

# Assessing the Potential of Metaphoricity of Verbs Using Corpus Data

Marco Del Tredici, Núria Bel Rafecas

Universitat Pompeu Fabra  
Roc Boronat, 138, Barcelona, Spain  
marco.deltredici@upf.edu, nuria.bel@upf.edu

## Abstract

The paper investigates the relation between metaphoricity and distributional characteristics of verbs, introducing POM, a corpus-derived index that can be used to define the upper bound of metaphoricity of any expression in which a given verb occurs. The work moves from the observation that while some verbs can be used to create highly metaphoric expressions, others can not. We conjecture that this fact is related to the number of contexts in which a verb occurs and to the frequency of each context. This intuition is modelled by introducing a method in which each context of a verb in a corpus is assigned a vector representation, and a clustering algorithm is employed to identify similar contexts. Eventually, the Standard Deviation of the relative frequency values of the clusters is computed and taken as the POM of the target verb. We tested POM in two experimental settings obtaining values of accuracy of 84% and 92%. Since we are convinced, along with (Shutova, 2015), that metaphor detection systems should be concerned only with the identification of highly metaphoric expressions, we believe that POM could be profitably employed by these systems to *a priori* exclude expressions that, due to the verb they include, can only have low degrees of metaphoricity.

**Keywords:** Distributional Semantics, Metaphoricity, Corpora

## 1. Introduction

Metaphor has been defined as the exploitation of a norm in language: words have one or few frequent contexts of use that are considered normal (norms), and some unfrequent contexts (exploitations) among which are metaphors, that are used to express new insights with a rhetorical effect (Hanks, 2004). Such an effect can have different degrees of strength, which correspond to the degrees of *metaphoricity* of the metaphor: the stronger the rhetorical effect, the higher the degree of metaphoricity.

The usefulness of considering metaphoricity when investigating and modelling metaphors has been stressed both from the theoretical point of view (Dunn, 2010; Hanks, 2004; Hanks, 2006; Nunberg, 1987) and in the field on NLP (Dunn, 2014; Dunn, 2013; Hovy et al., 2013; Mohler et al., 2015). (Shutova, 2015) in her vast review on metaphor processing systems suggests that real-world NLP applications should be concerned with the identification of metaphorical expressions with high degree of metaphoricity, i.e. those expressions that need to be interpreted differently from literal language and therefore processed with specific tools. At the opposite, they should not address low-metaphoricity expressions, since their meanings are already present in dictionaries and can be interpreted using standard word sense disambiguation techniques. From this point of view, the degree of metaphoricity is a key element to distinguish between metaphors that require specific tools to be processed and those that do not. In spite of this, the author notes that the majority of current systems in NLP still ignore metaphoricity, assuming instead a binary distinction between literal (0) and metaphoric (1), without considering all the possible values in the range between these two ends [0-1]. In general, such systems tend to label as metaphoric any expression that can not be considered strictly literal (e.g. 'see the point'). As a consequence, a large number of expressions are considered metaphoric, including those low-metaphoricity expressions that, according to Shutova, would not need to be interpreted differently from literal lan-

guage.

Also (Dunn, 2013) stresses the relevance of metaphoricity for metaphor processing systems. The author presents a thorough evaluation of four different systems, reporting fairly poor results for all of them and showing that a common weakness is the high number of false positives, i.e. of non-metaphoric expressions that are labelled as metaphoric. Dunn points out that low performances are partly due to ignoring the gradient nature of metaphor, thus confirming that implementing metaphoricity is not just advisable because more adherent to main theoretical positions on metaphor, but also because it is essential for the realization of effective computational systems for the modelling of this phenomenon.

At the best of our knowledge, (Dunn, 2014) is the only approach that explicitly addresses metaphoricity, introducing a system that assigns each input sentence a value of metaphoricity between 0 and 1.

