Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing.

Statistical Measures of the Semi-Productivity of Light Verb Constructions

Suzanne Stevenson and **Afsaneh Fazly** and **Ryan North**
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4
Canada
{suzanne,afsaneh,ryan}@cs.toronto.edu

Abstract

We propose a statistical measure for the degree of acceptability of light verb constructions, such as *take a walk*, based on their linguistic properties. Our measure shows good correlations with human ratings on unseen test data. Moreover, we find that our measure correlates more strongly when the potential complements of the construction (such as *walk*, *stroll*, or *run*) are separated into semantically similar classes. Our analysis demonstrates the systematic nature of the semi-productivity of these constructions.

1 Light Verb Constructions

Much research on multiword expressions involving verbs has focused on verb-particle constructions (VPCs), such as *scale up* or *put down* (e.g., Bannard et al., 2003; McCarthy et al., 2003; Villavicencio, 2003). Another kind of verb-based multiword expression is light verb constructions (LVCs), such as the examples in (1).

(1) a. Sara **took a stroll** along the beach.
 b. Paul **gave a knock** on the door.
 c. Jamie **made a pass** to her teammate.

These constructions, like VPCs, may extend the meaning of the component words in interesting ways, may be (semi-)productive, and may or may not be compositional. Interestingly, despite these shared properties, LVCs are in some sense the opposite of VPCs. Where VPCs involve a wide range of verbs in combination with a small number of particles, LVCs involve a small number of verbs in combination with a wide range of co-verbal elements.

An LVC occurs when a *light verb*, such as *take*, *give*, or *make* in (1), is used in conjunction with a complement to form a multiword expression. A verb used as a light verb can be viewed as drawing on a subset of its more general semantic features (Butt, 2003). This entails that most of the distinctive meaning of a (non-idiomatic) LVC comes from the complement to the light verb. This property can be seen clearly in the paraphrases of (1) given below in (2): in each, the complement of the light verb in (1a–c) contributes the main verb of the corresponding paraphrase.[1]

(2) a. Sara **strolled** along the beach.
 b. Paul **knocked** on the door.
 c. Jamie **passed** to her teammate.

The linguistic importance and crosslinguistic frequency of LVCs is well attested (e.g., Butt, 2003; Folli et al., 2003). Furthermore, LVCs have particular properties that require special attention within a computational system. For example, many LVCs (such as those in (1) above) exhibit compositional and semi-productive patterns, while others (such as *take charge*) may be more fixed. Thus, LVCs present the well-known problem with multiword expressions of determining whether and how they should be listed in a computational lexicon. Moreover, LVCs are divided into different classes of constructions, which have distinctive syntactic and semantic properties (Wierzbicka, 1982; Kearns, 2002). In general, there is no one "light verb construction" that can be dealt with uniformly in a computational system, as is suggested by Sag et al. (2002), and generally assumed by earlier computational work on these constructions (Fontenelle, 1993; Grefenstette and Teufel, 1995; Dras and Johnson, 1996). Rather there are different types of LVCs, each with unique properties.

In our initial computational investigation of light verb phenomena, we have chosen to focus on a particular class of semi-productive LVCs in English, exemplified by such expressions as *take a stroll*, *take a run*, *take a walk*, etc. Specifically, we investigate the degree to which we can determine, on the basis of corpus statistics, which words form a valid complement to a given light verb in this type of construction.

[1] The two expressions differ in aspectual properties. It has been argued that the usage of a light verb adds a telic component to the event in most cases (Wierzbicka, 1982; Butt, 2003); though see Folli et al. (2003) for telicity in Persian LVCs.

