	$ME \times 0.60$	0.37	0.47	32
Corpus 2	Monolingual	Int.	$\alpha \times 0.80$	0.41	0.35	34
		ME	$ME \times 1.10$	0.47	0.36	25
	Bilingual	Int.	$\alpha \times 0.50$	0.28	0.30	65
		ME	$ME \times 0.40$	0.30	0.26	27

parameter value at which the best performance occurred. As in the previous analysis where parsimony was held constant, the Intentional Model's best performance slightly exceeded the ME Model for the original corpus (Corpus 1 Monolingual: Intentional Model Lexicon $F = 0.47$, ME Model Lexicon $F = 0.44$). However, when the optimal level of parsimony is selected for each model and each corpus, the ME Model's lexicons outperformed the Intentional Model's lexicons in three out of four corpora.

3.3.2. Performance across the parsimony parameter space

F scores yielded by the foregoing search present one important weakness: Very similar values of the parsimony parameter can produce highly variable results, especially as illustrated in Fig. 1C. Thus, the results of the search for an optimal parsimony parameter are likely to depend on other variations in models' performance, not attributable to the specific value of the parsimony parameter. By looking at the overall trend across the parameter space, we can make a more stable estimate of model performance at varying levels of parsimony. Fig. 1 depicts the observed Lexicon F scores for each model for the four corpora and the lowess-smoothed (kernel size = 4) trend lines describing these performance data. The values along this curve estimate the model's expected performance at a given parameter value. As in the observed F scores, the ME Model's smoothed estimates exceeded the Intentional Model on all corpora except Corpus 1 Monolingual. Table 5 lists the best estimated Lexicon F scores for each model and the parameter values at which they occurred. The F scores of the intents corresponding to each best lexicon are also listed, reflecting strong agreement between these two metrics.

We also asked how dependent each model was on finding this optimized value of the parsimony parameter for maximizing performance. The area under a lowess-smoothed curve measures that model's overall accuracy across all values in the search space. For example, a model that achieves a high F score at the optimal parameter setting but shows very poor performance at all other values of the parameter will have a smaller area under the curve than a model with moderately high performance at all values of the parameter. This pattern is evident in Fig. 1A, where the Intentional Model's best lexicons (around $\alpha \times 0.80$ to $\alpha \times 1.30$) are slightly better than the ME Model's best lexicons. However,

