

Child Acquisition of Multiword Verbs: A Computational Investigation

Aida Nematzadeh, Afsaneh Fazly, and Suzanne Stevenson

Abstract Traditional theories of grammar, as well as computational modelling of language acquisition, have focused either on aspects of word learning, or grammar learning. Work on intermediate linguistic constructions (the area between words and combinatory grammar rules) has been very limited. Although recent usage-based theories of language learning emphasize the role of multiword constructions, much remains to be explored concerning the precise computational mechanisms that underlie how children learn to identify and interpret different types of multiword lexemes. The goal of the current study is to bring in ideas from computational linguistics on the topic of identifying multiword lexemes, and to explore whether these ideas can be extended in a natural way to the domain of child language acquisition. We take a first step toward computational modelling of the acquisition of a widely-documented class of multiword verbs, such as *take the train* and *give a kiss*, that children must master early in language learning. Specifically, we show that simple statistics based on the linguistic properties of these multiword verbs are informative for identifying them in a corpus of child-directed utterances. We present preliminary experiments demonstrating that such statistics can be used within a word learning model to learn associations between meanings and sequences of words.

A. Nematzadeh (✉)

Department of Computer Science, University of Toronto, Toronto, Canada
e-mail: aida@cs.toronto.edu

A. Fazly

School of Computer Science, Institute for Research in Fundamental Sciences (IPM),
P.O.Box 19395-5746, Tehran, Iran
e-mail: afsaneh.fazly@gmail.com

S. Stevenson

Department of Computer Science, University of Toronto, Toronto, Canada
e-mail: suzanne@cs.toronto.edu

1 Introduction

Traditional theories of grammar distinguish between lexical knowledge (the individual words that a speaker knows) and grammatical knowledge (the rules for combining words into meaningful utterances). However, there is a rich range of linguistic phenomena in the less explored area between words and combinatory rules/constraints. For example, a multiword lexeme such as *take the train* has an idiosyncratic semantics (“use a train as mode of transport”) that suggests its treatment as a lexical unit, since the meaning cannot be compositionally derived in a general manner.¹ But *take the train* also behaves as a syntactic phrase, undergoing various alternative means of expression (e.g., *took a train*, *take the fast train*, *take trains all over Europe*). Much research on language has thus focused on a range of multiword lexemes such as idioms, light verb constructions, noun compounds, and collocations (e.g., [15, 20, 22, 46, 48, 55, 76]). Psycholinguists have also shown the importance of co-occurrence and contingent frequency effects between words, and between words and syntactic patterns in the learning and processing of language (e.g., [5, 57, 70, 72]).

In theories of language acquisition in particular, especially usage-based accounts of language learning (which eschew complex innate linguistic knowledge), the role of multiword constructions has been emphasized (e.g., [40, 41, 74]). However, *computational modelling* of language acquisition has continued to focus on various aspects of word learning (e.g., [33, 37, 49, 65, 77]), or grammar learning (e.g., [17, 69]). Work on intermediate constructions has mostly been limited to identifying general properties of verb argument usages (e.g., [3, 4, 13, 18, 23, 61, 63]), rather than on multiword lexemes. Recent work by Borensztajn et al. [9] uses a probabilistic model (in the DOP framework) to show that a grammar learner can progress from highly lexicalized to multiword tree fragments, on the basis of statistical patterns in the kind of input children receive. Bannard and Matthews [7] further give evidence from human subjects that children are sensitive to the frequencies of multiword sequences. These studies provide evidence that children recognize and produce certain (e.g., high-frequency) multiword sequences in their input, but do not address what sort of cues (other than, e.g., frequency) a child might use to identify, and treat differentially, the various distinguished types of multiword lexemes suggested by linguistic analyses.

Thus in the study of child language acquisition, much remains to be explored concerning the precise computational mechanisms that underlie how children learn to identify different types of multiword lexemes—that is, how they recognize that an idiosyncratic semantics is associated with a sequence of words (rather than single words plus combinatory rules), and how the idiosyncratic meaning relates to the surface (lexical and syntactic) form of a particular combination. In contrast, there

¹A compositional approach to *take the train* would depend on knowledge of a very specialized meaning of *take* restricted to occur with a narrow range of objects, which is essentially an alternative lexicalization of the necessary knowledge. See Fazly et al. [31] for a computational approach to the restricted productivity of such expressions.

has been significant work in computational linguistics on this very topic, with the development of statistical measures, both for identifying multiword lexemes in a corpus, and for determining the syntactic and semantic behaviour of the particular type of multiword lexeme in question (e.g., [8, 19, 21, 25, 28, 30, 43, 50, 53, 67, 71, 75]). The goal of our research here is to explore whether this computational work on multiword lexemes can be extended in a natural way to the domain of child language acquisition, where an informative cognitive model must take into account the two issues of what kind of data the child is exposed to, and what kinds of processing of that data is cognitively plausible for a child.

In pursuing these questions, we focus in particular on the acquisition of multiword verbs, such as *take the train* and *give a kiss*. These constructions are a rich and productive source of predication which children must master in most languages, doing so at very young ages [41]. For example, consider the following conversation from the CHILDES database ([11], sarah130a.cha):

- *MOT: you're not gonna *take any toys* down to the beach today you know.
 *CHI: why?
 ...
 *MOT: we have to *take the train*.

Here, the mother uses the verb *take* first in its core literal meaning (in *take any toys*), and then within a multiword lexeme in which *take* has a non-literal meaning and combines with the particular argument to express the use of a mode of transportation (in *take the train*). The child's further responses within this conversation give no indication that she is puzzled by these very different usages of *take*. Yet they do pose a very significant puzzle for researchers: It has been noted that children learn highly frequent verbs (such as *take*) first (e.g., [41]), and yet it is precisely these verbs that are also the most polysemous, showing a wide range of metaphorical sense extensions in multiword lexemes, which children recognize and deal with effectively [16, 44, 73].

Research over the last few years has shown that the distinctions among literal and non-literal verb–argument combinations (such as *take the toys* versus *take the train* or *take a nap*) are in principle learnable based on statistics over usages of such expressions (e.g., [30, 75]). However, such work depends on very large amounts of data (from corpora on the order of 100 M words) and on sophisticated statistical and grammatical calculations over such data. The goal here is to determine what is learnable through the means available to a child—that is, on the basis of data in child-directed speech and using simpler, cognitively plausible calculations.