Our approach draws on a linguistic analysis, presented in Section 2, in which the complement of this type of LVC (e.g., *a walk* in *take a walk*) is—in spite of the presence of the determiner *a*—actually a verbal element (Wierzbicka, 1982; Kearns, 2002). Section 3 describes how this analysis motivates both a method for generalizing over verb classes to find potential valid complements for a light verb, and a mutual information measure that takes the linguistic properties of this type of LVC into account. In Section 4, we outline how we collect the corpus statistics on which we base our measures intended to distinguish "good" LVCs from poor ones. Section 5 describes the experiments in which we determine human ratings of potential LVCs, and correlate those with our mutual information measures. As predicted, the correlations reveal interesting class-based behaviour among the LVCs. Section 6 analyzes the relation of our approach to the earlier computational work on LVCs cited above. Our investigation is preliminary, and Section 7 discusses our current and future research on LVCs.

2 Linguistic Properties of LVCs

An LVC is a multiword expression that combines a light verb with a complement of type noun, adjective, preposition or verb, as in, respectively, *give a speech*, *make good (on)*, *take (NP) into account*, or *take a walk*. The light verb itself is drawn from a limited set of semantically general verbs; among the commonly used light verbs in English are *take*, *give*, *make*, *have*, and *do*. LVCs are highly productive in some languages, such as Persian, Urdu, and Japanese (Karimi, 1997; Butt, 2003; Miyamoto, 2000). In languages such as French, Italian, Spanish and English, LVCs are semi-productive constructions (Wierzbicka, 1982; Alba-Salas, 2002; Kearns, 2002).

The syntactic and semantic properties of the complement of an LVC determine distinct types of constructions. Kearns (2002) distinguishes between two usages of light verbs in LVCs: what she calls a true light verb (TLV), as in *give a groan*, and what she calls a vague action verb (VAV), as in *give a speech*. The main difference between these two types of light verb usages is that the complement of a TLV is claimed to be headed by a verb. Wierzbicka (1982) argues that although the complement in such constructions might appear to be a zero-derived nominal, its syntactic category when used in an LVC is actually a verb, as indicated by the properties of such TLV constructions. For example, Kearns (2002) shows that, in contrast to VAVs, the complement of a TLV usually cannot be definite (3), nor can it be the surface subject of a passive construction (4) or a fronted *wh*-element (5).

(3) a. Jan gave the speech just now.
 b. * Jan gave the groan just now.

(4) a. A speech was given by Jan.
 b. * A groan was given by Jan.

(5) a. Which speech did Jan give?
 b. * Which groan did Jan give?

Because of their interesting and distinctive properties, we have restricted our initial investigation to light verb constructions with TLVs, i.e. "LV a V" constructions, as in *give a groan*. For simplicity, we will continue to refer to them here generally as LVCs. The meaning of an LVC of this type is almost equivalent to the meaning of the verbal complement (cf. (1) and (2) in Section 1). However, the light verb does contribute to the meaning of the construction, as can be seen by the fact that there are constraints on which light verb can occur with which complement (Wierzbicka, 1982). For example, one can *give a cry* but not **take a cry*. The acceptability depends on semantic properties of the complement, and, as we explore below, may generalize in consistent ways across semantically similar (complement) verbs, as in *give a cry*, *give a moan*, *give a howl*; **take a cry*, **take a moan*, **take a howl*.

Many interesting questions pertaining to the syntactic and semantic properties of LVCs have been examined in the linguistic literature: How does the semantics of an LVC relate to the semantics of its parts? How does the type of the complement affect the meaning of an LVC? Why do certain light verbs select for certain complements? What underlies the (semi-)productivity of the creation of LVCs?

Given the crosslinguistic frequency of LVCs, work on computational lexicons will depend heavily on the answers to these questions. We also believe that computational investigation can help to precisely answer the questions as well, by using statistical corpus-based analysis to explore the range and properties of these constructions. While details of the underlying semantic representation of LVCs are beyond the scope of this paper, we address the questions of their semi-productivity.

3 Our Proposal

The initial goal in our investigation of semi-productivity is to find a means for determining how well particular light verbs and complements go together. We focus on the "LV a V" constructions because we are interested in the hypothesis that the complement to the LV is a verb, and think that the properties of this construction may place interesting restrictions on what forms a valid LVC.