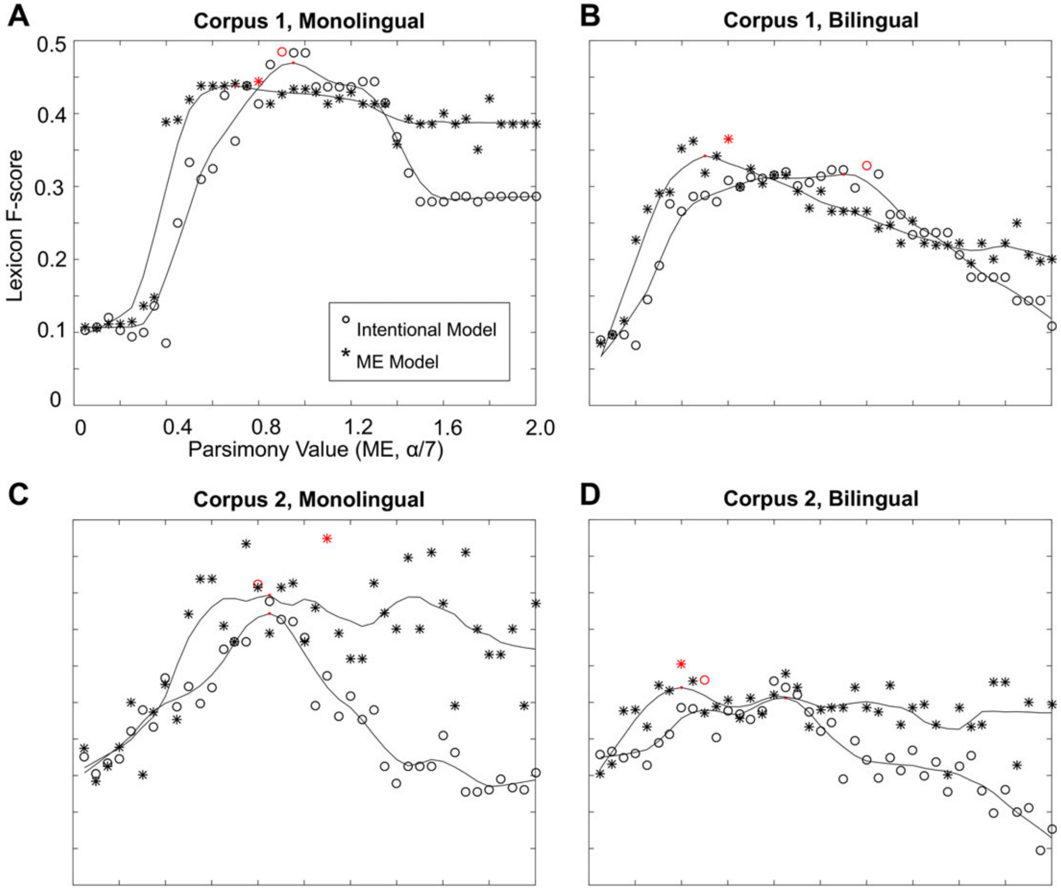

Fig. 1. *F* scores for the best lexicons produced by the Intentional and ME Models. Analogous parsimony values for each model are plotted together, where $ME = 1$ yields the same constraint as $\alpha = 7$, and values along the horizontal axis are multiples of these parameters. For each input corpus, the best observed *F* score is highlighted in red. Smoothed performance curves are depicted as continuous lines. Upper panels describe models' performance on Corpus 1 Monolingual (A) and Bilingual (B). Lower panels describe performance on Corpus 2 Monolingual (C) and Bilingual (D).

across the full range, the ME Model produces better lexicons overall (area = 0.70 vs. Intentional Model area = 0.61). This difference is further illustrated by the relatively constant performance of the ME Model from 0.50 to 2.0, while the Intentional Model has a clear maximum and tails with relatively low *F* scores.

3.4. Discussion

The ME Model we propose in this experiment displayed qualitatively similar performance to Frank et al.'s (2009) Intentional Model when held to a level of parsimony analogous to Experiment 1. The ME Model's revised prior probability formula is still

Table 5
Lexicon and corresponding intents F scores as well as area under the lexicon curve estimated for the Intentional Model (Int.) and ME Model based on the lowess-smoothed performance within the space of 0.05–2 times baseline ($\alpha = 7$, $ME = 1$)

Corpus	Language	Model	Parameter	Lexicon (F)	Intents (F)	Area
Corpus 1	Monolingual	Int.	$\alpha \times 0.95$	0.47	0.55	0.61
		ME	$ME \times 0.70$	0.44	0.54	0.70
	Bilingual	Int.	$\alpha \times 1.10$	0.32	0.41	0.46
		ME	$ME \times 0.50$	0.34	0.48	0.50
Corpus 2	Monolingual	Int.	$\alpha \times 0.85$	0.37	0.33	0.45
		ME	$ME \times 0.85$	0.40	0.29	0.66
	Bilingual	Int.	$\alpha \times 0.85$	0.26	0.27	0.36
		ME	$ME \times 0.40$	0.27	0.26	0.47

relatively conservative with regard to the mapping of multiple words to a given object, especially since the baseline parsimony parameter value was chosen for learning a monolingual corpus. The general similarity of results from each model reflects the choice of α and L to roughly match Experiment 1. This baseline test establishes the viability of applying parsimony at the mapping level rather than the whole lexicon level without severe detriment to model performance.

