Multi-Relational Hyperbolic Word Embeddings from Natural Language Definitions

Marco Valentino[1], **Danilo S. Carvalho**[2,3], **André Freitas**[1,2,3]
[1] Idiap Research Institute, Switzerland
[2] Department of Computer Science, University of Manchester, UK
[3] National Biomarker Centre, CRUK-MI, University of Manchester, UK
[1]`{firstname.lastname}@manchester.ac.uk`
[2]`{firstname.lastname}@idiap.ch`

Abstract

Natural language definitions possess a recursive, self-explanatory semantic structure that can support representation learning methods able to preserve explicit conceptual relations and constraints in the latent space. This paper presents a *multi-relational* model that explicitly leverages such a structure to derive *word embeddings* from definitions. By automatically extracting the relations linking defined and defining terms from dictionaries, we demonstrate how the problem of learning word embeddings can be formalised via a translational framework in *Hyperbolic space* and used as a proxy to capture the global semantic structure of definitions. An extensive empirical analysis demonstrates that the framework can help imposing the desired structural constraints while preserving the semantic mapping required for controllable and interpretable traversal. Moreover, the experiments reveal the superiority of the Hyperbolic word embeddings over the Euclidean counterparts and demonstrate that the multi-relational approach can obtain competitive results when compared to state-of-the-art neural models, with the advantage of being intrinsically more efficient and interpretable[1].

1 Introduction

A *natural language definition* is a statement whose core function is to describe the essential meaning of a word or a concept. As such, extensive collections of definitions (Miller, 1995; Zesch et al., 2008), such as the ones found in dictionaries or technical discourse, are often regarded as rich and reliable sources of information from which to derive textual embeddings (Tsukagoshi et al., 2021; Bosc and Vincent, 2018; Tissier et al., 2017; Noraset et al., 2017; Hill et al., 2016).

A fundamental characteristic of natural language definitions is that they are widely abundant, possessing a *recursive, self-explanatory* semantic structure which typically connects the meaning of terms composing the definition (*definiens*) to the meaning of the terms being defined (*definiendum*). This structure is characterised by a well-defined set of semantic roles linking the terms through explicit relations such as subsumption and differentiation (Silva et al., 2016) (see Figure 1). However, existing paradigms for extracting embeddings from natural language definitions rarely rely on such a structure, often resulting in poor interpretability and semantic control (Mikolov et al., 2013; Pennington et al., 2014; Reimers and Gurevych, 2019).

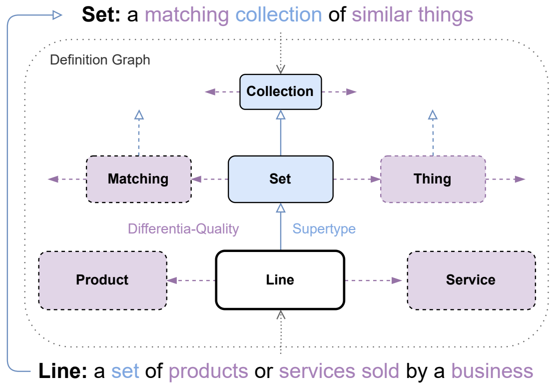

Figure 1: How can we inject the recursive, hierarchical structure of *natural language definitions* into word embeddings? This paper investigates *Hyperbolic manifolds* to learn *multi-relational* representations exclusively from definitions, formalising the problem via a translational framework to preserve the semantic mapping between concepts in latent space.

This paper investigates new paradigms to overcome these limitations. Specifically, we posit the following research question: *"How can we leverage and preserve the explicit semantic structure of natural language definitions for neural-based embeddings?"* To answer the question, we explore *multi-relational* models that can learn to explicitly map *definenda*, *definiens*, and their corresponding *semantic relations* within a continuous vector

[1]Code and data available at: https://github.com/neuro-symbolic-ai/multi_relational_hyperbolic_word_embeddings

Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics Volume 1: Long Papers, pages 23–34

space. Our aim, in particular, is to build an embedding space that can encode the structural properties of the relevant semantic relations, such as concept hierarchy and differentiation, as a product of geometric constraints and transformations. The multi-relational nature of such embeddings should be intrinsically interpretable, and define the movement within the space in terms of mapped relations and entities. Since *Hyperbolic manifolds* have been demonstrated to correspond to continuous approximations of recursive and hierarchical structures (Nickel and Kiela, 2017), we hypothesise them to be the key to achieve such a goal.