We begin by summarizing the motivation and approach to deriving simple statistics based on the linguistic properties of the multiword lexemes under study (first presented in [32]). We then present new experiments that show that such statistics can be informative in identifying such multiword lexemes in child-directed speech. Then we turn to a novel approach for incorporating these statistical measures into an existing model of word learning, to show further that such statistics can be used within a natural process of word learning to associate a single meaning with a sequence of words. In this way, we take a first step toward computational

modelling of acquisition of the kinds of multiword verbs that children must master early in language learning, shedding light on the mechanisms that could underlie a usage-based model of this process.

2 Multiword Lexemes with Basic Verbs

The highly frequent and highly polysemous verbs referred to above include what are called “basic” verbs—those that express physical actions or states central to human experience, such as *give*, *get*, *take*, *put*, *see*, and *stand*, among others. These verbs undergo metaphorical sense extensions of their core physical meanings that enable them to combine with various arguments to form multiword lexemes [15, 58, 59, 62]. We focus here on expressions in which a basic verb is combined with a noun in its direct object position to form either a literal combination (as in *take the toys*) or a multiword lexeme (such as *take the train*, *take a nap*). We refer to all such expressions (both literal and non-literal) as verb–noun combinations or verb–noun pairs, with the understanding that the verb is a basic verb.

Verb–noun combinations that form multiword lexemes are very frequent in many languages (e.g., [1, 20, 45, 46, 51, 54]). Such expressions show a range of semantic idiosyncrasy, where the semantics of the multiword lexeme is more or less related to the semantics of the verb and the noun separately [38, 66]. Thus, verb–noun combinations can be viewed as lying on a continuum (without completely clear boundaries) from entirely literal and compositional, to highly idiomatic. However, for convenience we can think of classes of constructions on this continuum, each identified by a particular way in which the verb and the noun component contribute to the meaning of the construction. Following [30], we consider four possible classes; these are listed below with an example from the child-directed speech used in our experiments along with some information about the semantic contribution of the components of expressions in that class:

1. Literal combination or LIT

- *Give (me) the lion*
 - *Give*: physical transfer of possession
 - *Lion*: a physical entity

2. Abstract combination or ABS

- *Give (her) time*
 - *Give*: abstract transfer or allocation
 - *Time*: an abstract meaning

3. Light verb construction or LVC

- *Give (the doll) a bath*
 - *Give*: convey/conduct an action
 - *Bath*: a predicative meaning

4. Idiomatic combination or IDM

- *Give (me) the slip*
 - *Give, slip*: no/highly abstract contribution

These classes are important in the context of child language acquisition because there is a clear connection between the linguistic properties of each class and the meaning of the expressions in the class. Such a relation can enable language learners to generalize their item-specific knowledge, for example by making predictions about the meanings of new expressions based on their likely class. For example, when a child hears a new expression such as *give a shout*, if she recognizes that this is likely an LVC, then she can infer that it roughly means the same thing as the noun—i.e., *shout*—which contributes the predicative meaning, and also infer any other properties holding of LVCs more generally.²

The four classes of expressions above have differing linguistic behaviours that can be cues to the underlying distinctions among the classes [30]. Specifically, expressions from each class exhibit particular lexical and syntactic behaviour that closely relate to the semantic properties of the class. We next elaborate on these properties and behaviours, and describe how they can form the basis for statistical measures for distinguishing the classes.

3 Linguistic Properties and the Usage-Based Measures

It has been shown that children are sensitive to the frequency of occurrence of multiword sequences (e.g., [7]). However, simple co-occurrence frequency of a verb and a noun (or measures of association between the two) do not suffice for accurate identification of multiword verb–noun lexemes [29]. We thus further hypothesize that children are also sensitive to the syntactic and semantic properties of each class of verb–noun combination. As a first step to examining this hypothesis, we need to verify whether information about such properties is available in the input children receive, and whether the available information is useful for determining the semantic class of a given combination. We note that there is some overlap in the properties exhibited by the various non-literal classes. We thus further simplify our task here by aiming to distinguish the non-literal expressions (those from ABS, LVC, IDM) from literal ones (LIT). There is only one instance of an IDM in our data, hence in our presentation of the measures here, we discuss the properties with respect to the ABS+LVC classes.

As noted earlier, computational linguistic studies have developed sophisticated statistical measures based on such properties, which have achieved success in identifying non-literal combinations when evaluated on large amounts of text corpus

²For example, adult competence with the language includes the knowledge that this refers to a single occurrence of a bounded ‘shouting’ action [12, 76].

data (e.g., [28, 30]). Given the hypothesized importance of simplicity in language learning (c.f. [60]), our goal here is to use simpler measures (tapping into similar properties) that are more cognitively plausible, and that are robust when used with smaller amounts of child-directed speech (CDS). We note that some of the measures explained in this section are taken and adapted for this purpose from Fazly [29]. The resulting measures fit into three groups based on the linguistic properties of the verb and the noun in a verb–noun combination: the degree of association of the verb and noun, the semantic properties of the noun, and the degree of syntactic fixedness of the expression.

3.1 Association of a Verb–Noun Pair

In a literal verb–noun combination, where the verb contributes its core physical semantics, a wide variety of nouns can occur as the noun component (e.g., one can *give an apple, a book, a car, a dog*, etc.). In contrast, in a non-literal combination, the verb has an abstract and/or metaphorical meaning and hence can combine with a set of nouns that is semantically, and somewhat idiosyncratically, restricted (e.g., *give a groan/cry/yell*, but not *give a gripe*, [31]). Moreover, the latter group of nouns often contribute a specific abstract meaning to the combinations they appear in, and hence may not occur as the direct object of other verbs as frequently as do concrete nouns. As a result, we expect the verb and the noun component in non-literal expressions to co-occur more often compared to the components of literal combinations [14, 27]. Below we explain two different measures capturing the marked frequency of a verb–noun pair.

Frequency. The simplest way to measure the association of a verb and a noun is by the frequency of co-occurrence of the verb–noun pair $\langle v, n \rangle$, as in:

$$\text{Cooc}(v, n) \doteq \text{freq}(v, n, \text{gr} = \text{dobj}) \quad (1)$$

where $\text{gr} = \text{dobj}$ indicates that the noun is the direct object of the verb. We assume that children are able to keep track of simple counts of such verb–noun pairs.

