Pablo Gamallo*

The role of syntactic dependencies in compositional distributional semantics

DOI 10.1515/cllt-2016-0038

Abstract: This article provides a preliminary semantic framework for Dependency Grammar in which lexical words are semantically defined as contextual distributions (sets of contexts) while syntactic dependencies are compositional operations on word distributions. More precisely, any syntactic dependency uses the contextual distribution of the dependent word to restrict the distribution of the head, and makes use of the contextual distribution of the head to restrict that of the dependent word. The interpretation of composite expressions and sentences, which are analyzed as a tree of binary dependencies, is performed by restricting the contexts of words dependency by dependency in a left-to-right incremental way. Consequently, the meaning of the whole composite expression or sentence is not a single representation, but a list of contextualized senses, namely the restricted distributions of its constituent (lexical) words. We report the results of two large-scale corpus-based experiments on two different natural language processing applications: paraphrasing and compositional translation.

Keywords: distributional similarity, compositional semantics, syntactic analysis, dependencies

1 Introduction

The main proposal of this paper is to put syntactic dependencies at the core of semantic composition. We propose a semantic space in which each syntactic dependency is associated with two binary operations: a *head operation* which builds the sense of the head word by considering the semantic restrictions (or selectional preferences) of the dependent one, and a *dependent operation* which results in a new sense of the dependent word by taking into account the selectional preferences of the head. Consider for instance the expression “drive a tunnel” in which the two words are related by the direct object dependency.

***Corresponding author: Pablo Gamallo,** CiTIUS – Centro Singular de Investigación en Tecnoloxías da Información, University of Santiago de Compostela, Galiza, Spain, E-mail: pablo.gamallo@usc.es

In that context, the head function makes use of the selectional preferences of the noun to select for the *digging* sense of the polysemous verb "drive". By contrast, in "read a passage", it is the dependent function that uses the preferences of the verb to activate one of the senses of the polysemous noun "passage", namely *a segment of a written work or speech*, instead of *a path or channel* or *the act of moving*. It follows that a syntactic dependency between two words carries two complementary selective functions, each one imposing its own selectional preferences. These two functions allow the two related words to mutually disambiguate or discriminate the sense of each other by co-selection (or co-discrimination).

Besides semantic composition by co-selection, we also define semantic interpretation as an incremental process. Interpretation is built up from left to right as each syntactic dependency is processed, following their combinatorial properties. For instance, to interpret "the bulldozer drove a passage", we identify two syntactic dependencies, subject and direct object, and define two sequential composition processes. First, the subject dependency uses the (inverse) selectional preferences of "bulldozer" to select the digging sense of the verb "drive". And then, the direct object relation makes use of the preferences required by "drive" (already disambiguated) to select the path/channel sense of the noun "passage".

Figure 1 illustrates the compositional and incremental process of building the sense of words by co-selection and left-to-right. Given the composite expression "a b c" and its dependency analysis depicted in the first row of the figure, two compositional processes are driven by the two dependencies (r_m and r_n) involved in the analysis. First, r_m is decomposed into two functions: h_m and d_m. The head function h_m takes the sense of the head word b and the selectional preferences of a, noted here as a^m, as input, and returns a new denotation of the head word: b_1. Similarly, the dependent function d_m takes as input the sense of the dependent word a and the selectional preferences b^m, and returns a new

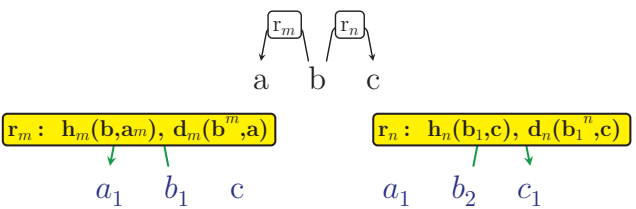

Figure 1: Syntactic analysis of the expression "a b c" and left-to-right construction of the word senses.

denotation of the dependent word: a_1. Next, the relation r_n between words b and c is also decomposed into the head and dependent functions: h_n and d_n. Function h_n combines the contextualized head b_1 with the selectional preferences c^n, and returns a more specific sense of the head: b_2. Finally, function d_n takes as input the sense of the dependent word c and the selectional preferences b_1^n, and builds a new contextualized sense of the dependent word: c_1. The subscript specifies the number of times a word has been combined with another. At the end of the process, we have not obtained one single sense for the whole expression "a b c", but one contextualized sense per word: a_1, b_2 and c_1. Notice that b_2 is the sense of the root and, then, it can be seen as the sense of the composite expression.

In our approach, words and their selectional preferences denote distributional representations, whereas syntactic dependencies are compositional operations on them. In Erk (2013), distributional representations stand for mental objects which are linked to intensions of logical expressions. Similarly, in Copestake and Herbelot (2012), distributions are also used as intensions, but they are linked to extensions, namely to the *ideal distribution* of a word, which consists of all the contexts in which a word could occur. We follow the Copestake and Heberlot's suggestion and so define word denotations as sets of contexts. Selectional preferences will also be defined as context sets, and semantic composition driven by syntactic dependencies will be just set intersections between context sets.

Dependencies have been traditionally considered as syntactic objects. They are at the centre of dependency-based grammars: e.g. the Tesnière's Dependency Grammar (Tesniére 1959), the Mel'čuk's Meaning-Text Theory (Kahane 2003), or Word Grammar (Hudson 2003). However, the meaning of dependencies has not been clearly defined yet. In our approach, we situate dependencies in the semantic space where they denote compositional operations. The main contribution of this article is to provide a semantic description of syntactic dependencies by taking into account compositional distributional semantics and incremental interpretation in a corpus-based approach.

This article is organized as follows. In the next Section 2, we describe our compositional semantic model where lexical words denote entities defined in distributional terms as sets of contexts, while dependencies are binary functions on those entities. Special attention will be paid to a particular application: compositional translation based on non-parallel corpora. Next, in Section 3 we define the incremental approach to interpretation. We analyze some examples from our corpus-based application for compositional and incremental translation. Then, Section 4 describes two large-scale corpus-based experiments: paraphrasing and compositional translation. In Section 5, we introduce related work

on compositional distributional approaches as well as on theories based on incremental interpretation. And finally, relevant conclusions are reported in Section 6.

2 The semantic space

We propose a simple semantic space with just two semantic types: *entities* and *binary relations*. Entities are denotations of lexical units and binary relations are denotations of syntactic dependencies (nominal subject, direct object, noun modification, prepositional object, ...). Dependencies combine entities to construct more specific entities in a compositional way.

Notice that we use the term *entity* to refer to basic objects, such as events or individuals, with no internal structure and which can not be defined in terms of other, more basic entities (Davidson 1969).

The semantics of determiners and verb specifiers are beyond the scope of the article. Therefore, grounding operations and quantification are not considered in our universe of interpretation.

The two semantic types, entities and dependencies, are defined in the following subsections.