Following these motivations and research hypotheses, we present a *multi-relational framework* for learning *word embeddings* exclusively from *natural language definitions*. Our methodology consists of two main phases. First, we build a *specialised semantic role labeller* to automatically extract multi-relational triples connecting definienda and definiens. This explicit mapping allows casting the learning problem into a *link prediction* task, which we formalise via a *translational* objective (Balazevic et al., 2019; Feng et al., 2016; Bordes et al., 2013). By specialising the translational framework in Hyperbolic space through *Poincaré manifolds*, we are able to jointly embed entities and semantic relations, imposing the desired structural constraints while preserving the explicit mapping for a controllable traversal of the space.

An extensive empirical evaluation led to the following conclusions:

1. Instantiating the multi-relational framework in Euclidean and Hyperbolic spaces reveals the explicit gains of Hyperbolic manifolds in capturing the global semantic structure of definitions. The Hyperbolic embeddings, in fact, outperform the Euclidean counterparts on the majority of the benchmarks, being also superior on one-shot generalisation experiments designed to assess the structural organisation and interpretability of the embedding space.

2. A comparison with distributional approaches and previous work based on autoencoders demonstrates the impact of the semantic relations on the quality of the embeddings. The multi-relational model, in fact, outperforms previous approaches with the same dimensions, while being intrinsically more interpretable and controllable.

3. The multi-relational framework is competitive with state-of-the-art Sentence-Transformers, having the advantage of requiring less computational and training resources, and possessing a significantly lower number of dimensions.

4. We conclude by performing a set of qualitative analyses to visualise the interpretable nature of the traversal for such vector spaces. We found that the multi-relational framework enables robust semantic control, clustering the closely defined terms according to the target semantic transformations.

To the best of our knowledge, we are the first to conceptualise and instantiate a multi-relational Hyperbolic framework for representation learning from natural language definitions, opening new research directions for improving the interpretability and structural control of neural embeddings.

2 Background

2.1 Natural Language Definitions

Natural language definitions possess a *recursive, self-explanatory* semantic structure. Such structure connects the meaning of terms composing the definition (*definiens*) to the meaning of the terms being defined (*definiendum*) through a set of semantic roles (see Table 1). These roles describe particular semantic relations between the concepts, such as subsumption and differentiation (Silva et al., 2016). Previous work has shown the possibility of automated categorisation of these semantic roles (Silva et al., 2018a), and leveraging those can lead to models with higher interpretability and better navigation control over the semantic space (Carvalho et al., 2023; Silva et al., 2019, 2018b).

It is important to notice that while the definitions are lexically indexed by their respective definienda, the terms they define are *concepts*, and thus a single lexical item (definiendum) can have multiple definitions. For example, the word "line" has the following two definitions, among others:

"*An infinitely extending one-dimensional figure that has no curvature.*"

"*A set of products or services sold by a business, or by extension, the business itself.*"

from which upon analysis, we can find the roles of **supertype** and differentia quality, as follows:

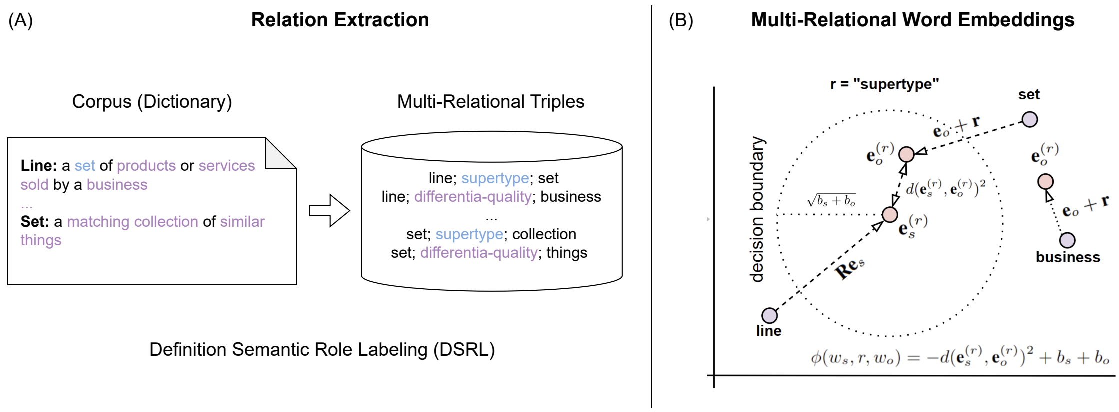

Figure 2: An overview of the multi-relational framework for learning word embeddings from definitions. The methodology consists of two main phases: (A) building a specialised *semantic role labeller* (DSRL) for the annotation of natural language definitions and the extraction of relations from large dictionaries; (B) formalising the learning problem as a *link prediction* task via a translational framework. The translational formulation acts as a proxy for minimising the distance between words that are connected in the definitions (e.g., *line* and *set*) while preserving the semantic relations for interpretable and controllable traversal of the space.

"*An infinitely extending one-dimensional* ***figure*** *that has no curvature.*"

"*A* ***set*** *of products or services sold by a business, or by extension, the business itself.*"

A definiendum can then be identified by the interpretation of its associated terms, categorised according to its semantic roles within the definition. A line which has "figure" as *supertype* is thus a different concept from a line which has "set" as *supertype*. The same can be applied for the other aforementioned roles: a line with supertype "set" and distinguished by the *differentia quality* "product" is different from a line distinguished by "point" on the same role. This is a recursive process, as each term in a definition is also representing a concept, which may be defined in the dictionary. This entails a hierarchical and multi-relational structure linking the terms in the definiendum and in the definiens.

2.2 Hyperbolic Embeddings

As the semantic roles induce multiple hierarchical and recursive structures (e.g., the `supertype` and `differentia quality` relation), we hypothesise that Hyperbolic geometry can play a crucial role in learning word embeddings from definitions. Previous work, in fact, have demonstrated that recursive and hierarchical structures such as trees can be represented in a continuous space via a d-dimensional Poincaré ball (Nickel and Kiela, 2017; Balazevic et al., 2019).

A Poincaré ball $(\mathbb{B}_c^d, g^{\mathbb{B}})$ of radius $1/\sqrt{c}, c > 0$ is a d-dimensional manifold equipped with the Riemannian metric $g^{\mathbb{B}}$. In such d-dimensional space, the distance between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{B}$ can be computed along a geodesic as follows:

$$d_{\mathbb{B}}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \tanh^{-1}\left(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|\right), \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm and $\oplus_c$ represents Mobiüs addition (Ungar, 2001):

$$\mathbf{x} \oplus \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2}, \quad (2)$$

with $\langle \cdot, \cdot \rangle$ representing the Euclidean inner product. A crucial feature of Equation 1 is that it allows determining the organisation of hierarchical structures locally, simultaneously capturing the hierarchy of entities (via the norms) and their similarity (via the distances) (Nickel and Kiela, 2017).

Remarkably, subsequent work has shown that this formalism can be extended for multi-relational graph embeddings via a translation framework (Balazevic et al., 2019), parametrising multiple Poincaré balls within the same embedding space (Section 3.2).

3 Methodology

We present a *multi-relational* model to learn *word embeddings* exclusively from *natural language definitions* that can leverage and preserve the *semantic relations* linking definiendum and definiens.

The methodology consists of two main phases: (1) building a specialised *semantic role labeller*

Role	Description
Supertype	An hypernym for the definiendum.
Differentia Quality	A quality that distinguises the definiendum from other concepts under the same supertype.
Differentia Event	an event (action, state or process) in which the definiendum participates and is essential to distinguish it from other concepts under the same supertype.
Event Location	the location (spatial or abstract) of a differentia event.
Event Time	the time in which a differentia event happens.
Origin Location	the definiendum's location of origin.
Quality Modifier	degree, frequency or manner modifiers that constrain a differentia quality.
Purpose	the main goal of the definiendum's existence or occurrence.
Associated Fact	a fact whose occurrence is/was linked to the definiendum's existence or occurrence.
Accessory Determiner	a determiner expression that doesn't constrain the supertype / differentia scope.
Accessory Quality	a quality that is not essential to characterize the definiendum.

