

# Acquiring Multiword Verbs: The Role of Statistical Evidence

Afsaneh Fazly, Aida Nematzadeh, and Suzanne Stevenson

Department of Computer Science

University of Toronto

{afsaneh, aida, suzanne}@cs.toronto.edu

## Abstract

In addition to words and grammar, young children learn a large number of multiword sequences that are semantically idiosyncratic and have particular syntactic behaviour, e.g., expressions formed from the combination of a verb and a noun, such as *take the train* and *give a kiss*. Given the high degree of polysemy of verbs that commonly participate in such constructions, an important question is what cues children use to identify (non-literal) multiword combinations. We provide evidence that certain statistical cues tapping into the properties of non-literal expressions are useful in separating these from literal combinations. Moreover, our experiments on naturally occurring child-directed data show that these cues are easily extractable from the input children receive.

## 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 suggest its treatment as a lexical unit, but also behaves as a syntactic phrase (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, and collocations (e.g., Cowie, 1981; Moon, 1998). Psycholinguists have also shown the importance of contingent frequency effects among words and syntactic patterns in the learning and processing of language (Nation et al., 2003; Sosa & MacFarlane, 2002).

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., Goldberg, 1995; Tomasello, 2003). However, *computational modelling* of language acquisition has continued to focus on various aspects of word learning (e.g., Li et al., 2007; Regier, 2005; Yu & Smith, 2006), or grammar learning (e.g., Clark, 2001; Sakas & Fodor, 2001), with work on intermediate constructions mostly limited to identifying general properties of verb argument usages (e.g., Alishahi & Stevenson, 2008; Chang, 2004; Dominey & Inui, 2004).

Thus there is a gap in the study of child language acquisition that has largely left unaddressed questions about the computational mechanisms that underlie how the child learns to identify multiword lexemes (i.e., recognizing that meaning is associated with a group of words, rather than single

words plus combinatory rules), and how she determines what syntactic manipulations of them are valid. In contrast, there has been significant work in computational linguistics on this very topic, with development of statistical measures for identifying multiword lexemes in a corpus, and for extracting their usage properties (e.g., Evert et al., 2004; Fazly et al., 2007). Our goal 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 (Goldberg, 1995). For example, consider the following conversation from the CHILDES database (Brown, 1973, 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, 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. 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 (Goldberg, 1995), 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 must recognize and deal with.

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 turn*) are in principle learnable based on statistics over usages of such expressions (Fazly & Stevenson, 2007; Venkatapathy & Joshi, 2005). However, such work depends on very large amounts of data (from corpora on the order of 100M 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. In this way, we take a first step toward computational modelling of acquisition of the various 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.

### 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 humans, 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. We focus here on expressions of the form  $V_{\text{basic}}+N_{\text{dobj}}$  (or simply  $V_{\text{b}}+N$ ), 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 turn*).

Multiword lexemes of the form  $V_{\text{b}}+N$  are very frequent in many languages (e.g., Cowie, 1981), and they 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. Thus,  $V_{\text{b}}+N$  combinations lie 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. Four possible  $V_{\text{b}}+N$  classes follow, with an indication of the semantic contribution of the components:

1. *give (me) the lion* (**literal combination or LIT**)
  - give: physical transfer of possession
  - NP: typically a physical entity
2. *give (her) time* (**abstract combination or ABS**)
  - give: abstract transfer or allocation
  - NP: often abstract meaning
3. *give (the doll) a bath* (**light verb construction or LVC**)
  - give: convey/conduct an action
  - NP: predicative meaning
4. *give (me) the slip* (**idiomatic combination or IDM**)
  - give, NP: no/highly abstract contribution

These classes are important in the context of child language acquisition because there is a clear connection between each class and the meaning of the expressions in the class. This relation could enable a language learner to make 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 they recognize that this is likely an LVC, then they can infer that it roughly means *shout*.<sup>1</sup>

<sup>1</sup>The LVC construction is associated with additional meanings, such as the boundedness of the action.

These classes of expressions have differing linguistic behaviours that can be cues to the underlying distinctions among the classes (Fazly & Stevenson, 2007). 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.