Conditional Probability. Although non-literal expressions are expected to co-occur more often compared to literal expressions, the co-occurrence of some literal expressions is also significant (e.g., *take the toy* in child-directed speech). However, the noun in a non-literal expression generally does not occur with as diverse a set of verbs as a noun in a literal expression. For example, *apple* can be used in many literal expressions with different verbs: *give the apple, take the apple, eat the apple*, and *wash the apple*, whereas *decision* only occurs in one non-literal verb–noun

combination: *make a decision*.³ In other words, while the verb in a LIT expression is typically thought of as selecting for a noun in direct object position, in a non-literal expression the noun can be viewed as selecting for a verb (e.g., [24, 43]). We measure this property by computing the conditional probability of a verb–noun pair given the noun (CProb).

$$\begin{aligned} \text{CProb}(v, n) &\doteq P(v|n, \text{gr} = \text{doobj}) \\ &= \frac{\text{freq}(v, n, \text{gr} = \text{doobj})}{\sum_{v'} \text{freq}(v', n, \text{gr} = \text{doobj})} \end{aligned} \quad (2)$$

This measure is still a very simple one for children, since it is composed of two frequency counts, although we should note that it does assume that children are able to keep track of the count of a noun as the direct object of any verb.⁴

3.2 *Semantic Properties of the Noun*

There is evidence that children are sensitive to the semantic differences between the nouns in a literal versus non-literal verb–noun combination [64]. For example, whereas the noun in a non-literal verb–noun combination is often non-referential, abstract, and/or predicative (as in *take time* and *give a hug*), the noun in a literal combination tends to be referential and concrete (as in *take the toys* and *give a banana*). Earlier work has used WordNet [35] to estimate non-referentiality and predicativeness by looking at the noun’s position in the taxonomy, and its morphological relation to a verb [30]. However, WordNet’s conceptual and lexical organization most likely does not reflect that of a child. Next, we explain two measures that instead aim to capture these properties with simple statistics over the surface behaviour of the noun.

Non-referentiality. Non-referential nouns (such as those in non-literal expressions) tend to appear in particular syntactic forms [42]—typically preceded by an indefinite determiner (such as *a/an*) or no determiner [34, 76]. Moreover, it has been shown that children indeed associate certain semantic properties with surface syntactic forms [10]. Here we assume that a noun is recognized as non-referential to the extent that it occurs in this preferred pattern of determiner use, i.e.:

³The choice of verb can vary among dialects of the language; for example, British speakers typically say *take a decision* instead of *make a decision* and *have a nap* instead of *take a nap*.

⁴Although it remains to be tested whether children actually do this, a construction grammar approach to language acquisition, as in Goldberg [41], supports this type of calculation, since the learner would keep track of which nouns can occur in which constructions.

$$\text{NRef}(n) \doteq P(pt_{\text{ntref}}|n) = \frac{\text{freq}(n, pt_{\text{ntref}})}{\text{freq}(n)} \quad (3)$$

where $pt_{\text{ntref}} = \langle \text{det}:a/\text{an}/\text{NULL } n \rangle$, $\text{freq}(n, pt_{\text{ntref}})$ is the frequency of occurrence of n in pattern pt_{ntref} , and the denominator estimates the frequency of n in any pattern. Note that we look at all occurrences of a noun irrespective of its grammatical relation to a verb; this is thus a simple relative frequency for a child to determine: of the instances she sees of this noun, what proportion are in this particular pattern.

Predicativeness. In a non-literal verb–noun combination, such as *make a decision*, the predicative meaning is contributed mainly by the noun component, i.e., *decision*. Moreover, in such expressions the noun is often morphologically related to a verb (e.g., *decision* as the nominalized form of *decide*). To capture this property, previous work has looked at whether the noun has a morphologically-related verb form [30]. We cannot assume that full knowledge of morphology is in place before a child starts learning about non-literal expressions. But it has been shown that young children can accurately predict whether a word is used as a verb or a noun in a given context [10]. We thus measure predicativeness of the noun n in a verb–noun pair as the relative frequency of the form n (e.g., *push* in *give a push*) being used as a verb (as in, e.g., *push the door*).

$$\text{Pred}(n) \doteq \frac{\text{freq}(n_V)}{\text{freq}(n_V) + \text{freq}(n_N)} \quad (4)$$

where $\text{freq}(n_V)$ is the frequency of the form n appearing as a verb, and $\text{freq}(n_N)$ is the frequency of the form n appearing as a noun.

3.3 Degree of Syntactic Fixedness

Young children show evidence of learning associations between a complex syntactic form and a specific semantic interpretation (e.g., [36, 70]). It is thus reasonable to assume that children can use the information about the surface syntactic behaviour of a verb–noun combination to identify its semantic class. Here we devise statistical measures that aim at capturing the differing syntactic behaviour of non-literal and literal combinations.

Non-literal expressions are known to have a fixed syntactic structure and not occur in a variety of forms [20, 26]. More specifically, ABS+LVC expressions, while allowing some variation, are relatively restricted compared to LIT expressions. For example, an LVC such as *give a shout* allows limited noun and determiner variation; e.g., *give some shouts* and *give the shout* are not as acceptable as *give a shout*. This is also true for ABS expressions. For example, *take a time* and *take times* are not recognized as acceptable variations of *take time*. In contrast, literal expressions are generally much more syntactically flexible, e.g., *take an apple*, *take the apple*, and *take three apples* are all acceptable.

Although there is some variation, most LVC and ABS expressions appear in the form $pt_{fixed} = \langle v \text{ det: } a/an/NULL \ n \rangle$. (Note that the noun is in the same pattern as for NRef above; the difference is that here the focus is on the degree to which the particular verb–noun combination leads to the use of that pattern for the noun.) Measures of this type of syntactic fixedness have required keeping track of probability distributions over a wide range of items and patterns [6, 30]. Here, we estimate the degree of syntactic fixedness of a target verb–noun combination with a much simpler measure—the relative frequency of the pair in the preferred pattern:

$$\begin{aligned} \text{Fixed}(v, n) &\doteq P(pt_{fixed} | v, n, \text{gr} = \text{doobj}) \\ &= \frac{\text{freq}(v, n, \text{gr} = \text{doobj}, pt_{fixed})}{\text{freq}(v, n, \text{gr} = \text{doobj})} \end{aligned} \quad (5)$$

Children appear to store specific information about the frequency of occurrence of multiword sequences in general (e.g., [7]), and about verb–argument structures in particular (e.g., [41, 74]). We thus expect the above calculations to be plausible for children.