In this work we also deal with metaphoricity but, rather than assigning a degree of metaphoricity to target sentences, we model the relation between metaphoricity and distributional characteristics of verbs.

We move from the observation that whereas some verbs can be used to create metaphors with high metaphoricity, others cannot. For instance, 'butcher' can be used to create metaphors with metaphoricity close to 1 - on the metaphoricity range [0-1] - (e.g. 'to *butcher* an ideal'<sup>1</sup>), while this is not possible for verbs like 'take' or 'imagine', which only occur in expressions whose maximum degree of metaphoricity is still close to 0<sup>2</sup>. We therefore say that verbs like 'butcher' have a high *Potential of Metaphoricity* (POM), while those like 'take' and 'imagine' have a low POM. It is worth stressing that verbs with high POM such as 'butcher' are not expected to occur only in metaphors

<sup>1</sup>All the examples are extracted from the British National Corpus (Burnard, 1998).

<sup>2</sup>We note that for some researchers, e.g. (Hanks, 2006), these verbs can not be used to create metaphoric expressions at all.

with high metaphoricity<sup>3</sup>, but just that they have the *potentiality* to do it. At the opposite, under no circumstances a metaphoric expression created with 'take' or 'imagine' can have a high degree of metaphoricity.

Hence, accordingly to what said, the semantic properties of a given verb can be leveraged not only to determine the degree of metaphoricity of a specific input sentence including that verb (Dunn, 2014), but also to predict the upper-bound of metaphoricity of *any* expression in which it occurs. We believe that such an information could be highly useful for systems that perform metaphor detection, since it would allow to *a priori* exclude metaphoric expressions that, being created with low-POM verbs (e.g. 'take'), can only have low degrees of metaphoricity. In this way, it would be possible to realize what (Shutova, 2015) suggests, that is: to label as metaphoric only those expressions that, having high metaphoricity, are *truly* figurative, and that therefore need to be interpreted differently from literal expressions, while ignoring slightly metaphoric expressions.

In this work we introduce a method to define the POM of a verb based on its distributional behaviour. We follow (Hanks, 2006) and conjecture that verbs that occur with high frequency in many contexts (e.g. 'take a decision', 'take a train', etc.) lose the potential to be used in sentences with high degrees of metaphoricity, while verbs that have just one, or very few, relatively high frequent contexts ('butcher an animal') and some very infrequent contexts ('butcher an ideal') have high POM. We computationally model this intuition by extracting the contexts of a target verb from a corpus, clustering them and computing the relative frequency of each cluster. Eventually, we compute the Standard Deviation (SD) of the relative frequency values and take the SD value obtained in this way as the POM of the target verb. The POM is therefore obtained without resorting to hand-crafted resources or knowledge bases, and is totally domain independent. Predictably, low SD values are considered indicative of low POM, while high values of SD indicate high POM.

We experimented this methodology in two settings. In the first one we leveraged the relation between metaphoricity and conventionalization: we calculated the POM of a set of verbs and predicted that verbs with low POM would mostly occur in conventional metaphors, whose senses could be found in a dictionary. At the opposite, we expected high-POM verbs to occur mainly occur in novel metaphors, that are not usually listed in dictionaries. We tested this hypothesis using WordNet, obtaining an accuracy of 84%.

In the second setting, we computed the POM of the verbs occurring in the dataset introduced by (Dunn, 2014) and then used it to predict human-based values of metaphoricity reported in the dataset. The accuracy of this experiment was 92%.

## 2. Background

Current NLP systems for the modelling of metaphor address two main tasks, metaphor identification and interpretation, and are based on three main theoretical frameworks.