Table 1: The complete set of Definition Semantic Roles (DSRs) considered in this work.

(DSRL) for the automatic annotation of natural language definitions from large dictionaries; (2) formalising the task of learning multi-relational word embeddings as a *link prediction* problem via a translational framework.

3.1 Definition Semantic Roles (DSRs)

Given a natural language definition $D = \{w_1, \dots, w_n\}$ including terms $w_1, \dots, w_n$ and semantic roles $SR = \{r_1, \dots, r_m\}$, we aim to build a DSRL that assigns one of the semantic roles in SR to each term in D. To this end, we explore the fine-tuning of different versions of BERT framing the task as a token classification problem (Devlin et al., 2019). To fine-tune the models, we adopt a publicly available dataset of ≈ 4000 definitions extracted from Wordnet, each manually annotated with the respective semantic roles[2] (Silva et al., 2016). Specifically, we annotate the definition sentences using BERT to annotate each token with a semantic role (i.e., supertype, differentia-quality, etc.). After the annotation, as we aim to learn word embeddings, we map back the tokens to the original words and use the associated semantic roles to construct multi-relational triples (Section 3.2).

[2]https://drive.google.com/drive/folders/12nJJHo7ryS6gVT-ukE-BsuHvAqPLUh3S?usp=sharing

	P	R	F1	Acc.
`bert-base-uncased`	0.76	0.80	0.78	0.86
`bert-large-uncased`	0.69	0.77	0.73	0.85
`distilbert`	0.76	0.79	0.77	0.86

Table 2: Micro-average results for the Definition Semantic Role Labeling (DSRL) task using different versions of BERT (Devlin et al., 2019).

Overall, we found `distilbert` (Sanh et al., 2019) to achieve the best trade-off between efficiency and accuracy (86%), obtaining performance comparable to `bert-base-uncased` while containing 40% less parameters. Therefore, we decided to employ `distilbert` for subsequent experiments. While more accurate DSRLs could be built via the fine-tuning of more recent Transformers, we regard this trade-off as satisfactory for current purposes.

Table 2 reports the detailed results achieved by different versions of BERT in terms of precision, recall, f1 score, and accuracy. To train the models, we adopted a k-fold cross-validation technique with $k = 5$, fine-tuning the models for 3 epochs in total via Huggingface[3].

3.2 Multi-Relational Word Embeddings

Thanks to the semantic annotation, it is possible to leverage the relational structure of natural language definitions for training word embeddings. Specifically, we rely on the semantic roles to cast the task into a *link prediction* problem. Given a set of definiendum-definition pairs, we first employ the DSRL to automatically annotate the definitions, and subsequently extract a set of subject-relation-object triples of the form (w_s, r, w_o), where w_s represents a defined term, r a semantic role, and w_o a term appearing in the definition of w_s with semantic role r. To derive the final set of triples for training, we remove the instances in which w_o represents a stop-word.

In order to train the word embeddings, the link prediction problem is formalised via a *translational objective function* $\phi(\cdot)$:

[3]https://huggingface.co/

$$\phi(w_s, r, w_o) = -d(\mathbf{e}_s^{(r)}, \mathbf{e}_o^{(r)})^2 + b_s + b_o$$
$$= -d(\mathbf{R}\mathbf{e}_s, \mathbf{e}_o + \mathbf{r})^2 + b_s + b_o, \quad (3)$$

where $d(\cdot)$ is a generic distance function, $\mathbf{e}_s$,$\mathbf{e}_o \in \mathbb{R}^d$ represent the embeddings of w_s and w_o respectively, and b_s, $b_o \in \mathbb{R}$ act as scalar biases for subject and object word. On the other hand, $\mathbf{r} \in \mathbb{R}^d$ is a translation vector encoding the semantic role r, while $\mathbf{R} \in \mathbb{R}^{d \times d}$ is a diagonal relation matrix. Therefore, the output of the objective function $\phi(w_s, r, w_o)$ is directly proportional to the similarity between $\mathbf{e}_s^{(r)}$ and $\mathbf{e}_o^{(r)}$, which represent the subject and object word embedding after applying a relation-adjusted transformation.