The ME Model's reliance on a preference for one-to-one mappings seems counterproductive to bilingual learning, but we reasoned that overcoming a prior bias to add one new name to each object based on specific evidence from the training corpus was a marginally smaller hurdle than overcoming a prior bias against adding twice as many words. Given the chance to adjust the *ME* parameter to the best level of constraint (as in the Intentional Model's optimization of α), the ME Model's learning of many-to-one word-to-object mappings can be improved in the bilingual corpora without losing its preference for new mappings to unnamed objects.

This experiment's results show that the ME Model yields better lexical learning than the Intentional Model across multiple levels of parsimony and when independently optimized. The gap between Corpus 1 Monolingual and Corpus 2 Monolingual narrowed with the ME Model for both the baseline parsimony ($ME = 1$) and the optimized values, signaling better generalization than the Intentional Model to new input. The ME Model's advantages for generalization were also evidenced by the overall results of the parameter space search. Across the search space for parsimony, the ME Model outperformed the Intentional Model for most parameter values. This finding confirms that our revised approach to parsimony generalizes to new language environments better than the Intentional Model, with or without adjustment of parsimony. Thus, when optimization is imperfect, not possible, or simply not an ecologically valid assumption about the learner, the ME Model is likely to produce better lexical learning for new data than the Intentional Model.

The search for the ME Model's optimal parsimony parameters for each corpus also confirmed our hypothesis that lexical learning was greatest with a weakened mutual exclusivity bias (especially under bilingual conditions) and that simulated learners in different word learning environments are best served by different degrees of this bias. Most strikingly, the values of the *ME* parameter that maximized Lexicon scores were mainly modulated by the number of languages (Bilingual: 0.50 and 0.40, Monolingual: 0.70 and 0.85), while the Intentional Model's optimal values of α were more closely related to the source corpus regardless of the language (Corpus 1: 0.95 and 1.10, Corpus 2: 0.85 and 0.85). In other words, a weaker mutual exclusivity bias (*ME*) best served lexical learning in all conditions, and decreased with the number of mappings to be learned, but the best lexicon size bias (α) depended on specific characteristics of the training corpora. Our simulation of mutual exclusivity bias is also consistent with the current behavioral literature and further explored in the General Discussion.

Interpretability is an important consideration for the ME Model, which could be improved upon in future research. Given the interpretive advantages of using valid generative models, additional work may focus on re-characterizing the ME Model's revised penalty prior along the lines of Johnson, Demuth, Frank, and Jones (2010) and Johnson, Demuth, and Frank's (2012) topic models to provide a generative, probabilistic interpretation for such penalties in bilingual acquisition.

4. General discussion

In this study, we set out to test the applicability of a Bayesian model of word learning to new language learning conditions, especially considering the large population of children who acquire two or more languages from infancy. For this evaluation, we used the Intentional Model of Frank et al. (2009) as a starting point because it has previously demonstrated significant advantages over several associative models and successfully simulated a number of empirical phenomena from early word learning in childhood. In addition to the training data used by Frank et al., we provided the model with two new learning conditions: (a) new monolingual input with similar characteristics to the Intentional Model's original training corpus and (b) bilingual input, in which approximately half of all training situations were in English and half in Spanish, roughly simulating the type of input a child would receive in a bilingual household. Both of these challenges improve the ecological validity by simulating variations in natural language environments faced by monolingual and bilingual children. We compared word learning (Lexicon) and referential intention (Intents) in the Intentional Model for all four corpora, and we introduced a revised version of the model (the ME Model), which improved lexical learning in three of the four conditions.

Experiment 1 revealed that a modest change to the Intentional Model's training data had seriously detrimental effects on the model's accuracy. When learning from the new monolingual corpus with slightly less information about correct mappings (more objects per situation, fewer presentations of each mapping), the model's Intents score dropped by nearly half. The Lexicon score also decreased by almost one-third, showing that this form of ambiguity seriously reduced the Intentional Model's ability to build accurate mappings between word types and objects. However, in comparison to the associative models (see Table 3), the Intentional Model still produced smaller, more parsimonious lexicons in both Monolingual conditions. The Bilingual versions of both corpora also yielded much lower scores than the Monolingual condition, particularly in the Lexicon test, but as in the Monolingual condition, the Intentional Model outperformed the IBM model on both the Lexicon *F* score and the relative parsimony of the best lexicon size. Thus, the general utility of a Bayesian approach to lexical learning is validated in Experiment 1.