The first one is the Conceptual Metaphor Theory (CMT) (Lakoff and Johnson, 1980), whereby a metaphor consists of a source-target mapping: metaphor modelling, thus, is performed discovering whether this mapping is present or not and finding the corresponding literal meaning (Shutova and Sun, 2013). The second theoretical framework is the Selectional Restrictions Hypothesis (Wilks, 1978), implemented for example in (Li and Sporleder, 2010). The basic idea here is that a metaphoric expression is characterized by the usage of a word that is not semantically related to the other words in the utterance, and that it is possible to detect metaphors through this semantic mismatching. Finally, the Abstractness Assumption (Turney et al., 2011) leverages the idea that metaphors occur when an abstract concept is explained using a more concrete one. Metaphor modelling, therefore, requires a measure of abstractness for target lexical items and their contexts.

Independently of the different conceptual framework adopted, the majority of the systems in literature model metaphor as a discrete property, ignoring the fact that several degrees of metaphoricity are possible. To our knowledge, the only work that explicitly addresses metaphoricity is (Dunn, 2014), which introduces a computationally-derived scalar measurement of metaphoricity and assigns to each input sentence a value between 0 (literal) and 1 (highly metaphoric). The author reports a correlation of 0.450 (Pearson's R,  $p < 0.01$ ) of the computational-derived measure with a human-based experimental measure of metaphoricity.

The goal of this paper is to introduce a method that leverages the distributional characteristics of a verb to define an index called POM, that can be used to predict the upper bound of metaphoricity of metaphoric expression in which the verb occurs. The basic motivation for this work is to overcome the current dichotomic view of metaphor in NLP proposing a method that predicts the different degrees of metaphoricity that a metaphoric expression can achieve.

## 3. Methodology

The POM of a verb is defined through the analysis of the contexts in which it occurs within a reference corpus (in our case the BNC) and its frequency within each context. We follow (Hanks, 2006) and conjecture that verbs that occur with similar relative frequency in many different contexts (e.g. 'take') have low POM, while verbs that have just one, or very few, relatively high frequent contexts and some very infrequent contexts (e.g. 'butcher') have high POM.

The identification of different contexts is hence the key elements of our method. Along with Hanks (2006) we consider the context of a verb as formed by the subject and/or the object with which it occurs. For the following examples:

- (i) invest *money*
- (ii) invest *cash*
- (iii) invest *time*

we consider (i) and (ii) as the same context of use of the verb 'invest', while (iii) as a different context.

---

<sup>3</sup>An example of a mild metaphor including 'butcher' is 'Croatian and Bosnian fascists *butchered* Serbs'.

In order to automatically identify similar contexts of a given verb, we followed a two-steps methodology: firstly, a vector representation of each context in which a target verb occurs was created. In the second step, a clustering algorithm was employed in order to identify similar vector representations and, therefore, similar contexts.

As for the realization of the first step, we initially extracted all the sentences in which a target verbs occurs in the British National Corpus. For each sentence we then selected the subject and object of the verb, and matched them with the corresponding vectorial representation, using the dependency based word embeddings (WE) introduced by (Levy and Goldberg, 2014). WE are low dimensional, dense and real-valued vectors which preserve syntactic and semantic information of words, and that have been proved to be efficient in several NLP tasks, such as detection of relational similarity (Mikolov et al., 2013b), word similarity tasks (Mikolov et al., 2013a) and contextual similarity (Melamud et al., 2015). When both subject and object were available in the same sentence, the context vector was defined by averaging them (Melamud et al., 2015). Otherwise, if one of the two was not present, the context vector would be equivalent to the available one. In the second step, we identified groups of similar contexts of the verb by clustering the context vectors obtained in phase 1. We used the Birch algorithm for its reliable performances with large sets of data (Zhang et al., 1996) and because the final number of clusters does not have to be previously defined: this is in line with the fact that the number of contexts of a verb is unknown. We used the scikit-learn implementation of the Birch algorithm<sup>4</sup>, whereby it is possible to experiment with different values for each parameter. Silhouette score (Rousseeuw, 1987), a widely employed metric for the interpretation and validation of clustering results, was employed as an external metric to evaluate the results obtained with the different settings and to select the best one.

