

# Compositional Semantics using Feature-Based Models from WordNet

**Pablo Gamallo**

Centro Singular de Investigación en  
Tecnoloxías da Información (CiTIUS)  
Universidade de Santiago de Compostela  
Galiza, Spain

pablo.gamallo@usc.es

**Martín Pereira-Fariña**

Centre for Argument  
Technology (ARG-tech)  
University of Dundee

Dundee, DD1 4HN, Scotland (UK)

m.z.pereirafarina@dundee.ac.uk

Departamento de Filosofía e Antropoloxía  
Universidade de Santiago de Compostela  
Pz. de Mazarelos, 15782, Galiza, Spain

martin.pereira@usc.es

## Abstract

This article describes a method to build semantic representations of composite expressions in a compositional way by using WordNet relations to represent the meaning of words. The meaning of a target word is modelled as a vector in which its semantically related words are assigned weights according to both the type of the relationship and the distance to the target word. Word vectors are compositionally combined by syntactic dependencies. Each syntactic dependency triggers two complementary compositional functions: the named *head function* and *dependent function*. The experiments show that the proposed compositional method performs as the state-of-the-art for subject-verb expressions, and clearly outperforms the best system for transitive subject-verb-object constructions.

## 1 Introduction

The principle of compositionality (Partee, 1984) states that the meaning of a complex expression is a function of the meaning of its constituent parts and of the mode of their combination. In the recent years, different distributional semantic models endowed with a compositional component have been proposed. Most of them define words as high-dimensional vectors where dimensions represent co-occurring context words. This distributional semantic representation makes it possible to combine vectors using simple arithmetic operations such as addition and multiplication, or more advanced compositional methods such as learning functional words as tensors and composing constituents through inner product operations.

Notwithstanding, these models are usually qualified as black box systems because they are usually not interpretable by humans. Currently, the field of interpretable computational models is gaining relevance<sup>1</sup> and, therefore, the development of more explainable and understandable models in compositional semantics is also an open challenge. In this field. On the other hand, distributional semantic models, given the size of the vectors, needs significant resources and they are dependent on particular corpus, which can generate some biases in their application to different languages.

Thus, in this paper, we will pay attention to compositional approaches which employ other kind of word semantic models, such as those based on the WordNet relationships; i.e., synsets, hypernyms, hyponyms, etc. Only in (Faruqui and Dyer, 2015) we can find a proposal for word vector representation using hand-crafted linguistic resources (WordNet, FrameNet, etc.), although a compositional frame is not explicitly adopted. Therefore, to the best of our knowledge, this is the first work using WordNet to build compositional semantic interpretations. Thus, in this article, we propose a method to compositionally build the semantic representation of composite expressions using a feature-based approach (Hadj Taieb et al., 2014): constituent elements are induced by WordNet relationships.

However, this proposal raises a serious problem: the semantic representation of two syntactically related words (e.g. the verb *run* and the noun *computer* in “*the computer runs*”) encodes incompatible information and there is no direct way of combining the features used to represent the meaning of the two words. On the one hand, the verb

<sup>1</sup><http://www.darpa.mil/program/explainable-artificial-intelligence>

*run* is related by synonymy, hypernym, hyponym and entailment to other verbs and, on the other, the noun *computer* is put in relation with other nouns by synonymy, hypernym, hyponym, and so on.

In order to solve this drawback, on the basis of previous work on dependency-based distributional compositionality (Thater et al., 2010; Erk and Padó, 2008), we distinguish between direct denotation and selectional preferences within a dependency relation. More precisely, when two words are syntactically related, for instance *computer* and the verb *run* by the subject relation, we build two contextualized senses: the contextualized sense of *computer* given the requirements of *run* and the contextualized sense of *run* given *computer*.

The sense of *computer* is built by combining the semantic features of the noun (its direct denotation) with the selectional preferences imposed by the verb. The features of the noun are built from the set of words linked to *computer* in WordNet, while the selectional preferences of *run* in the subject position are obtained by combining the features of all the nouns that can be the nominal subject of the verb (i.e. the features of *runners*). Then, the two sets of features are combined and the resulting new set represents the specific sense of the noun *computer* as nominal subject of *run*. The sense of the verb given the noun is built in a analogous way: the semantic features of the verb are combined with the (inverse) selectional preferences imposed by the noun, resulting in a new compositional representation of the verb *run* when it is combined with *computer* at the subject position. The two new compositional feature sets represent the contextualized senses of the two related words. During the contextualization process, ambiguous or polysemous words may be disambiguated in order to obtain the right representation.

For dealing with any sequence with  $N$  (lexical) words (e.g., “*the coach runs the team*”), the semantic process can be applied in two different ways: from left-to-right and from right-to-left. In the first case, it is applied  $N - 1$  times dependency-by-dependency in order to obtain  $N$  contextualized senses, one per lexical word. Thus, firstly, the subject dependency builds two contextualized senses: that of *run* given the noun *coach* and that of the noun given the verb. Then, the direct object dependency is applied on the already contextualized sense of the verb in order to contextualize it

again given *team* at the direct object position. This dependency also yields the contextualized sense of the object given the verb and its nominal subject (*coach+run*). At the end of the interpretation process, we obtain three fully contextualized senses. In the second case, from right-to-left, the semantic process process is applied in a similar way, being contextualized (and disambiguated) using the restrictions imposed by the verb and its nominal object (*run+team*). As in the first case, three slightly different word senses are also obtained.