The choice behind the translational formulation is dictated by a set of goals and research hypotheses. First, we hypothesise that the global multi-relational structure of dictionary definitions can be optimised locally via the extracted semantic relations (i.e., making words that are semantically connected in the definitions closer in the latent space). Second, the translational formulation allows for the joint embedding of words and semantic roles. This plays a crucial function as it enables the explicit parametrisation of multiple relational structures within the same vector space (i.e., with each semantic role vector acting as a geometrical transformation), and second, it allows for the explicit use of the semantic roles after training. By preserving the embeddings of the semantic relations, in fact, we aim to make the vector space intrinsically more interpretable and controllable.

Hyperbolic Model. Following previous work on multi-relational Poincaré embeddings (Balazevic et al., 2019), we specialise the general translational objective function in Hyperbolic space:

$$\phi_{\mathbb{B}}(w_s, r, w_o) = -d_{\mathbb{B}}(\mathbf{h}_s^{(r)}, \mathbf{h}_o^{(r)})^2 + b_s + b_o$$
$$= -d_{\mathbb{B}}(\mathbf{R} \otimes_c \mathbf{h}_s, \mathbf{h}_o \oplus_c \mathbf{r})^2 + b_s + b_o, \quad (4)$$

where $d_{\mathbb{B}}(\cdot)$ is the Poincaré distance, $\mathbf{h}_s$, $\mathbf{h}_o$, $\mathbf{r} \in \mathbb{B}_c^d$ are the hyperbolic embeddings of words and semantic roles, $\mathbf{R} \in \mathbb{R}^{d \times d}$ is a diagonal relation matrix, $\oplus$ and $\otimes$ represents Mobiüs addition (Equation 2) and matrix-vector multiplication (Ganea et al., 2018):

$$\mathbf{R} \otimes_c \mathbf{h} = \exp_0^c(\mathbf{R} \log_0^c(\mathbf{h})), \quad (5)$$

with $\log(\cdot)$ and $\exp(\cdot)$ representing the logarithmic and exponential maps for projecting a point into the Euclidean tangent space and back to the Poincaré ball.

Training & Optimization. The multi-relational model is optimised for link prediction via the Bernoulli negative log-likelihood loss (details in Appendix A). We employ Riemmanian optimization to train the Hyperbolic embeddings, enriching the set of extracted triples with random negative sampling. We found that the best results are obtained with 50 negative examples for each positive instance. In line with previous work on Hyperbolic embeddings (Balazevic et al., 2019) we set $c = 1$. Following guidelines for the development of word embeddings models (Bosc and Vincent, 2018; Faruqui et al., 2016), we perform model selection on the dev-set of word relatedness and similarity benchmarks (i.e., SimVerb (Gerz et al., 2016) and MEN (Bruni et al., 2014)).

4 Empirical Evaluation

4.1 Empirical Setup

To assess the quality of the word embeddings, we performed an extensive evaluation on word similarity and relatedness benchmarks in English: SimVerb (Gerz et al., 2016), MEN (Bruni et al., 2014), SimLex-999 (Hill et al., 2015), SCWS (Huang et al., 2012), WordSim-353 (Finkelstein et al., 2001) and RG-65 (Rubenstein and Goodenough, 1965), using WordNet (Fellbaum, 2010) as the main source of definitions[4]. In particular, we leverage the glosses in WordNet to extract the semantic roles via the methodology described in Section 3.1 and train the multi-relational word embeddings. While WordNet also provides a knowledge graph of linguistic relations, our goal is to test methods that are trained and evaluated *exclusively* on natural language definitions and that can more easily generalise to different dictionaries and definitions in a broader setting.

The multi-relational word embeddings are trained on a total of $\approx 400k$ definitions from which we are able to extract ≈ 2 million triples. In order to compare Euclidean and Hyperbolic spaces we train two different versions of the model by specialising the objective function accordingly (Equation 3). We experiment with varying dimensions for both Euclidean and Hyperbolic embeddings (i.e., 40, 80, 200, and 300), training the models for a total of 300 iterations. In line with previous work (Bosc and Vincent, 2018; Faruqui et al., 2016), we evaluate the models on downstream benchmarks

[4] https://github.com/tombosc/cpae/blob/master/data/dict_wn.json

Model	Dim	FT	PT	SV-d	MEN-d	SV-t	MEN-t	SL999	SCWS	353	RG
`Glove`	300	yes	no	12.0	54.8	7.8	57.0	19.8	46.8	44.4	57.5
`Word2Vec`	300	yes	no	35.2	62.3	36.4	59.9	34.5	54.5	**61.9**	65.7
`AE`	300	yes	no	34.9	42.7	32.5	42.2	35.6	50.2	41.4	64.8
`CPAE`	300	yes	no	42.8	48.5	34.8	49.2	39.5	54.3	48.7	67.1
`CPAE-P`	300	yes	yes	44.1	**65.1**	42.3	63.8	45.8	**60.4**	61.3	72.0
`bert-base`	768	no	yes	13.5	27.8	13.3	30.6	15.1	37.8	20.0	68.1
`bert-large`	1024	no	yes	16.1	23.4	14.4	26.8	13.4	35.7	19.8	60.7
`defsent-bert`	768	yes	yes	40.0	60.2	40.0	60.0	42.0	56.8	46.6	82.4
`defsent-roberta`	768	yes	yes	43.0	55.0	44.0	52.6	47.7	54.3	44.9	80.6
`distilroberta-v1`	768	no	yes	35.8	61.2	36.7	62.2	43.4	57.1	52.0	77.4
`mpnet-base-v2`	768	no	yes	45.9	64.9	42.5	**67.5**	49.5	58.6	56.5	81.3
`sentence-t5-large`	768	no	yes	**49.4**	63.1	50.2	66.3	57.3	56.1	51.8	**85.3**
Multi-Relational											
`Euclidean`	40	yes	no	39.1	62.9	35.7	65.4	36.3	58.2	52.1	80.9
`Euclidean`	80	yes	no	44.1	65.6	39.5	66.2	41.2	58.4	55.8	78.0
`Euclidean`	200	yes	no	47.3	67.0	41.0	67.6	43.4	60.6	55.4	78.1
`Euclidean`	300	yes	no	47.9	68.3	43.1	69.1	**44.7**	61.0	54.4	79.0
`Hyperbolic`	40	yes	no	36.7	66.2	34.3	66.4	31.8	57.7	49.9	75.5
`Hyperbolic`	80	yes	no	42.7	68.2	40.7	68.6	38.3	60.5	57.3	81.0
`Hyperbolic`	200	yes	no	48.8	71.9	44.7	73.2	40.7	62.5	62.5	**81.6**
`Hyperbolic`	300	yes	no	**50.6**	**72.6**	**45.4**	**74.2**	42.3	**63.0**	**63.3**	80.5

Table 3: Results on word similarity and relatedness benchmarks (Spearman's correlation). The column **FT** indicates whether the model is explicitly fine-tuned on natural language definitions, while **PT** indicates the adoption of a pre-training phase on external corpora.

Model	SV	MEN	SL999	353	RG
`Glove`	18.9	-	32.1	62.1	75.8
`Word2Vec`	-	72.2	28.3	**68.4**	-
`Our`	**45.4**	**74.2**	**42.3**	63.3	**80.5**

Table 4: Comparison with Hyperbolic word embeddings in the literature. The results for Glove and Word2Vec are taken from (Tifrea et al., 2018) and (Leimeister and Wilson, 2018) considering their best model.

comparing the predicted similarity between the pair of words to the ground truth via a Spearman's correlation coefficient.

4.2 Baselines

We evaluate a range of word embedding models on the same set of definitions (Bosc and Vincent, 2018). Specifically, we compare the proposed multi-relational embeddings against different paradigms adopted in previous work and state-of-the-art approaches. Here, we provide a characterisation of the models adopted for evaluation:

Distributional. We compare the multi-relational approach against distributional word embeddings (Mikolov et al., 2013; Pennington et al., 2014). Both `Glove` and `Word2Vec` have the same dimensionality as the multi-relational approach but are not designed to leverage or preserve explicit semantic relations during training.