2.1 Entities

Lexical units (i.e., content words) denote *entities*, which we define as set of contexts in a distributional model. Distributional contexts of words are semantic representations that can stand for intensions or mental concepts in the semantic space (Erk 2013) and may represent word extensions when they are taken as *ideal distributions* (i.e. all the contexts in which a word could occur with respect to some microworld) (Copestake and Herbelot 2012). As we can obtain a simple correspondence between ideal distributions and a first-order notion of extension, we may apply basic algebraic operations on them giving rise to a set-theoretic model. For example, contexts where "horse" is the subject of "run" will be a subset of all the contexts where "horse" occurs in the subject position, which in turn will be a subset of all the contexts in which "horse" occurs. We identify a context with the position of a word in a specific syntactic dependency: for instance $\langle N\_nsubj\_run \rangle$ is a context of "horse", and $\langle horse\_nsubj\_V \rangle$ is a context of "run". It means that the nominal subject of "run" is a context of "horse", while a context of the verb "run" is the noun "horse" in the subject position. The set of (ideal) contexts of "run" represents the entity *run*, while

horse is the set of contexts of "horse". Hereafter, denotations are noted in italics (*horse* and *run* are entities) and linguistic units with quotation marks ("horse" is a noun and "run" is a verb).

2.1.1 Word senses and selectional preferences

Given a word, we distinguish two types of context sets that can be associated with it: the *meaning* of a word (or its *sense* if the word has been contextualized), and the selectional preferences (or lexical restrictions) the word imposes on another one within a syntactic dependency.

The meaning of a word is the set of potential contexts in which such word could be used. Difference in meaning correlates with difference in distribution (Harris 1954). This idea was somehow inspired from the first linguists, such as Meillet (1921), who associated word meaning with word use at the beginning of the twentieth century.

The selectional preferences imposed by a word in a dependency are the contexts of all words that co-occur with it in that dependency. Word meanings are combined with selectional preferences to yield contextualized senses.

For instance, take again the words "run" and "horse" and their combination by means of the dependency *nsubj* (e.g. "a horse is running"). On the one hand, the entity *run* (the set of contexts in which "run" occurs) is combined with the preferences imposed by "horse" in that syntactic position, and noted $horse^{nsubj}$. In this situation, the entity $horse^{nsubj}$ represents the contexts of all verbs ("eat", "jump", ...) that may have "horse" as subject (except "run"). Both sets, *run* and $horse^{nsubj}$, are compatible and can be intersected since all are contexts of verbs. Their intersection is a not empty set that represents the sense of "run" in composite expressions such as "a horse is running":

$$run \cap horse^{nsubj} = run_1$$

It means that the contexts of the verbs the noun "horse" is the subject of ("eat", "jump", etc.) are used to restrict the set of contexts of "run". This restriction enables selecting only those contexts of "run" that activate one specific sense of the verb: the one referring to a *physical movement*. In other words, the combination of *run* with $horse^{nsubj}$ gives rise to a more specific running event, run_1, which is a subset of *run*.

On the other hand, the entity *horse* is combined with the preferences imposed by "run" as head of the *nsubj* relation. These preferences, noted run^{nsubj}, represent the contexts of all nouns ("dog", "car", "computer", ...) that may be in the subject position of the verb "run" (except "horse"). Both sets, *horse* and run^{nsubj}, are

compatible since all are contexts of nouns. Their intersection represents the contextualized sense of "horse" in expressions like "a horse is running":

$$horse \cap run^{nsubj} = horse_1$$

It results in a more elaborate denotation of the noun, $horse_1$, which is a subset of *horse*. Both, run_1 and $horse_1$, are contextualized senses of the generic meanings *run* and *horse*. Word meanings, word senses, and selectional preferences are the entities of our semantic space.

In a similar way, the contextualized senses of "coach" and "electric" in the composite "electric coach" are built by means of the following intersections:

$$coach \cap electric^{nmod} = coach_1$$

$$electric \cap coach^{nmod} = electric_1$$

Notice that the meaning of "coach", out of context, includes two opposite senses: *bus* and *trainer*. The selectional preferences imposed by the adjective as nominal modifier, $electric^{nmod}$, are the contexts of those nouns that can be modified by "electric" (e.g. "car", "device", etc.). They are combined with the contexts of the noun "coach" to build the new entity, $coach_1$, which is the subset of *coach* mostly referring to the *bus* sense, leaving most contexts related to the *trainer* sense out of the new entity.[1] On the other hand, $coach^{nmod}$ are the preferences imposed by the noun "coach". They represent the contexts of the adjectives that may modify the noun, and are used to select a contextualized sense of the adjective: $electric_1$.

Our definitions of *selectional preferences* and *word sense* are related to the Corpus Pattern Analysis (CPA) described in Hanks (2013) and Jezek and Hanks (2010). The authors state that the selectional preferences cannot be reduced to discrete categories such as Humans, Food, Artifact, Activity, etc. Lexical coertions are usually involved in word combinations and thus unexpected arguments are the rule and not the exception. In CPA, the ontology of lexical categories is a statistically based structure of collocational preferences, called "shimmering lexical sets". Each canonical member of a lexical set is recorded with statistical contextual information. Besides, in Jezek and Hanks (2010), it is assumed a mutual semantic conditioning between heads and dependents.

These ideas on selectional preferences and word sense are also close to the *discriminating* model (Schütze 1998), which does not make use of predefined and

1 Note that the new entity $coach_1$ also should contain contexts related to the *trainer* sense, which might be activated if the nominal expression is inserted in a larger linguistic context that clearly refers to that sense: "I bought an electric coach to train in my flat".

labeled senses from external resources such as WordNet (Fellbaum 1998). *Word sense discrimination* is based on unsupervised techniques which discriminate word senses by clustering similar words, i.e. by identifying words with similar context distribution. The distributional senses (i.e. word clusters) discovered by these techniques may not be equivalent to the traditional senses in a dictionary sense inventory. Indeed, they may not be related to discrete categories in a traditional ontology. For this reason, the evaluation of word sense discrimination is a difficult task (Navigli 2009). By contrast, supervised strategies for *word sense disambiguation* rely on text previously annotated with pre-defined sense labels (e.g. WordNet identifiers), and their objective is to learn classifiers to assign those sense labels to word instances.

Our approach uses an unsupervised strategy to discriminate senses. It is actually a compositional approach to word sense discrimination. However, to simplify, in this article the terms *disambiguation* and *discrimination* are used in the same way: to refer to those cases in which the selectional preferences tend to activate one of the senses of a polysemous word.

Word sense discrimination/disambiguation is a different task from *word specification* or *word restriction*, which means that the selectional preferences just specify the unambiguous meaning of a word or a previously discriminated sense. For instance, in "electric coach", co-selection performs both sense discrimination and sense specification. On the one hand, the adjective "coach" discriminates one of the senses of "coach" by activating the *bus* sense, and on the other hand, the noun "coach" just specifies the unambiguous meaning of the adjective. As we do not provide any formal definition of *sense*, the difference between discrimination and specification is established in an intuitive way. Their formal definition is beyond the scope of the article. When the difference between discrimination and specification is not relevant for our claims, we use the generic terms *selection* or *contextualization*. Similarly, co-selection may refer to different cases of contextualization: co-discrimination ("drive a passage"), co-specification ("electric engine"), or discrimination + specification ("electric coach").

There has been some criticism against the intersective method in compositional semantics. The intersection can fail if denotations of nouns, adjectives, and intransitive verbs are defined as predicates (from sets of individuals to true/false values) (Partee 2007). For instance, the meaning of "former president" is not a president any more, or the meaning of "fake gun" is not actually a gun. However, in the model we propose, $fake^{nmod}$ is a very large entity (the contexts of all nouns that are actually modified by "fake") which should share many contexts/properties (e.g. shape, color, etc.) with *gun* (the contexts of noun "gun"). Thus, in a semantic model based on context distribution,

a fake gun is actually a kind of gun and, thereby, the context intersection between $fake^{nmod}$ and *gun* should not be empty.

