Exploring terminological relations between multi-word terms in distributional semantic models

Yizhe Wang

University of Science and Technology of China

Béatrice Daille

Nantes Université

Nabil Hathout

Université de Toulouse

Abstract

A term is a lexical unit with specialized meaning in a particular domain. Terms may be simple (STs) or multi-word (MWTs). The organization of terms gives a representation of the structure of domain knowledge, which is based on the relationships between the concepts of the domain. However, relations between MWTs are often underrepresented in terminology resources. This work aims to explore distributional semantic models for capturing terminological relations between multi-word terms through lexical substitution and analogy. The experiments show that the results of the analogy-based method are globally better than those of the one based on lexical substitution and that analogy is well suited to the acquisition of synonymy, antonymy, and hyponymy while lexical substitution performs best for hypernymy.

Keywords: terminology relations, multi-word terms, analogy, lexical substitution, FastText, masked language modeling, transformer models, environment domain

1 Introduction

In terminology, the organization of terms is considered to mirror the knowledge structure of a domain (Zweigenbaum and Grabar 2000). This structure is based on the relations between the concepts of the domain which describe the connections that exist between terms, either simple terms (STs) or multi-word terms (MWTs)[1].

However, relations between terms are missing in many terminologies and there is a real need to identify and characterize them. Relations between terms can be identified by experts from existing resources or knowledge extracted from corpora. Much current research has focused on STs (Grabar and Hamon 2006; Zhang, Li, and Wang 2017; Zhu, Yan, and Wang 2017). Conversely, little work has been carried out on the acquisition of relations between MWTs (Hazem and Daille 2018). Most of the work on MWT relations concerns the exploitation of the internal structure of the MWTs using different types of linguistic information like syntactic information in Verspoor et al. (2003) or semantic information in Hazem and Daille (2018).

Our work focuses on evaluating the ability of distributional semantic models (DSMs) to capture basic terminological relations between MWTs in French in the domain of the environment. DSMs are applied on nominal MWTs made up of two lexical words. Distributional semantics (DS) (Lenci 2008), also known as vector space semantics, is a method for representing lexical meaning in NLP. Based on the distributional hypothesis (Harris 1954) that words with similar

[1] Simple terms, composed of a single graphic unit (e.g., *climate*), are usually distinguished from multi-word terms which contain several graphic units separated by spaces (e.g., *climate change*) (L'Homme 2004).

linguistic contexts tend to have similar meanings, DS represents lexical units by vectors produced by DSMs. These vectors encode the linguistic distribution of the words in the corpus. Words that have similar distributions are represented by vectors that are close. The application of distributional techniques to specialized corpora dates back to the 1990s (Morlane-Hondère and Fabre 2012). The current trend in NLP is to create DSMs from large corpora. However, large corpora in specialized domains are rare. In addition, many studies concern the medical domain (Paullada, Percha, and Cohen 2020; Chen et al. 2018; Bourigault 2002) while few works focus on the environmental domain (Hazem and Daille 2018; Bernier-Colborne and Drouin 2016).

In this paper, we explore the possibilities of both static and contextual distributional models to capture lexical-semantic relations between nominal multi-word terms of length 2, called biterms, with two methods. The first, based on lexical substitution, uses the predictions of a BERT masked language model[2]. The second captures lexical-semantic relations between biterms by means of analogy between term representations provided by a FastText static model. Our experiments were carried out with two main datasets. The first is a dataset composed of synonymous[3] biterms extracted from IATE[4], a terminology resource for translation. The second is a dataset of semantically related biterms created by semantic projection using various resources of the environment domain in French. Semantic projection is a method that extends relations between STs to the MWTs (e.g., the relation between *dry* and *wet* is extended to the MWTs *dry air* and *wet air*). It is often used to identify semantic relations between MWTs (Morin and Jacquemin 1999; Daille and Hazem 2014; Hazem and Daille 2018).

Section 2 presents a state of the art on the acquisition of semantic relations from DSMs and in particular by means of analogy and lexical substitution. Section 3 presents the experimental framework: the general resources selected to build our test data and carry out our experiments, the adaptation of analogy and lexical substitution methods to our data, the trained language models and the evaluation metrics. Section 4 describes the extraction of synonymous biterms from IATE and the experiments carried out on them. Section 5 presents the creation by semantic projection of a resource containing opposition and hierarchical relations and experiments aimed to compare our methods on the acquisition of these relations. Section 6 discusses and compares the results of the different methods. The paper concludes with a summary and a presentation of future avenues of research.

2 Identification of semantic relations in DSMs

Our study concerns the identification of terminological relations between biterms using DSMs. Most of the techniques proposed in the literature for the characterization of semantic relations are domain-independent (Lafourcade and Ramadier 2016). Since the basic terminological relations extended with synonymy do not show fundamental differences with the classic lexical semantic relations, many of the techniques designed to acquire the latter in the general domain can be used to identify relations between terms in specialized domains. In this section, we present a brief review of the literature on acquiring semantic relations using DSMs. We are particularly interested in works using analogy and lexical substitution.

[2]BERT-type models can also be considered as DSMs since their inner layers generate embeddings that encode several meaning aspects based on the distribution properties of words in texts (Mickus et al. 2020).

[3]Relations extracted from IATE include synonyms and near-synonyms.

[4]https://iate.europa.eu

2.1 Semantic relations acquisition using DSMs

The extraction of semantic relations using DSMs usually involves word embeddings (Gábor et al. 2018; Levy et al. 2015). One of the most direct methods uses the similarity between the representations of lexical units. This method is based on the assumption that words that are close in semantic space are often semantically related (Bernier-Colborne and Drouin 2016). Embeddings can also be used as features provided to classifiers to classify semantic relations (Gábor et al. 2018). Instead of using word embeddings directly, some works represent semantic relations by linear transformations of embeddings (Weeds et al. 2014; Fu et al. 2014; Roller, Erk, and Boleda 2014). This method is based on the assumption that some dimensions, or directions in the embeddings, are associated with particular relations (Vylomova et al. 2016). Due to the difficulties to identify semantic relations from the word embeddings alone (Bouraoui, Jameel, and Schockaert 2018), learning-based methods have also been proposed. Instead of representing the relations using word embeddings, relation vectors are learned by the model (Turney 2005; Hashimoto et al. 2015; Santos, Xiang, and Zhou 2015; Jameel, Bouraoui, and Schockaert 2017).

More recently, various contextual DSMs have been proposed, including ELMo (Peters, Neumann, Iyyer, et al. 2018) and BERT (Devlin et al. 2019). The relational knowledge captured by these models has been the subject of several studies (Qiao et al. 2022; Shi and Lin 2019; Xue et al. 2019). These models trained on large amounts of unlabeled text yield generalizable contextualized word embeddings which can be used to identify relations between lexical units. In addition, some of these models can be fine-tuned for relation identification tasks, reformulated as text classification tasks (Hou et al. 2020; Yao, Mao, and Luo 2019; Bouraoui, Camacho-Collados, and Schockaert 2020). The input of classifiers is a sentence containing two words of interest and the output the relation that connects them.

2.2 Semantic relation acquisition using lexical substitution

Lexical substitution can be used to extract semantic relations (Ferret 2021; Arefyev et al. 2020; Schick and Schütze 2020). The aim of lexical substitution is to propose substitutes that can replace a target word in a given context. For example, in "Depending on the case, the car is presented as clean, organic or green.", *car* can be replaced by its synonym *automobile* with no change to the meaning of the sentence. Other candidates including hyponyms like *sedan*, co-hyponyms like *truck* or hypernyms like *vehicle* could be substituted for *car* while keeping its meaning related to the original one. Lexical substitution task usually uses masked language models (MLM). These models are trained to predict the tokens that can be substituted for a special token <MASK> in a given sentence by considering its left and right contexts. The substitutes proposed by MLMs are generally semantically related to the target.

