Measuring the relative compositionality of verb-noun (V-N) collocations by integrating features

Sriram Venkatapathy[1]
Language Technologies Research Centre,
International Institute of Information
Technology - Hyderabad,
Hyderabad, India.
sriram@research.iiit.ac.in

Aravind K. Joshi
Department of Computer and
Information Science and Institute for
Research in Cognitive Science,
University of Pennsylvania,
Philadelphia, PA, USA.
joshi@linc.cis.upenn.edu

Abstract

Measuring the relative compositionality of Multi-word Expressions (MWEs) is crucial to Natural Language Processing. Various collocation based measures have been proposed to compute the relative compositionality of MWEs. In this paper, we define novel measures (both collocation based and context based measures) to measure the relative compositionality of MWEs of V-N type. We show that the correlation of these features with the human ranking is much superior to the correlation of the traditional features with the human ranking. We then integrate the proposed features and the traditional features using a SVM based ranking function to rank the collocations of V-N type based on their relative compositionality. We then show that the correlation between the ranks computed by the SVM based ranking function and human ranking is significantly better than the correlation between ranking of individual features and human ranking.

1 Introduction

The main goal of the work presented in this paper is to examine the relative compositionality of col-

---

[1]Part of the work was done at Institute for Research in Cognitive Science (IRCS), University of Pennsylvania, Philadelphia, PA 19104, USA, when he was visiting IRCS as a Visiting Scholar, February to December, 2004.

locations of V-N type using a SVM based ranking function. Measuring the relative compositionality of V-N collocations is extremely helpful in applications such as machine translation where the collocations that are highly non-compositional can be handled in a special way (Schuler and Joshi, 2004) (Hwang and Sasaki, 2005).

Multi-word expressions (MWEs) are those whose structure and meaning cannot be derived from their component words, as they occur independently. Examples include conjunctions like 'as well as' (meaning 'including'), idioms like 'kick the bucket' (meaning 'die'), phrasal verbs like 'find out' (meaning 'search') and compounds like 'village community'. A typical natural language system assumes each word to be a lexical unit, but this assumption does not hold in case of MWEs (Becker, 1975) (Fillmore, 2003). They have idiosyncratic interpretations which cross word boundaries and hence are a 'pain in the neck' (Sag et al., 2002). They account for a large portion of the language used in day-to-day interactions (Schuler and Joshi, 2004) and so, handling them becomes an important task.

A large number of MWEs have a standard syntactic structure but are non-compositional semantically. An example of such a subset is the class of non-compositional verb-noun collocations (V-N collocations). The class of non-compositional V-N collocations is important because they are used very frequently. These include verbal idioms (Nunberg et al., 1994), support-verb constructions (Abeille, 1988), (Akimoto, 1989), among others. The expression 'take place' is a MWE whereas 'take a gift' is not a MWE.

It is well known that one cannot really make a binary distinction between compositional and non-compositional MWEs. They do not fall cleanly into mutually exclusive classes, but populate the continuum between the two extremes (Bannard et al., 2003). So, we rate the MWEs (V-N collocations in this paper) on a scale from 1 to 6 where 6 denotes a completely compositional expression, while 1 denotes a completely opaque expression.

Various statistical measures have been suggested for ranking expressions based on their compositionality. Some of these are Frequency, Mutual Information (Church and Hanks, 1989) , distributed frequency of object (Tapanainen et al., 1998) and LSA model (Baldwin et al., 2003) (Schutze, 1998). In this paper, we define novel measures (both collocation based and context based measures) to measure the relative compositionality of MWEs of V-N type (see section 6 for details). Integrating these statistical measures should provide better evidence for ranking the expressions. We use a SVM based ranking function to integrate the features and rank the V-N collocations according to their compositionality. We then compare these ranks with the ranks provided by the human judge. A similar comparison between the ranks according to Latent-Semantic Analysis (LSA) based features and the ranks of human judges has been made by McCarthy, Keller and Caroll (McCarthy et al., 2003) for verb-particle constructions. (See Section 3 for more details). Some preliminary work on recognition of V-N collocations was presented in (Venkatapathy and Joshi, 2004).

We show that the measures which we have defined contribute greatly to measuring the relative compositionality of V-N collocations when compared to the traditional features. We also show that the ranks assigned by the SVM based ranking function correlated much better with the human judgement that the ranks assigned by individual statistical measures.

This paper is organized in the following sections **(1)** Basic Architecture, **(2)** Related work, **(3)** Data used for the experiments, **(4)** Agreement between the Judges, **(5)** Features, **(6)** SVM based ranking function, **(7)** Experiments & Results, and **(8)** Conclusion.

2   Basic Architecture

Every V-N collocation is represented as a vector of features which are composed largely of various statistical measures. The values of these features for the V-N collocations are extracted from the British National Corpus. For example, the V-N collocation 'raise an eyebrow' can be represented as [Frequency = 271, Mutual Information = 8.43, Distributed frequency of object = 1456.29, etc.]. A SVM based ranking function uses these features to rank the V-N collocations based on their relative compositionality. These ranks are then compared with the human ranking.

3   Related Work

(Breidt, 1995) has evaluated the usefulness of the Point-wise Mutual Information measure (as suggested by (Church and Hanks, 1989)) for the extraction of V-N collocations from German text corpora. Several other measures like Log-Likelihood (Dunning, 1993), Pearson's χ^2 (Church et al., 1991), Z-Score (Church et al., 1991) , Cubic Association Ratio (MI3), etc., have been also proposed. These measures try to quantify the association of two words but do not talk about quantifying the non-compositionality of MWEs. Dekang Lin proposes a way to automatically identify the non-compositionality of MWEs (Lin, 1999). He suggests that a possible way to separate compositional phrases from non-compositional ones is to check the existence and mutual-information values of phrases obtained by replacing one of the words with a similar word. According to Lin, a phrase is probably non-compositional if such substitutions are not found in the collocations database or their mutual information values are significantly different from that of the phrase. Another way of determining the non-compositionality of V-N collocations is by using 'distributed frequency of object' (DFO) in V-N collocations (Tapanainen et al., 1998). The basic idea in there is that "if an object appears only with one verb (or few verbs) in a large corpus we expect that it has an idiomatic nature" (Tapanainen et al., 1998).

Schone and Jurafsky (Schone and Jurafsky, 2001) applied Latent-Semantic Analysis (LSA) to the analysis of MWEs in the task of MWE discovery, by way

900

of rescoring MWEs extracted from the corpus. An interesting way of quantifying the relative compositionality of a MWE is proposed by Baldwin, Bannard, Tanaka and Widdows (Baldwin et al., 2003). They use LSA to determine the similarity between an MWE and its constituent words, and claim that higher similarity indicates great decomposability. In terms of compositionality, an expression is likely to be relatively more compositional if it is decomposable. They evaluate their model on English NN compounds and verb-particles, and showed that the model correlated moderately well with the Wordnet based decomposability theory (Baldwin et al., 2003).

McCarthy, Keller and Caroll (McCarthy et al., 2003) judge compositionality according to the degree of overlap in the set of most similar words to the verb-particle and head verb. They showed that the correlation between their measures and the human ranking was better than the correlation between the statistical features and the human ranking. We have done similar experiments in this paper where we compare the correlation value of the ranks provided by the SVM based ranking function with the ranks of the individual features for the V-N collocations. We show that the ranks given by the SVM based ranking function which integrates all the features provides a significantly better correlation than the individual features.

4 Data used for the experiments

The data used for the experiments is British National Corpus of 81 million words. The corpus is parsed using Bikel's parser (Bikel, 2004) and the Verb-Object Collocations are extracted. There are 4,775,697 V-N collocations of which 1.2 million are unique. All the V-N collocations above the frequency of 100 (n=4405) are taken to conduct the experiments so that the evaluation of the system is feasible. These 4405 V-N collocations were searched in Wordnet, American Heritage Dictionary and SAID dictionary (LDC,2003). Around 400 were found in at least one of the dictionaries. Another 400 were extracted from the rest so that the evaluation set has roughly equal number of compositional and non-compositional expressions. These 800 expressions were annotated with a rating from 1 to 6 by us-

ing guidelines independently developed by the authors. 1 denotes the expressions which are totally non-compositional while 6 denotes the expressions which are totally compositional. The brief explanation of the various ratings is as follows: **(1)** No word in the expression has any relation to the actual meaning of the expression. Example : "**leave a mark**". **(2)** Can be replaced by a single verb. Example : "**take a look**". **(3)** Although meanings of both words are involved, at least one of the words is not used in the usual sense. Example : "**break news**". **(4)** Relatively more compositional than (3). Example : "**prove a point**". **(5)** Relatively less compositional than (6). Example : "**feel safe**". **(6)** Completely compositional. Example : "**drink coffee**".

5 Agreement between the Judges

The data was annotated by two fluent speakers of English. For 765 collocations out of 800, both the annotators gave a rating. For the rest, at least one of the annotators marked the collocations as "don't know". Table 1 illustrates the details of the annotations provided by the two judges.

Ratings	6	5	4	3	2	1
Annotator1	141	122	127	119	161	95
Annotator2	303	88	79	101	118	76

Table 1: Details of the annotations of the two annotators

From the table 1 we see that annotator1 distributed the rating more uniformly among all the collocations while annotator2 observed that a significant proportion of the collocations were completely compositional. To measure the agreement between the two annotators, we used the Kendall's TAU (τ) (Siegel and Castellan, 1988). τ is the correlation between the rankings[1] of collocations given by the two annotators. τ ranges between 0 (little agreement) and 1 (full agreement). τ is defined as,

$$\tau = \frac{\sum_{i<j} sgn(x_i - x_j)\, sgn(y_i - y_j)}{\sqrt{(t_0 - t_1)\,(t_0 - t_2)}}$$

$$t_1 = \frac{\sum u_i\,(u_i - 1)}{2}, \quad t_2 = \frac{\sum v_i\,(v_i - 1)}{2}, \quad t_0 = \frac{n\,(n-1)}{2}$$

---

[1] computed from the ratings

901

where x_i's are the rankings of annotator1 and y_i's are the rankings of annotator2, n is the number of collocations, u_i is the number of values in the ith group of tied x values and v_i is the number of values in the ith group of tied y values.

We obtained a τ score of **0.61** which is highly significant. This shows that the annotators were in a good agreement with each other in deciding the rating to be given to the collocations. We also compare the ranking of the two annotators using Pearson's Rank-Correlation coefficient (r_s) (Siegel and Castellan, 1988). We obtained a r_s score of **0.71** indicating a good agreement between the annotators. A couple of examples where the annotators differed are **(1)** "perform a task" was rated 3 by annotator1 while it was rated 6 by annotator2 and **(2)** "pay tribute" was rated 1 by annotator1 while it was rated 4 by annotator2.

The 765 samples annotated by both the annotators were then divided into a training set and a testing set in several possible ways to cross-validate the results of ranking (section 8).

6 Features

Each collocation is represented by a vector whose dimensions are the statistical features obtained from the British National Corpus. The features used in our experiments can be classified as (1) Collocation based features and (2) Context based features.

6.1 Collocation based features

Collocation based features consider the entire collocation as an unit and compute the statistical properties associated with it. The collocation based features that we considered in our experiments are (1) Frequency, (2) Point-wise Mutual Information, (3) Least mutual information difference with similar collocations, (4) Distributed frequency of object and (5) Distributed frequency of object using the verb information.

6.1.1 Frequency (f)

This feature denotes the frequency of a collocation in the British National Corpus. Cohesive expressions have a high frequency. Hence, greater the frequency, the more is the likelihood of the expression to be a MWE.

6.1.2 Point-wise Mutual Information (m)

Point-wise Mutual information of a collocation (Church and Hanks, 1989) is defined as,

$$m(v, o) = \frac{f(v, o) * f(-, -)}{f(v, -) * f(-, o)}$$

where, v is the verb and o is the object of the collocation. The higher the Mutual information of a collocation, the more is the likelihood of the expression to be a MWE.

6.1.3 Least mutual information difference with similar collocations (μ)

This feature is based on Lin's work (Lin, 1999). He suggests that a possible way to separate compositional phrases from non-compositional ones is to check the existence and mutual information values of *similar collocations* (phrases obtained by replacing one of the words with a similar word). For example, 'eat apple' is a similar collocation of 'eat pear'.

For a collocation, we find the similar collocations by substituting the verb and the object with their *similar words*[2]. The similar collocation having the least mutual information difference is chosen and the difference in their mutual information values is noted.

If a collocation c has a set of similar collocations S, then we define μ as

$$\mu(v_c, o_c) = min_{s \in S} \left[abs \left(m(c) - m(s) \right) \right]$$

where $abs(x)$ returns the absolute value of x and v_c and o_c are the verb and object of the collocation c respectively. If similar collocations do not exist for a collocation, then this feature is assigned the highest among the values assigned in the previous equation. In this case, μ is defined as,

$$\mu(v, o) = max_{i,j} \left[\mu(v_i, o_j) \right]$$

where v and o are the verb and object of collocations for which similar collocations do not exist. The higher the value of μ, the more is the likelihood of the collocation to be a MWE.

[2]obtained from Lin's (Lin, 1998) automatically generated thesaurus (http://www.cs.ualberta.ca/~lindek/downloads.htm). We obtained the best results (section 8) when we substituted top-5 similar words for both the verb and the object. To measure the compositionality, semantically similar words are more suitable than synomys. Hence, we choose to use Lin's thesaurus (Lin, 1998) instead of Wordnet (Miller et al., 1990).

6.1.4 Distributed Frequency of Object (d)

The distributed frequency of object is based on the idea that "if an object appears only with one verb (or few verbs) in a large corpus, the collocation is expected to have idiomatic nature" (Tapanainen et al., 1998). For example, 'sure' in 'make sure' occurs with very few verbs. Hence, 'sure' as an object is likely to give a special sense to the collocation as it cannot be used with any verb in general. It is defined as,

$$d(o) = \frac{\sum_i^n f(v_i, o)}{n}$$

where n is the number of verbs occurring with the object (o), v_i's are the verbs cooccuring with o and $f(v_i, o) > T$. As the number of verbs (n) increases, the value of $d(o)$ decreases. Here, T is a threshold which can be set based on the corpus. This feature treats 'point finger' and 'polish finger' in the same way as it does not use the information specific to the verb in the collocation. Here, both the collocations will have the value $d("finger")$. The 3 collocations having the highest value of this feature are (1) come true, (2) become difficult and (3) make sure.

6.1.5 Distributed Frequency of Object using the Verb information (ρ)

Here, we have introduced an extension to the feature d such that the collocations like 'point finger' and 'polish finger' are treated differently and more appropriately. This feature is based on the idea that "a collocation is likely to be idiomatic in nature if there are only few other collocations with the same object and dissimilar verbs". We define this feature as,

$$\rho(v, o) = \frac{\sum_i^n f(v_i, o) \ * \ dist(v, v_i)}{n}$$

where $n < 50$ is the number of verbs occurring with o, v_i's are the verbs cooccuring with o and $f(v_i, o) > T$. $dist(v, v_i)$ is the distance between the verb v and v_i. It is calculated using the wordnet similarity measure defined by Hirst and Onge (Hirst and St-Onge, 1998). In our experiments, we considered top-50 verbs which co-occurred with the object o. We used a Perl package Wordnet::Similarity by Patwardhan[3] to conduct our experiments.

---

[3] http://www.d.umn.edu/~tpederse/similarity.html

6.2 Context based features

Context based measures use the context of a word/collocation to measure their properties. We represented the context of a word/collocation using a LSA model. LSA is a method of representing words/collocations as points in vector space.

The LSA model we built is similar to that described in (Schutze, 1998) and (Baldwin et al., 2003). First, 1000 most frequent content words (i.e., not in the stop-list) were chosen as "content-bearing words". Using these content-bearing words as column labels, the 50,000 most frequent terms in the corpus were assigned row vectors by counting the number of times they occurred within the same sentence as content-bearing words. Principal component analysis was used to determine the principal axis and we get the transformation matrix $A_{1000 X 100}$ which can be used to reduce the dimensions of the 1000 dimensional vectors to 100 dimensions.

We will now describe in Sections 6.2.1 and 6.2.2 the features defined using LSA model.

6.2.1 Dissimilarity of the collocation with its constituent verb using the LSA model (ι)

If a collocation is highly dissimilar to its constituent verb, it implies that the usage of the verb in the specific collocation is not in a general sense. For example, the sense of 'change' in 'change hands' would be very different from its usual sense. Hence, the greater the dissimilarity between the collocation and its constituent verb, the more is the likelihood that it is a MWE. The feature is defined as

$$\iota(c, v_c) = 1 - sim_{LSA}(c, v_c)$$

$$sim_{LSA}(c, v_c) = \frac{lsa(c) \ . \ lsa(v_c)}{|lsa(c)| \ * \ |lsa(v_c)|}$$

where, c is the collocation, v_c is the verb of the collocation and $lsa(x)$ is representation of x using the LSA model.

6.2.2 Similarity of the collocation to the verb-form of the object using the LSA model (ψ)

If a collocation is highly similar to the verb form of an object, it implies that the verb in the collocation does not contribute much to the meaning of the collocation. The verb either acts as a sort of

903

support verb, providing perhaps some additional aspectual meaning. For example, the verb 'give' in 'give a smile' acts merely as a support verb. Here, the collocation 'give a smile' means the same as the verb-form of the object i.e., 'to smile'. Hence, the greater is the similarity between the collocation and the verb-form of the object, the more is the likelihood that it is a MWE. This feature is defined as

$$\psi(c, vo_c) = \frac{lsa(c) \cdot lsa(vo_c)}{|lsa(c)| * |lsa(vo_c)|}$$

where, c is the collocation and vo_c is the verb-form of the object o_c. We obtained the verb-form of the object from the wordnet (Miller et al., 1990) using its 'Derived forms'. If the object doesn't have a verbal form, the value of this feature is 0. Table 2 contains the top-6 collocations according to this feature. All the collocations in Table 2 (except 'receive award' which does not mean the same as 'to award') are good examples of MWEs.

Collocation	Value		Collocation	Value
pay visit	0.94		provide assistance	0.92
provide support	0.93		give smile	0.92
receive award	0.92		find solution	0.92

Table 2: Top-6 collocations according to this feature

7   SVM based ranking function/algorithm

The optimal rankings on the training data is computed using the average ratings of the two users. The goal of the learning function is to model itself according to this rankings. It should take a ranking function f from a family of ranking functions F that maximizes the empirical τ (Kendall's Tau). τ expresses the similarity between the optimal ranking (r^*) and the ranking (r_f) computed by the function f. SVM-Light[4] is a tool developed by Joachims (Joachims, 2002) which provides us such a function. We briefly describe the algorithm in this section.

Maximizing τ is equivalent to minimizing the number of discordant pairs (the pairs of collocations which are not in the same order as in the optimal ranking). This is equivalent to finding the weight

[4]http://svmlight.joachims.org

vector $\bar{w}$ so that the maximum number of inequalities are fulfilled.

$$\forall (c_i, c_j) \in r^* : \bar{w} \ \phi(c_i) > \bar{w} \ \phi(c_j)$$

where c_i and c_j are the collocations, $(c_i, c_j) \in r^*$ if the collocation c_i is ranked higher than c_j for the optimal ranking r^*, $\phi(c_i)$ and $\phi(c_j)$ are the mapping onto features (section 6) that represent the properties of the V-N collocations c_i and c_j respectively and $\bar{w}$ is the weight vector representing the ranking function f_w.

Adding SVM regularization for margin maximization to the objective leads to the following optimization problem (Joachims, 2002).

$$minimize: \ V(\bar{w}, \bar{\epsilon}) = \frac{1}{2} \ \bar{w}\bar{w} + C \sum \epsilon_{i,j}$$

$$constr : \bar{w}(\phi(c_i) - \phi(c_j)) \geq 1 - \epsilon_{i,j}, \ \forall_{i,j} \ \epsilon_{i,j} \geq 0$$

where $\epsilon_{i,j}$ are the (non-negative) slack variables and C is the margin that allows trading-off margin size against training error. This optimization problem is equivalent to that of a classification SVM on pairwise difference vectors $\phi(c_i)$ - $\phi(c_j)$. Due to similarity, it can be solved using decomposition algorithms similar to those used for SVM classification (Joachims, 1999).

Using the learnt function $f_{\bar{w}^*}$ ($\bar{w}^*$ is the learnt weight vector), the collocations in the test set can be ranked by computing their values using the formula below.

$$rc(c_i) = \bar{w}^* \phi(c_i)$$

8   Experiments and Results

For training, we used 10% of the data and for testing, we use 90% of the data as the goal is to use only a small portion of the data for training (Data was divided in 10 different ways for cross-validation. The results presented here are the average results).

All the statistical measures show that the expressions ranked higher according to their decreasing values are more likely to be non-compositional. We compare these ranks with the human rankings (obtained using the average ratings of the users). To compare, we use Pearson's Rank-Order Correlation Coefficient (r_s) (Siegel and Castellan, 1988).

We integrate all the seven features using the SVM based ranking function (described in section 7). We

see that the correlation between the relative compositionality of the V-N collocations computed by the SVM based ranking function is significantly higher than the correlation between the individual features and the human ranking (Table 3).

Feature	Correlation		Feature	Correlation
f (f1)	0.129		ρ (f5)	0.203
m (f2)	0.117		ι (f6)	0.139
μ (f3)	0.210		ψ (f7)	*0.300*
d (f4)	0.111		Ranking f	**0.448**

Table 3: The correlation values of the ranking of individual features and the ranking of SVM based ranking function with the ranking of human judgements

In table 3, we also see that the contextual feature which we proposed, 'Similarity of the collocation to the verb-form of the object' (ψ), correlated significantly higher than the other features which indicates that it is a good measure to represent the semantic compositionality of V-N expressions. Other expressions which were good indicators when compared to the traditional features are 'Least mutual information difference with similar collocations' (μ) and 'Distributed frequency of object using the verb information' (ρ).
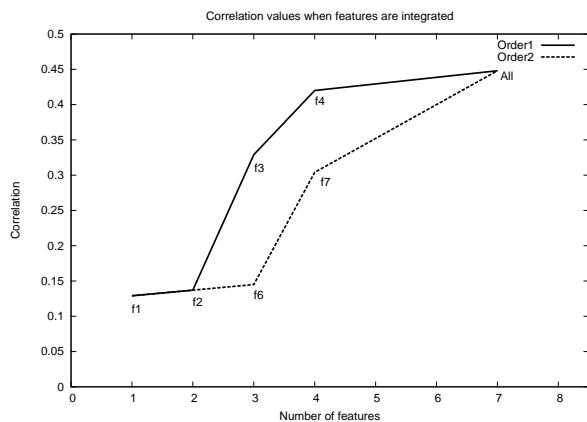

Figure 1: The change in r_s as more features are added to the ranking function

To observe the contribution of the features to the SVM based ranking function, we integrate the features (section 6) one after another (in two different ways) and compute the relative order of the collocations according to their compositionality. We see that as we integrate more number of relevant compositionality based features, the relative order correlates better (better r_s value) with the human ranking (Figure 1). We also see that when the feature 'Least mutual information difference with similar collocations' is added to the SVM based ranking function, there is a high rise in the correlation value indicating it's relevance. In figure 1, we also observe that the context-based features did not contribute much to the SVM based ranking function even though they performed well individually.

9 Conclusion

In this paper, we proposed some collocation based and contextual features to measure the relative compositionality of MWEs of V-N type. We then integrate the proposed features and the traditional features using a SVM based ranking function to rank the V-N collocations based on their relative compositionality. Our main results are as follows, **(1)** The properties 'Similarity of the collocation to the verb-form of the object', ' Least mutual information difference with similar collocations' and 'Distributed frequency of object using the verb information' contribute greatly to measuring the relative compositionality of V-N collocations. **(2)** The correlation between the ranks computed by the SVM based ranking function and the human ranking is significantly better than the correlation between ranking of individual features and human ranking.

In future, we will evaluate the effectiveness of the techniques developed in this paper for applications like Machine Translation. We will also extend our approach to other types of MWEs and to the MWEs of other languages (work on Hindi is in progress).

Acknowledgments

References

Anne Abeille. 1988. Light verb constructions and extraction out of np in a tree adjoining grammar. In *Pa-*

pers of the 24th Regional Meeting of the Chicago Linguistics Society.

Monoji Akimoto. 1989. Papers of the 24th regional meeting of the chicago linguistics society. In *Shinozaki Shorin*.

Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. 2003. An empirical model of multiword expression. In *Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*.

Colin Bannard, Timothy Baldwin, and Alex 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*.

Joseph D. Becker. 1975. The phrasal lexicon. In *Theoritical Issues of NLP, Workshop in CL, Linguistics, Psychology and AI, Cambridge, MA*.

Daniel M. Bikel. 2004. A distributional analysis of a lexicalized statistical parsing model. In *Proceedings of EMNLP*.

Elisabeth Breidt. 1995. Extraction of v-n-collocations from text corpora: A feasibility study for german. In *CoRR-1996*.

K. Church and Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th. Annual Meeting of the Association for Computational Linguistics, 1990*.

K. Church, W. Gale, P. Hanks, and D. Hindle. 1991. Parsing, word associations and typical predicate-argument relations. In *Current Issues in Parsing Technology. Kluwer Academic, Dordrecht, Netherlands, 1991*.

Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. In *Computational Linguistics - 1993*.

Charles Fillmore. 2003. An extremist approach to multiword expressions. In *A talk given at IRCS, University of Pennsylvania, 2003*.

G. Hirst and D. St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In *Fellbaum C., ed., Wordnet: An electronic lexical database. MIT Press*.

Young-Sook Hwang and Yutaka Sasaki. 2005. Context-dependent SMT model using bilingual verb-noun collocation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.

T. Joachims. 1999. Making large-scale svm learning practical. In *Advances in Kernel Methods - Support Vector Learning*.

T. Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL'98*.

Dekang Lin. 1999. Automatic identification of non-compositional phrases. In *Proceedings of ACL-99, College Park, USA*.

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

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to wordnet: an on-line lexical database. In *International Journal of Lexicography*.

G. Nunberg, I. A. Sag, and T. Wasow. 1994. Idioms. In *Language, 1994*.

I. A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multi-word expressions: a pain in the neck for nlp. In *Proceedings of CICLing , 2002*.

Patrick Schone and Dan Jurafsky. 2001. Is knowledge-free induction of multiword unit dictionary headwords a solved problem? In *Proceedings of EMNLP , 2001*.

William Schuler and Aravind K. Joshi. 2004. Relevance of tree rewriting systems for multi-word expressions. In *To be published*.

Hinrich Schutze. 1998. Automatic word-sense discrimination. In *Computational Linguistics*.

S. Siegel and N. John Castellan. 1988. In *Nonparametric Statistics of the Behavioral Sciences*. McGraw-Hill, NJ.

Pasi Tapanainen, Jussi Piitulaine, and Timo Jarvinen. 1998. Idiomatic object usage and support verbs. In *36th Annual Meeting of the Association for Computational Linguistics*.

Sriram Venkatapathy and Aravind K. Joshi. 2004. Recognition of multi-word expressions: A study of verb-noun (v-n) collocations. In *Proceedings of the International Conference on Natural Language Processing,2004*.

906