Thus, the output of phase 2 was, for every target verb, a set of clusters, where each cluster corresponded to a different context (e.g. vector representations of examples (i) and (ii) were clustered together, while the one in (iii) was assigned to a different cluster).

Finally, the Standard Deviation (SD) of the relative frequency values of clusters in the set was computed in order to assess the distributional characteristics of the verb. We took the SD value obtained in this way as the POM of the verb. Following our intuition, SD values were expected to be low for verbs occurring with high frequency in several contexts (e.g. 'take') and high for verbs occurring with high frequency in just one or few contexts ('butcher').

## 4. Experiments

We performed two experiment: in both of them, we computed the POM of a set of verbs and used it to predict the upper bound of metaphoricity of sentences in the datasets in which they occurred. Hence, for a verb  $x$  occurring in a set of sentences  $Y = \{y_1, y_2 \dots y_n\}$ , given the POM of  $x$  we would define an upper bound  $z$  and predict that no

expression in  $Y$  could have a value of metaphoricity higher than  $z$ .

### 4.1. Experiment 1

In the first experiment, we leveraged the relation between metaphoricity and conventionalization of metaphors (Shutova, 2015): while novel metaphors have high metaphoricity, conventional metaphors, being just one more kind of normal use of language (Hanks, 2006), have low metaphoricity. Thus, our intuition is that a verb that can only occur in expressions with low metaphoricity ('take') will mostly occur in conventional metaphors ('take a decision') which, being known senses of a word, are likely to be listed in a general dictionary. At the opposite, a verb that can occur in expressions with high metaphoricity ('butcher'), is also used in novel metaphors, which can not be found in dictionaries.

Given these premises, we randomly selected 100 verbs from the VU Amsterdam Metaphor Corpus (VUAMC)(Steen et al., 2010), a corpus derived from the BNC and manually annotated for metaphor, and calculated the POM of each verb. We then selected verbs whose POM values were either  $<1$  or  $>3$  (60 verbs overall), and extracted the sentences in VUAMC in which these verbs were labelled as metaphoric. Our prediction was that metaphoric sentences whose main verb had  $POM < 1$  would be conventional metaphors and that therefore the senses expressed by the verbs in these sentences could be found in WordNet (Fellbaum, 1998). At the opposite, we expected metaphoric sentences based on verbs with  $POM > 3$  to be novel metaphors, and therefore not included in WordNet. Overall, we experimented with 60 verbs occurring in 889 sentences, whereof 826 included verbs whose POM was  $<1$  and 63 verbs whose POM was  $>3$ . The overall accuracy of our prediction was 84%. In table 2 are listed detailed results for the two categories of verbs taken into consideration ( $POM < 1$  and  $POM > 3$ ).

POM	Number of metaphorical sentences	Number of verb senses in WordNet	Number of verb senses not in WordNet
$<1$	826	<b>730 (88.4%)</b>	96 (11.6%)
$>3$	63	18 (28.5%)	<b>45 (71.5%)</b>

Table 1: Results of the experiment performed on the dataset drawn from the VU Amsterdam Metaphor Corpus (VUAMC).

### 4.2. Experiment 2

For the second experiment we used the dataset specifically annotated for metaphoricity introduced by (Dunn, 2014). The dataset is composed of 60 sentences whose degree of metaphoricity is determined by the verb they include. Sentences belong to four domains: physical, mental, social and abstract. Five different verbs are present for each domain, and each verb occurs in three sentences. Sentences in the dataset had been labeled in several tasks (for more details, see (Dunn, 2014)). In one of these tasks,

<sup>4</sup><http://scikit-learn.org/stable/modules/clustering.html#birch>

participants were asked to judge each sentences as 'Not Metaphoric', 'Slightly Metaphoric', or 'Very Metaphoric'. The authors derived from the human judgements a measure of metaphoricity ranging for 0 to 1: this is the metaphoricity value that we wanted to predict using POM. In what follows, we will refer to this value as L1. In another task, participants had to judge each sentence either as 'Metaphoric' or 'Literal'. The percentage of 'Metaphoric' assigned to a target sentence is reported in label L2. As an example, for the following sentences:

(i) A lady on high heels clacked along, the type my mother says *invests* all of her brainpower in her looks.