Schick and Schütze (2020) use BERT-based lexical substitution to evaluate the influence of keyword frequency on the ability of the model to capture semantic properties without any task-specific fine-tuning. They focus on antonymy, hypernymy, co-hyponymy (and input corruption) in the general domain. For each relation, they extracted triples from WordNet as $<k, r, T>$ where k is a keyword, i.e., a WordNet entry; r is a relation; T is a set of target words related to k by the r relation and belonging to the model vocabulary. They also use manually-defined patterns that can express the relation between the keyword k (<W>) and a target word (_), such as "<W> is not _" for antonymy. They adopted an Attentive Mimicking method (Schick and Schütze 2019) to build word embeddings for rare words (Schick and Schütze 2019). With this method and a BERT-large model, the mean reciprocal rank (MRR) scores are of 0.529 for antonymy, 0.299 for hypernymy, and 0.227 for co-hyponymy. They also show that accuracy varies according to the frequency of the keywords and that the method under-performs for low frequency keywords[5].

MLM based lexical substitution is also used by Arefyev et al. (2020) to evaluate its ability to capture various types of semantic relations. They also propose two other methods that allow the prediction to consider the meaning of the masked word. One adds estimator of the proximity between the masked word and the substitutes. The other uses dynamic patterns such as "T and then <mask>" where T represents the masked word. For example, to query for the word *climate* in the context *the effects of climate change on humans*, they generate a context *the effects of climate and then <mask> change on humans* as input. They observe that most of the substitutes are synonyms and co-hyponyms (with percentages greater than 10% compared to all the proposed candidates) when the target word is a noun.

Our proposed method based on lexical substitution is related to some of these works but differs in its objectives: we aim at identifying the lexical relations between French biterms in the environment domain, while the works we have just presented focus on relations between English simple words in the general domain. We use contexts extracted from corpora as input instead of patterns that can express lexical semantic relations (Schick and Schütze 2020) or dynamic patterns. We also adopt the conditioning strategy (see Section 3.2) to provide the model with additional information on the masked word. Furthermore, we study all classical lexical relations while synonymy isn't considered by Schick and Schütze (2020) and antonymy isn't considered by Arefyev et al. (2020).

2.3 Analogy for semantic relation extraction

Analogy has also been used to extract semantic relations. Current research on analogy in word embeddings focuses on "proportional analogy" of the type $a : b :: c : d$ 'a is to b as c is to d'. The starting point for this research is the study by Mikolov et al. (2013) where they show that word2vec models capture various syntactic and semantic relations between words (such as *man-woman*, *adjective-adverb* and *opposite*) and that the similarity of the relations that connect two word pairs can be estimated by the difference between their semantic vectors: $V_a – V_b \approx V_c - V_d$. Following this work, several other teams have used the same method to acquire lexical, encyclopedic or specialized domain relations (Chaudhri et al. 2022; Chen et al. 2018; Gladkova, Drozd, and Matsuoka 2016). The existence of analogies in static embeddings has been theoretically justified by Allen and Hospedales (2019).

While most studies focus on analogy between words or between simple terms (Chen et al. 2018; Gladkova, Drozd, and Matsuoka 2016; Köper, Scheible, and im Walde 2015), Chaudhri et al. (2022) attempt to solve analogy equations between simple and multi-word terms in the biological domain in English. Their dataset is composed of quadruplets of terms such as *Carbon Phospholipid* : *fatty acid tail* :: *polar amino acid* : *polar side chain*. They are interested in relations that are specific to the biological field such as *a type of*. They solve analogy equations like $a : b :: c :?$ with a seq2seq[6] model (Sutskever, Vinyals, and Le 2014) with three encoders that model the input of a, b and c respectively. seq2vec[7] models such as ElMo and BERT have also been tested with a linear layer for the analogy output. They found that seq2vec ELMo is the model that provides the best raw accuracy (0.51).

Paullada et al. (2020) also focus on analogy between simple and multi-word terms in the biomedical domain with the aim of capturing domain-specific relations such as *Gene-Disease*.

[5]As CamemBERT model is trained with texts from the general domain, terms from specialised domains are most likely underrepresented. According to Schick and Schütze (2020), the fact that the terms have low frequencies could be a problem for the prediction of the model.

[6]A seq2seq model is a model that takes a sequence of items (words, letters, time series, etc.) as input and outputs another sequence of items.

[7]These models take a sequence as input and output a single vector.

They concatenated the words in the multi-word terms in order to generate their embeddings. In order to acquire fine-grained analogy, they created embeddings from a corpus made up of dependency-parsed sentences extracted from biomedical literature. Compared to a skip-gram model with negative sampling, the embeddings created from the syntactically parsed corpus do improve the retrieval of biomedical analogy.

Our study differs from the ones we just presented on several points. As we already pointed out, we work for the identification of terminological relations between French MWTs in the environment domain. Like Paullada et al. (2020), we adopt the vector offset[8] instead of seq2vec and seq2seq models (Chaudhri et al. 2022). However, our modeling of the MWTs presented in Section 4.3 differs from that of Paullada et al. (2020) who convert MWTs into simple terms. We used a FastText model (Bojanowski et al. 2017) trained on an indexed corpus where biterms and their parts are represented in the same vector space. Furthermore, we are interested in basic terminological relations between biterms and not in relations that are specific to a specialized domain.

3 Experimental framework

We make use of two methods to assess the ability of DSMs to capture lexical relations between biterms: (1) lexical substitution; (2) analogy. In this section, we present the general resources for the creation of test data, as well as the methods, models, and evaluation methods shared across our experiments.

3.1 Main resources

Our main resources are a corpus and various lexical relation databases.

3.1.1 Corpus

We created our test data and carried out our experiments using the French monolingual corpus PANACEA Environment (ELRA-W0065)[9] built in the framework of the PANACEA project[10]. The corpus is made up of 35,453 documents (50 million words) related to the environment domain. These documents have different levels of specialization and belong to various genres; they were crawled on the web from encyclopedias, blogs, government and non-governmental organization websites. Due to the heterogeneous nature of the environment domain, this corpus is more heterogeneous than a typical specialized corpus which normally only contains specialized texts (Bernier-Colborne 2017). The corpus has been pre-processed: we extracted the text from the original files and normalized the characters.

3.1.2 Lexical relation databases

Three test datasets were built from three French databases which provide lexical relations between simple terms and between simple words. These test datasets are: RefDicoenviro, RefIATE et RefDicosyn.

[8]Mikolov et al. (2013) show that vector offsets seem to mirror linguistic relations and can be used to perform analogical reasoning. They found that in an analogical quadruplet of words $a : a' :: b : b'$, the word b' could be identified using the offset between the vectors of a' and a.

[9]http://catalog.elra.info/en-us/repository/browse/ELRA-W0065/

[10]http://www.panacea-lr.eu/en/info-for-researchers/data-sets/monolingual-corpora

RefDicoenviro is made up of 830 pairs of simple terms (nouns or adjectives) linked by lexical relations and extracted from DiCoEnviro[11], a specialized dictionary of the environment domain. Relations between these term pairs fall into three categories (Bernier-Colborne and Drouin 2016):

- ANTI (116 pairs): opposite and contrastive relations (*froid*:*chaud* 'cold:hot'; *flore*:*faune* 'flora:fauna');
- HYP (191 pairs): hyponymy and hypernymy (*biomasse*:*combustible* 'biomass:fuel'; *consommation*:*surconsommation* 'consumption:over-consumption');
- QSYN (523 pairs): synonymy, quasi-synonymy and co-hyponymy (*diesel*:*gazole* 'diesel:gas oil'; *charbon*:*pétrole* 'coal:oil');

RefIATE is made up of 551 pairs of synonymous simple terms (nouns or adjectives) belonging to the environment domain and extracted form IATE, such as *rejet*:*déversement* 'discharge:spill'. IATE (Interactive Terminology for Europe) is a EU's terminology database made up of more of 8 million terms in the official languages of the European Union covering 20 domains. It is primarily intended for the translators working for the EU.

RefDicosyn is made up of 833,891 pairs of synonymous nouns and adjectives extracted from Dicosyn[12], such as *empreinte:impact* 'footprint:impact'. Dicosyn is an electronic dictionary of synonyms for French built from seven general language dictionaries.

3.2 Models for the lexical substitution and analogy methods

The meaning of expressions and therefore, the relation linking them depends on the contexts in which they appear (Depraetere 2019). For example, *changement du climat* 'climate change' and *réchauffement du climat* 'climate warming' may be synonymous in some contexts but not in others. For example, in (1) *changement du climat* is the synonym of *réchauffement du climat* but in (2) it is not a synonym as it means a climate cooling.