### Usage-based Measures

As a first attempt at distinguishing the different classes of  $V_{\text{b}}+N$  combinations, here we focus on separating literal combinations (items in the LIT class) from the non-literal ones (items in the ABS, LVC, and IDM classes). We are mainly concerned about two of the non-literal  $V_{\text{b}}+N$  classes, namely ABS and LVC, as these are more commonly found in child-directed speech (CDS). We examine some of the salient linguistic properties of these classes, and explain how these clues might be used by a language learner. For each property, we devise simple, frequency-based, statistical measures that draw on the discussed properties of non-literal  $V_{\text{b}}+N$ s.

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 data (e.g., Evert et al., 2004; Fazly et al., 2007). Given the hypothesized importance of simplicity in language learning (cf. Onnis et al., 2002), 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 CDS.

We elaborate here on three groups of relevant properties: the degree of entrenchment of a  $V_{\text{b}}+N$ , the semantic properties of its noun component, and the degree of syntactic fixedness of the combination.

### Frequency and Association

Non-literal  $V_{\text{b}}+N$ s often have an entrenched, collocational status, leading to generally higher frequencies of occurrence than for particular literal combinations (Evert, 2008). This is partly due to the fact that a non-literal  $V_{\text{b}}+N$  tends to be limited in terms of its noun possibilities, which are restricted by the particular metaphorical sense of the verb. In contrast, a literal  $V_{\text{b}}+N$  is a freer combination with a wider range of possibilities for the noun component. The frequency of a  $V_{\text{b}}+N$  may thus provide a clue to a language learner that helps them decide whether the combination is literal or non-literal.

Computational linguists have developed numerous co-occurrence measures that draw on sophisticated statistical calculations (Evert, 2008). In order to reflect the cognitive constraints on a child, here we look simply at the frequency of co-occurrence of  $v$  and  $n$ , as in:

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

where  $\text{gr} = \text{dobj}$  specifies that the grammatical relationship between  $v$  and  $n$  is direct object.

Although we generally expect non-literal combinations to have higher frequency than literal ones, some literal combinations also appear frequently in the speech children receive. However, the noun of a non-literal combination is often an “unusual” direct object that does not appear with many verbs, due to its abstract nature (e.g., *kiss*). In contrast, the noun of a literal combination is typically a concrete entity that may appear as the direct object of many different verbs. In addition to co-occurrence frequency, we thus also measure the strength of association between the verb and the noun components of a target combination, using the conditional probability of  $v$  given  $n$ . This measure is expected to determine the extent to which  $n$  is associated with  $v$  as opposed to other verbs:

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

where the denominator is the frequency of  $n$  appearing as the direct object of any verb. This measure requires the child merely to keep track of the same frequency as in (1) (which is also the numerator of (2)), but to compare it to the use of the noun with verbs overall (the denominator of (2)).

The association of a noun with particular basic verbs is clearly related to the semantic properties of the noun, which we turn to next.

### Semantic Properties of the Noun

There is evidence that children might be sensitive to the semantic differences between the noun in a literal versus non-literal  $V_b+N$  (Quochi, 2007). For example, whereas the noun in a non-literal  $V_b+N$  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*). The semantic properties of the noun in a  $V_b+N$  thus may provide a child language learner with an important cue as to whether the combination is literal or non-literal. Earlier work has used WordNet<sup>2</sup> to estimate non-referentiality and predicativeness by looking at the noun’s position in the taxonomy, and its morphological relation to a verb (Fazly & Stevenson, 2007). However, WordNet’s conceptual and lexical organization most likely does not reflect that of a child. Next, we explain two measures that instead capture these properties with simple statistics over the surface behavior of the noun.