Lastly, word sense disambiguation is out of the aim of this paper. Here, we only use WordNet for extracting semantic information from words, but not to identify word senses.

The article is organized as follow: In the next section (2), different approaches on ontological feature-based representations and compositional semantics are introduced and discussed. Then, sections 3 and 4 respectively describe our feature-based semantic representation and compositional strategy. In Section 5, some experiments are performed to evaluate the quality of the word models and compositional word vectors. Finally, relevant conclusions are reported in Section 6.

## 2 Related Work

Our approach relies on two tasks: to build feature-based representations using WordNet relations, and to build compositional vectors using the WordNet representations. In this section, we will examine work related to these two tasks.

### 2.1 Feature-Based Approaches

Tversky (1977), in order to define a similarity measure, assumes that any object can be represented as a collection (set) of features or properties. Therefore, a similarity metric is a feature-matching process between two objects. This consists of a linear combination of the measures of their common and distinctive features. It is worth noting that this is a non-symmetric measure.

In the particular case of semantic similarity metrics, each word or concept is featured by means of a set of words (Hadj Taieb et al., 2014). Framed into an ontology such as WordNet, these sets of words are obtained from taxonomic (hypernym, hyponym, etc.) and non-taxonomic (synsets, glosses, meronyms, etc.) properties (Meng et al., 2013), although these last ones are classified as secondary in many cases (Slimani, 2013). The

main objective of this approach is to capture the semantic knowledge induced by ontological relationships.

Our model is partly inspired by that defined in (Rodríguez and Egenhofer, 2003). It proposes that the set of properties that characterizes a word may be stratified into three groups: i) *synsets*; ii) *features* (e.g., meronyms, attributes, hyponym, etc.), and, iii) *neighbor concepts* (those linked via semantic pointers). Each one of these strata is weighted according to its contribution to the representation of the concept. The measure analyzes the overlapping among the three strata between the two terms under comparison.

## 2.2 Compositional Strategies

Several models for compositionality in vector spaces have been proposed in recent years, and most of them use bag-of-words as basic distributional representations of word contexts. The basic approach to composition, explored by Mitchell and Lapata (2008; 2009; 2010), is to combine vectors of two syntactically related words with arithmetic operations: addition and component-wise multiplication. The additive model produces a sort of union of word contexts, whereas multiplication has an intersective effect. According to Mitchell and Lapata (2008), component-wise multiplication performs better than the additive model. However, in (Mitchell and Lapata, 2009; Mitchell and Lapata, 2010), these authors explore weighted additive models giving more weight to some constituents in specific word combinations. For instance, in a noun-subject-verb combination, the verb is provided with higher weight because the whole construction is closer to the verb than to the noun. Other weighted additive models are described in (Guevara, 2010) and (Zanzotto et al., 2010). All these models have in common the fact of defining composition operations for just word pairs. Their main drawback is that they do not propose a more systematic model accounting for all types of semantic composition. They do not focus on the logical aspects of the functional approach underlying compositionality.

Other distributional approaches develop sound compositional models of meaning inspired by Montagovian semantics, which induce the compositional meaning of the functional words from examples adopting regression techniques commonly used in machine learning (Krishnamurthy and

Mitchell, 2013; Baroni and Zamparelli, 2010; Baroni, 2013; Baroni et al., 2014). In our approach, by contrast, compositional functions, which are driven by dependencies and not by functional words, are just basic arithmetic operations on vectors as in (Mitchell and Lapata, 2008). Arithmetic approaches are easy to implement and produce high-quality compositional vectors, which makes them a good choice for practical applications (Baroni et al., 2014).

Other compositional approaches based on Categorical Grammar use tensor products for composition (Grefenstette et al., 2011; Coecke et al., 2010). A neural network-based method with tensor factorization for learning the embeddings of transitive clauses has been introduced in (Hashimoto and Tsuruoka, 2015). Two problems arise with tensor products. First, they result in an information scalability problem, since tensor representations grow exponentially as the phrases grow longer (Turney, 2013). And second, tensor products did not perform as well as component-wise multiplication in Mitchell and Lapata’s (2010) experiments.

There are also works focused on the notion of sense contextualization, e.g., Dinu and Lapata (2010) work on context-sensitive representations for lexical substitution. Reddy et al. (2011) work on dynamic prototypes for composing the semantics of noun-noun compounds and evaluate their approach on a compositionality-based similarity task.