(1) *il a établi que le changement du climat était « sans équivoque » et que les émissions de gaz à effet de serre provenant des activités humaines étaient responsables de l'augmentation des températures depuis cent ans.* 'it's established that climate change was "unequivocal" and that greenhouse gas emissions from human activities were responsible for the increase in temperatures over the past 100 years.'

(2) *à quelle vitesse la réduction des concentrations atmosphériques de GES de courte durée entraînerait un changement du climat* 'how quickly would reductions in short-term atmospheric concentrations of GHGs lead to a change in climate'

The first method we propose aims at identifying lexical relations between biterms, using a masked language model (MLM) to perform a lexical substitution task. MLMs are suitable for this task because they are trained to predict tokens that may appear in a context in a given position signaled by the special token <MASK>. Therefore, their predictions are highly context dependent. In this study, we make use of two querying strategies: basic MLM and conditional MLM.

MLMs are trained to predict the token that may appear in the <MASK> position. For relation prediction, MLM method can be described formally as follows: let MWT_1 and MWT_2 be a pair of biterms that have the same syntactic structure, such that MWT_1 contains the lexical

[11] DiCoEnviro. Dictionnaire Fondamental de l'Environnement http://olst.ling.umontreal.ca/cgibin/dicoenviro/search.cgi.

[12] http://crisco.unicaen.fr/dicosyno/

words W_1 and W_3, and MWT_2, W_2 and W_3. S_1 is a context of MWT_1 and S_2 is a context of MWT_2. We first mask W_1 in S_1 and look for the rank of W_2 among the predictions of the MLM. Conversely, we mask W_2 in S_2 and look for the rank of W_1 among the predictions. If MWT_1 and MWT_2 have compositional meanings and given that they have the same syntactic structure, W_3 should contribute equally to the meaning of MWT_1 and MWT_2. Therefore, if W_2 or W_1 are among the first predictions of the two queries, we can predict that MWT_1 and MWT_2 are semantically related and that their relation is probably the same as that between W_1 and W_2. The method can be illustrated by the following example:

MWT pair: *préservation des forêts* 'forest preservation' and *protection des forêts* 'forest protection'

Target relation: *synonymy*

Masked word: *préservation*

Masked context: *l' aide financière à apporter pour la <mask> des forêts sera l' un des grands sujets abordés lors de la conférence* 'financial support for the <mask> of forests will be a major topic at the conference'

10 first predictions: *préserver, protection, conservation, restauration, reproduction, régénération, dégradation, durabilité, disponibilité, production*

Observation: *protection* appears at rank 2 in the prediction list.

Zhou et al. (2019) show that MLMs yield candidates that can be semantically very different from the masked word while still being well suited to the context. Qiang et al. (2019) propose to use BERT for lexical simplification by adding to the original sentence (in which no word is masked) as additional conditioning of the prediction of the masked tokens. Espinosa-Anke et al. (2021) use the same conditioning method to study collocations. More specifically, the conditioning method consists in using as input of the MLM the concatenation of the original sentence (where the target word is unmasked) and the masked one (where the target word is masked) in order to improve the prediction of the target word. We repeated the previous experiment with conditional queries. The input queries provided to the MLM with and without conditioning are illustrated below:

Query without conditioning: *l'aide financière à apporter pour la <mask> des forêts sera l'un des grands sujets abordés lors de la conférence.*

Query with conditioning: *l'aide financière à apporter pour la préservation des forêts sera l'un des grands sujets abordés lors de la conférence [SEP] l'aide financière à apporter pour la <mask> des forêts sera l'un des grands sujets abordés lors de la conférence.*

The MLM method for lexical substitution relies on two characteristics of the data: the biterms share one word and have the same syntactic structure. In order to generalize to biterms of any syntactic structure, we used analogy to identify the relations.

The detection of relations between biterms by analogy follows from the observation that if $W_1 : W_2 :: MWT_1 : MWT_2$ is an analogy then the relation between W_1 and W_2 is preserved between MWT_1 and MWT_2.

To solve these analogies, we use vector offsets. The solutions of an analogy equation $a : b :: c :?$ could be found among the words whose vector representation V_d is similar to the vector $V_c – V_a + V_b$. The solution to the equation in this case is: $\underset{d \in Voc}{\text{argmax}}\ similarity(V_d, V_c - V_a + V_b)$. In other words, we look in the vocabulary *Voc* for terms whose vector V_d is the closest to the vector $V_c - V_a + V_b$. This method is known as 3cosADD (Levy and Goldberg 2014).

Recall that we are interested in the relation that exists between MWT_1 and MWT_2 knowing the relation between W_1 and W_2. Let VW_1, VW_2, $VMWT_1$ and $VMWT_2$ be the vector representations of W_1, W_2, MWT_1 et MWT_2. We can choose MWT_1 or MWT_2 as the unknown biterm in the analogy equation. Thus, each quadruplet yields two analogy equations: $W_1 : W_2 :: MWT_1 :?$ or $W_1 : W_2 ::? : MWT_2$. If for example we choose MWT_2 as the unknown biterm, we seek to establish whether $VMWT_2$ is close to the expected vector $V_{\text{expected}} = VMWT_1 - VW_1 + VW_2$.

The results presented below are the average of the solutions obtained using the two equations. The following example illustrates the method by analogy:

Analogy equation: *empreinte :impact ::empreinte environnemental :?*

Unknow biterm: *impact environnemental*

5 first predictions: *impact environnemental, impact écologique, impact positif, effet environnemental, coût de dépollution*

Observation: *impact environnemental* appears at rank 1 in the prediction list

3.3 Distributional semantic models

We used two types of distributional semantic models: a contextual one, CamemBERT (Martin et al. 2020), and an static one, FastText (Bojanowski et al. 2017).

The CamemBERT-large model, pre-trained for the masked token prediction task is a language model for French. It has been pretrained on the French subcorpus of the OSCAR multilingual corpus. Its tokenizer is SentencePiece (Kudo and Richardson 2018). Its vocabulary is a mix of whole words and wordpieces. Words absent from the vocabulary are tokenized into subwords. For example, *piscicole* 'fish farming' is split into the wordpieces `pis`, `s`, and `cicole`.

FastText is a library for learning static models, but unlike Word2Vec and Glove, it computes embeddings for both the words and their n-grams of characters. This specificity can be exploited to build models that contain representations of the MWTs and of their parts and in which the representations of MWTs are independent of those of their components. We created a FastText model from the PANACEA corpus. The corpus and the terms were lemmatized before extracting the terms and training the model. Some studies like (Bullinaria and Levy 2012) have shown that lemmatization slightly improves the results of the distributional methods on some tasks.

3.4 Evaluation metrics

We evaluated our methods using the MRR score and the accuracy at TOP_1, TOP_5 and TOP_{10}, i.e., the proportion of queries with correct answers at the first position, among the first 5 candidates and among the first 10 candidates respectively. MRR is used to assess the performance of methods that take a query as input and that output an ordered list of responses. It looks at the rank of the first correct answer only.

$$\text{MRR} = \frac{1}{|W|} \times \sum_{i=1}^{|W|} \frac{1}{Rank_i}$$

where $|W|$ is the number of queries, $Rank_i$ is the rank of the first correct answer for the i-th query. The closer the MRR score is to 1, the better the model performs.

4 Acquisition of synonymy between biterms

Our first experiment concerns the acquisition of synonymy with lexical substitution and analogy. We start with synonymy because of the availability of resources. In order to test the methods, we need data composed of pairs of semantically related biterms from the environment domain. However, few terminology resources provide relations between terms, and especially between MWTs. Among the online resources dealing with the environment, IATE is the only one that provides a significant number of synonymous biterms that can be used as a dataset for our experiments.

We will first explain how we extracted the data from IATE. Then, we present the test data, the experiment and the results for the two methods, lexical substitution and analogy.

4.1 Extraction of synonymous biterms from IATE

In order to build a set of nominal biterm pairs, we first extracted 20,154 French biterms of the environment domain, like *séparateur magnétique* 'magnetic separator'. We then formed pairs of synonymous biterms using their unique identifier (synonymous terms in IATE have the same identifier). For example, the identifier of both *déferreur magnétique* and *séparateur magnétique* is *1427234*. We then filtered out the biterm pairs that don't occur in PANACEA. 1,064 biterm pairs were kept.