**Non-Referentiality:** The referential status of a noun is related to its abstractness, which in turn relates to the participation of the noun in certain syntactic forms (Grant, 2005). Most prominently, a non-referential (abstract) noun tends to be preceded by an indefinite determiner (such as *a/an*) or no determiner. Here we assume that a noun is non-referential to the extent that it prefers this pattern of determiner use over others. To estimate the non-referentiality of the noun in a target  $\langle v, n \rangle$ , we thus look into the likelihood of the occurrences

<sup>2</sup><http://wordnet.princeton.edu/>

of  $n$  in the pattern  $pt_{nref} = \langle \text{det}:a/an/\text{NULL } n \rangle$ :

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

where  $\text{freq}(n, pt_{nref})$  is the frequency of occurrence of  $n$  in pattern  $pt_{nref}$ , and the denominator estimates the frequency of  $n$  in any pattern.<sup>3</sup> Again, this is a simple relative frequency for the child to determine: of the instances she sees of this noun, what proportion are in this particular pattern.<sup>4</sup>

**Predicativeness:** In estimating predicativeness, we assume that a child is restricted to a simple observation of how frequently the noun form  $n$  is also observed as a verb (e.g., noting that *push* in *give a push* is also used frequently as a verb in CDS). That is:

$$\text{Pred}(n) \doteq \text{freq}(v'), \text{ where } \text{form}(v') = \text{form}(n) \quad (4)$$

### Degree of Syntactic Fixedness

Most non-literal combinations tend to be restricted with respect to their syntactic expression (Cowie, 1981; Everaert et al., 1995). For example, an LVC such as *give a push* tends to mainly appear with the indefinite determiner *a* and a singular noun *push*. An ABS combination such as *give a minute* also tends to be restricted in its syntactic patterns; e.g., most speakers would consider *give me minutes* or *give me the (next) minute* as odd. In contrast, a literal phrase such as *give a banana* is much more flexible with respect to the determiner introducing the noun and/or the number of the noun — e.g., one can *give another banana*, *give the banana*, or *give two bananas*. Children are sensitive to the syntactic behaviour of both words and constructions (e.g., Goldberg, 1995; Tomasello, 2003). It is thus plausible that children recognize the syntactic fixedness of non-literal  $V_b+N_s$ , and use it as a clue for identifying these combinations.

Most LVC and ABS expressions appear in the form  $pt_{pref} = \langle v \text{ det}:a/an/\text{NULL } n \rangle$ .<sup>5</sup> Measures of this type of syntactic fixedness have required keeping track of probability distributions over a wide range of items and patterns (Banard, 2007; Fazly & Stevenson, 2007). Here, we estimate the degree of syntactic fixedness of a target combination,  $\langle v, n \rangle$ , with a much simpler measure — the relative frequency of  $\langle v, n \rangle$  in the preferred pattern:

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

Given the item-based nature of children’s verb-argument learning (Tomasello, 2003), we expect this to be a reasonable calculation for children.

<sup>3</sup>To estimate these frequencies, we look at all occurrences of a noun irrespective of its grammatical relation to a verb.

<sup>4</sup>It has been shown that children learn the semantic category of a word using the syntactic patterns the word appears in (Brown, 1957).

<sup>5</sup>The noun is in the same pattern as for NRef; the difference is that here the focus is on the degree to which the particular  $V+N$  combination leads to the use of that pattern for the noun.

Table 1: Breakdown of experimental expressions (freq  $\geq 5$ ).

$V_b$	Total	LIT	ABS	LVC	ABS+LVC
<i>take</i>	108	77	18	13	31
<i>give</i>	92	75	7	10	17
<i>take and give</i>	200	152	25	23	48

## Experimental Setup

**Corpus.** We use a subset of the American English portion of the CHILDES database (MacWhinney, 2000), automatically parsed using the parser of Sagae et al. (2007). We exclude 16 of the corpora from our analysis either because they contain no child-directed transcription (and we limit our statistics to CDS), or because they contain data that is explicitly racially and/or socio-economically distinguished (and we are concerned that the language used might therefore show different statistical patterns). Our final corpus of child-directed speech contains about 600,000 utterances, with around 3.2 million words (including punctuation).