We have described five simple statistical measures that may be plausible for children to keep track of. In the remainder of the paper, we first present experiments that evaluate how well the measures can identify non-literal verb–noun combinations in child-directed speech, and then describe extensions to a word learning model that enable it to learn the meaning of such expressions by incorporating these statistical measures.

4 Evaluating the Statistical Measures

In this section, we present two types of experiments to determine the potential of our statistical measures to identify non-literal verb–noun combinations in child-directed speech. Each of our measures assigns a numerical score to the expressions that reflects one of the linguistic properties that may be useful to a child in determining which are literal and which are non-literal. To evaluate their effectiveness, we first (in Sect. 4.2) apply a hierarchical agglomerative clustering algorithm that uses the scores to separate all the experimental expressions into two clusters, and then see how closely those clusters correspond to the actual labels on the expressions as LIT, or as ABS+LVC. Since we assume that, in any learning situation, a combination of the cues might be at work, we use all five measures as input to the clustering algorithm.

The clustering results thus show the effectiveness of the measures working together to separate non-literal from literal combinations. We further analyze (in Sect. 4.3) each individual measure in its ability to separate literal and non-literal expressions, in order to better understand how relevant each measure is to the identification of multiword lexemes. We begin by presenting the details of the experimental data and evaluation methods.

4.1 Experimental Setup

Corpus. To gather input for our experiments, we use the American English section of the CHILDES database [52], removing 16 corpora that either lack child-directed speech (CDS) or belong to a special group with a particular language use (e.g., socio-economically distinguished). All the data are automatically parsed with the parser of Sagae et al. [68]. Because we are interested in what is learnable from input a child is exposed to, the statistics for all experiments are extracted from CDS. The size of the CDS portion of the corpus is about 600,000 utterances, which contain nearly 3.2 million words (including punctuation).

Experimental expressions. In this work we focus on two basic verbs, *take* and *give*, because they are highly polysemous and frequently used in verb–noun combinations [15]. We extract verb–noun combinations that contain these verbs from the CDS portion of the data. The final expression list that is used in the experiments includes those verb–noun pairs with a frequency of at least 5. In some experiments, we further restrict the data to higher-frequency verb–noun combinations, i.e., those occurring at least 10 times. Dealing with low-frequency items is important in modeling child language acquisition, and here we vary the relatively low cutoff to see if it helps to have more items. The final list of expression types was annotated by a native English speaker with four classes: LIT, ABS, LVC, and IDM. Note that we consider expression types, not tokens. Thus, if a verb–noun combination had usages that fall into more than one class, the annotator chose the class that seemed to reflect the predominant usage.⁵ Invalid expressions (due to parsing errors) and the single instance of an IDM were removed from the expression list. Table 1 presents the number of expressions in each class, as well as the total number of non-literal expressions (ABS+LVC).

Evaluation. To evaluate the clustering experiments, we assign to each resulting cluster a label (either LIT or ABS+LVC), which is the label of the majority of items in the cluster, and calculate accuracy (*Acc*) and completeness (*Comp*) as measures of the goodness of the cluster. Accuracy gives the proportion of expressions in a cluster that have the same label as the cluster; completeness gives the proportion of all expressions with the same label as the cluster that are actually placed in that cluster. (Note that *Acc* is similar to precision, and *Comp* to recall.)

Recall that our measures are designed such that each is expected to be higher for the non-literal expressions than for the literal ones. In evaluating the measures individually, we can thus use each measure to rank the expressions and see whether ABS+LVC expressions are generally ranked higher than LIT ones. We do this for

⁵For example, the verb–noun pair *give-hand* may occur as an ABS usage (*give me a hand cleaning up*) or as a LIT usage (*give me Mr. PotatoHead's hand* or *give me your pretty hands*). In most cases of such potential ambiguity, the annotator had a clear intuition of which would be the predominant usage, since the alternative would be odd to find in CDS. In some cases, such as *give-hand*, the actual corpus usages were examined to determine the most frequent class.

Table 1 A detailed breakdown of the experimental expressions

V_b	Total	LIT	ABS	LVC	ABS+LVC
198 expressions with $freq \geq 5$					
<i>take</i>	108	77	18	13	31 (29%)
<i>give</i>	90	73	7	10	17 (19%)
<i>take and give</i>	198	150	25	23	48 (24%)
98 expressions with $freq \geq 10$					
<i>take</i>	57	38	8	11	19 (33%)
<i>give</i>	41	30	4	7	11 (27%)
<i>take and give</i>	98	68	12	18	30 (31%)

take and *give* expressions separately, and for all expressions together. We use a standard evaluation metric, namely average precision (*AvgPrec*), which reflects the goodness of a measure in placing expressions from the target classes (ABS and LVC) before those from the other class (LIT), and is calculated as the average of *precision* scores at different thresholds.

We also compare the performance of each measure against a baseline which reflects how hard the task is. We randomly assign a value between 0 and 1 to each expression in a set, generating a random ranked list. We repeat this process 1,000 times and report the average of the *AvgPrec* values for each of these random lists as our baseline. We also calculate the relative error rate reduction (*ERR*) of each measure over the random baseline. To calculate *ERR* for a measure, we divide the difference between the error rates of the measure and the baseline by that of the baseline.

4.2 Measures in Combination: Clustering

Results of the clustering experiments are shown in Table 2. We can see that *Acc* for non-literal expressions is high only for the higher-frequency expressions (compare C_2 in each panel of the table). We also see that literal expressions are better separated than non-literal ones since their *Comp* score is much higher (compare C_1 and C_2 for each panel of the table). Looking closely at the number of expressions of different labels (LIT, LVC, and ABS) in each cluster, it is clear that ABS expressions are more mixed with LIT expressions compared to LVC ones. Consequently, the measures are better in separating LVC from LIT than ABS from LIT.