We notice that the biterms that share their identifier are not always synonyms. We therefore removed the biterm pairs containing plural inflection (*pollution des eaux* 'water pollution' and *pollution d'eau*), synonymous derivation (*réchauffement planétaire* 'global warming' and *réchauffement de la planète*), or involving terms of length 3 (*déplacement domicile-travail* 'travel from home to work'). The remaining set, DataIATE, contains 928 biterm pairs.

In order to further select the data for the experiments, we performed an additional analysis upon DataIATE. The results are summarized in Table 1. DataIATE includes 563 biterm pairs that share one lexical word and have identical syntactic structure, such as *environnement de travail* 'work environment' and *milieu de travail* 'workplace' (they share the noun *travail* 'work' and have the same Noun Preposition Noun structure). These biterm pairs are referred to as DataIATE_MLM in the following.

Table 1: Features of set of synonymous biterm pairs extracted from IATE

Set features	Denomination	Nb
biterm pairs	DataIATE	928
biterm pairs with identical structure and one common lexical element	DataIATE_MLM	563
biterm pairs with a frequency greater than 5 in PANACEA	DataIATE_FastText	599

4.2 Acquisition of synonymy between biterms using a masked language model

Test dataset for the MLM experiments. The MLM method presented in Section 3.2 requires pairs of biterms with the same syntactic structure and one common lexical element. The dataset we used is DataIATE_MLM from which we removed the pairs which contain lexical elements not included in the model's vocabulary. As explained in Section 3.3, out-of-vocabulary words are split in wordpieces and cannot be retreived. 396 pairs remain in the dataset once these pairs are removed.

Besides, we now need contexts to feed the MLM model. For each biterm that occurs in the 396 pairs of DataIATE_MLM, we extracted 100 contexts or less from PANACEA taking into account some quality criteria of good contexts as defined by Kilgarriff et al.[13] (2008) (some biterms occur less than 100 times in PANACEA). The quality criteria that we applied are the following: (*i*) the context length has 10 to 100 words; (*ii*) the contexts are sentences; (*iii*) the contexts contain the biterm under consideration and at least one other term of the environment domain. We used the set of the French terms of the environment domain in IATE for the verification of the third condition.

We removed from the dataset the biterms without context that meets the three context quality criteria, such as *taxe d'émission* 'emission tax'. We end up with a final test set made up of 317 biterm pairs and 24,265 contexts.

Experiment. The objective of the experiment being the acquisition of lexical semantic relations between biterms of the environment domain, the rank of the semantically related terms should not be considered relative to all the tokens of the model, but only to the simple terms of the domain in IATE. However, as IATE has a low coverage of the simple terms of the domain, we have extended the model vocabulary with the lexical elements that occur in the biterms in DataIATE, for a total of 784 words. Moreover, we exclude the masked term from the predictions when computing the rank.

For this experiment, we used the CamemBERT-large model.

Results. Table 2 gives the results of the basic and conditional queries. Results are improved with the conditional queries. MRR scores increase from 0.302 to 0.374, meaning that on average, the correct answer is in the 3rd or 4th rank in the first setting and in the 2nd and 3rd in the second. Accuracy scores are also improved.

Table 2: Results of basic MLM and conditional MLM predictions of synonymy between biterms

		Accuracy		
	MRR	TOP_1	TOP_5	TOP_{10}
Basic MLM	0.302	0.189	0.416	0.532
Conditional MLM	0.374	0.253	0.502	0.613

Qualitative analysis. We analyzed the 10 first predictions of 100 conditional queries randomly selected. Most of the predictions are semantically linked to the masked term, synonyms and term variants of which derivational variants being the most common. For example, when we mask *écologique* 'environmental' in a context of *tourisme écologique* 'ecological tourism', the 10 first predictions are *écologie, vert, environnemental, éthique, biologique, énergétique, durable, responsable, naturel, urbain.* These results are consistent with the observation made by Ferret (2021) and Arefyev et al. (2020). For some queries, variants or synonyms of the target term

[13] Meyer (2001) considers contexts containing valuable information as knowledge rich contexts (KRCs). These KRCs contain both a term of interest in a particular domain and a knowledge pattern (KP) that shows how the term of interest is related to other terms in the domain. Barrière (2004) expands the definition of KPs to include semantic information. Hmida et al. (2015) propose to use collocations by completing KPs to extract KRCs. Although contexts containing KPs and collocations are often rich in information, the expressions for a given relation are varied and the number of patterns and collocations is limited.

appear at the first ranks. For example, when we mask *habitation* 'home' in the context *habitation individuelle* 'individual house', the target term *maison* 'house' appears at rank 71, but its synonym *logement* 'housing' appears at rank 2.

We also noticed that better results were obtained for biterm pairs in which W1 and W2 are synonyms. To confirm this observation, we repeated the experiment on 25 biterm pairs in which the simple terms W1 and W2 are synonyms in IATE, resulting in 2,432 queries. The MRR score obtained for the conditional queries is 0.561 (50% improvement). This result confirms the benefit of combining distributional and compositional approaches to the identification of synonymous biterms (Daille 2017). The compositional method consists in generating synonyms of a biterm by substituting one of the biterm components by one of its synonyms.

4.3 Identification of synonymy between biterms by means of analogy

Test dataset for analogy. The test dataset for the analogy experiments is DataIATE_FastText (see Section 4.1). We built $W_1 : W_2 :: MWT_1 : MWT_2$ quadruplets where (*i*) MWT_1 and MWT_2 share a lexical element W_3; (*ii*) W_1 is the other lexical element in MWT_1 and W_2 is the other lexical element in MWT_2; (*iii*) W_1 and W_2 are synonymous.

We built three sets of quadruplets from three French relation databases presented in Section 3.1: RefIATE, RefDicosyn and RefDicoenviro_qsyn (simple term pairs classified as QSYN in RefDicoenviro). Quadruplet_RefIATE contains 20 quadruplets, Quadruplet_RefDicoenviro_qsyn 63 quadruplets and Quadruplet_RefDicosyn 156 quadruplets.

Experiment. Because we look for relations between biterms, the rank is calculated with respect to the 5,002 nominal biterms extracted from IATE that occur at least 5 times in PANACEA (DataIATE_FastText).

We used a FastText model to acquire the analogies because we need a model able to represent in the same space but independently biterms and their components. The representations of the former should not be computed compositionally from the representations of their constituents because the analogy equation would then always be trivial.

In order to directly generate representations for the biterms and their components in the same vector space, we annotated the corpus so that they are indexed separately. For example, a biterm like *air froid* 'cold air' yields three tokens `air` 'air', `froid` 'cold' and `air_froid` 'cold_air'. We also forced the model not to split the words into subwords by setting the parameter `-maxn` to 0. The performance of the FastText model can be significantly impacted by its hyper-parameters (Levy, Goldberg, and Dagan 2015). However, optimizing the average accuracy over a set of diverse relations may not be relevant (Gladkova, Drozd, and Matsuoka 2016). For this reason, we did not try to optimize them. The hyper-parameters we used are: `min_count=3`, `maxn=0`, `window size=5` , `model=skipgram`, `epoch=20`, `lr=0.05`. The other parameters are set to their default values.

Results. Table 3 gives the results obtained with the method by analogy. The MRR scores are high, all above 0.6 showing that analogy is efficient for the detection of synonymy. Quadruplets included in quadruplet_RefIATE yield the best results with a MRR score of 0.744. This good result may be due to the fact that the simple terms and the biterms in quadruplet_RefIATE come from the same database, i.e., IATE. Another interesting outcome is the similar results obtained with the quadruplets built from specialized environment databases and from general language dictionaries. Several lexical units that belong to the general lexicon have also terminological meanings in the environment domain such as *impact* 'impact', *effet* 'effect'. For example, *effet apaisant* 'calming effect' (general language); *effet sur l'environnement* 'environmental effect' (environment domain).

Qualitative analysis. We examined the first five predictions of all queries where the unknown is MWT2. For 73.6% of the quadruplets the target term is one of the first five predictions. When the target term does not appear at TOP_5, one of its synonyms or variants does. For example, for *effet : incidence :: effet sur l'environnement : incidence sur l'environnement* 'effect : impact :: environmental effect : environmental impact', the target term *incidence sur l'environnement* only appears at rank 3,698, but its derivational variant *incidence environnemental* is the first candidate.

Table 3: Synonymy acquisition using analogy between FastText representations with data extracted from IATE

		Accuracy		
	MRR	TOP_1	TOP_5	TOP_{10}
Quadruplet_RefIATE	**0.744**	**0.650**	**0.875**	**0.900**
Quadruplet_RefDicoenviro_qsyn	0.624	0.548	0.723	0.746
Quadruplet_RefDicosyn	0.612	0.522	0.728	0.766

5 Acquiring other types of lexical relations

We are also interested in acquiring other types of lexical relations between MWTs, in particular opposition relations and hierarchical relations which are considered to be essential terminological relations (L'Homme 2020). However, because no terminological resource provides such relations for the MWTs of the environmental domain, we created a dataset of such relations by semantic projection. We noticed that more than 80% of pairs of synonymous biterms extracted from IATE share a lexical word. Semantic projection seems therefore suitable for the identification of lexical relations if we admit that the pairs of biterms that have one of their constituents in an opposition or hierarchical relation also share this characteristic.

In this section, we present the creation by semantic projection of a dataset of four lexical relations between nominal biterms. The resource is then used in experiments same as those presented in Section 4.

5.1 Generation of semantically related biterms by semantic projection

The creation of the dataset involves several steps. We first extracted nominal term candidates that contain two lexical words from the PANACEA corpus using TermSuite (Cram and Daille 2016). Among these, we have kept only those whose frequency in the PANACEA corpus is above 5. We then formed pairs of biterm candidates using semantic projection.

We applied the inference rule in (1) to the pairs of STs from RefDicoenviro in order to form pairs of candidate biterms, without any additional restriction on the order of the constituents nor on their syntactic structure. More formally, let MWT_1 and MWT_2 be two biterm candidates such that $voc(MWT_1) = \{ST_1, W_1\}$ and $voc(MWT_2) = \{ST_2, W_2\}$ where $voc(x)$ is the set of the lexical words of x. For example, voc(protection de la flore 'flora protection') = {*flore* 'flora', *protection* 'protection'}. If ST_1 and ST_2 are two STs connected by a relation R in RefDicoenviro and if $W_1 = W_2$, then we predict that MWT_1 and MWT_2 are also connected by R[14]. Semantic projection can be stated as (1) where M is the set of biterms that occur in PANACEA, S the

[14]Provided that the meaning of the two biterms is compositional.

set of STs and L is the set of the lexical words that occur in PANACEA.

$$\forall MWT_1 \in M, \forall MWT_2 \in M, \text{such that } \exists ST_1 \in S, \exists ST_2 \in S, \exists W \in L, \exists R, \tag{1}$$
$$[voc(MWT_1) = \{ST_1, W\} \wedge voc(MWT_2) = \{ST_2, W\} \wedge R(ST_1, ST_2)] \Rightarrow R(MWT_1, MWT_2)$$

For symmetric relations (e.g., synonymy, antonymy), when the semantic projection yields both $R(MWT_1, MWT_2)$ and $R(MWT_2, MWT_1)$, we only kept one pair in the dataset.

We then checked that the candidates are terms by looking up in the online terminology banks TERMIUM Plus[15], Le Grand Dictionnaire[16] and IATE[17]. We consider any candidate that belong to one of these banks and to a field falling under the environmental domain as a term of the environment domain. As a result, we ended up with 80 pairs of biterms linked by ANTI, 51 pairs linked by HYP and 100 pairs linked by QSYN. The dataset is small because many of the terms extracted by TermSuite are too specific (e.g., *conservation des papillons* 'butterfly conservation') to be present in one of the banks. We will call these 231 quadruplets DataProjSem[18].

The biterm pairs in DataProjSem have then been manually checked by three annotators in order to assess whether the relation between the STs is preserved between the biterms. The inter-annotation agreement (Fleiss' kappa) is 0.69, which is fair. An adjudication phase was then performed. Table 4 presents an excerpt of the validated quadruplets. The relation between the STs and the biterms is recorded in the column Rel; column Anno indicates whether the relation is preserved (1) or not (0).

Table 4: Excerpt of the validated quadruples created by semantic projection

ST1	ST2	MWT1	MWT2	rel	anno
froid	chaud	air froid	air chaud	ANTI	1
'cold'	'hot'	'cold air'	'hot air'		
combustible	gaz	stockage de combustible	stockage de gaz	HYP	1
'fuel'	'gas'	'fuel storage'	'gas storage'		
recyclage	récupération	recyclage des matériaux	matériaux de récupération	QSYN	0
'recycling'	'salvage'	'material recycling'	'material salvage'		

In what follows, we call DataProj the subset of quadruplets where the relation is preserved. The number of these quadruplets for the three types of relations is presented in Table 4. We noticed that in all of them, the two biterms have the same syntactic structure, even though we did not impose any constraint on their syntactic structure during the projection.

Table 5: Number of quadruplets in DataProj for the three types of relation.

ANTI	HYP	QSYN	Total
69	26	85	180

[15]https://www.btb.termiumplus.gc.ca/
[16]http://www.granddictionnaire.com/
[17]https://iate.europa.eu/
[18]The dataset is available online: https://github.com/YizWang/List-of-semantically-linked-MWTs

5.2 Acquiring the other lexical relations by means of masked language models

Test dataset used in the MLM experiments. The MLM experiments are evaluated by means of a test set made up of the quadruplets of the DataProj where both STs belong to the vocabulary of the model. This subset contains 61 pairs of ANTI, 16 pairs of hypernymy, 16 pairs of hyponymy and 71 pairs of QSYN.

Experimentation. We performed an MLM experiment like the one presented in Section 4.2 using 100 contexts (or less) extracted from the PANACEA corpus for each biterm that appears in one of the quadruplets of the test set. As before, we used the CamemBERT-large model.

Results and discussion. In this experiment, the rank of the semantically related terms is established with respect to a vocabulary made up of the STs in DiCoEnviro that occur in the PANACEA corpus and in the model vocabulary (a total of 796 STs). Table 6 and Table 7 show the results obtained for basic and conditional queries respectively. We see in Table 6 that basic MLM models acquire QSYN relations with an MRR of 0.54, meaning that on average the correct answer is at first or second position. On the other hand, ANTI and HYP are more difficult to identify. Table 7 shows that the conditional method improves MRR and accuracy significantly especially for hypernymy with a 69% increase. As before, we manually analyzed the first ten predictions of 100 randomly chosen conditional queries. The conclusions are the same as for synonymy: for many queries, variants or synonyms of the masked term appear at the first ranks.

Table 6: Results of the basic MLM prediction for each type of lexical relation

		Accuracy		
Basic MLM	MRR	TOP_1	TOP_5	TOP_{10}
ANTI	0.316	0.160	0.498	0.687
Hypernymy	0.360	0.229	0.513	0.635
Hyponymy	0.398	0.289	0.506	0.645
QSYN	**0.541**	**0.385**	**0.739**	**0.839**

Table 7: Results of the conditional MLM prediction for each type of lexical relation

		Accuracy		
Conditional MLM	MRR	TOP_1	TOP_5	TOP_{10}
ANTI	0.358	0.195	0.567	0.747
Hypernymy	**0.614**	**0.448**	0.814	**0.938**
Hyponymy	0.449	0.286	0.652	0.874
QSYN	0.604	0.421	**0.843**	0.928

5.3 Acquiring the other semantic relations by means of analogy

Test dataset for analogy. We used DataProj as test set for the analogy experiment.

Experiments. We carried out the same experiment as in Section 4.3.

Results and discussion. As in the previous experiment, each quadruplet generates two analogy equations. For the symmetrical relations ANTI and QSYN, the results presented below are the average of the two tests. For hyponymy and hypernymy, the results are computed separately. In this experiment, the rank of the solution is calculated relative to a set of 1,679 nominal biterms made up of 386 biterms that occur in DataProjSem (see Section 5.1) and 1,354 related nominal biterms extracted from DiCoEnviro, and which occur in PANACEA.