3.1 Generalizing over Verb Classes

As noted above, there are constraints in an "LV a V" construction on which complements can occur with particular light verbs. Moreover, similar potential complements pattern alike in this regard—that is, semantically similar complements may have the same pattern of co-occurrence across different light verbs. Since the complement is hypothesized to be a verbal element, we look to verb classes to capture the relevant semantic similarity. The lexical semantic classes of Levin (1993) have been used as a standard verb classification within the computational linguistics community. We thus propose using these classes as the semantically similar groups over which to compare acceptability of potential complements with a given light verb.[2]

Our approach is related to the idea of substitutability in multiword expressions. Substituting pieces of a multiword expression with semantically similar words from a thesaurus can be used to determine productivity—higher degree of substitutability indicating higher productivity (Lin, 1999; McCarthy et al., 2003).[3] Instead of using a thesaurus-based measure, Villavicencio (2003) uses substitutability over semantic verb classes to determine potential verb-particle combinations.

Our method is somewhat different from these earlier approaches, not only in focusing on LVCs, but in the precise goal. While Villavicencio (2003) uses verb classes to generalize over verbs and then confirms whether an expression is attested, we seek to determine how good an expression is. Specifically, we aim to develop a computational approach not only for characterizing the set of complements that can occur with a given light verb in these LVCs, but also to quantify the acceptability.

In investigating light verbs and their combination with complements from various verb semantic classes, we expect that these LVCs are not fully idiosyncratic, but exhibit systematic behaviour. Most importantly, we hypothesize that they show class-based behaviour—i.e., that the same light verb will show distinct patterns of acceptability with complements across different verb classes. We also explore whether the light verbs themselves show different patterns in terms of how they are used semi-productively in these constructions.

We choose to focus on the light verbs *take*, *give*, and *make*. We choose *take* and *give* because they seem similar in their ability to occur in a range of LVCs, and yet they have almost the opposite semantics. We hope that the latter will reveal interesting patterns in occurrence with the different verb classes. On the other hand, *make* seems very different from both *take* and *give*. It seems much less restrictive in its combinations, and also seems difficult to distinguish in terms of light versus "heavy" uses. We expect it to show different generalization behaviour from the other two light verbs.

3.2 Devising an Acceptability Measure

Given the experimental focus, we must devise a method for determining acceptability of LVCs. One possibility is to use a standard measure for detecting collocations, such as pointwise mutual information (Church et al., 1991). "LV a V" constructions are well-suited to collocational analysis, as the light verb can be seen as the first component of a collocation, and the string "a V" as the second component. Applying this idea to potential LVCs, we calculate pointwise mutual information, I(lv; aV).

In addition, we use the linguistic properties of the "LV a V" construction to develop a more informed measure. As noted in Section 2, generally only the indefinite determiner *a* (or *an*) is allowed in this type of LVC. We hypothesize then that for a "good" LVC, we should find a much higher mutual information value for "LV a V" than for "LV [det] V", where [det] is any determiner other than the indefinite. While I(lv; aV) should tell us whether "LV a V" is a good *collocation* (Church et al., 1991), the difference between the two, I(lv; aV) - I(lv; detV), should tell us whether the collocation is an LVC.

To summarize, we assume that:

- if $I(lv; aV) \ll 0$ then
  "LV a V" is likely not a good collocation;
- if $I(lv; aV) - I(lv; detV) \ll 0$ then
  "LV a V" is likely not a true LVC.