We performed two-way clustering on the assumption that a two-way distinction would be easier for the measures than a three-way distinction. However, the poor performance on ABS expressions may be due to a weakness of the measures, or may be due to a need for three clusters to capture the pattern in the data. We thus also performed a three-way clustering to examine the goodness of measures in dividing expressions into ABS, LVC, and LIT classes (see Table 3). According to the results, ABS expressions do not form a separate cluster, and are again mixed in

Table 2 Two-way clustering results. C_i represents Cluster i ; *Label* is the majority class in the cluster; *Acc* and *Comp* are explained in the text

	LVC	ABS	LIT	<i>Label</i>	<i>Acc</i> (%)	<i>Comp</i> (%)
On 198 expressions with $freq \geq 5$						
C_1	1	13	122	LIT	90	81
C_2	22	12	28	ABS+LVC	55	71
On 98 expressions with $freq \geq 10$						
C_1	1	9	65	LIT	87	96
C_2	17	3	3	ABS+LVC	87	67

Table 3 Three-way clustering results. C_i represents Cluster i ; *Label* is the majority class in the cluster; *Acc* and *Comp* are explained in the text

	LVC	ABS	LIT	<i>Label</i>	<i>Acc</i> (%)	<i>Comp</i> (%)
On 198 expressions with $freq \geq 5$						
C_1	3	7	27	LIT	73	18
C_2	1	13	122	LIT	90	81
C_3	19	5	1	LVC	76	83

with the LIT and LVC clusters. Future work will need to verify whether this is due to an inconsistent annotation of the ABS expressions, or because our measures do not adequately capture properties of this class. Interestingly, however, a three-way clustering results in forming a more coherent LVC class: compare *Acc* and *Comp* for C_3 in Table 3 with those for C_2 in the top panel of Table 2.

4.3 Performance of the Individual Measures

We test the performance of each measure, for *take* and *give* expressions separately, and for all the expressions with *take* and *give*. The results in Table 4 show that all measures perform better than the baseline (at separating non-literal expressions from literal ones), with CProb, Pred, and Fixed having the best performance. These results suggest that simple statistical measures that draw on specific linguistic properties of non-literal verb–noun combinations—measures which are plausible for children to keep track of—can indeed be effective in recognizing non-literal expressions.

We also observe that, in general, our measures perform better on the expressions composed with *take* than the expressions with *give*. A possible explanation is that the *give* expressions are more complicated, because *give* more often occurs in a double object construction (in comparison to *take*). It remains to be tested whether children also show more difficulty in learning *give* expressions.

Looking at performance on higher-frequency expressions, we see that all measures show an improvement. However, note that only for two of the measures

Table 4 Performance (*AvgPrec*) of the individual measures. The numbers in parentheses show the *ERR* of the measures for *take* and *give* expressions combined

Measure	<i>take</i>	<i>give</i>	<i>take</i> and <i>give</i>
On 198 expressions with <i>freq</i> ≥ 5			
Baseline	0.28	0.19	0.24
Cooc	0.53	0.38	0.51 (0.35)
CProb	0.65	0.47	0.56 (0.42)
NRef	0.49	0.33	0.40 (0.21)
Pred	0.59	0.54	0.59 (0.46)
Fixed	0.66	0.44	0.56 (0.40)
On 98 expressions with <i>freq</i> ≥ 10			
Baseline	0.33	0.27	0.31
Cooc	0.57	0.41	0.54 (0.34)
CProb	0.71	0.57	0.64 (0.48)
NRef	0.62	0.49	0.55 (0.35)
Pred	0.68	0.59	0.67 (0.52)
Fixed	0.84	0.56	0.71 (0.58)

(NRef and Fixed) the gain in performance is substantially more than the increase in the baseline performance. These two measures summarize the syntactic behaviour of a word or a combination by examining all their usages. For higher-frequency expressions (with more usages), it is possible that the evidence available for these measures is more reliable, resulting in better performance.

5 Embedding the Measures into a Word Learning Model

The results presented so far suggest that simple statistics over the usages of a verb–noun combination (and its components) have the potential to provide useful cues for a child to identify non-literal expressions. We need to explore further how children learning the vocabulary of their native language might use such statistical cues to recognize that certain combinations of words in their input actually form multiword lexemes. We investigate this issue by incorporating (some of) the statistical measures into the operations of an existing computational model of early word learning in children, namely, that of [33].

We first give a brief overview of the original word learning model in Sect. 5.1 (we refer the interested reader to [33] for a full explanation of this model). When processing a multiword lexeme, such as *take a nap*, the original model finds a meaning for each individual word (*take*, *a*, *nap*) just as it does for a literal combination of words, such as *take any toys*. There is no mechanism for the model

to associate a single meaning with the sequence of words *take a nap*.⁶ We thus add a preprocessing step, described in Sect. 5.2, in which the model draws on statistics collected thus far to decide whether a given sequence of words in the input utterance should be considered as a multiword lexeme. Section 5.3 presents an evaluation of the new model with respect to the acquisition of multiword lexemes of the form verb–noun.

5.1 The Original Word Learning Model

We use the model of Fazly et al. [33], which is a probabilistic incremental model of cross-situational word learning in children. The input to the model is a list of pairs of an utterance (what the child hears, represented as a set of words) and a scene (what the child perceives or conceptualizes, represented as a set of meaning symbols), as in:⁷

Utterance: *Joe is happily eating an apple*

Scene: JOE, IS, HAPPILY, EAT, A, APPLE

The model incrementally learns a meaning for each word in the input as a probability distribution over all meaning symbols, $P(m|w)$, referred to as the *meaning probability* of the word, as in:



Prior to receiving any usages of a given word, the model assumes that all symbols have equal probability as its meaning. The model then updates the meanings of words by processing each utterance–scene pair in two steps.

As the first step in processing an input utterance–scene pair, the model, like children, must determine which meaning symbol in the scene is associated with

⁶The original model of Fazly et al. treats utterances as unordered bags of words, ignoring syntactic information. Syntax is arguably a valuable source of knowledge in word learning in children (e.g., [39, 56]). In a preliminary study, Alishahi and Fazly [2] also show that the word learning model can potentially benefit from knowledge of syntactic categories. Such information might be necessary for the acquisition of multiword lexemes, and should be further investigated in the future.

⁷Following Fazly et al. [33] we assume that words such as *a* and *is* also have corresponding meaning symbols in the scene. Such words are often considered by linguists to mainly have a grammatical function. However, it is reasonable to assume that language learners perceive some aspects of their meaning (e.g., definite/indefinite for a determiner such as *a*, and state/action for the verb *be*) from the scene.

each word in the utterance. (Note that the input does not indicate which meaning goes with which word.) This process is called the *alignment* of words and meaning symbols. Alignment is probabilistic, so that each word is aligned more or less strongly with each meaning, according to the model's partially-learned knowledge of meaning probabilities as calculated thus far. Specifically, the probability of aligning a meaning symbol and a word in the current input is proportional to the current meaning probability of that meaning symbol for the word, and is disproportional to the meaning probabilities of the meaning symbol and the other words in the utterance. That is, a word w and meaning m are strongly aligned if $P(m|w)$ is relatively high and $P(m|w')$ is relatively low for other words w' in the utterance.