**Experimental expressions.** From the CDS portion of our corpus, we extracted  $V_b$ +Ns with a frequency of 5 or higher, formed from either of the two basic verbs *take* and *give* (in all inflected forms). A native speaker of English (an author of this paper) annotated each expression (*not* the individual instances of the expression) as literal (LIT), abstract (ABS), light verb construction (LVC), or idiom (IDM). Expressions for which different instances may belong to different categories were given a single category according to the annotator’s assessment of the predominant usage. Erroneous expressions (and the single idiom in the set) were then removed. Table 1 shows the number of expressions in each individual category, as well as the total number of non-literal expressions (ABS and LVC combined).

**Evaluation.** Each of our statistical 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 the effectiveness of these measures, we 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. To evaluate the clusters, we assign to each 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 that have the same label as the cluster that are actually placed in that cluster.

The clustering results show the effectiveness of the measures working together to separate non-literal from literal

combinations. We further analyze each individual measure to better understand how relevant it is to the acquisition of multiword lexemes. Recall that the measures are designed such that each is expected to be higher for the multiword (non-literal) expressions than for literal ones. 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 expressions using *take*, using *give*, and using either *take* or *give*. We use a standard evaluation metric, namely average precision (*AvgPrec*), to summarize the performance of each measure at this task. *AvgPrec* reflects the goodness of a measure in placing expressions from the target classes (ABS and LVC) before those from the other (LIT), and is calculated as the average of *precision* scores at different thresholds. That is, for a given measure, we consider a threshold at each unique value assigned by that measure, and calculate the proportion of ABS and LVC expressions with a score higher than the threshold; *AvgPrec* is the average of these proportions over all 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 1000 times and report the average of the *AvgPrec* values for each of these random lists as our baseline.

## Results

We first look at the results of clustering the combined set of experimental expressions using all five measures. We then analyze the goodness of each individual measure in separating the two types of expressions, both on the combined set, as well as on *take* and *give* expressions separately.

### Clustering the Expressions

Table 2 shows the results of clustering on the combined set of expressions in two conditions: considering all 200 expressions, and considering only the 98 high-frequency ones — those with a frequency 10 and higher. (Number of expressions of each category in the high-frequency sets are: LIT: 38, ABS: 8, and LVC: 11 for *take*, and LIT: 30, ABS: 4, and LVC: 7 for *give*.) In both conditions, the clustering algorithm is successful at identifying literals (*Acc* of 86% and *Comp* of 92% and 94%). The clustering has a harder time identifying ABS and LVC expressions. The *Acc* results are mixed (68% on all expressions and 83% on high-frequency expressions). However, *Comp* is rather low in both conditions, indicating that there are a good number of ABS and/or LVC expressions which are mixed with literals (those that are put in the LIT cluster). A closer look shows that most of the non-literal expressions that are mixed with the literal ones are of type ABS. Thus, our measures seem to mainly capture properties characteristic of LVCs.

### Ranking Using the Individual Measures

Table 3 shows the performance of the individual measures and that of the baseline in a variety of conditions. We first

Table 2: Clustering results (*Acc* and *Comp*).  $C_i$  represents Cluster  $i$ , Label is the label assigned to a cluster, which is the label of the majority class in the cluster.

On 200 expressions with $freq \geq 5$						
	LVC	ABS	LIT	Label	<i>Acc</i>	<i>Comp</i>
$C_1$	5	18	140	LIT	86%	92%
$C_2$	18	7	12	ABS+LVC	68%	52%

  

On 98 expressions with $freq \geq 10$						
	LVC	ABS	LIT	Label	<i>Acc</i>	<i>Comp</i>
$C_1$	1	9	64	LIT	86%	94%
$C_2$	17	3	4	ABS+LVC	83%	67%

Table 3: Performance (*AvgPrec*) of the individual measures on expressions with  $freq \geq 5$ , and on those with  $freq \geq 10$ .

On all expressions ( $freq \geq 5$ )			
Measure	<i>take</i>	<i>give</i>	<i>take and give</i>
Baseline	.29 ± .04	.19 ± .04	.24 ± .03
Cooc	.53	.38	.51
CProb	.65	.47	.56
NRef	.50	.32	.40
Pred	.60	.57	.62
Fixedness	.57	.31	.43

  

On high-frequency expressions ( $freq \geq 10$ )			
Measure	<i>take</i>	<i>give</i>	<i>take and give</i>
Baseline	.33 ± .06	.27 ± .07	.31 ± .05
Cooc	.57	.41	.54
CProb	.71	.57	.64
NRef	.63	.47	.56
Pred	.67	.66	.68
Fixedness	.86	.49	.66

look into the performances on all expressions (top panel of the table); performances on the high-frequency expressions (bottom panel) have a similar trend. All our measures outperform the baseline, with the best measures overall being Pred and CProb. Fixedness performs well on *take*, but not as well on *give*, perhaps because of the more complex syntactic behaviour of *give* (e.g., its common appearance in the double-object syntactic construction). NRef is the measure with the poorest performance, indicating that it is important to examine the syntactic behaviour of a noun within the context of a  $V_b+N$  combination (especially for expressions with *take*).

The results show that, generally, all measures are good at separating non-literal expressions from literal ones. To further examine whether the measures differ with respect to their identification of ABS and LVC classes, we look at the *AvgPrec* when only one of the classes (ABS or LVC) is considered as target. This shows whether both classes are identified with similar precision, or whether there is a difference between the two. In all cases (for all five measures, on *take* expressions and on *give* expressions), *AvgPrec* values for LVC as target are better (in most cases, much better) than those for ABS as tar-

get, even though LVC and ABS have similar baselines (results are not shown).

Frequency has been shown to have a facilitating effect in the acquisition of linguistic constructions by children. Here, we examine the performance of our measures on high-frequency expressions to see whether we observe a similar facilitating effect on the measures (bottom panel of Table 3). Only two of the measures (Fixedness and NRef) show a notable improvement on high-frequency expressions.<sup>6</sup> These two measures summarize the syntactic behaviour of a word or a combination by examining all their usages. For high-frequency expressions (with more usages), it is possible that the evidence available for these measures is more reliable, resulting in better performance.

## Discussion

Our results show that, for expressions with both *take* and *give*, two measures achieve the best performance: Pred (the frequency of usages of the noun as a verb) and CProb (the proportion of the usages of the noun as a direct object, with this particular verb). Although the expressions under study are verb-based, the best-performing measures are thus ones that draw on very simple frequencies regarding the noun.

The success of Pred at identifying non-literal expressions is an indication that, for many such expressions, the noun component appears frequently as a verb in the input children receive. Because the basic verbs in these  $V_b+N_s$  can take on a wide range of meanings, it is not surprising that the predicativeness of the noun would be such a highly salient cue in indicating whether the verb is being used literally or metaphorically. The formulation of Pred is thus a very simple cue that children could use to help them identify multiword lexemes.

The good performance of CProb suggests that many non-literal expressions exhibit collocational behaviour — that is, the frequency of co-occurrence of the two components is markedly high. However, CProb consistently outperforms raw co-occurrence frequency (Cooc), showing that it is not simply collocational behaviour that is key to identifying these multiword lexemes. Although we normally think of verbs as selecting for nouns, it is clear from the behaviour of CProb that, given the polysemy of basic verbs, the noun tends to select for an appropriate verb in these combinations.

The performance of our individual measures is generally better on expressions with *take* than on those with *give*. (This is also true of the clustering results, though these results were not reported.) This is the case even on high-frequency expressions with comparable baselines for the two expression sets. These results predict that children may have a harder time learning expressions involving *give* than *take*. The Fixedness measure especially performs well on *take*, but poorly on *give*, suggesting that the more complex syntactic constructions that *give* appears in (e.g., the double object construction) may cause children difficulty.

<sup>6</sup>*AvgPrec* of the other measures is better on high-frequency expressions; however, the baseline performance is also higher on these.

Our results (both for clustering and for the individual measures) are also generally better on high-frequency expressions, reinforcing that expressions with more usages might be easier to learn. However, the best-performing measures work well even on low-frequency items, indicating that children could form hypotheses about these expressions even with very little data. Our ongoing work focuses on both the above issues — the verb used and the expression frequencies — by investigating the order of acquisition of different expressions by children of different ages.

The results also reveal that the measures are better at separating LVCs from LITs, indicating that our measures mainly tap into properties of LVCs. More research is needed to better understand the distinguishing properties of ABS expressions.