Table 8 shows that overall, the method by analogy performs better than MLM methods. Analogy is better at identifying ANTI and QSYN between biterms showing close MRR scores: 0.72 for ANTI, 0.793 for QSYN. These results are consistent with those of Ferret (2021). QSYN and ANTI are more easily captured than hierarchical relations, in particular hypernymy. However, if the MRR result for hyponymy is the lowest, accuracy scores are surprisingly the best at TOP_5 with a MRR score of 0.962 and at TOP_{10} with a MRR score of 1.

Table 8: Results of the analogy experiment for each type of relation

		Accuracy		
Conditional MLM	MRR	TOP_1	TOP_5	TOP_{10}
ANTI	0.720	0.590	0.926	0.967
Hypernymy	0.613	0.423	0.808	0.923
Hyponymy	0.579	0.346	**0.962**	**1.000**
QSYN	**0.793**	**0.697**	0.937	0.958

6 Discussion

In this paper, we proposed two DSM-based methods for acquiring lexical relations between biterms. These methods have been evaluated on two datasets: synonyms extracted from IATE and biterm pairs constructed by semantic projection and linked by ANTI, HYP and QSYN. First, it is worth noting that the two datasets are complementary since the intersection of DataIATE and DataProj's QSYN pairs contains only 9 pairs of biterms. This complementarity is probably due to the nature of IATE and DiCoEnviro. IATE's synonyms are intended for translation and writing while DiCoEnviro was created to describe and structure terms in the environment domain. In addition, there is a slight difference in the relations themselves. Some relations labelled as synonymy in DataIATE are hypernymic[19] such as *vie sauvage:animaux sauvages* 'wildlife:wild animals'; similarly, QSYN contains co-hyponymy relations such as *réseau ferroviaire:réseau routier* 'railroad network:road network'.

Our two methods perform similarly for synonymy i.e., on both dataset, analogy works better than lexical substitution. However, the two methods give better results on the QSYN pairs than on the synonymous biterms extracted from IATE, possibly because QSYN biterm pairs created by semantic projection are more constrained: they share the same syntactic structure and the two STs they contain are semantically related. This shows that the semantic compositionality of MWTs is captured by MLM and somehow modeled by analogy.

The main difference between our two methods is that MLM predictions are highly context-dependent while analogy predictions are context-independent because the FastText model is not contextual. The latter is therefore more suitable for identifying lexical relations. These observations confirm those of Peters et al. (2018) who show that contextual language models

[19] In some cases, hypernymy can be equated with quasi-synonymy since they are interchangeable in certain contexts (Polguère 2016).

under-perform on semantic relation identification with respect to static embedding models when evaluated on analogy tasks.

We also observe that the results of the analogy-based method are better for symmetrical relations than for asymmetrical relations. Two reasons may explain this difference. The first is that we use a symmetrical distance (i.e., cosine) which might be less suitable for asymmetrical relations. One possible solution could be to use a directional similarity measure such as WeedsPrec (Weeds and Weir 2003)(Weeds and Weir 2003). For example, Lenci and Benotto (2012) show that WeedsPrec perform better than cosine for hypernym identification with an Average Precision score of 0.4 vs 0.23. The second reason is that the difference between the embeddings of two words does not accurately represent the relation that exists between them. The relation is only approximately identified(Vylomova et al. 2015). This approximation disadvantages asymmetrical relations more than symmetrical ones.

The results obtained with the MLM-based method can be compared with those of Schick and Schütze (2020). The methods used and the relations considered are similar. The main difference is that our study is conducted within a specialized domain. Schick and Schütze's (2020) MRR scores are better than ours for antonymy but lower for hypernymy: 0.570 and 0.462 respectively while ours are 0.358 and 0.614. Antonymy is difficult to capture by MLM-based method. The good results obtained by Schick and Schütze (2020) for antonymy could probably be explained by the fact that they expressed the lexical relations using predefined patterns whereas the entries of our model are contexts extracted from corpora. Their patterns guide the prediction of antonyms more accurately than the contexts because they contain explicit negative expressions. Conversely, contexts extracted from corpora provide more information about the masked word for the prediction of synonymy and hypernymy, in part because they are longer.

The results of the analogy-based method can be compared to those of Chaudhri et al. (2022) who use a seq2seq model and a seq2vec model complemented by a linear layer to solve analogy equations. These methods are relatively complex compared to ours, which is simply based on vector offset. In addition, the work of Chaudhri et al. (2022) focuses on domain-specific relations in biology, which are different from the lexical semantic relations we are interested in. Once these differences are known, we can nevertheless broadly compare their methods to ours. The best Top 1 accuracy score obtained by Chaudhri et al. (2022) for analogy equations containing MWTs is 0.51 (with the seq2vec ELMo model). It is significantly lower than ours which is 0.697. Consequently, there is potentially a strong potential for progress that would justify carrying out the same tasks using the methods we propose.

Our results can also be compared with those of Hazem and Daille (2018) for synonymy. They propose different methods to capture synonymy between MWTs in the domain of wind energy, a field close to that of the environment. Hazem and Daille (2018) obtain a MAP of 0.349 for French MWTs using a semantic projection method and a word2vec model in order to take advantage of distributional relations. This value is close to the one we obtained on the dataset extracted from IATE using the lexical substitution based method with a MRR of 0.374. However, this score is much lower than that of the analogy method whose MRR is 0.744. On the other hand, we also use semantic projection to discover pairs of MWTs that are in a QSYN relation (QSYN is mainly composed of synonyms). The accuracy with this second setting is 0.60. It is much higher than the MAP score of 0.25 they report. One reason that may explain the difference in performance between their methods and ours is the fact that the MWTs we consider are in the same relations as the simple terms they contain, whereas Hazem and Daille (2018) use relations other than synonymy to infer MWT synonymy.

We also saw in Section 5.2 that of the conditional MLM method gives better results for hypernymy than for synonymy. Because conditional MLM queries yield answers that are semantically similar to the masked term, one may expect conditional MLM to favor synonymy. However, we saw that hypernym MRR score is the one that increases the most. This is partly due to the fact that the contexts of the hypernymy pairs are less informative that the ones of the synonymy pairs. We analyzed the 641 contexts used for predicting the 16 hypernymy pairs in DataProj which are included in the model vocabulary. We notice that these contexts are relatively poor: there are short and contain little linguistic knowledge. These queries are therefore the ones that benefit the most from the conditional strategy.

We also check whether the analogy-based and MLM-based methods give good results for the same quadruplets or whether they are complementary. We analyze the results at TOP_5 of the queries generated by the 147 quadruplets of the DataProj where both STs belong to the vocabulary of the model (Table 9). We observe that the two methods are quite complementary. Of the 10 quadruplets for which the solution provided by the analogy-based method is beyond the 5^{th} position, 6 have a solution that is in the TOP_5 for the lexical substitution method. Conversely, of the 66 quadruplets for which the solution provided by the substitution-based method is beyond the 5^{th} position, 62 have a solution that is in the TOP_5 for the analogy method. However, the size of our dataset is not large enough to enable us to effectively characterize the quadruplets on which the models fail.

Table 9: Comparison of the methods based on analogy (ANA) and lexical substitution (MLM) at Top 5 on the results of 147 queries

	ANA >5	ANA ≤ 5	MLM >5	MLM ≤ 5
ANA >5	10	0	4	6
ANA ≤ 5	0	137	62	75
MLM >5	4	62	66	0
MLM ≤ 5	6	75	0	81

The polysemous nature of scientific vocabulary is a well-known phenomenon. The environment domain which encompasses several sub-domains results in domain specific meanings of certain terms, Terms such as *gaz* ‘gaz’ are used across scientific domains such as energy or chemistry having different meanings.

Polysemy penalizes the methods based on static models, but also the MLM-based ones. For example, the meaning of *air frais* ‘fresh air’ in (3) is ‘outdoor air’ while it is ‘cold air’ in (4). When we mask *frais* in (3), the first three predictions are *neuf* ‘new’, *comprimé* ‘compressed’ and *extérieur* ‘external’. When we do the same in (4), the first three predictions are *froid* ‘cold’, *chaud* ‘hot’ and *glacial* ‘icy’. This shows that the MLM predictions can be improved by limiting the effects of term ambiguity.