As the second step, the meaning probabilities of the words in the current utterance are updated according to the accumulated (probabilistic) evidence from prior co-occurrences of words and meaning symbols (reflected in the alignment probabilities). This evidence is collected by maintaining a running total of the alignment probabilities over all input pairs encountered so far, yielding an accumulated frequency of co-occurrence of a word–meaning pair, weighted by the strength of alignment between the two each time they are observed together. Meaning probabilities for current words are then re-calculated from these incrementally-accumulated alignment probabilities.

5.2 *Learning the Verb–Noun Multiword Lexemes*

The approach described above learns a separate meaning probability distribution for each word. To enable the model to learn a meaning distribution for a verb–noun combination such as *give a kiss*, the model must be able to identify the expression as a single unit of meaning. To achieve this, we add an input pre-processing step to the original model and slightly modify the way alignment probabilities are calculated.

We assume that upon receiving an utterance–scene pair containing any verb–noun combination (literal or non-literal), a learner (here the model) simultaneously considers two possible interpretations: That the verb–noun combination is a multiword lexeme, or that the combination is literal. That is, when the original model receives an input such as:

U : *give me a kiss*
 S : GIVE, ME, A, KISS

our modified model will also consider the alternative interpretation in which the verb and noun form a single unit of meaning:

U' : *give-kiss me a*
 S' : ME, A, GIVE-KISS

This alternative interpretation is created by merging the verb and the noun into a single word (*give-kiss*), and by creating a new meaning symbol for the associated

event (GIVE-KISS). We assume that the learner has a certain confidence in either of these interpretations given what has been learned about words and meanings in the input thus far. Specifically, the learner calculates a probability $\text{prob}_{\text{mwl}}(v, n)$ which reflects its confidence that the verb–noun combination in the utterance is a non-literal multiword lexeme, as in (U' – S') above. This probability combines the two statistical measures, namely CProb and Pred, which were the best in separating literal and non-literal expressions in our earlier experiments.⁸ More formally, $\text{prob}_{\text{mwl}}(v, n)$ is computed as in:

$$\text{prob}_{\text{mwl}}(v, n) = \alpha * \text{CProb}(v, n) + (1 - \alpha) * \text{Pred}(n)$$

where α is set to 0.5, weighting the evidence from the two statistical measures equally. Thus, the interpretation that a verb–noun combination is a multiword lexeme, as in (U' – S') above, is assigned a confidence score equal to $\text{prob}_{\text{mwl}}(v, n)$, and the other interpretation, as in (U – S) above, is given the confidence score of $1 - \text{prob}_{\text{mwl}}(v, n)$.

Whenever there is a verb–noun pair in an utterance, we calculate separate alignment probabilities over two possible utterance–scene pairs corresponding to the two interpretations. The two sets of alignment probabilities are then combined, using $\text{prob}_{\text{mwl}}(v, n)$ as a weight, to get a single alignment probability for each word and meaning symbol in the input pair:

$$\begin{aligned} \text{align}(w|m) &= \text{prob}_{\text{mwl}}(v, n) * \text{align}_1(w|m) \\ &+ (1 - \text{prob}_{\text{mwl}}(v, n)) * \text{align}_2(w|m) \end{aligned}$$

Note that for a w – m pair that occurs only in one interpretation (e.g., *give-kiss*–GIVE-KISS), its alignment would be zero in the other interpretation. This means that the learner aligns each word and meaning symbol to the extent that it is confident that the corresponding interpretation is accurate. The modified alignment probabilities are then used to calculate the meaning probabilities as in the original model.

5.3 Experiments on the Modified Word Learner

We expect the modified word learning model to learn a single meaning for non-literal verb–noun pairs but not for literal ones. That is, we expect a meaning probability such as $P(\text{GIVE-KISS}|\textit{give-kiss})$ to be high, since *give-kiss* is a multiword lexeme that expresses a kissing event. By contrast, $P(\text{GIVE-PRESENT}|\textit{give-present})$

⁸We did not incorporate the Fixed measure into this probability, because this measure needs to consider the usage pattern across several occurrences, and many of the experimental items in this corpus have frequency of only 1 or 2.

Table 5 The number and percentage of verb–noun combinations in each class that are learned correctly: i.e., as literal for the LIT class, and as non-literal for the ABS and LVC classes

Class	Size	Learned correctly	
		Number	Percentage (%)
LIT	115	105	91
ABS	24	8	33
LVC	32	24	75

should be low, since *give a present* is literal with individual associations of *give* to GIVE and *present* to PRESENT.

We use the same data as in Fazly et al. [33]: 180,499 utterance–scene pairs, where the utterances are taken from the Manchester corpus in the CHILDES database [52], and the scene representations are automatically constructed using an input-generation lexicon containing a symbol as the meaning of each word. Because the Manchester corpus is British English and some American English verb–noun multiword lexemes with *take* occur with other basic verbs in British English, we only consider the verb–noun combinations with *give* in the current experiments. Since children can learn meanings of very low frequency words, we do not apply a frequency cut-off, but rather consider all verb–noun combinations with *give* in the corpus. The number of LIT, ABS, and LVC expressions used in our experiments is shown in Table 5.

In Fazly et al. [33], a word–meaning pair is considered learned if the probability of the correct meaning given the word is above 0.7. This is a somewhat arbitrary cut-off, but to be consistent we use the same threshold. We say that a verb–noun combination with *verb* and *noun* is “learned as a multiword lexeme” if the probability $P(\text{VERB-NOUN}|\text{verb-noun})$ is above this threshold—that is, the combination of the verb and noun words are associated with a single (correct) meaning. We say that a verb–noun combination is “learned correctly” if the combination is non-literal and is learned as a multiword lexeme, or the combination is literal and is **not** learned as a multiword lexeme. To evaluate the model’s ability in learning multiword lexemes, we look at the proportion of expressions from each class that are learned correctly; see Table 5.