Subsequently, in Experiment 2, we sought to leverage the success of the Intentional Model's Bayesian architecture in a revised model that more explicitly implemented children's mutual exclusivity (*ME*) biases, based on the important role that ME plays in early word learning (Markman, 1994). The ME Model's new prior probability term shifted focus from total lexicon size to the average number of mappings per object in the lexicon. When we compared the ME Model's performance to the Intentional Model using the same set of input corpora and a prior probability (as set by the *ME* parameter) roughly equivalent to that of the Intentional Model, the ME Model showed relatively little difference from the Intentional Model. The principal difference between the ME Model and Intentional Model, however, was the potentially unique effect of adjusting each model's parsimony parameter, a manipulation we addressed by searching the parameter spaces for each model in parallel. In this search, Experiment 2 also revealed an overall advantage for the ME Model in lexical learning and an ecologically plausible relationship between the optimal conservatism of the mutual exclusivity bias and the corpora variations that were generated for the preceding experiment. The results of Experiment 2 highlight the sensitivity of Bayesian models to the parsimony assumption, and demonstrate the flexibility of the ME Model to adapt to varying language environments.

The findings outlined thus far provide insights into several important directions: (a) the constraints which are thought to optimize models of monolingual learning may not directly generalize to all language environments, (b) it may be possible to successfully simulate both monolingual and bilingual developmental trajectories with a single model when performance data for each type of input are considered and balanced, and (c) an adequate model of language learning requires prior assumptions that are resilient to differing demands of variable language environments. In what follows, we provide an analysis of the Intentional and ME Models' performance across the four corpora against the extant empirical literature on early word learning.

One ostensible contributor to the performance of the Intentional and ME Models in the Bilingual conditions was the decreased number of exposures for every word–object mapping: Having a greater number of word types to map per object, each word type was matched with its respective object less frequently than in the Monolingual conditions. Although finding the optimal *ME* parameter produced a considerable improvement over the baseline parsimony in Corpus 1 Bilingual, even the best Lexicon performance in this condition (*F* score around 0.34) was considerably lower than achieved in the analogous Monolingual corpus. Optimizing the *ME* parameter for bilingual learning did not erase the bilingual deficit in either corpus, indicating that the model was not able to learn as quickly with approximately half the amount of input per language.

The effect of reduced input (per language) in the ME Model is commensurate with current views of frequency effects in bilingualism. The Frequency-Lag Hypothesis (Gollan, Montoya, Cera, & Sandoval, 2008) proposes that splitting a lifetime of language input between two languages results in lower subjective frequency of words for bilinguals relative to monolinguals, as evidenced by bilinguals' significant deficits in picture naming speed for low-frequency words relative to their monolingual counterparts. In the shorter term, bilingual children's receptive vocabularies, indeed, lag relative to their monolingual counterparts (Bialystok, Luk, Peets, & Yang, 2010), possibly attributable to relative deficits in input frequency. However, both of these effects are described in developmental terms, looking at cumulative learning over several months or years. The present simulation uses only several hundred utterances and a limited subset of potential referent objects to train the models, representing mere hours of parent–child interaction. One remaining question is the extent to which the short-term *inferential* deficits exhibited by the model would reflect longer term *developmental* outcomes, given a greater amount of training material. At this stage, our prediction based on the model is that in an experimental paradigm using a comparable training period of about six hundred referential utterances, children who receive bilingual input would acquire many fewer new words than their monolingual counterparts.

Numerous empirical studies have demonstrated the important role a mutual exclusivity bias plays in monolingual language acquisition (e.g., Halberda, 2003; Liittschwager & Markman, 1994; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003); however, most computational models have skirted this problem by limiting training data to one-to-one word–object pairs or producing large, unnaturalistic lexicons (see Frank et al., 2009, for a comparison of the Intentional Model against other major word learning models).

By varying the *ME* parameter, we demonstrated that the ME Model's implementation of bias at the level of words mapped per object makes better use of the sparse input than the Intentional Model. Thus, despite the reduced co-occurrence frequencies of word–object pairs in the Bilingual input and the increased number of competitor objects in Corpus 2, an optimally adjusted value of the *ME* parameter exists by which lexical learning is improved over baseline parsimony ($ME = 1$) and over the Intentional Model. This metalinguistic sensitivity to mutual exclusivity and adjustment of the bias has also been observed in children. Between 16 and 18 months, bilingual children significantly attenuate their mutual exclusivity bias compared to monolingual peers (Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010), and this shift in bias appears to be predicated on acquisition of translation equivalents between their two languages (Byers-Heinlein & Werker, 2013). In Bayesian terms, when learning two languages results in a sufficiently low posterior probability due to violations of mutual exclusivity (by recognizing translation equivalents), infants adjust their prior biases accordingly, loosening this constraint, or, in the case of multilinguals, abandoning it completely (Byers-Heinlein & Werker, 2009), which raises the posterior probability for their lexicon.

We observed that in both corpora, the model showed the best performance with a bias parameter in the prior that decreased inversely to the number of languages being learned. The ME Model's optimal *ME* parameters follow a pattern very similar to the 18-month-old infants in Byers-Heinlein and Werker's (2009) study. Bilingual infants' increases in looking time at novel objects were only 67% that of monolingual infants (Bilingual: 0.08 increase in the proportion of looking time, Monolingual: 0.12 increase). This ratio compares favorably with the degrees of bias that optimized the ME Model's lexical learning, wherein the best *ME* parameters for bilingual corpora were 50%–70% their monolingual equivalents.

Future modeling work should focus on extending the versatility of the Bayesian framework to accommodate a broader set of learning conditions and elaborating its implications for the underlying mechanisms and processes in word learning. The present ME Model retains the advantages of the Bayesian framework for inference and offers two novel insights: Bilingualism is a tractable problem under this architecture, and a weakened mutual exclusivity bias may optimize bilingual learning with only minimal loss for monolingual performance. The mechanism for the occurrence and timing of this metalinguistic awareness is, however, unspecified in the model. At present, we can only infer, based on the modeling data obtained, that *ME* serves as a key modulating variable in bilingual word learning in childhood, and children must make an additional inference about *which* learning context they are in and therefore adjust the mutual exclusivity bias accordingly. This problem of context detection has been addressed in statistical learning by Qian et al. (2012) through a framework of hierarchical Bayesian inference. The potential application of this hierarchical approach to word learning would involve selection of a monolingual or bilingual context and the consequent adjustment of the prior term.

Future studies may also adapt the present architecture to offer greater insight into longitudinal development patterns. At present, the model produces only cumulative results over whatever amount of input it is provided. A model which could leverage known (or believed) mappings learned previously toward the interpretation of new corpora with new mappings would offer researchers the ability to track simulated development patterns over time, with sensitivity to longitudinal effects such as the vocabulary spurt (Goldfield & Reznick, 1990; see Mayor & Plunkett, 2010; Li, Zhao, & MacWhinney, 2007, for existing models) or U-shaped developmental functions (Strauss, 1982).

Acknowledgments

We thank Elizabeth Carlson for assistance with simulations and data analysis, Michael C. Frank for assistance with the Intentional Model and the original training corpus and for valuable feedback on a previous version of this manuscript, and Janet Van Hell for service as SVR's summer mentor in 2012. Support to this project was made possible by a grant from the National Science Foundation (BCS-1057855).

References

Au, T., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: To honor or not to honor? *Child Development*, *61*(5), 1474–1490. https://doi.org/10.2307/1130757.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13* (4), 525–531. https://doi.org/10.1017/s1366728909990423

Bonawitz, E., & Griffiths, T. (2010). Deconfounding hypothesis generation and evaluation in Bayesian models. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 2260–2265). Austin, TX: Cognitive Science Society.

Brown, P., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*(2), 263–311.

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815–823. https://doi.org/10.1111/j.1467-7687.2009.00902.x.

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010.

Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, *92*(1), 25–45. https://doi.org/10.1016/j.jecp.2005.03.007.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063. doi: 10.1111/j.1551-6709.2010.01104.x.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.1111/j.1467-8624.1993.tb02933.x.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585. https://doi.org/10.1111/j.1467-9280.2009.02335.x.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183. https://doi.org/10.1017/S0305000900013167.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001.

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, 23–34. https://doi.org/10.1016/S0.

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009.x.

Johnson, M. H., Demuth, K., & Frank, M. C. (2012). Exploiting social information in grounded language learning via grammatical reductions. In *Proceedings of the Association for Computational Linguistics* (pp. 883–891). Stroudsburg, PA: Association for Computational Linguistics.

Johnson, M. H., Demuth, K., Frank, M. C., & Jones, B. K. (2010). Synergies in learning words and their referents. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 23, pp. 1018–1026). Red Hook, NY: Curran Associates, Inc.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*(4), 581–612. https://doi.org/10.1080/15326900701399905.

Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, *30*(6), 955–968. https://doi.org/10.1037/0012-1649.30.6.955.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77. https://doi.org/10.1016/0364-0213(90)90026-S.

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, *92*, 199–227. https://doi.org/10.1016/0024-3841(94)90342-5.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275. https://doi.org/10.1016/S0010-0285(03)00034-3.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1–31. https://doi.org/10.1037/a0018130.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. doi:10. 1037/a0029872.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00228.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010.

Strauss, S. (1982). *U-shaped behavioral growth*. New York: Academic Press.

Tenenbaum, J. B., & Xu, F. (2000). Word learning as Bayesian inference. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the Cognitive Science Society* (pp. 517–522). Ann Arbor, MI: Cognitive Science Society.

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288–297. https://doi.org/10.1111/j.1467-7687.2007.00590.x.

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in crosssituational statistical learning. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 22nd annual conference of the Cognitive Science Society* (pp. 517–522). Austin, TX: Cognitive Science Society. doi:10.1.1.218.7089

Appendix

Table A1
Experiment 1 best lexicons for Intentional Model

Corpus 1				Corpus 2			
Monolingual		Bilingual		Monolingual		Bilingual	
Word	Object	Word	Object	Word	Object	Word	Object
Bear	Bear	Hand	Hand	The	Cheese	Hotdog	Hotdog
Hiphop	Mirror	Davids	Lamb	Bear	Bear	Maza	Dough
Bottle	Bear	Gatito	Kitty	Blocks	Baby	Rosy	Rabbit
On	Ring	David	Book	Brush	Box	Gofres	Waffles
Lamb	Lamb	Sombrero	Hat	Hair	Brush	A	Face
Laugh	Cow	Hat	Hat	Brush	Brush	Brush	Brush
Bunnyrabbit	Bunny	Sheep	Sheep	Red	Truck	Ido	Book
Baby	Book	Bigbird	Bird	Waffles	Waffles	Ruff	Pig
Birdie	Duck	Bunnies	Mirror	Alphabet	Alphabet	You	Hotdog
Bird	Duck	Hiphop	Mirror	Ruff	Pig	You	Face
Ring	Ring	Meow	Baby	You	Face	Guau	Pig
Moocow	Cow	Bunnyrabbit	Bunny	The	Face	Esta	Cheese
Kittycat	Kitty	Mhmm	Hand	The	Hotdog	Esta	Face
Book	Book	Vaca	Cow	Hotdog	Hotdog	Doggy	Dog
Meow	Baby	Bottle	Bear	Flashlight	Flashlight	Car	Car
Bunnies	Mirror	Encontrar	Cow	Bang	Brush	Grande	Bear
Hand	Hand	Cerdo	Pig	Rosy	Doll	You	Box
Sheep	Sheep	Libro	Book	Happens	Blocks	Oh	Face
Pig	Pig	Abelardo	Bird	The	Dog	Cepilla	Box
Oink	Pig	Oink	Pig	Doggy	Dog	Bloques	Blocks
Mhmm	Hand			Cheese	Pepperoni	Queso	Pepperoni
Bigbird	Bird			Doors	Car	Bebe	Baby
Hat	Hat			Rabbit	Rabbit	Bang	Box
Put	Ring			Dough	Dough	The	Alphabet
				Joey	Book	The	Dog
						The	Face