(3) *les bâtiments sont mis en surpression avec un apport d’**air frais***
‘the buildings are put under overpressure with a fresh air supply’

(4) *cette tempête est née de la dépression qui s’ est formée vendredi sur l’ atlantique à environ 1 000km des açores , dépression provoquée par le contraste entre l’ **air frais** de l’ océan et l’ air chaud qui remonte d’ afrique*
‘this storm is born from the depression which formed on Friday on the Atlantic at approximately 1 000 km of the Azores, depression caused by the contrast between the fresh air of the ocean and the hot air which goes up of Africa’

7 Conclusion

In this paper, we explored the capacity of distributional semantic models to capture lexical semantic relations between MWTs. We concentrated on nominal biterms which are well represented in terminological databases. We used masked language models and analogy in order to predict the lexical relations under consideration. For our experiments, we used two datasets: the first is composed of synonymous biterms extracted from IATE, the second is created by semantic projection using various terminological resources and covering a wider range of lexical relations.

The results show that analogy capture synonymy, antonymy and hyponymy better than masked language models. This result suggests that analogy does, to some extent, capture the semantic composition encoded in the static embeddings; conversely, compositional generalization of Transformer-based models seems weaker. The best results are obtained with analogy, with an MRR of 0.793 for synonymy, of 0.720 for antonymy, of 0.613 for hypernymy and of 0.579 for hyponymy.

We plan to follow this work by improving the compositional generalization capabilities of Transformer models given the compositional nature of MWT meaning. We also plan to use methods based on generative models in which lexical substitution becomes a natural language generation task(Lee et al. 2021). In addition, we intend to use methods that are more complex, including context classification and graph models that represent relational information (Jinling Xu et al. 2021; Peng et al. 2017).

Funding

This work has been funded by the French National Research Agency (ANR) under the ADDICTE (ANR-17-CE23-0001) and the DELICES (ANR-19-CE38-0005) projects.

References

Allen, Carl, and Timothy Hospedales. 2019. "Analogies Explained: Towards Understanding Word Embeddings." In *International Conference on Machine Learning*, 223–31. Long Beach, California, USA: PMLR.

Arefyev, Nikolay, Boris Sheludko, Alexander Podolskiy, and Alexander Panchenko. 2020. "A Comparative Study of Lexical Substitution Approaches Based on Neural Language Models." *ArXiv Preprint ArXiv:2006.00031.*

Barrière, Caroline. 2004. "Knowledge-Rich Contexts Discovery." In *Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI*, 187–201. London, Ontario, Canada: Springer.

Bernier-Colborne, Gabriel. 2017. "Aide à l'identification de Relations Lexicales Au Moyen de La Sémantique Distributionnelle et Son Application à Un Corpus Bilingue Du Domaine de l'environnement." *PhD Diss., Université de Montréal.*

Bernier-Colborne, Gabriel, and Patrick Drouin. 2016. "Évaluation des modèles sémantiques distributionnels : le cas de la dérivation syntaxique (Evaluation of distributional semantic models : The case of syntactic derivation)." In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. volume 2 : TALN (Articles longs)*, 125–38. Paris, France: AFCP - ATALA.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. “Enriching Word Vectors with Subword Information.” *Transactions of the Association for Computational Linguistics* 5: 135–46.

Bouraoui, Zied, Jose Camacho-Collados, and Steven Schockaert. 2020. “Inducing Relational Knowledge from BERT.” In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:7456–63. New York, New York, United States.

Bouraoui, Zied, Shoaib Jameel, and Steven Schockaert. 2018. “Relation Induction in Word Embeddings Revisited.” In *Proceedings of the 27th International Conference on Computational Linguistics*, 1627–37. Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Bourigault, Didier. 2002. “UPERY : Un Outil D’analyse Distributionnelle Étendue Pour La Construction D’ontologies à Partir de Corpus.” In *Actes de La 9e Conférence Sur Le Traitement Automatique Des Langues Naturelles. Articles Longs*, 75–84. Nancy, France: ATALA.

Bullinaria, John A, and Joseph P. Levy. 2012. “Extracting Semantic Representations from Word Co-Occurrence Statistics: Stop-Lists, Stemming, and SVD.” *Behavior Research Methods* 44 (3): 890–907.

Chaudhri, Vinay K., Justin Xu, Han Lin Aung, and Sajana Weerawardhena. 2022. “A Corpus of Biology Analogy Questions as a Challenge for Explainable AI.” In *Bridging Human Intelligence and Artificial Intelligence*, edited by Mark V. Albert, Lin Lin, Michael J. Spector, and Lemoyne S. Dunn, 327–37. Educational Communications and Technology: Issues and Innovations. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-84729-6_23.

Chen, Zhiwei, Zhe He, Xiuwen Liu, and Jiang Bian. 2018. “Evaluating Semantic Relations in Neural Word Embeddings with Biomedical and General Domain Knowledge Bases.” *BMC Medical Informatics and Decision Making* 18 (2): 53–68.

Cram, Damien, and Béatrice Daille. 2016. “Terminology Extraction with Term Variant Detection.” In *Proceedings of ACL-2016 System Demonstrations*, 13–18. Berlin, Germany: Association for Computational Linguistics.

Daille, Béatrice. 2017. *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Vol. 19. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Daille, Béatrice, and Amir Hazem. 2014. “Semi-Compositional Method for Synonym Extraction of Multi-Word Terms.” In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, 1202–7. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/679_Paper.pdf.

Depraetere, Ilse. 2019. “Meaning in Context and Contextual Meaning: A Perspective on the Semantics-Pragmatics Interface Applied to Modal Verbs.” *Anglophonia. French Journal of English Linguistics*, no. 28.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Espinosa Anke, Luis, Joan Codina-Filba, and Leo Wanner. 2021. “Evaluating Language Models for the Retrieval and Categorization of Lexical Collocations.” In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1406–17. Online: Association for Computational Linguistics.

Ferret, Olivier. 2021. “Exploration des relations sémantiques sous-jacentes aux plongements contextuels de mots (Exploring semantic relations underlying contextual word embeddings).” In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, 26–36. Lille, France: ATALA.

Fu, Ruiji, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. “Learning Semantic Hierarchies via Word Embeddings.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1199–1209. Baltimore, Maryland: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-1113.

Gábor, Kata, Davide Buscaldi, Anne-Kathrin Schumann, Behrang QasemiZadeh, Haifa Zargayouna, and Thierry Charnois. 2018. “SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers.” In *Proceedings of The 12th International Workshop on Semantic Evaluation*, 679–88. New Orleans, Louisiana.

Gladkova, Anna, Aleksandr Drozd, and Satoshi Matsuoka. 2016. “Analogy-Based Detection of Morphological and Semantic Relations with Word Embeddings: What Works and What Doesn’t.” In *Proceedings of the NAACL Student Research Workshop*, 8–15. San Diego, California.

Grabar, Natalia, and Thierry Hamon. 2006. “Terminology Structuring through the Derivational Morphology.” In *International Conference on Natural Language Processing (in Finland)*, 652–63. Berlin Heidelberg: Springer.

Harris, Zellig S. 1954. “Distributional Structure.” *Word* 10 (2–3): 146–62.

Hashimoto, Kazuma, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. 2015. “Task-Oriented Learning of Word Embeddings for Semantic Relation Classification.” In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, 268–78. Beijing, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/K15-1027.

Hazem, Amir, and Béatrice Daille. 2018. “Word Embedding Approach for Synonym Extraction of Multi-Word Terms.” In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 297–303. Miyazaki, Japan: European Language Resources Association (ELRA).

Hmida, Firas, Emmanuel Morin, and Béatrice Daille. 2015. “Extraction de Contextes Riches En Connaissances En Corpus Spécialisés.” In *Actes de La 22e Conférence Sur Le Traitement Automatique Des Langues Naturelles. Articles Courts*, 109–15. Caen, France: ATALA.

Hou, Jiaqi, Xin Li, Haipeng Yao, Haichun Sun, Tianle Mai, and Rongchen Zhu. 2020. “Bert-Based Chinese Relation Extraction for Public Security.” *IEEE Access* 8: 132367–75.

Jameel, Shoaib, Zied Bouraoui, and Steven Schockaert. 2017. “Modeling Semantic Relatedness Using Global Relation Vectors.” *ArXiv Preprint ArXiv:1711.05294*.

Kilgarriff, Adam, Miloš Husák, Katy McAdam, Michael Rundell, and Pavel Rychlý. 2008. “GDEX: Automatically Finding Good Dictionary Examples in a Corpus.” In *Proceedings of the XIII EURALEX International Congress*, 425–32. Barcelona, Spain: Documenta Universitaria.

Köper, Maximilian, Christian Scheible, and Sabine Schulte im Walde. 2015. “Multilingual Reliability and ‘Semantic’ Structure of Continuous Word Spaces.” In *Proceedings of the 11th International Conference on Computational Semantics*, 40–45. London, UK: Association for Computational Linguistics.

Kudo, Taku, and John Richardson. 2018. “SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing.” In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 66–71. Brussels, Belgium: Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-2012.

Lafourcade, Mathieu, and Lionel Ramadier. 2016. “Semantic Relation Extraction with Semantic Patterns Experiment on Radiology Reports.” In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 4578–82. Portorož, Slovenia: European Language Resources Association (ELRA).

Lenci, Alessandro. 2008. “Distributional Semantics in Linguistic and Cognitive Research.” *Italian Journal of Linguistics* 20 (1): 1–31.

Lenci, Alessandro, and Giulia Benotto. 2012. “Identifying Hypernyms in Distributional Semantic Spaces.” In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop* on Semantic Evaluation (SemEval 2012), 75–79. Montréal, Canada: Association for Computational Linguistics.

Levy, Omer, and Yoav Goldberg. 2014. “Linguistic Regularities in Sparse and Explicit Word Representations.” In Proceedings of the Eighteenth Conference on Computational Natural Language Learning, 171–80. Ann Arbor, Michigan: Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-1618.

Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. “Improving Distributional Similarity with Lessons Learned from Word Embeddings.” *Transactions of the Association for Computational Linguistics* 3: 211–25.

Levy, Omer, Steffen Remus, Chris Biemann, and Ido Dagan. 2015. “Do Supervised Distributional Methods Really Learn Lexical Inference Relations?” In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 970–76. Denver, Colorado: Association for Computational Linguistics.

L’Homme, Marie-Claude. 2004. *La Terminologie: Principes et Techniques*. Montréal: Pum.

L'Homme, Marie-Claude. 2020. *Lexical Semantics for Terminology: An Introduction*. Vol. 20. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Martin, Louis, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. "CamemBERT: A Tasty French Language Model." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7203–19. Online: Association for Computational Linguistics.

Meyer, Ingrid. 2001. "Extracting Knowledge-Rich Contexts for Terminography." In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin and Marie-Claude L'Homme, 279–302. Amsterdam / Philadelphia: John Benjamins.

Mickus, Timothee, Denis Paperno, Mathieu Constant, and Kees van Deemter. 2020. "What Do You Mean, BERT?" In *Proceedings of the Society for Computation in Linguistics 2020*, 279–90. New York, New York: Association for Computational Linguistics.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. "Linguistic Regularities in Continuous Space Word Representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–51. Atlanta, Georgia: Association for Computational Linguistics.

Morin, Emmanuel, and Christian Jacquemin. 1999. "Projecting Corpus-Based Semantic Links on a Thesaurus." In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, 389–96. USA: Association for Computational Linguistics. https://doi.org/10.3115/1034678.1034739.

Morlane-Hondère, François, and Cécile Fabre. 2012. "Le Test de Substituabilité à l'épreuve Des Corpus: Utiliser l'analyse Distributionnelle Automatique Pour l'étude Des Relations Lexicales." In *3e CMLF*, 1:1001–15. France: EDP Sciences.

Paullada, Amandalynne, Bethany Percha, and Trevor Cohen. 2020. "Improving Biomedical Analogical Retrieval with Embedding of Structural Dependencies." In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing*, 38–48. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.bionlp-1.4.

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2227–37. New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202.

Peters, Matthew E., Mark Neumann, Luke Zettlemoyer, and Wen-tau Yih. 2018. "Dissecting Contextual Word Embeddings: Architecture and Representation." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1499–1509. Brussels, Belgium: Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1179.

Polguère, Alain. 2016. *Lexicologie et Sémantique Lexicale: Notions Fondamentales*. Montréal: Presses de l'Université de Montréal.

Qiang, Jipeng, Yun Li, Yi Zhu, Yunhao Yuan, and Xindong Wu. 2019. "A Simple BERT-Based Approach for Lexical Simplification." *ArXiv* abs/1907.06226.

Qiao, Bo, Zhuoyang Zou, Yu Huang, Kui Fang, Xinghui Zhu, and Yiming Chen. 2022. "A Joint Model for Entity and Relation Extraction Based on BERT." *Neural Comput. Appl.* 34 (5): 3471–81.

Roller, Stephen, Katrin Erk, and Gemma Boleda. 2014. "Inclusive yet Selective: Supervised Distributional Hypernymy Detection." In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1025–36. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Santos, Cicero Nogueira dos, Bing Xiang, and Bowen Zhou. 2015. "Classifying Relations by Ranking with Convolutional Neural Networks." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 626–34. Beijing, China: Association for Computational Linguistics. https://doi.org/10.3115/v1/P15-1061.

Schick, Timo, and Hinrich Schütze. 2019. "Attentive Mimicking: Better Word Embeddings by Attending to Informative Contexts." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 489–94. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1048.

———. 2020. "Rare Words: A Major Problem for Contextualized Embeddings and How to Fix It by Attentive Mimicking." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:8766–74. New York, USA: AAAI Press.

Shi, Peng, and Jimmy Lin. 2019. "Simple BERT Models for Relation Extraction and Semantic Role Labeling." *ArXiv Preprint ArXiv:1904.05255*.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. "Sequence to Sequence Learning with Neural Networks." In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 3104–12. Montreal, Canada: MIT Press.

Turney, Peter D. 2005. "Measuring Semantic Similarity by Latent Relational Analysis." In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 1136–41. IJCAI'05. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Verspoor, Cornelia M, Cliff Joslyn, and George J Papcun. 2003. "The Gene Ontology as a Source of Lexical Semantic Knowledge for a Biological Natural Language Processing Application." In *SIGIR Workshop on Text Analysis and Search for Bioinformatics*, 51–56. Toronto, Canada.

Vylomova, Ekaterina, Laura Rimell, Trevor Cohn, and Timothy Baldwin. 2016. "Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences for Lexical Relation Learning." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1671–82. Berlin, Germany: Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1158.

Weeds, Julie, Daoud Clarke, Jeremy Reffin, David Weir, and Bill Keller. 2014. "Learning to Distinguish Hypernyms and Co-Hyponyms." In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2249–59. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Weeds, Julie, et David Weir. 2003. “A General Framework for Distributional Similarity”. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, 81-88. EMNLP ’03. USA: Association for Computational Linguistics.

Xue, Kui, Yangming Zhou, Zhiyuan Ma, Tong Ruan, Huanhuan Zhang, and Ping He. 2019. “Fine-Tuning BERT for Joint Entity and Relation Extraction in Chinese Medical Text.” In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 892–97. San Diego, CA, USA: IEEE.

Yao, Liang, Chengsheng Mao, and Yuan Luo. 2019. “KG-BERT: BERT for Knowledge Graph Completion.” *ArXiv Preprint ArXiv:1909.03193*.

Zhang, Li, Jun Li, and Chao Wang. 2017. “Automatic Synonym Extraction Using Word2Vec and Spectral Clustering.” In *2017 36th Chinese Control Conference (CCC)*, 5629–32. Dalian, China: IEEE.

Zhou, Wangchunshu, Tao Ge, Ke Xu, Furu Wei, and Ming Zhou. 2019. “BERT-Based Lexical Substitution.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3368–73. Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1328.

Zhu, Yongjun, Erjia Yan, and Fei Wang. 2017. “Semantic Relatedness and Similarity of Biomedical Terms: Examining the Effects of Recency, Size, and Section of Biomedical Publications on the Performance of Word2vec.” *BMC Medical Informatics and Decision Making* 17 (1): 1–8.

Zweigenbaum, Pierre, and Natalia Grabar. 2000. “Liens Morphologiques et Structuration de Terminologie.” In *IC 2000 : Ingénierie Des Connaissances*, 325–34. Toulouse: Irit.