The results in Table 5 show that the model performs very well on the LVC and LIT expressions (75 % and 91 %, respectively), but only a small proportion (33 %) of the ABS expressions are learned correctly. A closer look at the results shows that many of the non-literal expressions with a low frequency of 1 are not learned correctly. This includes 46 % of LVC expressions with frequency 1, and 85 % of ABS expressions with frequency 1. This finding is in line with what has been observed in children: that children are faster at producing more familiar (frequent) multiword sequences [7]. It remains to be tested whether children also are unable to learn some of these MWEs (as MWEs) from a single exposure.

6 Conclusions

Our results confirm that simple statistical measures that draw on linguistic properties of non-literal expressions are useful in identifying them. The best measure for *give* and *take* expressions is Pred, i.e., the normalized frequency of the usages of the noun as a verb. The success of this measure indicates that the predicativeness of the noun is a salient property of non-literal verb–noun combinations. The goodness of CProb in identifying non-literal expressions suggests that the verb–noun pair in such expressions is more entrenched compared to literal ones and exhibits collocational behaviour. However, collocational behaviour alone is not a very good indicator of non-literal expressions; the CProb measure consistently outperforms Cooc (which only quantifies the entrenchment of the verb–noun pair). The key difference between these two measures is that in CProb, we also measure the degree that the noun selects for the appropriate verb. The Fixed measure which looks at a specific syntactic pattern for non-literal expressions performs as well as CProb for all expressions, but is the best measure for expressions having frequency of at least ten, for which there is sufficient evidence of typical syntactic usage.

Our measures are generally better for higher-frequency expressions. However, two of the best measures (Pred and CProb) perform well on both expressions with frequency of at least 5 and higher-frequency expressions, suggesting that children might be able to learn verb–noun combinations even with very little input. Our results also show that the performance of our measures is better for *take* expressions compared to *give*. The Fixed measure especially performs well on *take*, but less well on *give*, suggesting that the more complex syntactic constructions that *give* appears in (e.g., the double object construction) may cause children difficulty.

We also integrate our measures into a word learning model, and show that the new model can successfully learn the meaning of many LVC expressions. Future work will need to further investigate why it is harder for the model to learn the meaning of ABS expressions. In the experiments presented in this article, we have focused on a small number of verb–noun combinations (namely, 117) formed around one particular verb (i.e., *give*). To better understand the generalizability of our findings, future research will need to extend these experiments to other verbs (e.g., *take*) and to other types of multiword lexemes (e.g., noun compounds).

Another limitation of the model is that it learns word meanings by mapping each word to a distinct ‘concept’ (e.g., *give-kiss* must be mapped to GIVE-KISS). In the future, we need to use a richer semantic representation where each concept is comprised of finer-grained semantic primitives. The use of such a representation would enable the model to determine semantic similarities among words (e.g., the similarity between the meaning of the expression *give-kiss* and that of the verb *kiss*), which would further allow it to make generalizations across different types of lexical items.

References

1. Alba-Salas, J. (2002). *Light verb constructions in Romance: A syntactic analysis*. Ph.D. thesis, Cornell University.
2. Alishahi, A., & Fazly, A. (2010). Integrating syntactic knowledge into a model of cross-situational word learning. In *Proceedings of CogSci'2010*, Portland.
3. Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science: A Multidisciplinary Journal*, 32(5), 789–834.
4. Alishahi, A., & Stevenson, S. (2011). Gradual acquisition of verb selectional preferences in a Bayesian model. In Poibeau et al. (2011).
5. Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. ISSN 0749-596X.
6. Bannard, C. (2007). A measure of syntactic flexibility for automatically identifying multiword expressions in corpora. In *Multiword Expression'07: Proceedings of the Workshop on a Broader Perspective on Multiword Expressions* (pp. 1–8). Prague: Association for Computational Linguistics.
7. Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248.
8. Bannard, C., Baldwin, T., & Lascarides, A. (2003). A statistical approach to the semantics of verb-particles. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment* (pp. 65–72), Sapporo.
9. Borensztajn, G., Zuidema, W., & Bod, R. (2009). Children's grammars grow more abstract with age – evidence from an automatic procedure for identifying the productive units of language. *Topics in Cognitive Science*, 1(1), 175–188.
10. Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal Psychology*, 55(1), 1–5.
11. Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.
12. Butt, M. (1997). Aspectual complex predicates, passives and dispositionability. In *Talk Held at the 1997 Meeting of the Linguistics Association of Great Britain (LAGB'97)*, University of Essex. <http://ling.uni-konstanz.de/pages/home/butt/>.
13. Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the ACL'04 Workshop on Psycho-Computational Models of Human Language Acquisition* (pp. 17–24), Geneva.
14. Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 115–164). Hillsdale: Erlbaum.
15. Claridge, C. (2000). *Multiword verbs in early modern english*. *Language and Computers* 32. New York: Rodopi.
16. Clark, E. V. (1996). Early verbs, event-types, and inflections. In C. E. Johnson & J. H. V. Gilbert (Eds.), *Children's language* (Vol. 9, pp. 61–73). Mahwah: Erlbaum.
17. Clark, A. (2001). Unsupervised induction of stochastic context free grammars with distributional clustering. In *Proceedings of Conference on Computational Natural Language Learning* (pp. 105–112), Toulouse.
18. Connor, M., Fisher, C., & Roth, D. (2011). Starting from scratch in semantic role labeling: Early indirect supervision. In Poibeau et al. (2011).
19. Cook, P., & Stevenson, S. (2006). Classifying particle semantics in English verb-particle constructions. In *Proceedings of the COLING-ACL Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties* (pp. 45–53), Sydney.
20. Cowie, A. P. (1981). The treatment of collocations and idioms in learner's dictionaries. *Applied Linguistics*, II(3), 223–235.
21. Deane, P. (2005). A nonparametric method for extraction of candidate phrasal terms. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)* (pp. 605–613), Ann Arbor.