(ii) I wanted to find out what was left over when I *subtracted* my professional identity from who I was.

L1 value for sentence (i) was 0.571, while for sentence (ii) it was 0.286; L2 value was is 83.33% for both (i) and (ii). The values reported for the sentences above show that L1 and L2 could be inconsistent: especially for sentence (ii), the low level of metaphoricity reported in L1 (0.286) doesn't seem to be in line with the high percentage of L2 (83.33%). We thus compared the results in L1 and L2 of each sentence by converting the L2 percentage in a value in the range [0-1] (e.g. 83.33% = 0.833) and considered for our task only sentences for which the difference between L1 and L2 was less than 0.30<sup>5</sup>. As a result, 11 sentences were eliminated from the dataset (49 sentences left).

For our experiment, we computed the POM of the verbs in the dataset and then we used it to predict the upper-bound of metaphoricity as reported in L1 of any sentence including that verb. Since it was not possible to predict the exact value of metaphoricity reported in L1, the range of metaphoricity [0-1] was split in three sub-ranges [0-0.33], [0.34-0.67], [0.68-1], and each sub-range was considered as a class in the dataset. Our prediction was that verbs with  $POM < 1$  would correspond to values of L1 in the first range, i.e. that the upper bound of metaphoricity of expressions including such verbs would be 0.33. For  $1 < POM < 3$  we predicted an upper bound of 0.67; for  $POM > 3$  the upper bound was 1. For example, since 'see' had  $POM = 0.71$ , our prediction was that no sentence including it would have an upper bound of metaphoricity higher than 0.33. At the contrary, given the  $POM = 4.27$  of the verb 'obey', we predicted that a metaphoric expression including this verb could achieve the highest upper bound of metaphoricity (1). The overall accuracy of our prediction was 92%. In table 1 the results of our predictions in terms of precision, recall and f-score are reported.

## 5. Discussion

The results of both experiments confirmed our intuition that the distributional characteristics of a verb can be used to predict the upper bound of metaphoricity of metaphoric expressions in which they occur.

Regarding the first experiment, the majority of expressions (88.4%) in the dataset drawn from VUAMC including verbs with  $POM < 1$  (e.g. 'the money *come* from my

Upper bound	Precision	Recall	F-Score
0.33	1	1	1
0.67	.8	1	.9
1	1	.87	.93

Table 2: Results of the experiment performed on the dataset introduced by (Dunn, 2014).

mother') are correctly considered conventional metaphors whose senses are available in WordNet. Such a prediction was not confirmed for few sentences (11.6%), and even in these cases the error analysis showed that they were not novel metaphor, but idiomatic expressions (e.g. 'come to light') or colloquialisms ('come on!'). As for verbs with  $POM > 3$ , our prediction that they were to be found in novel metaphors was conformed in 71.5% of the cases. For the wrongly classified sentences (28.5%), we could identify two main reasons for the errors: the first one is that, as said above (see 1), verbs with high metaphoricity are not expected to occur *only* in novel metaphors, but also in more conventionalized metaphoric expressions, that can therefore be found in WordNet. Secondly, the error analysis showed that errors were mainly due to the fact that WordNet, being a very large lexical base, includes also some novel metaphors, like 'The huge waves *buried* the small boat and it sank shortly thereafter' and 'She *perched* her hat on her head'.