2.1.2 Subtypes of entities

We consider that lexical words belonging to different syntactic categories (e.g. "horse", "run", "electric") have their own types of contexts, which are defined in different and incompatible distributional spaces. Contexts of nouns differ from contexts of verbs which, in turn, are different from contexts of adjectives and adverbs. According to these differences, we distinguish three subtypes of entities: *individuals*, which are defined in the space of nominal contexts; *processes*, defined as verbal contexts; and *qualities* consisting of contexts of adjectives and adverbs. Therefore, the entities denoted by nouns like "horse" or "John", as well as the root head of expressions like "horse running in the park" or "the man who is running in the park" are all individuals. Similarly, the entities denoted by verbs like "run" or "eat" as well as the root head of expressions such as "eat meat", "is eating meat", "John is eating", or "John is eating meat", are all processes. Qualities are the entities denoted by "electric" or "slowly" and the head of "very good", "difficult to do", and so on. The upper-level ontology of entity subtypes we have introduced is close to the main *kinds* defined in Aristotle's Categories (Studtmann 2014).

This is in accordance with the semantic categories proposed by Cognitive Grammar (Langacker 1991). This theory distinguishes three basic semantic types according to the modes of organizing denotations: *things* are denoted by nouns and nominals, *processes* are denoted by verbs and clauses, and finally *atemporal relations* are associated with adjectives and adverbs.[2] These three basic categories are defined according to their different ways of organizing denotations. A study on denotations according to their various modes of grammatical organization is beyond the scope of the paper. For more details, see (Gamallo 2003).

2.2 Dependency-based compositional functions

In the semantic space, a dependency is associated with two binary functions: both the *head* and the *dependent* functions. The head function takes as input the

2 Besides adjectives and adverbs, prepositions also denote atemporal relations in Cognitive Grammar. However, in our model prepositions will introduce syntactic dependencies.

meaning/sense of the head and the selectional preferences associated with the dependent word, and it results in a more restricted/contextualized sense of the head. The dependent function takes as input the meaning/sense of the dependent word and the selectional preferences imposed by the head, and it yields a contextualized sense of the dependent.

Let us consider a syntactic dependency, *nsubj* (nominal subject), which denotes two compositional functions in the semantic space represented by the following binary λ-expressions:

$$\lambda y \lambda x \, nsubj_h(x, y) \tag{1}$$

$$\lambda x \lambda y \, nsubj_d(x, y) \tag{2}$$

where x and y are variables for entities: x stands for the denotation of the head while y represents the denotation of the dependent. The semantic type of any compositional function (derived from a binary dependency) is $\langle e, \langle e, e \rangle \rangle$, where e is the atomic type for entities. The first argument of a compositional function is the entity used to contextualize the sense of the second argument. If we apply the head function represented by the λ-expression in (1) to the individual *horse*, and the dependency function represented in 2 to the process *run*, we obtain the following unary functions:

$$\lambda x \; nsubj_h(x, horse^{nsubj}) \tag{3}$$

$$\lambda y \; nsubj_d(run^{nsubj}, y) \tag{4}$$

where $horse^{nsubj}$ is the result of unifying the contexts of all those verbs related with "horse" via the subject dependency, while run^{nsubj} is the result of unifying the contexts of all those nouns related with "run" at the subject position, more formally:

$$horse^{nsubj} = \bigcup_{P \in Horse} P \tag{5}$$

$$run^{nsubj} = \bigcup_{I \in Run} I \tag{6}$$

In eq. (5), P represents a *process* (or set of contexts denoted by a verb) and ***Horse*** stands for a very specific set of sets, namely the set of entities denoted by verbs co-occurring with "horse" in the subject position. In eq. (6), I represents an *individual* (or set of contexts denoted by a noun) and ***Run*** stands for the set of entities denoted by nouns co-occurring with "run" at the subject position.

In more intuitive terms, the unary head function represented in 3 stands for the inverse selectional preferences that the noun "horse" imposes on any verb in the subject position, while the dependent function in 4 can be seen as the

selectional preferences imposed by the verb "run" on its nominal subject. Both functions are of type $\langle e, e \rangle$. They take an entity as argument and return a more elaborate entity. In the case of 3, it takes a process and returns a process restricted by the nominal subject. For 4, it takes an individual and returns an individual specified by the verb. If these two functions are applied to *run* and *horse*, respectively, we obtain:

$$nsubj_h(run, horse^{nsubj}) = \quad run \cap horse^{nsubj} = \quad run_1 \tag{7}$$

$$nsubj_d(run^{nsubj}, horse) = \quad horse \cap run^{nsubj} = \quad horse_1 \tag{8}$$

In each combination, we make the intersection of two sets of contexts.[3] The final set resulting from the head function, run_1 (see eq. (7) above), represents a contextualized process, which is the denotation of the head verb in composite expressions such as "a horse is running", "horses ran", etc. This contextualized process is a subset of that denoted by "run": $run_1 \subseteq run$. The set resulting from the dependent function (eq. (8)) represents a contextualized individual, which is a subset of the entity denoted by "horse": $horse_1 \subseteq horse$.

Notice that, in approaches to computational semantics inspired by Combinatory Categorial Grammar (Steedman 1996) and Montagovian semantics (Montague 1970), the interpretation process for composite expressions such as "horses are running" 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 "red" would denote functions while "horses" and "car" would be their arguments. By contrast, we do not define verbs and adjectives (or adverbs) as functional artifacts driving the compositional process. In our compositional approach, dependencies are the active functions that control and rule the selectional requirements imposed by the two related words. Dependencies, instead of relational words, are then conceived of as the main functional operations taking part in composition. In fact, our unary predicates can be seen as semantic structures very similar to the functions denoted by

3 Before intersecting both sets, $horse^{nsubj}$ must be updated by removing the contexts that were only provided by "run", while the contexts only provided by "horse" should also be removed from run^{nsubj}.

adjectives or intransitive verbs in the standard compositional approaches. For instance, $\lambda x\ nmod_h(x, red^{nsubj})$ is a function from entities to entities, ($\langle e, e\rangle$), more precisely from individuals to individuals, as the traditional denotation of adjective "red". However, in our model this function is not directly associated with the adjective but to the lexico-syntactic pattern in which the adjective is assigned the role of dependent (within the *nmod* syntactic dependency). In addition, we consider that the same dependency, *nmod*, also enables a *from qualities to qualities* function, such as $\lambda y\ nmod_h(car^{nsubj}, y)$, which refers to the preferences imposed by "car" to the adjectives that may modify it.

This way, two syntactically dependent expressions are no longer interpreted as a rigid "predicate-argument" structure, where the predicate is the active function imposing the semantic preferences on a passive argument, which matches such preferences. On the contrary, each constituent word imposes its selectional preferences on the other one 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 2008; Gamallo et al. 2005). Not only verbs or adjectives are taken as predicates selecting different types of nouns, but also nouns select for different types of verbs and adjectives.

The combination of a verbal process with a nominal entity via a syntactic dependency actually represents the assignment operation of an argument to a particular semantic/thematic role of the (content of the) verb.