22. Devereux, B. J. & Costello, F. J. (2011). Learning to interpret novel noun-noun compounds: Evidence from category learning experiments. In Poibeau et al. (2011).
23. Dominey, P. F., & Inui, T. (2004). A developmental model of syntax acquisition in the construction grammar framework with cross-linguistic validation in English and Japanese. In *Proceedings of the ACL'04 Workshop on Psycho-Computational Models of Human Language Acquisition* (pp. 33–40), Geneva.
24. Dras, M. (1995). Automatic identification of support verbs: A step towards a definition of semantic weight. In *Proceedings of the Eighth Australian Joint Conference on Artificial Intelligence* (pp. 451–458). Singapore: World Scientific.
25. Dras, M., & Johnson, M. (1996). Death and lightness: Using a demographic model to find support verbs. In *Proceedings of the Fifth International Conference on the Cognitive Science of Natural Language Processing* (pp. 165–172), Dublin.
26. Everaert, M., van der Linden, E. -J., Schenk, A., & Schreuder, R. (Eds.). (1995). *Idioms: Structural and psychological perspectives*. Hillsdale: Lawrence Erlbaum Associates.
27. Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook*. Berlin: Mouton de Gruyter. Article 58.
28. Evert, S., Heid, U., & Spranger, K. (2004). Identifying morphosyntactic preferences in collocations. In *Proceedings of the 4th Int'l Conference on Language Resources and Evaluation* (pp. 907–910), Lisbon.
29. A. Fazly. (2007). *Automatic acquisition of lexical knowledge about multiword predicates*. Ph.D. in Computer Science, University of Toronto.
30. Fazly, A., & Stevenson, S. (2007). Distinguishing subtypes of multiword expressions using linguistically-motivated statistical measures. In *Multiword Expression'07: Proceedings of the Workshop on a Broader Perspective on Multiword Expressions* (pp. 9–16), Prague. Association for Computational Linguistics.
31. Fazly, A., Stevenson, S., & North, R. (2007). Automatically learning semantic knowledge about multiword predicates. *Journal of Language Resources and Evaluation*, 41(1), 61–89.
32. Fazly, A., Nematzadeh, A., & Stevenson, S. (2009). Acquiring multiword verbs: The role of statistical evidence. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam.
33. Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34, 1017–1063.
34. Fellbaum, C. (1993). *The determiner in English idioms* (pp. 271–295). Hillsdale: Lawrence Erlbaum Associates.
35. Fellbaum, C. (Ed.). (1998). *WordNet, an electronic lexical database*. Cambridge/London: MIT Press.
36. Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5(1), 55–64.
37. Frank, M., Goodman, N., & Tenenbaum, J. B. (2007). A Bayesian framework for cross-situational word-learning. In *Advances in Neural Information Processing Systems*. Cambridge/London: MIT
38. Gentner, D., & France, I. M. (2004). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 343–382). San Mateo: Kaufmann.
39. Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17(8), 684–691.
40. Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
41. Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
42. Grant, L. E. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, 10(4), 429–451.

43. Grefenstette, G., & Teufel, S. (1995). Corpus-based method for automatic identification of support verbs for nominalization. In *Proceedings of the 7th Meeting of the European Chapter of the Association for Computational Linguistics (EACL'95)* (pp. 98–103), Dublin.
44. Israel, M. How children get constructions. In M. Fried & J. -O. Ostman (Eds.), *Pragmatics in construction grammar and frame semantics*. John Benjamins. (submitted)
45. Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional? *Lexicology*, 3(1), 273–318.
46. Kearns, K. (2002). Light verbs in English. unpublished manuscript. <http://www.ling.canterbury.ac.nz/people/kearns.html>.
47. Krott, A., Gagne, C., & Nicoladis, E. (2009). How the parts relate to the whole: Frequency effects on childrens interpretations of novel compounds. *Journal of Child Language*, 36(01), 85–112.
48. Kytö, M. (1999). *Collocational and idiomatic aspects of verbs in Early Modern English* (pp. 167–206). Amsterdam/Philadelphia: John Benjamins Publishing Company.
49. Xiaowei, P. Li, & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581–612.
50. Lin, D. (1999). Automatic identification of non-compositional phrases. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 317–324), College Park. Association for Computational Linguistics.
51. Lin, T. -H. (2001). *Light verb syntax and the theory of phrase structure*. Ph.D. thesis, University of California, Irvine.
52. MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. The Database* (3rd ed., Vol. 2). Mahwah: Lawrence Erlbaum Associates.
53. McCarthy, D., Keller, B., & Carroll, J. (2003). Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment* (pp. 73–80), Sapporo.
54. Miyamoto, T. (2000). *The light verb construction in Japanese: The role of the verbal noun*. Amsterdam/Philadelphia: John Benjamins.
55. Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. New York: Oxford University Press.
56. Naigles, L., & Kako, E. T. (1993). First contact in verb acquisition: Defining a role for syntax. *Child Development*, 64, 1665–1687.
57. Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, 86, 314–329.
58. Newman, J. (1996). *Give: A cognitive linguistic study*. Berlin/New York: Mouton de Gruyter.
59. Newman, J., & Rice, S. (2004). Patterns of usage for English SIT, STAND, and LIE: A cognitively inspired exploration in corpus linguistics. *Cognitive Linguistics*, 15(3), 351–396.
60. Onnis, L., Roberts, M., & Chater, N. (2002). Simplicity: A cure for overgeneralizations in language acquisition. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 720–725), Fairfax.
61. Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Proceeding of the 32nd Annual Meeting of the Cognitive Science Society*, Austin.
62. Pauwels, P. (2000). *Put, set, lay and place: A cognitive linguistic approach to verbal meaning*. Munich: Lincom Europa.
63. Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37(3), 607–642.
64. Quochi, V. (2007). *A usage-based approach to light verb constructions in Italian: Development and use*. Ph.D. thesis, Universit'a di Pisa.
65. Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29, 819–865.
66. Riehemann, S. (2001). *A constructional approach to idioms and word formation*. Ph.D. thesis, Stanford University, Stanford.

67. Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'02)* (pp. 1–15), Mexico City, Mexico.
68. Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2007). High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the ACL'07 Workshop on Cognitive Aspects of Computational Language Acquisition*, Prague.
69. Sakas, W., & Fodor, J. D. (2001). The structural triggers learner. In S. Bertolo (Eds.), *Language acquisition and learnability*, (172–233). Cambridge: Cambridge University Press.
70. Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, 24, 777–803
71. Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1), 143–177.
72. Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language*, 83, 227–236.
73. Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb 'go'. *Journal of Child Language*, 29, 783–811.
74. Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
75. Venkatapathy, S., & Joshi, A. (2005). Measuring the relative compositionality of verb-noun (V-N) collocations by integrating features. In *Proceeding of HLT-EMNLP'05* (pp. 899–906), Vancouver.
76. Wierzbicka, A. (1982). Why can you Have a Drink when you can't *Have an Eat? *Language*, 58(4), 753–799.
77. Yu, C., & Smith, L. B. (2006). Statistical cross-situational learning to build word-to-world mappings. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver.