Automatically Distinguishing Literal and Figurative Usages of Highly Polysemous Verbs

Afsaneh Fazly and Ryan North and Suzanne Stevenson
Department of Computer Science
University of Toronto
{afsaneh,ryan,suzanne}@cs.toronto.edu

Abstract

We investigate the meaning extensions of very frequent and highly polysemous verbs, both in terms of their compositional contribution to a light verb construction (LVC), and the patterns of acceptability of the resulting LVC. We develop compositionality and acceptability measures that draw on linguistic properties specific to LVCs, and demonstrate that these statistical, corpus-based measures correlate well with human judgments of each property.

1 Introduction

Due to a cognitive priority for concrete, easily visualizable entities, abstract notions are often expressed in terms of more familiar and concrete things and situations (Newman, 1996; Nunberg et al., 1994). This gives rise to a widespread use of metaphor in language. In particular, certain verbs easily undergo a process of metaphorization and meaning extension (e.g., Pauwels, 2000; Newman and Rice, 2004). Many such verbs refer to states or acts that are central to human experience (e.g., *sit*, *put*, *give*); hence, they are often both highly polysemous and highly frequent. An important class of verbs prone to metaphorization are **light verbs**, on which we focus in this paper.

A light verb, such as *give*, *take*, or *make*, combines with a wide range of complements from different syntactic categories (including nouns, adjectives, and prepositions) to form a new predicate called a **light verb construction** (LVC). Examples of LVCs include:

1. (a) Azin *took a walk* along the river.
(b) Sam *gave a speech* to a few students.
(c) Joan *takes care* of him when I am away.
(d) They *made good* on their promise to win.
(e) You should always *take* this *into account*.

The light verb component of an LVC is "semantically bleached" to some degree; consequently, the semantic content of an LVC is assumed to be determined primarily by the complement (Butt, 2003). Nevertheless, light verbs exhibit meaning variations when combined with different complements. For example, *give* in *give (someone) a present* has a literal meaning, i.e., "transfer of possession" of a THING to a RECIPIENT. In *give a speech*, *give* has a figurative meaning: an abstract entity (*a speech*) is "transferred" to the audience, but no "possession" is involved. In *give a groan*, the notion of transfer is even further diminished.

Verbs exhibiting such meaning variations are widespread in many languages. Hence, successful NLP applications—especially those requiring some degree of semantic interpretation—need to identify and treat them appropriately. While figurative uses of a light verb are indistinguishable on the surface from a literal use, this distinction is essential to a machine translation system, as Table 1 illustrates. It is therefore important to determine automatic mechanisms for distinguishing literal and figurative uses of light verbs.

Moreover, in their figurative usages, light verbs tend to have similar patterns of cooccurrence with semantically similar complements (e.g., Newman, 1996). Each similar group of complement nouns can even be viewed as a possible meaning extension for a light verb. For example, in *give advice*, *give orders*, *give a speech*, etc., *give* contributes a notion of

Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition, pages 38–47,
Ann Arbor, June 2005. ©2005 Association for Computational Linguistics

Sentence in English	Intermediate semantics	Translation in French
Azin gave Sam a book.	(e1/give :agent (a1/“Azin”) :theme (b1/“book”) :recepient (s1/“Sam”))	Azin a donné un livre à Sam. Azin gave a book to Sam.
Azin gave the lasagna a try.	(e2/give-a-try≈try :agent (a1/“Azin”) :theme (l1/“lasagna”))	Azin a essayé le lasagne. Azin tried the lasagna.

Table 1: Sample sentences with literal and figurative usages of *give*.

“abstract transfer”, while in *give a groan*, *give a cry*, *give a moan*, etc., *give* contributes a notion of “emission”. There is much debate on whether light verbs have one highly abstract (underspecified) meaning, further determined by the context, or a number of identifiable (related) subsenses (Pustejovsky, 1995; Newman, 1996). Under either view, it is important to elucidate the relation between possible interpretations of a light verb and the sets of complements it can occur with.

This study is an initial investigation of techniques for the automatic discovery of meaning extensions of light verbs in English. As alluded to above, we focus on two issues: (i) the distinction of literal versus figurative usages, and (ii) the role of semantically similar classes of complements in refining the figurative meanings.

In addressing the first task, we note the connection between the literal/figurative distinction and the degree to which a light verb contributes compositionally to the semantics of an expression. In Section 2, we elaborate on the syntactic properties that relate to the compositionality of light verbs, and propose a statistical measure incorporating these properties, which places light verb usages on a continuum of meaning from literal to figurative. Figure 1(a) depicts such a continuum in the semantic space of *give*, with the literal usages represented as the core.

The second issue above relates to our long-term goal of dividing the space of figurative uses of a light verb into semantically coherent segments, as shown in Figure 1(b). Section 3 describes our hypothesis on the class-based nature of the ability of potential complements to combine with a light verb. At this point we cannot spell out the different figurative meanings of the light verb associated with such classes. We take a preliminary step in proposing a statistical measure of the acceptability of a combination of a light verb and a class of complements, and explore the extent to which this measure can reveal class-based behaviour.

Subsequent sections of the paper present the corpus extraction methods for estimating our compositionality and acceptability measures, the collection of human judgments to which the measures will be compared, experimental results, and discussion.

2 Compositionality of Light Verbs

2.1 Linguistic Properties: Syntactic Flexibility

We focus on a broadly-documented subclass of light verb constructions, in which the complement is an activity noun that is often the main source of semantic predication (Wierzbicka, 1982). Such complements are assumed to be indefinite, non-referential **predicative nominals** (PNs) that are often morphologically related to a verb (see the complements in examples (1a–c) above). We refer to this class of light verb constructions as “LV+PN” constructions, or simply LVCs.

There is much linguistic evidence that semantic properties of a lexical item determine, to a large extent, its syntactic behaviour (e.g., Rappaport Hovav and Levin, 1998). In particular, the degree of compositionality (decomposability) of a multiword expression has been known to affect its participation in syntactic transformations, i.e., its syntactic flexibility (e.g., Nunberg et al., 1994). English “LV+PN” constructions enforce certain restrictions on the syntactic freedom of their noun components (Kearns, 2002). In some, the noun may be introduced by a definite article, pluralized, passivized, relativized, or even *wh*-questioned:

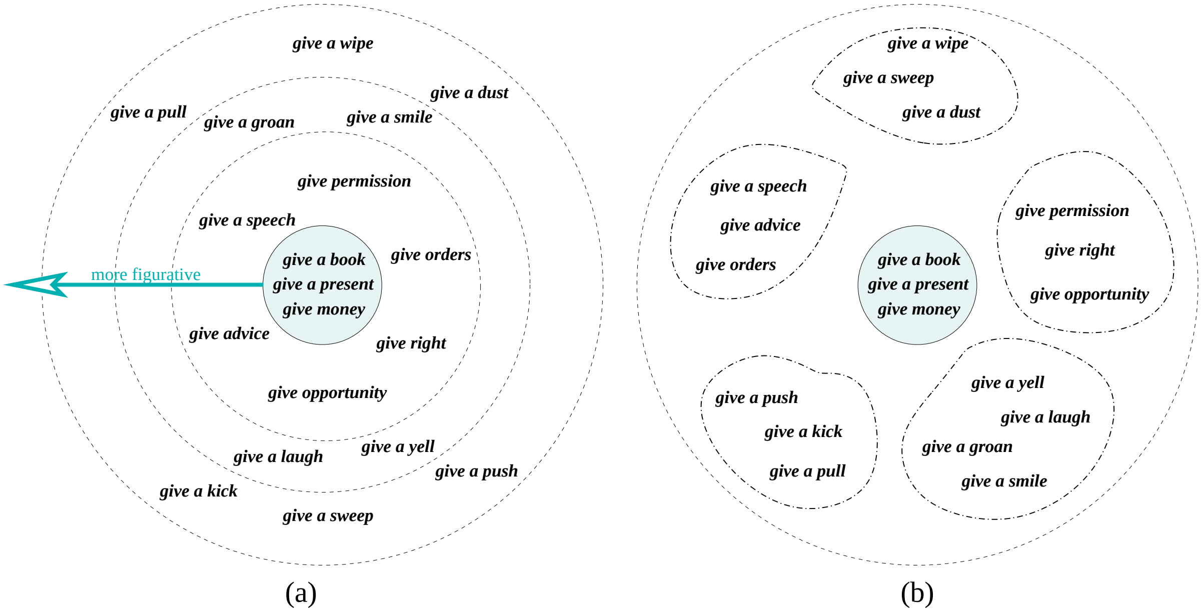

Figure 1: Two possible partitionings of the semantic space of *give*.

2. (a) Azin *gave a speech* to a few students.
   (b) Azin *gave the speech* just now.
   (c) Azin *gave* a couple of *speeches* last night.
   (d) *A speech* was *given* by Azin just now.
   (e) *Which speech* did Azin *give*?

Others have little or no syntactic freedom:

3. (a) Azin *gave a groan* just now.
   (b) * Azin *gave the groan* just now.
   (c) ? Azin *gave* a couple of *groans* last night.
   (d) * *A groan* was *given* by Azin just now.
   (e) * *Which groan* did Azin *give*?

Recall that *give* in *give a groan* is presumed to be a more abstract usage than *give* in *give a speech*. In general, the degree to which the light verb retains aspects of its literal meaning—and contributes them compositionally to the LVC—is reflected in the degree of syntactic freedom exhibited by the LVC. We exploit this insight to devise a statistical measure of compositionality, which uses evidence of syntactic (in)flexibility of a potential LVC to situate it on a scale of literal to figurative usage of the light verb: i.e., the more inflexible the expression, the more figurative (less compositional) the meaning.

2.2 A Statistical Measure of Compositionality

Our proposed measure quantifies the degree of syntactic flexibility of a light verb usage by looking at its frequency of occurrence in any of a set of relevant syntactic patterns, such as those in examples (2) and (3). The measure, $\textsc{Comp}(LV,N)$, assigns a score to a given combination of a light verb (LV) and a noun (N):

$$\begin{aligned}&\textsc{Comp}(LV,N) = \\ &\ \textsc{Assoc}(LV;N) + \\ &\ \textsc{Diff}\ (\textsc{Assoc}(LV;N,PS_{pos}),\textsc{Assoc}(LV;N,PS_{neg}))\end{aligned}$$

That is, the greater the association between LV and N, and the greater the difference between their association with positive syntactic patterns and negative syntactic patterns, the more figurative the meaning of the light verb, and the higher the score.

The strength of the association between the light verb and the complement noun is measured using pointwise mutual information (PMI) whose standard formula is given here:[1]

$$\begin{aligned}\textsc{Assoc}(LV;N) &= \log\frac{\Pr(LV,N)}{\Pr(LV)\ \Pr(N)} \\ &\approx \log\frac{n\,f(LV,N)}{f(LV)\ f(N)}\end{aligned}$$

where n is an estimate of the total number of verb and object noun pairs in the corpus.

[1] PMI is subject to overestimation for low frequency items (Dunning, 1993), thus we require a minimum frequency of occurrence for the expressions under study.

PS_{pos} represents the set of syntactic patterns preferred by less-compositional (more figurative) LVCs (e.g., as in (3a)), and PS_{neg} represents less preferred patterns (e.g., those in (3b–e)). Typically, these patterns most affect the expression of the complement noun. Thus, to measure the strength of association between an expression and a set of patterns, we use the PMI of the light verb, and the complement noun appearing in all of the patterns in the set, as in:

$$\begin{aligned}\textsc{Assoc}(LV;N,PS_{pos}) &= \mathrm{PMI}\,(LV;N,PS_{pos}) \\ &= \log\frac{\Pr(LV,N,PS_{pos})}{\Pr(LV)\,\Pr(N,PS_{pos})} \\ &\approx \log\frac{n\,f(LV,N,PS_{pos})}{f(LV)\,f(N,PS_{pos})}\end{aligned}$$

in which counts of occurrences of N in syntactic contexts represented by PS_{pos} are summed over all patterns in the set. $\textsc{Assoc}(LV;N,PS_{neg})$ is defined analogously using PS_{neg} in place of PS_{pos}.

DIFF measures the difference between the association strengths of the positive and negative pattern sets, referred to as $\textsc{Assoc}_{pos}$ and $\textsc{Assoc}_{neg}$, respectively. Our calculation of ASSOC uses maximum likelihood estimates of the true probabilities. To account for resulting errors, we compare the two confidence intervals, $[\textsc{Assoc}_{pos} \pm \Delta\textsc{Assoc}_{pos}]$ and $[\textsc{Assoc}_{neg} \pm \Delta\textsc{Assoc}_{neg}]$, as in Lin (1999). We take the minimum distance between the two as a conservative estimate of the true difference:

$$\begin{aligned}&\textsc{Diff}(\textsc{Assoc}(LV;N,PS_{pos}),\ \textsc{Assoc}(LV;N,PS_{neg})) \approx \\ &\quad(\textsc{Assoc}_{pos} - \Delta\textsc{Assoc}_{pos}) \\ &\quad-(\textsc{Assoc}_{neg} + \Delta\textsc{Assoc}_{neg})\end{aligned}$$

Taking the difference between confidence intervals lessens the effect of differences that are not statistically significant. (The confidence level, $1-\alpha$, is set to 95% in all experiments.)

3 Acceptability Across Semantic Classes

3.1 Linguistic Properties: Class Behaviour

In this aspect of our work, we narrow our focus onto a subclass of "LV+PN" constructions that have a PN complement in a stem form identical to a verb, preceded (typically) by an indefinite determiner (as in (1a–b) above). Kearns (2002), Wierzbicka (1982), and others have noted that the way in which LVs combine with such PNs to form acceptable LVCs is semantically patterned—that is, PNs with similar semantics appear to have the same trends of cooccurrence with an LV.

Our hypothesis is that semantically similar LVCs—i.e., those formed from an LV plus any of a set of semantically similar PNs—distinguish a figurative subsense of the LV. In the long run, if this is true, it could be exploited by using class information to extend our knowledge of acceptable LVCs and their likely meaning (cf. such an approach to verb particle constructions by Villavicencio, 2003).

As steps to achieving this long-term goal, we must first devise an acceptability measure which determines, for a given LV, which PNs it successfully combines with. We can even use this measure to provide evidence on whether the hypothesized class-based behaviour holds, by seeing if the measure exhibits differing behaviour across semantic classes of potential complements.

3.2 A Statistical Measure of Acceptability

We develop a probability formula that captures the likelihood of a given LV and PN forming an acceptable LVC. The probability depends on both the LV and the PN, and on these elements being used in an LVC:

$$\begin{aligned}&\textsc{Acpt}(LV,PN) \\ &\quad= \Pr(LV,PN,LVC) \\ &\quad= \Pr(PN)\,\Pr(LVC|PN)\,\Pr(LV|PN,LVC)\end{aligned}$$

The first factor, $\Pr(PN)$, reflects the linguistic observation that higher frequency words are more likely to be used as LVC complements (Wierzbicka, 1982). We estimate this factor by $f(PN)/n$, where n is the number of words in the corpus.

The probability that a given LV and PN form an acceptable LVC further depends on how likely it is that the PN combines with *any* light verbs to form an LVC. The frequency with which a PN forms LVCs is estimated as the number of times we observe it in the prototypical "LV a/an PN" pattern across LVs. (Note that such counts are an overestimate, since we cannot determine which usages are indeed LVCs vs. literal uses of the LV.) Since these counts consider the PN only in the context of an indefinite determiner, we normalize over counts of "a/an PN" (noted as *aPN*) to form the conditional probability estimate of the second factor:

$$\Pr(LVC|PN) \approx \frac{\sum_{i=1}^{v} f(LV_i, aPN)}{f(aPN)}$$

where v is the number of light verbs considered.

The third factor, $\Pr(LV|PN, LVC)$, reflects that different LVs have varying degrees of acceptability when used with a given PN in an LVC. We similarly estimate this factor with counts of the given LV and PN in the typical LVC pattern: $f(LV, aPN)/f(aPN)$.

Combining the estimates of the three factors yields:

$$\textsc{Acpt}(LV, PN) \approx$$

$$\frac{f(PN)}{n} \times \frac{\sum_{i=1}^{v} f(LV_i, aPN)}{f(aPN)} \times \frac{f(LV, aPN)}{f(aPN)}$$

4 Materials and Methods

4.1 Light Verbs

Common light verbs in English include *give*, *take*, *make*, *get*, *have*, and *do*, among others. We focus here on two of them, i.e., *give* and *take*, that are frequently and productively used in light verb constructions, and are highly polysemous. The WordNet polysemy count (number of different senses) of *give* and *take* are 44 and 42, respectively.

4.2 Experimental Expressions

Experimental expressions—i.e., potential LVCs using *give* and *take*—are drawn from two sources. The development and test data used in experiments of compositionality (bncD and bncT, respectively) are randomly extracted from the BNC (BNC Reference Guide, 2000), yielding expressions covering a wide range of figurative usages of *give* and *take*, with complements from different semantic categories. In contrast, in experiments that involve acceptability, we need figurative usages of "the same type", i.e., with semantically similar complement nouns, to further examine our hypothesis on the class-based behaviour of light verb combinations. Since in these LVCs the complement is a predicative noun in stem form identical to a verb, we form development and test expressions by combining *give* or *take* with verbs from selected semantic classes of Levin (1993), taken from Stevenson et al. (2004).

4.3 Corpora

We gather estimates for our COMP measure from the BNC, processed using the Collins parser (Collins, 1999) and TGrep2 (Rohde, 2004). Because some LVCs can be rare in classical corpora, our ACPT estimates are drawn from the World Wide Web (the subsection indexed by AltaVista). In our comparison of the two measures, we use web data for both, using a simplified version of COMP. The high level of noise on the web will influence the performance of both measures, but COMP more severely, due to its reliance on comparisons of syntactic patterns.

Web counts are based on an exact-phrase query to AltaVista, with the number of pages containing the search phrase recorded as its frequency.[2] The size of the corpus is estimated at 3.7 billion, the number of hits returned in a search for *the*. These counts are underestimates of the true frequencies, as a phrase may appear more than once in a web page, but we assume all counts to be similarly affected.

4.4 Extraction

Most required frequencies are simple counts of a word or string of words, but the syntactic patterns used in the compositionality measure present some complexity. Recall that PS_{pos} and PS_{neg} are pattern sets representing the syntactic contexts of interest. Each pattern encodes several syntactic attributes: v, the voice of the extracted expression (active or passive); d, the type of the determiner introducing N (definite or indefinite); and n, the number of N (singular or plural). In our experiments, the set of patterns associated with less-compositional use, PS_{pos}, consists of the single pattern with values active, indefinite, and singular, for these attributes. PS_{neg} consists of all patterns with at least one of these attributes having the alternative value.

While our counts on the BNC can use syntactic mark-up, it is not feasible to collect counts on the web for some of the pattern attributes, such as voice. We develop two different variations of the measure, one for BNC counts, and a simpler one for

[2] All searches were performed March 15–30, 2005.

	give		*take*	
Human Ratings	bncD	bncT	bncD	bncT
‘low’	20	10	36	19
‘medium’	35	16	9	5
‘high’	24	10	27	10
Total	79	36	72	34

Table 2: Distribution of development and test expressions with respect to human compositionality ratings.

web counts. We thus subscript COMP with abbreviations standing for each attribute in the measure: COMP$_{vdn}$ for a measure involving all three attributes (used on BNC data), and COMP$_{d}$ for a measure involving determiner type only (used on web data).

5 Human Judgments

5.1 Judgments of Compositionality

To determine how well our proposed measure of compositionality captures the degree of literal/figurative use of a light verb, we compare its scores to human judgments on compositionality. Three judges (native speakers of English with sufficient linguistic knowledge) answered yes/no questions related to the contribution of the literal meaning of the light verb within each experimental expression. The combination of answers to these questions is transformed to numerical ratings, ranging from 0 (fully non-compositional) to 4 (largely compositional). The three sets of ratings yield linearly weighted Kappa values of .34 and .70 for *give* and *take*, respectively. The ratings are averaged to form a consensus set to be used for evaluation.[3]

The lists of rated expressions were biased toward figurative usages of *give* and *take*. To achieve a spectrum of literal to figurative usages, we augment the lists with literal expressions having an average rating of 5 (fully compositional). Table 2 shows the distribution of the experimental expressions across three intervals of compositionality degree, ‘low’ (ratings ≤ 1), ‘medium’ ($1 <$ ratings < 3), and ‘high’ (ratings ≥ 3). Table 3 presents sample expressions with different levels of compositionality ratings.

	Sample Expressions	
Human Ratings	*give*	*take*
‘low’	*give a squeeze*	*take a shower*
‘medium’	*give help*	*take a course*
‘high’	*give a dose*	*take an amount*

Table 3: Sample expressions with different levels of compositionality ratings.

5.2 Judgments of Acceptability

Our acceptability measure is compared to the human judgments gathered by Stevenson et al. (2004). Two expert native speakers of English rated the acceptability of each potential “LV+PN” construction generated by combining *give* and *take* with candidate complements from the development and test Levin classes. Ratings were from 1 (unacceptable) to 5 (completely natural; this was capped at 4 for test data), allowing for “in-between” ratings as well, such as 2.5. On test data, the two sets of ratings yielded linearly weighted Kappa values of .39 and .72 for *give* and *take*, respectively. (Interestingly, a similar agreement pattern is found in our human compositionality judgments above.) The consensus set of ratings was formed from an average of the two sets of ratings, once disagreements of more than one point were discussed.