Incorrect mappings are shaded.

Table A2
Experiment 2 best lexicons for ME Model

Corpus 1				Corpus 2			
Monolingual		Bilingual		Monolingual		Bilingual	
Word	Object	Word	Object	Word	Object	Word	Object
Laugh	Cow	Who	Boy	Bang	Box	Ruff	Pig
Dada	Woman	Papa	Woman	Blocks	Blocks	Guau	Pig
Set	Face	And	Book	Rabbit	Rabbit	Queso	Pepperoni
Bottle	Bear	Sombrero	Hat	Doors	Car	Hotdog	Hotdog
Bear	Bear	Viejito	Man	Dough	Dough	The	Hotdog
Lamb	Lamb	Bird	Duck	Baby	Baby	Bang	Brush
Sheep	Sheep	Bunnyrabbit	Bunny	Waffles	Waffles	Gofres	Waffles
Ring	Ring	Hat	Hat	Rosy	Doll	Cepilla	Box
Book	Book	Hiphop	Mirror	Brush	Brush	Brush	Brush
Bunnyrabbit	Mirror	Bunnies	Mirror	Oscar	Oscar	Brush	Box
Mommy	Man	Birdie	Duck	Red	Truck	Doggy	Dog
Put	Ring	Libro	Book	Alphabet	Alphabet	The	Face
On	Ring	On	Ring	The	Cheese	Oh	Face
Baby	Rattle	Bigbird	Bird	The	Face	Joey	Book
Looking	Eyes	Meow	Baby	Hotdog	Hotdog	Bebe	Baby
Courtney	Boy	Oso	Bear	Flashlight	Flashlight	Grande	Bear
Who	Girl	Oink	Pig	Bear	Bear	Lanterna	Alphabet
Bunnies	Bunny	Dododo	Duck	The	Dog	You	Brush
Kittycat	Kitty	Sister	Girl	Joey	Book	You	Face
Hat	Hat	Mhmm	Hand	You	Face	Mu–ecita	Rabbit
Moocow	Cow	Hand	Hand	You	Box	Manejar	Truck
Hand	Hand	Put	Ring	Ruff	Pig	Who	Doll
Mhmm	Hand	Bottles	Face	Cheese	Pepperoni	Oscar	Oscar
Meow	Baby	Sheep	Sheep	Ernie	Ernie	Drop	Cheese
Bigbird	Bird	Gatito	Kitty	Doggy	Dog	Pedazos	Cheese
Pig	Pig	Davids	Lamb			Maza	Dough
Oink	Pig	Ring	Ring			Mas	Dough
Bird	Duck	Bottle	Eyes			Take	Flashlight
Birdie	Duck	Encontrar	Cow			Pepperoni	Pepperoni
		Vaca	Cow			Algo	Dog
		Cerdo	Pig			Rosy	Rabbit
		Abelardo	Bird			Conejo	Rabbit
						Blocks	Blocks
						Rosy	Doll
						Oso	Bear
						Doors	Car
						Car	Car

Incorrect mappings are shaded.