Overall, our results confirm that many statistical cues relevant to the identification of non-literal expressions are available in the input children receive, and are informative. Moreover, the statistical cues in this study are very simple (using only simple frequencies) and thus easily accessible to young children. Future work will need to explore how to embed these measures into a model of word learning, to show in detail how children might identify and learn these types of multiword expressions.

## References

- Alishahi, A., & Stevenson, S. (2008). A computational model for early argument structure acquisition. *Cognitive Sci.*, 32(5), 789–834.
- Bannard, C. (2007). A measure of syntactic flexibility for automatically identifying mwes in corpora. In *Proc. of the ACL wkshp on a broader perspective on mwes* (pp. 1–8).
- Brown, R. (1957). *J. of Abnormal Psych.*, 55(1), 1–5.
- Brown, R. (1973). *A first language: The early stages*. Harvard Univ. Press.
- Chang, N. (2004). Putting meaning into grammar learning. In *Proc. of the ACL'04 wkshp on psycho-computational models of human lang. acquisition* (pp. 17–24).
- Clark, A. (2001). Unsupervised induction of stochastic context free grammars with distributional clustering. In *Proc. of CoNLL* (pp. 105–112).
- Cowie, A. P. (1981). The treatment of collocations and idioms in learner's dictionaries. *Applied Ling.*, II(3), 223–235.
- 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 *Proc. of the ACL'04 wkshp on psycho-computational models of human lang. acquisition* (pp. 33–40).
- Everaert, M., Linden, E.-J. van der, Schenk, A., & Schreuder, R. (Eds.). (1995). *Idioms: Structural and psychological perspectives*. Lawrence Erlbaum Associates, Publishers.
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An International Handbook*. Mouton de Gruyter. (Article 58)
- Evert, S., Heid, U., & Spranger, K. (2004). Identifying morphosyntactic preferences in collocations. In *Proc. of the 4th int'l conf. on lang. resources and evaluation* (pp. 907–910).
- Fazly, A., & Stevenson, S. (2007). Distinguishing subtypes of multiword expressions using linguistically-motivated statistical measures. In *Proc. of the ACL'07 wkshp on a broader perspective on multiword expressions* (pp. 9–16).
- Fazly, A., Stevenson, S., & North, R. (2007). Automatically learning semantic knowledge about multiword predicates. *J. of Lang. Resources and Evaluation*, 41(1), 61–89.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. UChicago Press.
- Grant, L. E. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, 10(4), 429–451.
- Li, P., Xiaowei, & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Sci.*, 31, 581–612.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). MahWah, NJ: Lawrence Erlbaum Associates.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach* (R. W. Bailey, N. Osselton, & G. Stein, Eds.). Oxford Univ. Press.
- 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. *J. of Experimental Child Psych.*, 86, 314–329.
- Onnis, L., Roberts, M., & Chater, N. (2002). Simplicity: A cure for overgeneralizations in language acquisition. In *Proc. of the 24th annual conf. of the Cognitive Sci. Society* (pp. 720–725).
- Quochi, V. (2007). *A usage-based approach to light verb constructions in Italian: Development and use*. Unpublished doctoral dissertation, Universit'a di Pisa.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Sci.*, 29, 819–865.
- Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2007). High-accuracy annotation and parsing of CHILDES transcripts. In *Proc. of the ACL'07 wkshp on cognitive aspects of computational lang. acquisition*.
- Sakas, W., & Fodor, J. D. (2001). The structural triggers learner. In S. Bertolo (Ed.), *Language acquisition and learnability* (pp. 172–233). Cambridge Univ. Press.
- Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Lang.*, 83, 227–236.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard Univ. Press.
- Venkatapathy, S., & Joshi, A. (2005). Measuring the relative compositionality of verb-noun (V-N) collocations by integrating features. In *Proc. of HLT-EMNLP'05* (pp. 899–906).
- Yu, C., & Smith, L. B. (2006). Statistical cross-situational learning to build word-to-world mappings. In *Proc. of the 28th annual conf. of the Cognitive Sci. Society*.