6 Experimental Results

To evaluate our compositionality and acceptability measures, we compare them to the relevant consensus human ratings using the Spearman rank correlation coefficient, r_s. For simplicity, we report the absolute value of r_s for all experiments. Since in most cases, correlations are statistically significant ($p \ll .01$), we omit p values; those r_s values for which p is marginal (i.e., $.01 \leq p \leq .10$) are subscripted with an “m” in the tables. Correlation scores in boldface are those that show an improvement over the baseline, PMI$_{\text{LVC}}$.

The PMI$_{\text{LVC}}$ measure is an informed baseline, since it draws on properties of LVCs. Specifically, PMI$_{\text{LVC}}$ measures the strength of the association between a light verb and a noun appearing in syntactic patterns preferred by LVCs, i.e., PMI$_{\text{LVC}}$$=$ PMI$(LV;N,PS_{pos})$. Assuming that an acceptable LVC forms a detectable collocation, PMI$_{\text{LVC}}$ can be interpreted as an informed baseline for degree of acceptability. PMI$_{\text{LVC}}$ can also

[3] We asked the judges to provide short paraphrases for each expression, and only use those expressions for which the majority of judges expressed the same sense.

			PMI$_{LVC}$	COMP$_{vdn}$
LV	Data Set	n	r_s	r_s
give	bncT	36	.62	.57
	bncDT	114	.68	**.70**
	bncDT/a	79	.68	**.75**
take	bncT	34	.51	**.59**
	bncDT	106	.52	**.61**
	bncDT/a	68	.63	**.72**

Table 4: Correlations (r_s; n = # of items) between human compositionality ratings and COMP measure (counts from BNC).

	Levin class:	18.1,2	30.3	43.2
LV		n=35	n=18	n=35
give	% fair/good ratings	51	44	54
	log of mean ACPT	-6	-4	-5
take	% fair/good ratings	23	28	3
	log of mean ACPT	-4	-3	-6

Table 5: Comparison of the proportion of human ratings considered "fair" or "good" in each class, and the $\log_{10}$ of the mean ACPT score for that class.

be considered as a baseline for the degree of compositionality of an expression (with respect to the light verb component), under the assumption that the less compositional an expression, the more its components appear as a fixed collocation.

6.1 Compositionality Results

Table 4 displays the correlation scores of the human compositionality ratings with COMP$_{vdn}$, our compositionality measure estimated with counts from the BNC. Given the variety of light verb usages in expressions used in the compositionality data, we report correlations not only on test data (bncT), but also on development and test data combined (bncDT) to get more data points and hence more reliable correlation scores. Compared to the baseline, COMP$_{vdn}$ has generally higher correlations with human ratings of compositionality.

There are two different types of expressions among those used in compositionality experiments: expressions with an indefinite determiner *a* (e.g., *give a kick*) and those without a determiner (e.g., *give guidance*). Despite shared properties, the two types of expressions may differ with respect to syntactic flexibility, due to differing semantic properties of the noun complements in the two cases. We thus calculate correlation scores for expressions with the indefinite determiner only, from both development and test data (bncDT/a). We find that COMP$_{vdn}$ has higher correlations (and larger improvements over the baseline) on this subset of expressions. (Note that there are comparable numbers of items in bncDT and bncDT/a, and the correlation scores are highly significant—very small *p* values—in both cases.)

To explore the effect of using a larger but noisier corpus, we compare the performance of COMP$_{vdn}$ with COMP$_d$, the compositionality measure using web data. The correlation scores for COMP$_d$ on bncDT are .41 and .35, for *give* and *take*, respectively, compared to a baseline (using web counts) of .37 and .32. We find that COMP$_{vdn}$ has significantly higher correlation scores (larger r_s and much smaller *p* values), as well as larger improvements over the baseline. This is a confirmation that using more syntactic information, from less noisy data, improves the performance of our compositionality measure.[4]

6.2 Acceptability Results

We have two goals in assessing our ACPT measure: one is to demonstrate that the measure is indeed indicative of the level of acceptability of an LVC, and the other is to explore whether it helps to indicate class-based patterns of acceptability.