So far, all the cited works are based on bag-of-words to represent vector contexts and, then, word senses. However, there are a few works using vector spaces structured with syntactic information. Thater et al. (2010) distinguish between *first-order* and *second-order* vectors in order to allow two syntactically incompatible vectors to be combined. This work is inspired by that described in (Erk and Padó, 2008). Erk and Padó (2008) propose a method in which the combination of two words,  $a$  and  $b$ , returns two vectors: a vector  $a'$  representing the sense of  $a$  given the selectional preferences imposed by  $b$ , and a vector  $b'$  standing for the sense of  $b$  given the (inverse) selectional preferences imposed by  $a$ . A similar strategy is reported in Gamallo (2017). Our approach is an attempt to join the main ideas of these syntax-based models (namely, second-order vectors, selectional preferences and two returning words per combi-

nation) in order to apply them to WordNet-based word representations.

### 3 Semantic Features from WordNet

A word meaning is described as a feature-value structure. The *features* are the words with which the target word is related to in the ontology (e.g., in WordNet, hypernym, hyponym, etc.) and the *values* correspond to weights computed taking into account two parameters: the relation type and the edge-counting distance between the target word and each word feature (i.e. the number of relations required to achieve the feature from the target word) (Rada et al., 1989).

The algorithm to set the feature values is the following. Given a target word  $w_1$  and the feature set  $F$ , where  $w_i \in F$  if  $w_i$  is a word semantically related to  $w_1$  in WordNet, the weight for the relation between  $w_1$  and  $w_i$  is computed by equation 1:

$$weight(w_1, w_i) = \sum_{j=1}^R \frac{1}{length(w_1, w_i, r_j)} \quad (1)$$

where  $R$  is the number of different semantic relations (e.g. synonymy/synset, hyperonymy, hyponymy, etc) that WordNet defines for the part-of-speech of the target word. For instance, nouns have five different relations, verbs four and adjectives just two.  $length(w_1, w_i, r_j)$  is the length of the path from the target word  $w_1$  to its feature  $w_i$  in relation  $r_j$ .  $length(w_1, w_i, r_j) = 1$  when  $r_j$  stands for the synonymy relationship, i.e. when  $w_1$  and  $w_i$  belong to the same synset;  $length(w_1, w_i, r_j) = 2$  if  $w_i$  is at the first level within the hierarchy associated to relation  $r_j$ .

For instance, the length value of a direct hypernym is 2 because there is a distance of two arcs with regard to the target word: the first arc goes from the target word to a synset and the second one is the hyperonymy relation between the direct hypernym and the synset. The length value increases in one unit as the hierarchy level goes up, so at level 4, the length score is 5 and then the partial weight is  $1/5 = 0.2$ . For some non-taxonomic relations, namely meronymy, holonymy and coordinates, there is only one level in WordNet, but the distance is 3 since the target word and the word feature (part, whole or coordinate term) are separated by a synset and a hypernym.

As a feature word  $w_i$  may be related to the target  $w_1$  via different semantic relations (with-

out distinguishing between different word senses), the final weight is the addition of all partial weights. For instance, take the noun *car*. It is related to *automobile* through two different relationships: they belong to the same synset and the latter is a direct hypernym of the former, so  $weight(car, automobile) = 1/1 + 1/2 = 1.5$ .

To compute compositional operations on words, the feature-value structure associated to each word is modeled as a vector, where features are dimensions, words are objects, and weights the values for each object/dimension position.

## 4 Compositional Semantics

### 4.1 Syntactic Dependencies As Compositional Functions

Our approach is also inspired in (Erk and Padó, 2008). Here, semantic composition is modeled in terms of function application driven by binary dependencies. A dependency is associated in the semantic space with two compositional functions on word vectors: the head and the dependent functions. To explain how they work, let us take the direct object relation (*doobj*) between the verb *run* and the noun *team* in the expression “*run a team*”. The head function,  $doobj_{\uparrow}$ , combines the vector of the head verb  $r\vec{u}n$  with the selectional preferences imposed by the noun, which is also a vector of WordNet features, and noted  $team^{\circ}$ . This combination is performed by component-wise multiplication and results in a new vector  $r\vec{u}n_{doobj_{\uparrow}}$ , which represents the contextualized sense of *run* given *team* in the *doobj* relation:

$$doobj_{\uparrow}(r\vec{u}n, team^{\circ}) = r\vec{u}n \odot team^{\circ} = r\vec{u}n_{doobj_{\uparrow}}$$

To build the (inverse) selectional preferences imposed by the dependent word *team* as direct object on the verb, we require a reference corpus to extract all those verbs of which *team* is the direct object. The selectional preferences of *team* as direct object of a verb, and noted  $team^{\circ}$ , is a new vector obtained by component-wise addition of the vectors of all those verbs (e.g. *create*, *support*, *help*, etc) that are in *doobj* relation with the noun *team*:

$$team^{\circ} = \sum_{\vec{v} \in \mathbf{T}} \vec{v}$$

where  $\mathbf{T}$  is the vector set of verbs having *team* as direct object (except *run*).  $\mathbf{T}$  is thus included in the

subspace of verb vectors. Component-wise addition has an union effect.

Similarly, the dependent function,  $dobj_{\downarrow}$ , combines the noun vector  $\vec{team}$  with the selectional preferences imposed by the verb, noted  $\vec{run}^{\circ}$ , by component-wise multiplication. Such a combination builds the new vector of  $\vec{team}_{dobj_{\downarrow}}$ , which stands for the contextualized sense of *team* given *run* in the *dobj* relation:

$$dobj_{\downarrow}(\vec{run}^{\circ}, \vec{team}) = \vec{team} \odot \vec{run}^{\circ} = \vec{team}_{dobj_{\downarrow}}$$

The selectional preferences imposed by the head word *run* to its direct object are represented by the vector  $\vec{run}^{\circ}$ , which is obtained by adding the vectors of all those nouns (e.g. *company*, *project*, *marathon*, etc) which are in relation *dobj* with the verb *run*:

$$\vec{run}^{\circ} = \sum_{\vec{v} \in \mathbf{R}} \vec{v}$$

where  $\mathbf{R}$  is the vector set of nouns playing the direct object role of *run* (except *team*).  $\mathbf{R}$  is included in the subspace of nominal vectors.

Each multiplicative operation results in a compositional vector of a contextualized word. Component-wise multiplication has an intersective effect. The vector standing for the selectional preferences restricts the vector of the target word by assigning weight 0 to those WordNet features that are not shared by both vectors. The new compositional vector as well as the two constituents all belong to the same vector subspace (the subspace of nouns, verbs, or adjectives).

Notice that, in approaches to computational semantics inspired by Combinatory Categorical Grammar (Steedman, 1996) and Montagovian semantics (Montague, 1970), the interpretation process for composite expressions such as “*run a team*” or “*electric coach*” relies on rigid function-argument structures: relational expressions, like verbs and adjectives, are used as predicates while nouns and nominals are their arguments. In the composition process, each word is supposed to play a rigid and fixed role: the relational word is semantically represented as a selective function imposing constraints on the denotations of the words it combines with, while non-relational words are in turn seen as arguments filling the constraints imposed by the function. For instance, *run*

and *electric* would denote functions while *team* and *coach* would be their arguments.

By contrast, we deny the rigid “predicate-argument” structure. In our compositional approach, dependencies are the active functions that control and rule the selectional requirements imposed by the two related words. Thus, each constituent word imposes its selectional preferences on the other within a dependency-based construction. This is in accordance with non-standard linguistic research which assumes that the words involved in a composite expression impose semantic restrictions on each other (Pustejovsky, 1995; Gamallo et al., 2005; Gamallo, 2008).

## 4.2 Recursive Compositional Application

In our approach, the consecutive application of the syntactic dependencies found in a sentence is actually the process of building the contextualized sense of all the lexical words which constitute it. Thus, the whole sentence is not assigned to a unique meaning (which could be the contextualized sense of the *root* word), but one sense per lemma, being the sense of the *root* just one of them.

This incremental process may have two directions: from left-to-right and vice versa (i.e., from right-to-left). Figure 1 illustrates the incremental process of building the sense of words dependency-by-dependency from left-to-right. Thus, given the composite expression “*the coach runs the team*” and its dependency analysis depicted in the first row of the figure, two compositional processes are driven by the two dependencies involved in the analysis (*nsubj* and *dobj*). Each dependency is decomposed into two functions: head ( $nsubj_{\uparrow}$  and  $dobj_{\uparrow}$ ) and dependent ( $nsubj_{\downarrow}$  and  $dobj_{\downarrow}$ ) functions.<sup>2</sup> The first compositional process applies, on the one hand, the head function  $nsubj_{\uparrow}$  on the denotation of the head verb ( $\vec{run}$ ) and on the selectional preferences required by *coach* ( $\vec{coach}^{\circ}$ ), in order to build a contextualized sense of the verb:  $\vec{run}_{nsubj_{\uparrow}}$ . On the other hand, the dependent function  $nsubj_{\downarrow}$  builds the sense of *coach* as nominal subject of *run*:  $\vec{coach}_{nsubj_{\downarrow}}$ . Then, the contextualized head vector is involved in the compositional process driven by *dobj*. At this level of semantic composition, the selectional preferences imposed on the noun *team*

<sup>2</sup>We do not consider the meaning of determiners, auxiliary verbs, or tense affixes. Quantificational issues associated to them are also beyond the scope of this work.

stand for the semantic features of all those nouns which may be the direct object of *coach+run*. At the end of the process, we have not obtained one single sense for the whole expression, but one contextualized sense per lexical word:  $\vec{coach}_{n_{subj}\downarrow}$ ,  $\vec{run}_{n_{subj}\uparrow+do_{bj}\uparrow}$  and  $\vec{team}_{do_{bj}\downarrow}$ .

In other case, from right-to-left, the verb *run* is first restricted by *team* at the direct object position, and then by its subject *coach*. In addition, this noun is now restricted by the selectional preferences imposed by *run* and *team*, that is, it is combined with the semantic features of all those nouns that may be the nominal subject of *run+team*.

## 5 Experiments

We have performed several similarity-based experiments using the semantic word model defined in Section 3 and the compositional algorithm described in 4.<sup>3</sup> First, in Subsection 5.1, we evaluate just word similarity without composition. Then, in Subsection 5.2, we evaluate the simple compositional approach by making use of a dataset with similar noun-verb pairs (NV constructions). Finally, the recursive application of compositional functions is evaluated in Subsection 5.3, by making use of a dataset with similar noun-verb-noun pairs (NVN constructions).

In all experiments, we made use of datasets suited for the task at hand, and compare our results with those obtained by the best systems for the corresponding dataset. Moreover, in order to build the selectional preferences of the syntactically related words, we used the British National Corpus (BNC). Syntactic analysis on BNC was performed with the dependency parser DepPattern (Gamallo and González, 2011; Gamallo, 2015), previously PoS tagged with Tree-Tagger (Schmid, 1994).

### 5.1 Word Similarity

Recently, the use of word similarity methods has been criticised as a reliable technique for evaluating distributional semantic models (Batchkarov et al., 2016), given the small size of the datasets and the limitation of context information as well. However, given this procedure still is widely accepted, we have performed two different kinds of experiments: rating by similarity and synonym detection with multiple-choice questions.

<sup>3</sup>Both the software and the semantic word model are freely available at <http://fegalaz.usc.es/~gamallo/resources/CompWordNet.tar.gz>.

#### 5.1.1 Rating by Similarity

In the first experiment, we use the WordSim353 dataset (Finkelstein et al., 2002), which was constructed by asking humans to rate the degree of semantic similarity between two words on a numerical scale. This is a small dataset with 353 word pairs. The performance of a computational system is measured in terms of correlation (Spearman) between the scores assigned by humans to the word pairs and the similarity Dice coefficient assigned by our system (*WN*) built with the WordNet-based model space.

Table 1 compares the Spearman correlation obtained by our model, *WN*, with that obtained by the corpus-based system described in (Halawi et al., 2012), which is the highest score reached so far on that dataset. Even if our results were clearly outperformed by that corpus-based method, *WN* seems to behave well if compared with the state-of-the-art knowledge-based (unsupervised) strategy reported in (Agirre et al., 2009).

<i>Systems</i>	$\rho$	<i>Method</i>
<i>WN</i>	0.69	knowledge
(Hassan and Mihalcea, 2011)	0.62	knowledge
(Agirre et al., 2009)	0.66	knowledge
(Halawi et al., 2012)	<b>0.81</b>	corpus

Table 1: Spearman correlation between the WordSim353 dataset and the rating obtained by our knowledge-based system *WN* and the state-of-the-art for both knowledge and corpus-based strategies.

#### 5.1.2 Synonym Detection with Multiple-Choice Questions

In this evaluation task, a target word is presented with four synonym candidates, one of them being the correct synonym of the target. For instance, for the target *deserve*, the system must choose between *merit* (the correct one), *need*, *want*, and *expect*. Accuracy is the number of correct answers divided by the total number of words in the

<i>Systems</i>	<i>Noun</i>	<i>Adj</i>	<i>Verb</i>	<i>All</i>
<i>WN</i>	<b>0.85</b>	<b>0.85</b>	<b>0.75</b>	0.80
(Freitag et al., 2005)	0.76	0.76	0.64	0.72
(Zhu, 2015)	0.71	0.71	0.63	0.69
(Kiela et al., 2015)	-	-	-	<b>0.88</b>

Table 2: Accuracy of three systems on the WBST test (synonym detection on nouns, adjectives, and verbs)

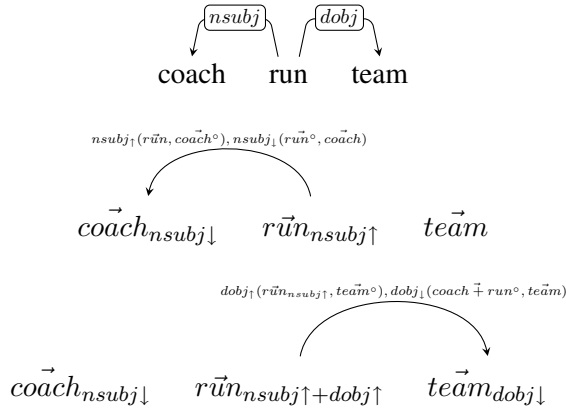


Figure 1: Syntactic analysis of the expression “the coach runs the team” and left-to-right construction of the word senses.

dataset.

The dataset is an extended TOEFL test, called the WordNet-based Synonymy Test (WBST) proposed in (Freitag et al., 2005). WBST was produced by generating automatically a large set of TOEFL-like questions from the synonyms in WordNet. In total, this procedure yields 9,887 noun, 7,398 verb, and 5,824 adjective questions, a total of 23,509 questions, which is a very large dataset. Table 2 shows the results. In this case, the accuracy obtained by *WN* for the three syntactic categories is close to state-of-the-art corpus-based method for this task (Kiela et al., 2015), which is a neural network trained with a huge corpus containing 8 billion words from English Wikipedia and newswire texts.

## 5.2 Noun-Verb Composition

The first experiment aimed at evaluating our compositional strategy uses the test dataset by Mitchell and Lapata (2008), which comprises a total of 3,600 human similarity judgments. Each item consists of an intransitive verb and a subject noun, which are compared to another noun-verb pair (NV) combining the same noun with a synonym of the verb that is chosen to be either similar or dissimilar to the verb in the context of the given subject. For instance, “child stray” is related to “child roam”, being *roam* a synonym of *stray*. The dataset was constructed by extracting NV composite expressions from the British National Corpus (BNC) and verb synonyms from WordNet. In order to evaluate the results of the tested systems, Spearman correlation is computed between individual human similarity scores and the systems’ predictions.

In this experiment, we compute the similarity between the contextualized heads of two NV composites and between their contextualized dependent expressions. For instance, we compute the similarity between “eye flare” vs “eye flame” by comparing first the verbs *flare* and *flame* when combined with *eye* in the subject position (head function), and by comparing how (dis)similar is the noun *eye* when combined with both the verbs *flare* and *flame* (dependent function). In addition, as we are provided with two similarities (head and dep) for each pair of compared expressions, it is possible to compute a new similarity score by averaging the results of head and dependent functions (head+dep).

Table 3 shows the Spearman’s correlation values ( $\rho$ ) obtained by the three versions of *WN*: only head function (head), only dependent function (dep) and average of both (head+dep). The latter score value is comparable to the state-of-the-art system for this dataset, reported in (Erk and Padó, 2008). It is also very similar to the most recent results described in (Dinu et al., 2013), where the authors made use of the compositional strategy defined in (Baroni and Zamparelli, 2010).

Systems	$\rho$
WN (head+dep)	<b>0.29</b>
WN (head)	0.26
WN (dep)	0.14
(Erk and Padó, 2008)	0.27
(Dinu et al., 2013)	0.26

Table 3: Spearman correlation for intransitive expressions using the benchmark by Mitchell and Lapata (2008)

### 5.3 Noun-Verb-Noun Composition

The last experiment consists in evaluating the quality of compositional vectors built by means of the consecutive application of head and dependency functions associated with nominal subject and direct object. The experiment is performed on the dataset developed in (Grefenstette and Sadrzadeh, 2011a). The dataset was built using the same guidelines as Mitchell and Lapata (2008), using transitive verbs paired with subjects and direct objects: NVN composites.

Given our compositional strategy, we are able to compositionally build several vectors that somehow represent the meaning of the whole NVN composite expression. In order to know which is the best compositional strategy and be exhaustive and complete, we evaluate all of them; i.e., both left-to-right and right-to-left strategies. Thus, take again the expression “*the coach runs the team*”. If we follow the left-to-right strategy (noted *nv-n*), at the end of the compositional process, we obtain two fully contextualized senses:

**nv-n\_head** The sense of the head *run*, as a result of being contextualized first by the preferences imposed by the subject and then by the preferences required by the direct object. We note *nv-n\_head* the final sense of the head in a NVN composite expression following the left-to-right strategy.

**nv-n\_dep** The sense of the object *team*, as a result of being contextualized by the preferences imposed by *run* previously combined with the subject *coach*. We note *nv-n\_dep* the final sense of the direct object in a NVN composite expression following the left-to-right strategy.

If we follow the right-to-left strategy (noted *n-vn*), at the end of the compositional process, we obtain two fully contextualized senses:

**n-nv\_head** The sense of the head *run* as a result of being contextualized first by the preferences imposed by the object and then by the subject.

**n-nv\_dep** The sense of the subject *coach*, as a result of being contextualized by the preferences imposed by *run* previously combined with the object *team*.

<i>Systems</i>	$\rho$
WN ( <i>nv-n_head+dep</i> )	0.35
WN ( <i>nv-n_head</i> )	0.34
WN ( <i>nv-n_dep</i> )	0.13
WN ( <i>n-vn_head+dep</i> )	<b>0.50</b>
WN ( <i>n-vn_head</i> )	0.35
WN ( <i>n-vn_dep</i> )	0.44
WN ( <i>n-vn+nv-n</i> )	0.47
(Grefenstette and Sadrzadeh, 2011b)*	0.28
(Van De Cruys et al., 2013)*	0.37
(Tsubaki et al., 2013)*	0.44
(Milajevs et al., 2014)	0.46
(Polajnar et al., 2015)	0.35
(Hashimoto et al., 2014)	0.48
(Hashimoto and Tsuruoka, 2015)	0.48
Human agreement	0.75

Table 4: Spearman correlation for transitive expressions using the benchmark by Grefenstette and Sadrzadeh (2011)

Table 4 shows the Spearman’s correlation values ( $\rho$ ) obtained by all the different versions built from our model *WN*. The best score was achieved by averaging the head and dependent similarity values derived from the *n-vn* (right-to-left) strategy. Let us note that, for NVN composite expressions, the left-to-right strategy seems to build less reliable compositional vectors than the right-to-left counterpart. Besides, the combination of the two strategies (*n-vn+nv-n*) does not improve the results of the best one (*n-vn*).<sup>4</sup> The score values obtained by the different versions of the right-to-left strategy outperform other systems for this dataset (see results reported below in the table). Our best strategy ( $\rho = 0.50$ ) also outperforms the neural network strategy described in (Hashimoto and Tsuruoka, 2015), which achieved 0.48 without considering extra linguistic information not included in the dataset. The ( $\rho$ ) scores for this task are reported for averaged human ratings. This is due to a disagreement in previous work regarding which metric to use when reporting results. We mark with asterisk those systems reporting ( $\rho$ ) scores based on non-averaged human ratings.

## 6 Conclusions

In this paper, we described a compositional model based on WordNet features and dependency-based functions on those features. It is a recursive proposal since it can be repeated from left-to-right or from right-to-left and the sense of each constituent word is performed in a recursive way.

<sup>4</sup>*n-vn+nv-n* is computed by averaging the similarities of both *n-vn\_head+dep* and *nv-n\_head+dep*



Our compositional model tackles the problem of *information scalability*. This problem states that the size of semantic representations should not grow exponentially, but proportionally; and, information must not be lost using fixed size of compositional vectors. In our approach, however, even if the size of the compositional vectors is fixed, there is no information loss since each word of the composite expression is associated to a compositional vector representing its context-sensitive sense. In addition, the compositional vectors do not grow exponentially since their size is fixed by the vector space: they are all first-order (or direct) vectors. Finally, the number of vectors increases in proportion to the number of constituent words found in the composite expression. So, both points are successfully solved.

In future work, we will try to design a compositional model based on word semantic representations combining WordNet-based features with syntactic-based distributional contexts as well as extend our model to full sentences instead of the simple ones described in this paper.

## Acknowledgements

We would like to thank the anonymous reviewers for helpful comments and suggestions. This work was supported by a 2016 BBVA Foundation Grant for Researchers and Cultural Creators, and by project TELPARES (FFI2014-51978-C2-1-R) and grant TIN2014-56633-C3-1-R, Ministry of Economy and Competitiveness. It has received financial support from the Consellería de Cultura, Educación e Ordenación Universitaria (Postdoctoral Training Grants 2016 and Centro Singular de Investigación de Galicia accreditation 2016-2019, ED431G/08) and the European Regional Development Fund (ERDF).

## References

Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Paşca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 19–27, Stroudsburg, PA, USA. Association for Computational Linguistics.

Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space.

In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP'10, pages 1183–1193, Stroudsburg, PA, USA.

Marco Baroni, Raffaella Bernardi, and Roberto Zamparelli. 2014. Frege in space: A program for compositional distributional semantics. *LiLT*, 9:241–346.

Marco Baroni. 2013. Composition in distributional semantics. *Language and Linguistics Compass*, 7:511–522.

Miroslav Batchkarov, Thomas Kober, Jeremy Reffin, Julie Weeds, and David Weir. 2016. A critique of word similarity as a method for evaluating distributional semantic models. In *Proceedings of the ACL Workshop on Evaluating Vector Space Representations for NLP*, Berlin, Germany.

B. Coecke, M. Sadrzadeh, and S. Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis*, 36(1-4):345–384.

Georgiana Dinu and Mirella Lapata. 2010. Measuring distributional similarity in context. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1162–1172, Cambridge, MA, October. Association for Computational Linguistics.

G. Dinu, N. Pham, and M. Baroni. 2013. General estimation and evaluation of compositional distributional semantic models. In *ACL 2013 Workshop on Continuous Vector Space Models and their Compositionality (CVSC 2013)*, pages 50–58, East Stroudsburg PA.

Katrin Erk and Sebastian Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of EMNLP*, Honolulu, HI.

Manaal Faruqui and Chris Dyer. 2015. Non-distributional word vector representations. In *Proceedings of ACL*.

Lev Finkelstein, Evgeniy Gabilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppín. 2002. Placing search in context: the concept revisited. *ACM Trans. Inf. Syst.*, 20(1):116–131.

Dayne Freitag, Matthias Blume, John Byrnes, Edmond Chow, Sadik Kapadia, Richard Rohwer, and Zhiqiang Wang. 2005. New experiments in distributional representations of synonymy. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 25–32.

Pablo Gamallo and Isaac González. 2011. A grammatical formalism based on patterns of part-of-speech tags. *International Journal of Corpus Linguistics*, 16(1):45–71.