In order to capture these two conditions in a single measure, we combine them by using a linear approximation to the two lines given by $I(lv; aV) = 0$ and $I(lv; aV) - I(lv; detV) = 0$. The most straightforward line approximating the combined effect of these two conditions is:

$$2 \times I(lv; aV) - I(lv; detV) = 0$$

We hypothesize that this combined measure—i.e., $2 \times I(lv; aV) - I(lv; detV)$—will correlate bet-

[2] We also need to compare generalizability over semantic noun classes to further test the linguistic hypothesis. We initially performed such experiments on noun classes in WordNet, but, due to the difficulty of deciding an appropriate level of generalization in the hierarchy, we left this as future work.

[3] Note that although Lin characterizes his work as detecting non-compositionality, we agree with Bannard et al. (2003) that it is better thought of as tapping into productivity.

Development Classes		
Levin #	Name	Count
10.4.1*	*Wipe* Verbs, Manner	30
17.1	*Throw* Verbs	30
51.3.2*	*Run* Verbs	30
Test Classes		
Levin #	Name	Count
18.1,2	*Hit* and *Swat* Verbs	35
30.3	*Peer* Verbs	18
43.2*	Sound Emission	35
51.4.2	Motion (non-vehicle)	10

Table 1: Levin classes used in our experiments. A '*' indicates a random subset of verbs in the class.

Determiner	Search Strings
Indefinite	*give/gives/gave a cry*
Definite	*give/gives/gave the cry*
Demons.	*give/gives/gave this/that cry*
Possessive	*give/gives/gave my/.../their cry*

Table 2: Searches for light verb *give* and verb *cry*.

ter with human ratings of the LVCs than the mutual information of the "LV a V" construction alone.

For I(lv; detV), we explore several possible sets of determiners standing in for "det", including *the*, *this*, *that*, and the possessive determiners. We find, contrary to the linguistic claim, that *the* is not always rare in "LV a V" constructions, and the measures excluding *the* perform best on development data.[4]

4 Materials and Methods

4.1 Experimental Classes

Three Levin classes are used for the development set, and four classes for the test set, as shown in Table 1. Each set of classes covers a range of LVC productivity with the light verbs *take*, *give*, and *make*, from classes in which we felt no LVCs were possible with a given LV, to classes in which many verbs listed seemed to form valid LVCs with a given LV.

4.2 Corpora

Even the 100M words of the British National Corpus (BNC Reference Guide, 2000) do not give an acceptable level of LVC coverage: a very common LVC such as *take a stroll*, for instance, is attested only 23 times. To ensure sufficient data to detect less common LVCs, we instead use the Web as our corpus (in particular, the subsection indexed by the Google search engine, http://www.google.com).

Using the Web to overcome data sparseness has been attempted before (Keller et al., 2002); however, there are issues: misspellings, typographic errors, and pages in other languages all contribute to noise in the results. Moreover, punctuation is ignored in Google searches, meaning that search results can cross phrase or sentence boundaries. For instance, an exact phrase search for "take a cry" would return a web page which had the text *It was too much to take. A cry escaped his lips.* When searching for an unattested LVC, these noisy results can begin to dominate. In ongoing work, we are devising some automatic clean-up methods to eliminate some of the false positives.

On the other hand, it should be pointed out that not all "good" LVCs will appear in our corpus, despite its size. In this view we differ from Villavicencio (2003), who assumes that if a multiword expression is not found in the Google index, then it is not a good construction. As an example, consider *The clown took a cavort across the stage*. The LVC seems plausible; however, Google returns no results for "took a cavort". This underlines the need for determining plausible (as opposed to attested) LVCs, which class-based generalization has the potential to support.

4.3 Extraction

To measure mutual information, we gather several counts for each potential LVC: the frequency of the LVC (e.g., *give a cry*), the frequency of the light verb (e.g., *give*), and the frequency of the complement of the LVC (e.g., *a cry*). To achieve broader coverage, counts of the light verbs and the LVCs are collapsed across three tenses: the base form, the present, and the simple past. Since we are interested in the differences across determiners, we search for both the LVC ("give [det] cry") and the complement alone ("[det] cry") using all singular determiners. Thus, for each LVC, we require a number of LVC searches, as exemplified in Table 2, and analogous searches for "[det] V".

All searches were performed using an exact string search in Google, during a 24-hour period in March, 2004. The number of results returned is used as the frequency count. Note that this is an underestimate, since an LVC may occur than once in a single web page; however, examining each document to count the actual occurrences is infeasible, given the number of possible results. The size of the corpus (also

[4] Cf. *I took the hike that was recommended*. This finding supports a statistical corpus-based approach to LVCs, as their usage may be more nuanced than linguistic theory suggests.

needed in calculating our measures) is estimated at 5.6 billion, the number of hits returned in a search for "the". This is also surely an underestimate, but is consistent with our other frequency counts.

NSP is used to calculate pointwise mutual information over the counts (Banerjee and Pedersen, 2003).

5 Experimental Results

In these initial experiments, we compare human ratings of the target LVCs to several mutual information measures over our corpus counts, using Spearman rank correlation. We have two goals: to see whether these LVCs show differing behaviour according to the light verb and/or the verb class of the complement, and to determine whether we can indeed predict acceptability from corpus statistics. We first describe the human ratings, then the correlation results on our development and test data.

5.1 Human Ratings

We use pilot results in which two native speakers of English rated each combination of "LV a V" in terms of acceptability. For the development classes, we used integer ratings of 1 (unacceptable) to 5 (completely natural), allowing for "in-between" ratings as well, such as 2.5. For the test classes, we set the top rating at 4, since we found that ratings up to 5 covered a larger range than seemed natural. The test ratings yielded linearly weighted Kappa values of .72, .39, and .44, for *take*, *give*, and *make*, respectively, and .53 overall.[5]

To determine a consensus rating, the human raters first discussed disagreements of more than one rating point. In the test data, this led to 6% of the ratings being changed. (Note that this is 6% of ratings, not 6% of verbs; fewer verbs were changed, since for some verbs both raters changed their rating after discussion.) We then simply averaged each pair of ratings to yield a single consensus rating for each item.

In order to see differences in human ratings across the light verbs and the semantic classes of their complements, we put the (consensus) human ratings in bins of low (ratings < 2) , medium (ratings ≥ 2, < 3), and high (ratings ≥ 3). (Even a score of 2 meant that an LVC was "ok".) Table 3 shows the distribution of medium and high scores for each of the light verbs and test classes. We can see that some classes generally allow more LVCs across the light verbs (e.g., 18.1,2) than others (e.g, 43.2). Furthermore, the light verbs show very different patterns of acceptability for different classes—e.g., *give* is fairly good with 43.2, while *take* is very bad, and the pattern is reversed for 51.4.2. In general, *give* allows more LVCs on the test classes than do the other two light verbs.

Class #	N	take	give	make
18.1,2	35	8 (23%)	15 (43%)	8 (23%)
30.3	18	5 (28%)	5 (28%)	3 (17%)
43.2	35	1 (3%)	11 (31%)	9 (26%)
51.4.2	10	7 (70%)	2 (20%)	1 (10%)

Table 3: Number of medium and high scores for each LV and class. N is the number of test verbs.

5.2 Correlations with Statistical Measures

Our next step is to see whether the ratings, and the patterns across light verbs and classes, are reflected in the statistical measures over corpus data. Because our human ratings are not normally distributed (generally having a high proportion of values less than 2), we use the Spearman rank correlation coefficient r to compare the consensus ratings to the mutual information measures.[6]

As described in Section 3.2, we use pointwise mutual information over the "LV a V" string, as well as measures we developed that incorporate the linguistic observation that these LVCs typically do not occur with definite determiners. On our development set, we tested several of these measures and found that the following had the best correlations with human ratings:

- **MI:** I(lv; aV)
- **DiffAll:** 2 $\times$ I(lv; aV) - I(lv; detV)

where I(lv; detV) is the mutual information over strings "LV [det] V", and det is any determiner other than *a*, *an*, or *the*. Note that **DiffAll** is the most general of our combined measures; however, some verbs are not detected with other determiners, and thus **DiffAll** may apply to a smaller number of items than **MI**.

We focus on the analysis of these two measures on test data, but the general patterns are the same

[5] Agreement on the development set was much lower (linearly weighted Kappa values of .37, .23, and .56, for *take*, *give*, and *make*, respectively, and .38 overall), due to differences in interpretation of the ratings. Discussion of these issues by the raters led to more consistency in test data ratings.

[6] Experiments on the development set to determine a threshold on the different measures to classify LVCs as good or not showed promise in their coarse match with human judgments. However, we set this work aside for now, since the correlation coefficients are more informative regarding the fine-grained match of the measures to human ratings, which cover a fairly wide range of acceptability.

LV	Class #	MI r	MI (p)	MI N	DiffAll r	DiffAll (p)	DiffAll N
take	18.1,2	.52	(<.01)	34	.51	(<.01)	33
	30.3	.53	(.02)	18	.59	(.02)	15
	43.2	.24	(.20)	31	.32	(.10)	27
	51.4.2	.68	(.03)	10	.65	(.04)	10
	all	.53	(<.01)	93	.52	(<.01)	85
give	18.1,2	.26	(.14)	33	.30	(.10)	32
	30.3	.33	(.20)	17	.27	(.33)	15
	43.2	.38	(.03)	33	.58	(<.01)	25
	51.4.2	.09	(.79)	10	-.13	(.71)	10
	all	.28	(.01)	93	.33	(<.01)	82
make	18.1,2	.51	(<.01)	34	.49	(<.01)	34
	30.3	.16	(.52)	18	-.11	(.68)	17
	43.2	-.12	(.52)	34	-.19	(.29)	33
	51.4.2	-.08	(.81)	10	-.20	(.58)	10
	all	.36	(<.01)	96	.26	(.01)	94

Table 4: Spearman rank correlation coeffcents r, with p values and number of items N, between the mutual information measures and the consensus human ratings, on unseen test data.

on the development set. Table 4 shows the correlation results on our unseen test LVCs. We get reasonably good correlations with the human ratings across a number of the light verbs and classes, indicating that these measures may be helpful in determining which light verb plus complement combinations form valid LVCs. In what follows, we examine more detailed patterns, to better analyze the data.

First, comparing the test correlations to Table 3, we find that the classes with a low number of "good" LVCs have poor correlations. When we examine the correlation graphs, we see that, in general, there is a good correlation between the ratings greater than 1 and the corresponding measure, but when the rating is 1, there is often a wide range of values for the corpus-based measure. One cause could be noise in the data, as mentioned earlier—that is, for bad LVCs, we are picking up too many "false hits", due to the limitations of using Google searches on the web. To confirm this, we examine one development class (10.4.1, the *Wipe* manner verbs), which was expected to be bad with *take*. We find a large number of hits for "take a V" that are not good LVCs, such as "take a strip [of tape/of paper]", "take a pluck[-and-play approach]". On the other hand, some examples with unexpectedly high corpus measures are LVCs the human raters were simply not aware of ("take a skim through the manual"), which underscores the difficulty of human rating of a semi-productive construction.

Second, we note that we get very good correlations with *take*, somewhat less good correlations with *give*, and generally poor correlations with *make*. We had predicted that *take* and *give* would behave similarly (and the difference between *take* and *give* is less pronounced in the development data). We think one reason *give* has poorer correlations is that it was harder to rate (it had the highest proportion of disagreements), and so the human ratings may not be as consistent as for *take*. Also, for a class like 30.3, which we expected to be good with *give* (e.g., *give a look*, *give a glance*), we found that the LVCs were mostly good only in the dative form (e.g., *give her a look*, *give it a glance*). Since we only looked for exact matches to "LV a V", we did not detect this kind of construction.

We had predicted that *make* would behave differently from *take* and *give*, and indeed, except in one case, the correlations for *make* are poorer on the individual classes. Interestingly, the correlation overall attains a much better value using the mutual information of "LV a V" alone (i.e., the **MI** measure). We think that the pattern of correlations with *make* may be because it is not necessarily a "true light verb" construction in many cases, but rather a "vague action verb" (see Section 2). If so, its behaviour across the complements may be somewhat more arbitrary, combining different uses.

Finally, we compare the combined measure **DiffAll** to the mutual information, **MI**, alone. We hypothesized that while the latter should indicate a collocation, the combined measure should help to focus on LVCs in particular, because of their linguistic property of occurring primarily with an indefinite determiner. On the individual classes, when considering correlations that are statistically significant or marginally so (i.e., at the confidence level of 90%), the **DiffAll** measure overall has somewhat stronger correlations than **MI**. Over all complement verbs together, **DiffAll** is roughly the same as **MI** for *take*; is somewhat better for *give,* and is worse for *make*.[7]

Better performance over the individual classes indicates that when applying the measures, at least to *take* and *give,* it is helpful to separate the data according to semantic verb class. For *make,* the appropriate approach is not as clear, since the results on the individual classes are so skewed. In general, the results confirm our hypothesis that semantic verb classes are highly relevant to measuring the acceptability of LVCs of this type. The results also indicate the need to look in more detail at the properties of different light verbs.

6 Related Work

Other computational research on LVCs differs from ours in two key aspects. First, the work has looked at *any* nominalizations as complements of potential light verbs (what they term "support verbs") (Fontenelle, 1993; Grefenstette and Teufel, 1995; Dras and Johnson, 1996). Our work differs in focusing on verbal nouns that form the complement of a particular type of LVC, allowing us to explore the role of class information in restricting the complements of these constructions. Second, this earlier work has viewed *all* verbs as possible light verbs, while we look at only the class of potential light verbs identified by linguistic theory.

The difference in focus on these two aspects of the problem leads to the basic differences in approach: while they attempt to find probable light verbs for nominalization complements, we try to find possible (verbal) noun complements for given light verbs. Our work differs both practically, in the type of measure used, and conceptually, in the formulation of the problem. For example, Grefenstette and Teufel (1995) used some linguistic properties to weed out potential light verbs from lists sorted by raw frequency, while Dras and Johnson (1996) used frequency of the verb weighted by a weak predictor of its prior probability as a light verb. We instead use a standard collocation detection measure (mutual information), the terms of which we modify to capture linguistic properties of the construction.

More fundamentally, our proposal differs in its emphasis on possible class-based generalizations in LVCs that have heretofore been unexplored. It would be interesting to apply this idea to the broader classes of nominalizations investigated in earlier work. Moreover, our approach could draw on ideas from the earlier proposals to detect the light verbs automatically, since the precise set of LVs differs crosslinguistically—and LV status may indeed be a continuum rather than a discrete distinction.

7 Conclusions and Future Work

Our results demonstrate the benefit of treating LVCs as more than just a simple collocation. We exploit linguistic knowledge particular to the "LV a V" construction to devise an acceptability measure that correlates reasonably well with human judgments. By comparing the mutual information with indefinite and definite determiners, we use syntactic patterns to tap into the distinctive underlying properties of the construction.

Furthermore, we hypothesized that, because the complement in these constructions is a verb, we would see systematic behaviour across the light verbs in terms of their ability to combine with complements from different verb classes. Our human ratings indeed showed class-based tendencies for the light verbs. Moreover, our acceptability measure showed higher correlations when the verbs were divided by class. This indicates that there is greater consistency within a verb class between the corpus statistics and the ability to combine with a light verb. Thus, the semantic classes provide a useful way to increase the performance of the acceptability measure.