Regarding the latter, Stevenson et al. (2004) found differing overall levels of (human) acceptability for different Levin classes combined with *give* and *take*. This indicates a strong influence of semantic similarity on the possible LV and complement combinations. Our ACPT measure also yields differing patterns across the semantic classes. Table 5 shows, for each light verb and test class, the proportion of acceptable LVCs according to human ratings, and the log of the mean ACPT score for that LV and class combination. For *take*, the ACPT score generally reflects the difference in proportion of accepted expressions according to the human ratings, while for *give*, the measure is less consistent. (The three development classes show the same pattern.) The ACPT measure thus appears to reflect the differing patterns of acceptability across the classes, at least

[4] Using the automatically parsed BNC as a source of less noisy data improves performance. However, since these constructions may be infrequent with any particular complement, we do not expect the use of cleaner but more plentiful text (such as existing treebanks) to improve the performance any further.

LV	Levin Class	n	PMI$_{\text{LVC}}$ r_s	ACPT r_s
give	18.1,2	35	.39$_{\text{m}}$	**.55**
	30.3	18	.38$_{\text{m}}$	**.73**
	43.2	35	.30$_{\text{m}}$	**.34$_{\text{m}}$**
take	18.1.2	35	.57	**.61**
	30.3	18	.55	**.64**
	43.2	35	.43	**.47**

Table 6: Correlations (r_s; n = # of items) between acceptability measures and consensus human ratings (counts from web).

Human Ratings	LV	n	PMI$_{\text{LVC}}$ r_s	ACPT r_s	COMP$_d$ r_s
accept.	*give*	88	.31	**.42**	**.40**
(Levin)	*take*	88	.58	**.61**	.56
compos.	*give*	114	.37	.21$_{\text{m}}$	**.41**
(bncDT)	*take*	106	.32	.30	**.35**

Table 7: Correlations (r_s; n = # of items) between each measure and each set of human ratings (counts from web).

for *take*.

To get a finer-grained notion of the degree to which ACPT conforms with human ratings, we present correlation scores between the two, in Table 6. The results show that ACPT has higher correlation scores than the baseline—substantially higher in the case of *give*. The correlations for *give* also vary more widely across the classes.

These results together indicate that the acceptability measure may be useful, and indeed taps into some of the differing levels of acceptability across the classes. However, we need to look more closely at other linguistic properties which, if taken into account, may improve the consistency of the measure.

6.3 Comparing the Two Measures

Our two measures are intended for different purposes, and indeed incorporate differing linguistic information about LVCs. However, we also noted that PMI$_{\text{LVC}}$ can be viewed as a baseline for both, indicating some underlying commonality. It is worth exploring whether each measure taps into the different phenomena as intended. To do so, we correlate COMP with the human ratings of acceptability, and ACPT with the human ratings of compositionality, as shown in Table 7. (The formulation of the ACPT measure here is adapted for use with determiner-less LVCs.) For comparability, both measures use counts from the web. The results confirm that COMP$_d$ correlates better than does ACPT with compositionality ratings, while ACPT correlates best with acceptability ratings.

7 Discussion and Concluding Remarks

Recently, there has been increasing awareness of the need for appropriate handling of multiword expressions (MWEs) in NLP tasks (Sag et al., 2002). Some research has concentrated on the automatic acquisition of semantic knowledge about certain classes of MWEs, such as compound nouns or verb particle constructions (VPCs) (e.g., Lin, 1999; McCarthy et al., 2003; Villavicencio, 2003). Previous research on LVCs, on the other hand, has primarily focused on their automatic extraction (e.g., Grefenstette and Teufel 1995; Dras and Johnson 1996; Moirón 2004; though see Stevenson et al. 2004).

Like most previous studies that focus on semantic properties of MWEs, we are interested in the issue of compositionality. Our COMP measure aims to identify a continuum along which a light verb contributes to the semantics of an expression. In this way, our work combines aspects of earlier work on VPC semantics. McCarthy et al. (2003) determine a continuum of compositionality of VPCs, but do not distinguish the contribution of the individual components. Bannard et al. (2003), on the other hand, look at the separate contribution of the verb and particle, but assume that a binary decision on the compositionality of each is sufficient.