- Pablo Gamallo, Alexandre Agustini, and Gabriel Lopes. 2005. Clustering Syntactic Positions with Similar Semantic Requirements. *Computational Linguistics*, 31(1):107–146.
- Pablo Gamallo. 2008. The meaning of syntactic dependencies. *Linguistik OnLine*, 35(3):33–53.
- Pablo Gamallo. 2015. Dependency parsing with compression rules. In *International Workshop on Parsing Technology (IWPT 2015)*, Bilbao, Spain.
- Pablo Gamallo. 2017. The role of syntactic dependencies in compositional distributional semantics. *Corpus Linguistics and Linguistic Theory*.
- Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011a. Experimental support for a categorical compositional distributional model of meaning. In *Conference on Empirical Methods in Natural Language Processing*.
- Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011b. Experimenting with transitive verbs in a disccat. In *Workshop on Geometrical Models of Natural Language Semantics (EMNLP-2011)*.
- Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. 2011. Concrete sentence spaces for compositional distributional models of meaning. In *Proceedings of the Ninth International Conference on Computational Semantics, IWCS '11*, pages 125–134.
- Emiliano Guevara. 2010. A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics, GEMS '10*.
- M. A. Hadj Taieb, M. Ben Aouicha, and A. Ben Hamadou. 2014. Ontology-based approach for measuring semantic similarity. *Engineering Applications of Artificial Intelligence*, 36:238–261.
- Guy Halawi, Gideon Dror, Evgeniy Gabrilovich, and Yehuda Koren. 2012. Large-scale learning of word relatedness with constraints. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*, pages 1406–1414.
- Kazuma Hashimoto and Yoshimasa Tsuruoka. 2015. Learning embeddings for transitive verb disambiguation by implicit tensor factorization. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 1–11, Beijing, China, July. Association for Computational Linguistics.
- Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. 2014. Jointly learning word representations and composition functions using predicate-argument structures. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1544–1555, Doha, Qatar, October. Association for Computational Linguistics.
- S. Hassan and R. Mihalcea. 2011. Semantic relatedness using salient semantic analysis. In *Proceedings of AAAI Conference on Artificial Intelligence*.
- Douwe Kiela, Felix Hill, and Stephen Clark. 2015. Specializing word embeddings for similarity or relatedness. In Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton, editors, *EMNLP*, pages 2044–2048. The Association for Computational Linguistics.
- Jayant Krishnamurthy and Tom Mitchell, 2013. *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, chapter Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models, pages 1–10. Association for Computational Linguistics.
- Lingling Meng, Runqing Huang, and Junzhong Gu. 2013. A review of semantic similarity measures in wordnet. *International Journal of Hybrid Information Technology*, 6(1).
- Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver. 2014. Evaluating neural word representations in tensor-based compositional settings. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *EMNLP*, pages 708–719. ACL.
- Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT*, pages 236–244.
- Jeff Mitchell and Mirella Lapata. 2009. Language models based on semantic composition. In *Proceedings of EMNLP*, pages 430–439.
- Jeff Mitchell and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1439.
- Richard Montague. 1970. Universal grammar. *theoria. Theoria*, 36:373–398.
- Barbara Partee. 1984. Compositionality. In Frank Landman and Frank Veltman, editors, *Varieties of Formal Semantics*, pages 281–312. Dordrecht: Foris.
- Tamara Polajnar, Laura Rimell, and Stephen Clark. 2015. An exploration of discourse-based sentence spaces for compositional distributional semantics. In *Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics*, pages 1–11, Lisbon, Portugal, September. Association for Computational Linguistics.
- James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge.

- R. Rada, H. Mili, E. Bicknell, and M. Blettner. 1989. Development and application of a metric on semantic nets. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(1):17–30.
- Siva Reddy, Ioannis P. Klapaftis, Diana McCarthy, and Suresh Manandhar. 2011. Dynamic and static prototype vectors for semantic composition. In *Fifth International Joint Conference on Natural Language Processing, IJCNLP 2011, Chiang Mai, Thailand, November 8-13, 2011*, pages 705–713.
- M. Andrea Rodríguez and Max J. Egenhofer. 2003. Determining semantic similarity among entity classes from different ontologies. *IEEE Trans. Knowl. Data Eng.*, 15(2):442–456.
- H. Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*.
- Thabet Slimani. 2013. Description and evaluation of semantic similarity measures approaches. *International Journal of Computer Applications*, 80(1):25–33.
- Mark Steedman. 1996. *Surface Structure and Interpretation*. The MIT Press.
- Stefan Thater, Hagen Fürstenaу, and Manfred Pinkal. 2010. Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 948–957, Stroudsburg, PA, USA.
- Masashi Tsubaki, Kevin Duh, Masashi Shimbo, and Yuji Matsumoto. 2013. Modeling and learning semantic co-compositionality through prototype projections and neural networks. In *EMNLP*, pages 130–140. ACL.
- Peter D. Turney. 2013. Domain and function: A dual-space model of semantic relations and compositions. *Journal of Artificial Intelligence Research (JAIR)*, 44:533–585.
- A. Tversky. 1977. Features of similarity. *Psychological Review*, 84(4).
- Tim Van De Cruys, Thierry Poibeau, and Anna Korhonen. 2013. A Tensor-based Factorization Model of Semantic Compositionality. In *Conference of the North American Chapter of the Association of Computational Linguistics (HTL-NAACL)*, pages 1142–1151, Atlanta, United States, June.
- Fabio Massimo Zanzotto, Ioannis Korkontzelos, Francesca Fallucchi, and Suresh Manandhar. 2010. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, pages 1263–1271.
- Peng Zhu. 2015. N-grams based linguistic search engine. *International Journal of Computational Linguistics Research*, 6(1):1–7.