The correlations are far from perfect, however. In addition to noise in the data, one problem may be that these classes are too coarse-grained. Exploration is needed of other possible verb (and noun) classes as the basis for generalizing the complements of these constructions. However, we must also look to the measures themselves for improving our techniques. Several linguistic properties distinguish these constructions, but our measures only drew on one. In ongoing work, we are exploring methods for incorporating other linguistic behaviours into a measure for these constructions, as well as for LVCs more generally.

We are widening this investigation in other directions as well. Our results reveal interesting differences among the light verbs, indicating that the set of light verbs is itself heterogeneous. More research is needed to determine the properties of a broader

[7] The development data is similar to the test data in favouring **DiffAll** over **MI** across the individual classes. Over all development verbs together, **DiffAll** is somewhat better than **MI** for *take,* is roughly the same for *give,* and is somewhat worse for *make*.

range of light verbs, and how they influence the valid combinations they form with semantic classes.

Finally, we plan to collect more extensive rating data, but are concerned with the difficulty found in judging these constructions. Gathering solid human ratings is a challenge in this line of investigation, but this only serves to underscore the importance of devising corpus-based acceptability measures in order to better support development of accurate computational lexicons.

Acknowledgments

We thank Ted Pedersen (U. of Minnesota), Diana Inkpen (U. of Ottawa), and Diane Massam (U. of Toronto) for helpful advice and discussion, as well as three anonymous reviewers for their useful feedback. We gratefully acknowledge the support of NSERC of Canada.

References

J. Alba-Salas. 2002. *Light Verb Constructions in Romance: A Syntactic Analysis*. Ph.D. thesis, Cornell University.

S. Banerjee and T. Pedersen. 2003. The design, implementation, and use of the Ngram Statistic Package. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*.

C. Bannard, T. Baldwin, and A. Lascarides. 2003. A statistical approach to the semantics of verb-particles. In *Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, p. 65–72.

BNC Reference Guide. 2000. *Reference Guide for the British National Corpus (World Edition)*. http://www.hcu.ox.ac.uk/BNC, second edition.

M. Butt. 2003. The light verb jungle. http://www.ai.mit.edu/people/jimmylin/papers /Butt03.pdf.

K. Church, W. Gale, P. Hanks, and D. Hindle. 1991. *Using Statistics in Lexical Analysis*, p. 115–164. Lawrence Erlbaum.

M. Dras and M. Johnson. 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*, Dublin, Ireland.

R. Folli, H. Harley, and S. Karimi. 2003. Determinants of event type in Persian complex predicates. *Cambridge Working Papers in Linguistics*.

T. Fontenelle. 1993. Using a bilingual computerized dictionary to retrieve support verbs and combinatorial information. *Acta Linguistica Hungarica*, 41(1–4):109–121.

G. Grefenstette and S. Teufel. 1995. A corpus-based method for automatic identification of support verbs for nominalisations. In *Proceedings of EACL*, p. 98–103, Dublin, Ireland.

S. Karimi. 1997. Persian complex verbs: Idiomatic or compositional? *Lexicology*, 3(1):273–318.

K. Kearns. 2002. Light verbs in English. http://www.ling.canterbury.ac.nz/kate /lightverbs.pdf.

F. Keller, M. Lapata, and O. Ourioupina. 2002. Using the Web to overcome data sparseness. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, p. 230–237, Philadelphia, USA.

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

D. Lin. 1999. Automatic identification of non-compositional phrases. In *Proceedings of ACL-99*, p. 317–324.

D. McCarthy, B. Keller, and J. Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*.

T. Miyamoto. 2000. *The Light Verb Construction in Japanese: the role of the verbal noun*. John Benjamins.

I. Sag, T. Baldwin, F. Bond, A. Copestake, and D. Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, p. 1–15.

A. Villavicencio. 2003. Verb-particle constructions in the world wide web. In *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their use in Computational Linguistics Formalisms and Applications*.

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