Previous studies determine compositionality by looking at the degree of distributional similarity between an expression and its component words (e.g., McCarthy et al., 2003; Bannard et al., 2003; Baldwin et al., 2003). Because light verbs are highly polysemous and frequently used in LVCs, such an approach is not appropriate for determining their contribution to the semantics of an expression. We instead examine the degree to which a light verb usage is "similar" to the prototypical LVC, through a statistical comparison of its behaviour within different syntactic patterns. Syntactic flexibility and semantic compositionality are known to be strongly correlated for many types of MWEs (Nunberg et al., 1994). We thus intend to extend our approach to include other polysemous verbs with metaphorical extensions.

Our compositionality measure correlates well with the literal/figurative spectrum represented in human judgments. We also aim to determine finer-grained distinctions among the identified figurative usages of a light verb, which appear to relate to the semantic class of its complement. Semantic class knowledge may enable us to elucidate the types of relations between a light verb and its complement such as those determined in the work of Wanner (2004), but without the need for the manually labelled training data which his approach requires. Villavicencio (2003) used class-based knowledge to extend a VPC lexicon, but assumed that an unobserved VPC is not acceptable. We instead believe that more robust application of class-based knowledge can be achieved with a better estimate of the acceptability of various expressions.

Work indicating acceptability of MWEs is largely limited to collocational analysis using PMI-based measures (Lin, 1999; Stevenson et al., 2004). We instead use a probability formula that enables flexible integration of LVC-specific linguistic properties. Our ACPT measure yields good correlations with human acceptability judgments; indeed, the average increase over the baseline is about twice as high as that of the acceptability measure proposed by Stevenson et al. (2004). Although ACPT also somewhat reflects different patterns across semantic classes, the results clearly indicate the need for incorporating more knowledge into the measure to capture class-based behaviour more consistently.

The work presented here is preliminary, but is the first we are aware of to tie together the two issues of compositionality and acceptability, and relate them to the notion of class-based meaning extensions of highly polysemous verbs. Our on-going work is focusing on the role of the noun component of LVCs, to determine the compositional contribution of the noun to the semantics of the expression, and the role of noun classes in influencing the meaning extensions of light verbs.

References

Baldwin, T., Bannard, C., Tanaka, T., and Widdows, D. (2003). An empirical model of multiword expression decomposability. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 89–96.

Bannard, C., Baldwin, T., and 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*, pages 65–72.

BNC Reference Guide (2000). *Reference Guide for the British National Corpus (World Edition)*, second edition.

Butt, M. (2003). The light verb jungle. Workshop on Multi-Verb Constructions.

Collins, M. (1999). *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.

Dras, M. and 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*.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

Grefenstette, G. and Teufel, S. (1995). Corpus-based method for automatic identification of support verbs for nominalization. In *Proceedings of the 7th Meeting of the EACL*.

Kearns, K. (2002). Light verbs in English. manuscript.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press.

Lin, D. (1999). Automatic identification of non-compositional phrases. In *Proceedings of the 37th Annual Meeting of the ACL*, pages 317–324.

McCarthy, D., Keller, B., and 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*.

Moirón, M. B. V. (2004). Discarding noise in an automatically acquired lexicon of support verb constructions. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*.

Newman, J. (1996). *Give: A Cognitive Linguistic Study*. Mouton de Gruyter.

Newman, J. and 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.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language*, 70(3):491–538.

Pauwels, P. (2000). *Put, Set, Lay and Place: A Cognitive Linguistic Approach to Verbal Meaning*. LINCOM EUROPA.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.

Rappaport Hovav, M. and Levin, B. (1998). Building verb meanings. In Butt and Geuder, editors, *The Projection of Arguments: Lexical and Computational Factors*, pages 97–134. CSLI Publications.

Rohde, D. L. T. (2004). TGrep2 User Manual.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and 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)*, pages 1–15.

Stevenson, S., Fazly, A., and North, R. (2004). Statistical measures of the semi-productivity of light verb constructions. In *Proceedings of the ACL-04 Workshop on Multiword Expressions: Integrating Processing*, pages 1–8.

Villavicencio, A. (2003). Verb-particle constructions and lexical resources. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64.

Wanner, L. (2004). Towards automatic fine-grained semantic classification of verb-noun collocations. *Natural Language Engineering*, 10(2):95–143.

Wierzbicka, A. (1982). Why can you Have a Drink when you can't *Have an Eat? *Language*, 58(4):753–799.