2.3 A case study: Translating polysemous words

Most practical applications and test cases of compositional distributional semantics have turned around phrase and sentence paraphrasing (Mitchell and Lapata 2008, 2010). However, little attention has been paid to compositional translation of polysemous words. For instance, the verb "run" can be translated into Spanish by very different verbs on the basis of the contextual words it is combined. It can be translated by "correr" in "the horse is running", by "circular" (*to travel*) in "the bus runs along the highway", by "ejecutar" (*to execute*) in "the computer runs a program", or even by "dirigir" (*to manage*) in "the manager runs the company". Polysemy is probably the main source of problems for machine translation systems.

To perform compositional translation, we implemented a corpus-based strategy to build distributional representations of words. Given that we cannot work with ideal distributions, the distributional representation of a word is a vector computed from the occurrences of that word in a given corpus

(Grefenstette 1996). In distributional semantics computational models, each word is defined as a context vector, and each position in the vector represents a specific context of the word whose value is the frequency (or some statistical weight) of the word in that context. We are moving from a formal semantics approach relying on set-theoretic algebra into a corpus-based strategy based on linear algebra. Set unions are implemented as component-wise vector addition and set intersections as component-wise vector multiplication. The translation of a word is the result of selecting the nearest neighbor in a compositional bilingual vector space. More precisely, bilingual vectors are derived from both a bilingual dictionary used to define word contexts and non-parallel corpora used to obtain bilingual word co-occurrences with those contexts. Consequently, to build bilingual word vectors, first we employ the traditional approach to extract translation equivalents from non-parallel texts (Fung and Yee 1998; Rapp 1999; Gamallo 2007; Gamallo and Pichel 2008) and then, bilingual vectors are combined using the compositional model we have defined above in the current paper's section. The vectors were built by making use of a non-parallel corpus that consists of an English part containing the first 200M words from ukWaC corpus (Baroni et al. 2009). The Spanish part was derived from a 2014 dump file of the Spanish Wikipedia[4]: about 480M word tokens. Not only the English and Spanish parts are not translations of each other, but also they are not comparable. Finally, compositional translation of a given compound expression in the source language (English) is performed by searching its nearest neighbor vector (similarity score), among a set of candidates in the target language (Spanish).

Table 1 shows a small arbitrarily selected sample of English expressions containing polysemous words and their translations into Spanish using our compositional strategy (second column) and Google Translator (third column). For our strategy, first the input expressions are analyzed with the dependency-based parser DepPattern (Gamallo and González 2011), and then we apply compositional translation. Yet, our approach presents some limitations to be considered: only lexical words are translated and no inflection is performed. In addition, translation is made from lemmas to lemmas: we just translate lemmas of the source language by lemmas of the target one. However, differences in word order between the two languages are taken into account. Asterisk (*) is used to mark wrong translated words.

The input expressions in Table 1 are constituted by at least one ambiguous English word ("run", "coach", "drive", "hire") which are translated by different

4 http://dumps.wikimedia.org/eswiktionary.

Table 1: Samples of translations from English into Spanish of expressions containing polysemous words ("run", "coach", "drive", "hire").

English expression	Compositional translation	Google translate
the horse is running	caballo corer	el caballo se está ejecutando*
the car is running	coche circular	el coche está funcionando*
run a company	dirigir empresa	dirigir una empresa
run the marathon	correr maratón	correr el maratón
run in the park	correr en parque	correr en el parque
run a program	ejecutar programa	ejecutar un programa
drive a tunnel	excavar túnel	conducir* túnel
drive a coach	conducir autobús	conducir un coche*
electric coach	autobús eléctrico	entrenador* eléctrica
hire a coach	contratar entrenador	contratar a un entrenador
hire a house	alquiler casa	contratar* a una casa

words into Spanish according to the context. For instance, the verb "drive" is translated by "excavar" (*to dig*) in the context of "tunnel", and by "conducir" (*to lead/guide*) in the context of "coach". Notice that in this last example, there is *co-discrimination* of senses, since "coach" is also polysemous and is disambiguated in the context of "drive": it is translated by "autobús" (*bus*) instead of by "entrenador" (*trainer*). Another example with sense co-discrimination of two polysemous words is "hire a coach".

Notice that Google Translate tends to fail when one of the two following situations happens: i) the composite expression consists of two polysemous words; ii) the composite expression is not frequent and might not been found in the parallel corpus used for training. Google Translate is a statistical machine translation (SMT) engine (Koehn 2009). SMT systems constantly learn translations by pattern matching. As more content is analyzed from a parallel corpus, the engine learns more patterns, phrases or co-occurrences. Therefore, when the input phrase is not frequent and contains at least one ambiguous word, the system may fail. Our model, by contrast, does not rely on a parallel corpus but on any multilingual source of text corpora (e.g. the entire Web is a huge non-parallel multilingual corpus). In addition, it is able to predict the sense of a word in context even if the input composite expression has never occurred in the non-parallel corpus.

Yet, a critical problem of our strategy is efficiency and scalability. Our system takes about one minute in translating a single expression with just one dependency and takes polynomial time as the input size grows. A Big Data environment taking advantage of multi-core processors and distributed computing will be required to (at least partially) solve these two problems.

3 Dependencies and incremental interpretation

One of the basic assumptions on semantic interpretation made in frameworks such as Dynamic Logic (Groenendijk and Stokhof 1991), Discourse Representation Theory (Kamp and Reyle 1993), and Situation Semantics (Barwise 1987), is that the meaning of a sentence is dependent of the meaning of the previous sentence in the discourse, and modifies itself the meaning of the following sentence. Sentence meaning does not exist out of discursive unfolding. Meaning is incrementally constructed at the same time as discourse information is processed.

We assume that incrementality is true not only at the inter-sentence level but also at the inter-word level, i.e., between dependent words. In order for a sentence-level interpretation to be attained, dependencies must be established between individual constituents as soon as possible. This claim is assumed by a great variety of research (Kempson et al. 2001, 1997; Milward 1992; Costa et al. 2001; Schlesewsky and Bornkessel 2004). The incremental hypothesis states that information is built up on a left-to-right word-by-word basis in the interpretation process (Kempson et al. 2001). The meaning of an utterance is progressively built up as the words come in. The sense of a word is provided as part of the context for processing each subsequent word. Incremental processing assumes that humans interpret language without reaching the end of the input sentence, that is, they are able to assign a sense to the initial left fragment of an utterance. This hypothesis has received a large experimental support in the psycholinguistic community over the years (McRae et al. 1997; Tanenhaus and Carlson 1989; Truswell et al. 1994).

For instance, to interpret "the computer runs a program", it is required to interpret "the computer runs" as a fragment that restricts the type of nouns that can appear at the direct object position: "program", "script", "software", etc. In the same way "the manager runs" restricts the entities that a manager is used to run: "companies", "firms", etc. However, a left-to-right interpretation process cannot be easily assumed by a standard compositional approach. In a Montagovian approach, any transitive verb (or verb used as transitive) denotes a binary function, $\lambda y \lambda x\, run(x, y)$, which is first applied to the noun at the direct object position in order to build an intransitive verb, for instance $\lambda x\, run(x, program)$ ("run the program"). Then, this function is applied to the noun at the subject position to build a proposition (e.g. "the computer runs the program"). The standard compositional model does not provide any semantic interpretation for the subject + verb expression of a transitive construction: for instance "the computer runs" within the sentence "the computer runs the program". Hence, it is unable to simulate how the expression "the computer runs ..." restricts the type of nouns appearing at the direct object position of the verb.

By contrast, in our incremental compositional strategy, "the computer runs" within a sentence like "the computer runs word2vec" is a grammatical expression referring to two contextualized entities: the sense of "run" given "computer" as subject, and the sense of "computer" as nominal subject of "run". The contextualized sense of "run" helps interpreting "runs word2vec" with the sense previously activated by "computer": i.e. *running as operating a machine or a program*. Even if we have no information on the meaning of "word2vec", the disambiguated sense of the verb leads us to interpret that noun as a kind of software at the direct object position. Our incremental model is based on the semantic interpretation of composite expressions dependency-by-dependency from left-to-right. The incremental interpretation of "the computer runs word2vec" is illustrated in Figure 2.[5]

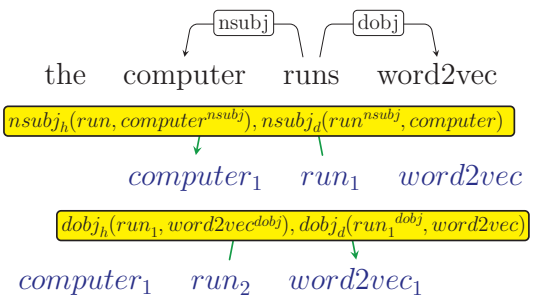

Figure 2: Syntactic analysis of the expression *"the computer runs word2vec"* and left-to-right construction of the word senses.

At the end of the left-to-right interpretation process, we obtain three contextualized senses, one per lexical word. Consequently, we get not only the compositional entity of the root word (here the verb "run"), but also the compositional entities associated with each constituent word. The contextualized sense of the root, run_2, represents the particular process the composite expression is referring to. The three verbal entities (*run*, run_1, and run_2) can be perceived as different degrees of sense specification of the running process denoted by the verb "run": $run_2 \subseteq run_1 \subseteq run$. In terms of sense discrimination, run_1 discriminates one specific sense from those presupposed by the generic meaning, *run*. By contrast,

5 We are considering neither determiners nor verb tense, whose semantic interpretation would consist in grounding nouns and verbs to a specific situation and, then, to specific individuals and events.

run_2 does not discriminate any new sense with regard to run_1. Both entities are referring to the same kind of running process, but run_2 is just a more specific process than run_1. The other two contextualized senses built by the interpretation process are the individuals $computer_1$ and $word2vec_1$.

Finally, once we have reached the last word by incremental interpretation, it is still possible to update the sense of the previous words by going backwards in the process of function application. Sense updating is performed by applying again the functions right-to-left to the first words, namely those appearing to the left of the root word. This right-to-left updating process is performed by using the current sense and selectional preferences of the root word. This way, the dependent function of $nsubj_d$ can be applied again on $computer_1$ by considering the selectional restrictions imposed by the root run_2. It returns $computer_2$, whose sense represents a very specific individual: *a computer running word2vec*. Both $computer_2$ and $word2vec_1$ are contextualized individuals that may be involved in further linguistic phenomena such as coreference linking. Consequently, they can be retrieved by anaphorical pronouns in further sentences at the discourse level.

Our proposed model is then able to simulate the incremental semantic interpretation dependency by dependency. The meaning of words is gradually elaborated as they are syntactically integrated in new dependencies. Therefore, syntactic analysis and semantic interpretation are merged into the same incremental process of information growth. This incremental procedure also allows us to take into account the influence of word order in the construction of meaning, in particular in those languages (e.g. Spanish and Portuguese) where the word order is not so rigid as in English. Furthermore, this syntactic-semantic parsing strategy is able to deal with the *garden path* sentence effect. It occurs when the sentence has a phrase or word with an ambiguous meaning that the reader interprets in a certain way, and when she/he reads the whole sentence there is a difference in what has been read and what was expected. Both left-to-right and right-to-left contextualization processes simulate the way a reader builds the meaning of composite expressions and sentences.

3.1 A case study: Incremental translation

As in the previous section, we find that machine translation might be an interesting application for our model. Our method is able to simulate the process of translating dependency by dependency from English into Spanish in a compositional way. We apply now compositional translation to expressions consisting of a sequence of dependencies. For instance, take the expression “the coach

ran a team". The process starts by proposing translation candidates to the two words related by the subject dependency: "the coach ran". As the two related words, "coach" and "run", are still ambiguous within this dependency, we get several translation candidates with very close similarity scores. These are the top five translation candidates for "the coach ran":

- autobús circular (*the bus travelled on the road*)
- autobús funcionar (*the bus was working*)
- entrenador dirirgir (*the trainer led something*)
- autobús dirigir (*the bus led something*)
- entrenador correr (*the trainer was running*)

In the next dependency, the denotation of "ran" in the context of "coach" (and noted run_1) is combined with that of "team". This results in a more contextualized entity, run_2, which is now clearly translated into Spanish by "dirigir" (*to lead*) in the context of "team". The new entity, run_2, constraints in turn "coach" to be translated by "entrenador" (*trainer*) instead of "autobús" (*bus*).

Table 2 shows some English expressions containing at least two syntactic dependencies. They are translated using our incremental left-to-right strategy

Table 2: Samples of translations from English into Spanish of expressions containing two or more dependencies.

English expression	Incremental translation	Google Translate
the coach ran a team	entrenador dirigir equipo	el entrenador corrió* un equipo
the manager runs the company	director dirigir empresa	el gerente dirige la empresa
the handsome coach is running a marathon	apuesto entrenador correr maratón	el entrenador guapo corre el maratón
the electric coach runs over a cat	autobús eléctrico atropellar gato	el entrenador* eléctrica corre* sobre un gato
the electric coach is turning	autobús eléctrico girar	el sofá* eléctrico está convirtiendo*
the computer is running word2vec	ordenador ejecutar word2vec	el equipo está funcionando* word2vec
the computer runs the program	ordenador ejecutar programa	el equipo se* ejecuta el programa
a man hired an electric coach	hombre alquilar autobús eléctrico	un hombre contrató* a un entrenador* eléctrica
the company hired a man	empresa contratar hombre	la empresa contrató a un hombre
the manager of the company fired an employee	gerente de empresa despedir empleado	el gerente de la empresa despedido un empleado
the terrorist fired an employee	terrorista disparar empleado	el terrorista disparó un empleado

including right-to-left updating. Results are compared to those obtained with Google Translate. As already mentioned, this system faces difficulties translating composite expressions with more than one polysemous word (in "the electric coach runs over a cat", the noun "coach" and the verb "run" are polysemous), or with infrequent words (as "word2vec" in "the computer is running word2vec").

4 Corpus-based experiments and evaluation

As the few examples of our case study introduced in the previous sections are not enough to evaluate the proposed method, we performed two larger scale experiments. First, in Subsection 4.1, we compare our strategy to build compositional vectors to that defined in Baroni and Zamparelli (2010), the state-of-the-art in the field according to the experiments reported in Dinu et al. (2013b). We used as gold standard the test dataset described in Mitchell and Lapata (2008). Next, in Subsection 4.2, we evaluate our translation strategy against a gold standard we have elaborated.

4.1 Mitchell and Lapata benchmark

In this experiment, we build a monolingual vector space and compute similarity between composite expressions. The English test dataset by Mitchell and Lapata (2008) 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 hereafter) 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. To evaluate the results of the tested systems, Spearman correlation is computed between individual human similarity scores and the systems' predictions.

As the objective of the experiment is to compute the similarity between pairs of NV composite expressions, we are able to compare the similarity not only between the contextualized heads of two composite expressions but also 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 compositional functions (*head* and *dependent*) for each pair of compared expressions, it is possible to compute a new similarity measure by averaging *head* and *dependent: DEP (head + dep)* strategy.

Table 3 shows the Spearman's correlation values (ρ) obtained by our method *DEP* and *Baroni@Zamparelli* (Baroni and Zamparelli 2010). For the latter, we used the software DISSECT (Dinu et al. 2013a).[6] The ρ score reached by our *DEP (head + dep)* strategy is 0.16, which is higher than using only head-based similarity (*head* in second row) or dependency-based similarity (*dep* in third row). This shows that the similarity obtained by combining the head and dependent functions is more accurate than that obtained by using only one type of compositional function.

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

Systems	ρ
DEP (head + dep)	0.16
DEP (head)	0.10
DEP (dep)	0.13
Baroni@Zamparelli	0.06

The three similarity strategies based on our algorithm, *DEP*, outperform the *Baroni@Zamparelli* system (0.06). All the score values were obtained on the basis of our relatively small English corpus (200M tokens). However, Dinu et al. (2013b) reported $\rho = 0.26$, which was obtained by the *Baroni@Zamparelli* system using a much larger corpus of about 2.8 billion tokens. Besides, this large corpus contains the BNC, which was used by Mitchell and Lapata to build the dataset.

It would be interesting to prepare or have access to test datasets mainly based on frequent but ambiguous words (in any corpus), which would allow us to more easily evaluate different systems on manageable corpora. This could prevent us from performing evaluations where the best systems are those that were applied on the largest corpora, even if those systems are not always provided with the best algorithm.

6 http://clic.cimec.unitn.it/composes/toolkit/introduction.html

The difference between *DEP* and *Baroni@Zamparelli* is statistically highly significant (Wilcoxon signed-rank test, $p < 0.0001$).

4.2 Translation of ADJ-NOUN compounds

To evaluate our compositional strategy for translation in a bilingual vector space, we built a test dataset with 607 ADJ NOUN (AN) English compounds associated with their corresponding NA or AN Spanish translations. To create this dataset, we identified all unambiguous nouns and adjectives from the dictionary, and selected the AN constructions occurring at least 10 times in the corpus whose constituents belong to the list of unambiguous adjectives and nouns. Then, we used the dictionary to translate them into Spanish expressions and all translations were revised and manually corrected.

Even if our dataset is constituted by AN English expressions that are all translated by NA or AN Spanish compounds, this is not always the case for other composite expressions. Several problems can arise when we translate a composite expression: *fertile translations* in which the target compound has more words than the source term ("cow milk" / *leche de vaca*); non-compositional expressions that can be translated by just one word ("dry humor" / *ironÃa*); intercategorial translations where, in some contexts, nouns (e.g. "coast") are translated by adjectives (*costero*). In fact, bilingual dictionaries do not make intercategorial translations.

In order to deal with all these potential cases, the source compound will be compared against a very large list of candidates including single words and compounds with different morphological and syntactic properties. Thus, for each English compound expression to be translated, the set of translation candidates contains the following three types of Spanish candidates:

- The compounds built with an English-Spanish dictionary and the appropriate translation templates. More precisely, it consists in decomposing the English composite term into atomic components, translating these components into Spanish and recomposing the translated components into Spanish composite terms. If the atomic English components are ambiguous, we obtain several Spanish candidates. If they are not ambiguous (as in our dataset), only one Spanish candidate is generated. This strategy is used in other compositional translation approaches (Grefenstette 1999; Tanaka and Baldwin 2003; Delpech et al. 2012; Morin and Daille 2012).
- The top-100 multiwords derived from the 10 most similar Spanish nouns to the English compound. This strategy is used to take into account fertile

translations in which the target compound has more words than the source term, or for intercategorial translations.
– The set of all Spanish single nouns. This is useful for non-compositional expressions which can be translated by just one word.

In total, each English compound is assigned about 50M translation candidates and, among all of them, only one candidate is correct.

We compared our strategy with one baseline: a corpus-based strategy in which the translation of a compound expression is obtained by selecting the most frequent compound candidate in a given corpus. This corpus-based strategy only considers those translation candidates which are also compound expressions. Hence, single nouns are not taken into account as candidate translations. This strategy actually follows the basic method described in Grefenstette (1999). Table 4 shows the accuracy obtained by our system and the baseline. Accuracy of our system, *DEP*, reaches 89%. If we consider the two nearest neighbors (instead of just the nearest one), accuracy achieves 93%. The corpus-based strategy fails because most of the candidates are well-formed expressions that can be found in a corpus. For instance, consider that the Spanish composite *leche materna* ("breast milk") is part of the top-100 Spanish candidate translations derived from "cow milk". Given that *leche materna* is a well-formed and frequent expression, it could be more frequent in the given corpus than *leche de vaca* ("cow milk") and thereby oddly taken as the equivalent translation of "cow milk".

Table 4: Accuracy obtained by the two strategies to translate AN English expressions into Spanish. The test dataset contains 607 AN expressions with unambiguous words.

Systems	Accuracy
DEP	0.89
Corpus-based	0.08

Notice that a basic dictionary-based strategy would reach 100% accuracy because, in this artificial dataset, all compounds are fully compositional and the constituent words are not ambiguous (according to our dictionary): each English word has only one Spanish translation. Therefore, for this dataset, contextualization is not required. Our strategy, however, is based on contextualization and behaves as if all expressions would contain ambiguous words. In a

dataset with ambiguous words, our method should keep similar accuracy while that of the basic dictionary-based strategy would drop dramatically.

A large-scale translation test with composite expressions containing ambiguous words should be performed. However, to the best of our knowledge, no test datasets with English-Spanish composite expressions (ambiguous or not) are available.[7]

5 Related work

Our model relies on two different approaches: compositional distributional semantics and incremental (or dynamic) semantic interpretation.

5.1 Compositional distributional semantics

Several models for compositionality in vector spaces have been proposed in recent years. The most basic approach to composition, explored by Mitchel and Lapata (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 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) and 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, mostly based on Combinatory Categorial Grammar and typed functional application inspired by Montagovian semantics (Baroni and Zamparelli

7 Our test dataset is freely available at http://fegalaz.usc.es/dataset-en-es.tgz

2010; Coecke et al. 2010; Grefenstette et al. 2011; Krishnamurthy and Mitchell 2013; Baroni 2013; Baroni et al. 2014). The functional approaches relying on Categorial Grammar distinguish the words denoting atomic types, which are represented as vectors, from those that denote compound functions applying on vectors. By contrast, in our compositional approach, function application is not driven by function words such as adjectives or verbs, but by binary dependencies. Our semantic space does not map the syntactic structure of Combinatory Categorial Grammar but that of Dependency Grammar.

Some of the approaches cited above induce the compositional meaning of the functional words from examples adopting regression techniques commonly used in machine learning (Baroni and Zamparelli 2010; Krishnamurthy and Mitchell 2013; Baroni 2013; Baroni et al. 2014). In our approach, by contrast, functions associated with dependencies are just basic arithmetic operations on vectors, as in the case of the arithmetic approaches to composition described above (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 Categorial Grammar use tensor products for composition (Coecke et al. 2010; Grefenstette et al. 2011). 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 simple component-wise multiplication in Mitchell and Lapata's experiments (Mitchell and Lapata 2010).

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 as in our approach. Thater et al. (2010) distinguish between *first-order* and *second-order* vectors in order to allow two syntactically incompatible vectors to be combined. The notion of second-order vector is close to our concept of *selectional preferences*. However, there are important differences between both approaches. In Thater et al. (2010), the combination of a first-order with a second-order vector returns a second-order vector, which can be combined with other second-order vectors. This could require the resort to third-order (or n-order) vectors at further levels of vector composition. By contrast, in our approach, any vector combination always returns a first-order vector. This simplifies the compositional process at any level of analysis.

The work by Thater et al. (2010) 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*. The main problem is that this approach does not propose any compositional model. Its objective is to simulate word sense disambiguation, but not to model semantic composition at any level of analysis. Thater et al. (2010) took up the basic idea from Erk and Padó (2008) of exploiting selectional preference information for contextualization and disambiguation. However, they did not borrow the idea of splitting the output of a word combination into two different vectors (one per word). As far as we know, no fully and coherent compositional approach has been proposed from the interesting idea of returning two contextualized vectors per combination. Our approach is an attempt to join the main ideas of these syntax-based models (namely, selectional preferences as second-order vectors and two returning senses per combination) into an entirely compositional model.

5.2 Incremental interpretation in dynamic syntax

Dynamic Syntax was introduced in Kempson et al. (1997, 2001). In this approach, a model of natural language understanding in which the development of an interpretation of a string is defined as an incremental left-to-right process of constructing a logical form representing one possible content attributable to the string. The denotation of an expression is defined as its context change potential. More precisely, interpretation is built up from left to right as each individual word is processed, following the combinatorial properties of the words as specified by their logical type. Words are assumed to project expressions in some logical language, and it is these that combine together to result in a logical form corresponding to the interpretation of the sentence. In addition, Dynamic Syntax relies on (partial) constitutive analysis.

The main differences between Dynamic Syntax and our interpretation strategy are the following:

- Dynamic Syntax projects tree structures from words on the basis of their sub-categorization properties. Words can have very complex function types. By contrast, in our approach the interpretable structures are projected from syntactic dependencies. Only dependencies are functional operations. Words just help dependencies to build compositional structures.
- Interpretation in Dynamic Syntax is a goal-oriented procedure. It tries to reach a final tree to be interpreted as a proposition. Meanwhile, interpretation relies on partial constituent structures being part of that previously presupposed full tree. By contrast, our approach is not goal-directed and,

thereby, partial structures are not required. Each dependency is interpretable without considering a full tree driving the interpretation process.
– Dynamic Syntax is just a left-to-right process. By contrast, in our approach, we also defined a short right-to-left scanning which updates the denotation of the first words in the sequence by using the contextualized information of the root word.

6 Conclusions

In this paper, syntactic dependencies were endowed with a combinatorial meaning. The fact of characterizing dependencies as compositional devices has important consequences on the way in which the process of semantic interpretation is considered. First, dependencies denote binary functions on entities (defined as sets of contexts), while lexical words denote entities. Second, the interpretation of a composite expression is not a single representation, but the contextualized denotation of each constituent (lexical) word. Third, the compositional process is performed in an incremental way dependency by dependency from left-to-right. It starts with very ambiguous and generic entities associated with the constituent words before composition, and results in less ambiguous entities associated with the contextualized words. At the end of the process, the words to the left of the root are updated using the contextualized sense of the root in a right-to-left strategy.

And fourth, as syntactic dependencies are conceived here as semantic operations, we situate syntax at the center of the semantic interpretation process. Syntax is not an autonomous module which is independent of semantics. Rather, it is described as a particular semantic level (Langacker 1991). In fact, there is evidence that disambiguating a syntactic structure (e.g. pp-attachment) and enriching weak lexical specifications of content (word sense disambiguation) are processes subject to the same psychological constraints (Sperber and Wilson 1995). Even if in our experiments the semantic interpretation has been performed after the syntactic analysis in dependencies, we claim that syntactic analysis and semantic interpretation should be merged into the same incremental process of information growth.

Substantial problems still remain unsolved. For instance, there is no clear borderline between compositional and non-compositional expressions (collocations, idioms, ...). Let us suppose that we represent word senses as vectors derived from large corpora (as in our experiments). It seems to be obvious that vectors of full compositional compounds should be built by means of compositional operations and predictions based on their constituent vectors. It is also

evident that vectors of entirely frozen expressions should be totally derived from corpus co-occurrences of the whole expressions without considering internal constituency. However, there are many expressions, in particular collocations (such as "save time", "go mad", "heavy rain", ...) which can be considered as both compositional and non-compositional. In those cases, it is not clear which is the best method to build their distributional representation: predicted vectors by compositionality or corpus-observed vectors of the whole expression?

Another problem that has not been considered is how to represent the semantics of some grammatical words, namely determiners and auxiliary verbs (i.e., noun and verb specifiers). For this purpose, we think a different functional approach would be required, probably closer to the work described by Baroni, who defines functions as linear transformations on vector spaces (Baroni et al. 2014).

In future work, we will address and go into detail about the idea of performing incremental translation with dependency-by-dependency processing. On the basis of this idea, we think that it could be possible to develop a new paradigm for machine translation which would not be based on parallel corpora. For this purpose, it will be necessary to better define how to generate translation candidates at whatever level of composition.

Finally, in order to get more efficiency and scalability, it will be required to integrate the system into a Big Data architecture with distributed databases and multi-core processors.

Acknowledgments: We would like to thank the anonymous reviewers for helpful comments and suggestions.

Funding: This work is funded by Project TELPARES, Ministry of Economy and Competitiveness (FFI2014-51978-C2-1-R), and the program "Ayuda Fundación BBVA a Investigadores y Creadores Culturales 2016".

References

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

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

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The wacky wide web: A collection of very large linguistically processed webcrawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

Baroni, Marco & 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, 1183–1193. Stroudsburg, PA, USA.

Barwise, Jon. 1987. *Recent developments in situation semantics*. Language and Artificial Intelligence. Berlin: Springer-Verlag.

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

Copestake, Ann & Aurelie Herbelot. 2012. Lexicalised compositionality. In http://www.cl.cam.ac.uk/ah433/lc-semprag.pdf.

Costa, F., V. Lombardo, P. Frasconi & G. Soda. 2001. Wide coverage incremental parsing by learning attachment preferences. In *Conference of the Italian Association for Artificial Intelligence (AIIA)*.

Davidson, Donald. 1969. *The individuation of events*, 216–234. Dordrecht: Springer Netherlands. ISBN 978-94-017-1466–2.

Delpech, Estelle, Béatrice Daille, Emmanuel Morin & Claire Lemaire. 2012. Extraction of domain-specific bilingual lexicon from comparable corpora: Compositional translation and ranking. In *COLING 2012, 24th International Conference on Computational Linguistics, Mumbai, India*, 745–762.

Dinu, G., N. Pham & M. Baroni. 2013a. Dissect: Distributional semantics composition toolkit. In *ACL 2013 Workshop on Continuous Vector Space Models and their Compositionality (CVSC 2013)*, 31–36. East Stroudsburg, PA.

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

Erk, Katrin. 2013. Towards a semantics for distributional representations. In *IWCS-2013*.

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

Fellbaum, C. 1998. A semantic network of English: The mother of all WordNets. *Computer and the Humanities* 32. 209–220.

Fung, Pascale & Lo Yuen Yee. 1998. An IR approach for translating new words from nonparallel, comparable texts. In *Coling'98*, 414–420. Montreal, Canada.

Gamallo, Pablo. 2003. Cognitive characterisation of basic grammatical structures. *Pragmatics and Cognition* 11(2). 209–240.

Gamallo, Pablo. 2007. Learning bilingual lexicons from comparable English and Spanish Corpora. In *Machine Translation SUMMIT XI*. Copenhagen, Denmark.

Gamallo, Pablo. 2008. The meaning of syntactic dependencies. *Linguistik OnLine* 35(3). 33–53.

Gamallo, Pablo, Alexandre Agustini & Gabriel Lopes. 2005. Clustering syntactic positions with similar semantic requirements. *Computational Linguistics* 31(1). 107–146.

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

Gamallo, Pablo & José Ramom Pichel. 2008. Learning Spanish-Galician translation equivalents using a comparable corpus and a bilingual dictionary. *LNCS* 4919. 413–423.

Grefenstette, Gregory. 1996. Evaluation techniques for automatic semantic extraction: Comparing syntactic and window based approaches. In B. Boguraev & J. Pustejovsky (eds.), *Corpus processing for lexical acquisition*, 205–216. Cambridge, MA: The MIT Press.

Grefenstette, Gregory. 1999. The World Wide Web as a resource for example-based machine translation tasks. In *Translating and the Computer 21: Proceedings of the 21st International Conference on Translating and the Computer*.

Grefenstette, Edward, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke & Stephen Pulman. 2011. Concrete sentence spaces for compositional distributional models of meaning. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '11, 125–134.

Groenendijk, J. & M. Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14. 39–100.

Guevara, Emiliano. 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.

Hanks, Patrick. 2013. *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.

Harris, Zellig. 1954. Distributional structure. *Word* 10(23). 146–162.

Hudson, Richard. 2003. The psychological reality of syntactic dependency relations. In *MTT 2003*. Paris.

Jezek, Elisabetta & Patrick Hanks. 2010. What lexical sets tell us about conceptual categories. Lexis [Online], 4 | 2010, Online since 14 April 2010. http://lexis.revues.org/555 (accessed 16 January 2017), DOI: 10.4000/lexis.555.

Kahane, Sylvain. 2003. Meaning-text theory. In V. Ágel et al. (eds.), *Dependency and valency: An international handbook of contemporary research*. Berlin: De Gruyter.

Kamp, H. & U. Reyle. 1993. *From discourse to logic: Introduction to model-theoretic semantics of natural language. Formal logic and discourse representation theory*. Dordrecht: Kluwer Academic Publisher.

Kempson, R., W. Meyer-Viol & D. Gabbay. 1997. Language understanding: A procedural perspective. In C. Retore (ed.), *First international conference on logical aspects of computational linguistics*, 228–247. Lecture Notes in Artificial Intelligence Vol. 1328. Springer Verlag.

Kempson, R., W. Meyer-Viol & D. Gabbay. 2001. *Dynamic syntax: The flow of language understanding*. Oxford: Blackwell.

Koehn, Philipp. 2009. *Statistical machine translation*. Cambridge: Cambridge University Press.

Krishnamurthy, Jayant & Tom Mitchell. 2013. *Proceedings of the workshop on continuous vector space models and their compositionality*, chap. Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models, 1–10. Association for Computational Linguistics.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar: Descriptive applications*, vol. 2. Stanford: Stanford University Press.

McRae, K., T.R. Ferreti & L. Amoyte. 1997. Thematic roles as verb-specific concepts. In M. MacDonald (ed.), *Lexical representations and sentence processing*, 137–176. Sussex, UK: Psychology Press.

Meillet, Antoine. 1921. *Linguistique historique et linguistique générale*. Paris: La Société Linguistique de Paris.

Milward, David. 1992. Dynamics, dependency grammar and incremental interpretation. In *14th Conference on Computational Linguistics (Coling92)*, 1095–1099. Nantes.

Mitchell, Jeff & Mirella Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of* ACL-08: *HLT*, 236–244.

Mitchell, Jeff & Mirella Lapata. 2009. Language models based on semantic composition. In *Proceedings of EMNLP*, 430–439.

Mitchell, Jeff & Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science* 34(8). 1388–1439.
Montague, Richard. 1970. Universal grammar. theoria. *Theoria* 36. 373–398.
Morin, Emmanuel & Béatrice Daille. 2012. Revising the compositional method for terminology acquisition from comparable corpora. In *COLING 2012, 24th International Conference on Computational Linguistics, Mumbai, India*, 1797–1810.
Navigli, Roberto. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys* 41(2). 1–69.
Partee, Barbara. 2007. Private adjectives: Subsective plus coercion. In R. Bäuerle, U. Reyle & T. E. Zimmermann (eds.), *Presuppositions and discourse*. Amsterdam: Elsevier.
Pustejovsky, James. 1995. *The generative lexicon*. Cambridge: MIT Press.
Rapp, Reinhard. 1999. Automatic identification of word translations from unrelated English and German Corpora. In *ACL'99*, 519–526.
Schlesewsky, M. & I. Bornkessel. 2004. On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua* 114. 1213–1234.
Schütze, Hinrich. 1998. Automatic word sense discrimination. *Computational Linguistics* 24(1). 97–124.
Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and cognition*, 2nd edn. Oxford: Blackwell.
Steedman, Mark. 1996. *Surface structure and interpretation*. Cambridge, MA: The MIT Press.
Studtmann, Paul. 2014. Aristotle's categories. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy*. Summer 2014 edn.
Tanaka, Takaaki & Timothy Baldwin. 2003. Noun-noun compound machine translation a feasibility study on shallow processing. In *Proceedings of the ACL* 2003 *Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 17–24. Sapporo, Japan.
Tanenhaus, M.K. & G.N. Carlson. 1989. Lexical structure and language comprehension. In W. Marslen-Wilson (ed.), *Lexical representation and process*, 530–561. Cambridge, MA: The MIT Press.
Tesniére, Lucien. 1959. *Eléments de syntaxe structurale*. Paris: Klincksieck.
Thater, Stefan, Hagen Fürstenau & Manfred Pinkal. 2010. Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 948–957. Stroudsburg, PA, USA.
Truswell, J.C., M.K. Tanenhaus & S.M. Garnsey. 1994. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33. 285–318.
Turney, Peter D. 2013. Domain and function: A dual-space model of semantic relations and compositions. *Journal of Artificial Intelligence Research (JAIR)* 44. 533–585.
Zanzotto, Fabio Massimo, Ioannis Korkontzelos, Francesca Fallucchi & Suresh Manandhar. 2010. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, 1263–1271.

Supplemental Material: The online version of this article (DOI: 10.1515/cllt-2016-0038) offers supplementary material, available to authorized users.