These results are in line with those obtained on the dataset introduced by (Dunn, 2014), with which the accuracy of our prediction is 92%. The test set is fairly small and hence, despite the good results reported in Table 2, it does not allow to fully assess the reliability of our method. However, we observe that using POM it is possible to perform a rather precise classification, correctly assigning, for example, an upper bound of metaphoricity of 0.33 to 'see' and 'enjoy', whose highest values of metaphoricity in the dataset are 0.091 and 0.167, and an upper bound of 0.67 to 'jump' and 'compute', that in the dataset have maximum values of metaphoricity of 0.6 and 0.375 respectively. Even if further evaluation is needed, these initial results indicate that the index can be employed to define different thresholds of metaphoricity, thus making it adaptable to the specific aims of different systems.

## 6. Conclusions

We introduced a method to compute the *Potential of Metaphoricity* (POM) of verbs, i.e. an index that, for a given verb, can be used to predict the upper bound of metaphoricity of all the expressions in which it occurs. The work moves from the basic idea that not all the verbs can be used to create metaphor with strong rhetoric effect, and implements this idea by leveraging the distributional characteristics of verbs in corpus. We believe such a method could be profitably employed by metaphor detection systems to filter out expression created with low POM verbs. This would allow to detect only metaphorical expressions that, having high degree of metaphoricity, need to be interpreted differently from literal language and

<sup>5</sup>The threshold was empirically defined by the authors.

therefore processed with specific tools, while ignoring low-metaphoricity expressions which can be processed using standard word sense disambiguation techniques.

## 7. Acknowledgements

This work was funded with the support of the IULA- UPF PhD fellowship program.

## 8. Bibliographical References

- Burnard, L. (1998). *The British National Corpus*.
- Dunn, J. (2010). Gradient semantic intuitions of metaphorical expressions. *Metaphor and Symbol*, 26(1):53–67.
- Dunn, J. (2013). What metaphor identification systems can tell us about metaphor-in-language. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 1–10.
- Dunn, J. (2014). Measuring metaphoricity. In *ACL (2)*, pages 745–751.
- Fellbaum, C. (1998). ed. wordnet: an electronic lexical database. *MIT Press, Cambridge MA*, 1:998.
- Hanks, P. (2004). The syntagmatics of metaphor and idiom. *International Journal of Lexicography*, 17(3):245–274.
- Hanks, P. (2006). Metaphoricity is gradable. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, 171:17.
- Hovy, D., Srivastava, S., Jauhar, S. K., Sachan, M., Goyal, K., Li, H., Sanders, W., and Hovy, E. (2013). Identifying metaphorical word use with tree kernels. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 52–57. Citeseer.
- Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press. *Chicago, IL*.
- Levy, O. and Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 302–308.
- Li, L. and Sporleder, C. (2010). Using gaussian mixture models to detect figurative language in context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 297–300. Association for Computational Linguistics.
- Melamud, O., Dagan, I., and Goldberger, J. (2015). Modeling word meaning in context with substitute vectors. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics, Denver, USA*.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mikolov, T., Yih, W.-t., and Zweig, G. (2013b). Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pages 746–751.
- Mohler, M., Tomlinson, M., and Rink, B. (2015). Cross-lingual semantic generalization for the detection of metaphor. *Computational Linguistics and Intelligent Text Processing*. Springer.
- Nunberg, G. (1987). Poetic and prosaic metaphors. In *Proceedings of the 1987 workshop on Theoretical issues in natural language processing*, pages 198–201. Association for Computational Linguistics.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- Shutova, E. and Sun, L. (2013). Unsupervised metaphor identification using hierarchical graph factorization clustering. In *HLT-NAACL*, pages 978–988.
- Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*.
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., and Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing.
- Turney, P. D., Neuman, Y., Assaf, D., and Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*, pages 680–690.
- Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence*, 11(3):197–223.
- Zhang, T., Ramakrishnan, R., and Livny, M. (1996). Birch: an efficient data clustering method for very large databases. In *ACM Sigmod Record*, volume 25, pages 103–114. ACM.