Autoencoders. This paradigm employs encoder-decoder architectures to learn word representations from natural language definitions. In particular, we compare our approach to an autoencoder-based model specialised for natural language definitions known as CPAE (Bosc and Vincent, 2018), which adopts LSTMs paired with a consistency penalty. Differently from our approach, CPAE requires initialisation with pre-trained word vectors to achieve the best results (i.e., CPAE-P).

Sentence-Transformers. Finally, we compare our model against Sentence-Transformers (Reimers and Gurevych, 2019). Here, we use Sentence-Transformers to derive embeddings for the target definienda via the encoding of the corresponding definition sentences in the corpus. As the main function of definitions is to describe the meaning of words, semantically similar words tend to possess similar definitions; therefore we expect Sentence-Transformers to organise the latent space in a semantically coherent manner when using definition sentences as a proxy for the word embeddings. We experiment with a diverse set of models ranging from BERT (Devlin et al., 2019) to the current state-of-the-art on semantic similarity benchmarks[5] (Ni et al., 2022; Song et al., 2020; Liu et al., 2019) and models trained directly on definition sentences (e.g., Defsent (Tsukagoshi et al., 2021)). While the evaluated Transformers do not require fine-tuning on the word similarity benchmarks, they are employed after being extensively pre-trained on large corpora and specialised in sentence-level semantic tasks. Moreover, the overall size of the resulting embeddings is significantly larger than the proposed multi-relational approach.

4.3 Word Embeddings Benchmarks

In this section, we discuss and analyse the quantitative results obtained on the word similarity and relatedness benchmarks (see Table 3).

Firstly, an internal comparison between Euclidean and Hyperbolic embeddings supports the central hypothesis that Hyperbolic manifolds are particularly suitable for encoding the recursive and hierarchical structure of definitions. As the dimensions of the embeddings increase, the quantitative analysis demonstrates that the Hyperbolic model can achieve the best performance on the majority of the benchmarks.

When compared to the distributional baselines, the multi-relational Hyperbolic embeddings clearly outperform both `Glove` and `Word2Vec` trained on the same set of definitions. Similar results can be observed when considering the autoencoder paradigm (apart from CPAE-P on SL999). Since the size of the embeddings produced by the models is comparable (i.e., 300 dimensions), we attribute the observed results to the encoded semantic relations, which might play a crucial role in imposing structural constraints during training.

Finally, the multi-relational model produces embeddings that are competitive with state-of-the-art Transformers. While the Hyperbolic approach can clearly outperform BERT on all the downstream tasks, we observe that Sentence-Transformers become increasingly more competitive when considering larger models that are fine-tuned on semantic similarity tasks and definitions (e.g., `sentence-t5-large` (Ni et al., 2022) and `defsent` (Tsukagoshi et al., 2021)). However, it is important to notice that the multi-relational embeddings not only require a small fraction of the Transformers' computational cost – e.g, T5-large (Raffel et al., 2020) is pre-trained on the C4 corpus ($\approx$ 750GB) while the multi-relational embeddings are only trained on WordNet glosses ($\approx$ 19MB), a difference of 4 orders of magnitude – but are also intrinsically more interpretable thanks to the explicit encoding of the semantic relations (see Section 4.5 and 5).

[5] `https://www.sbert.net/docs/pretrained_models.html`

4.4 Hyperbolic Word Embeddings

In addition to the previous analysis, we performed a comparison with existing Hyperbolic word embeddings in the literature (Table 4). In particular, we compare the proposed multi-relational model with Poincare Glove (Tifrea et al., 2018) and Hyperbolic Word2Vec (Leimeister and Wilson, 2018). The results show that our approach can outperform both models on the majority of the benchmarks, remarking the impact of the multi-relational approach and definitional model on the quality of the representation.

4.5 Multi-Relational Representation

To contrast the capacity of different geometric spaces to learn multi-relational representations, we design an additional experiment that tests the ability to encode out-of-vocabulary definienda (i.e., words never seen during training). In particular, we aim to quantitatively measure the precision in encoding the semantic relations by approximating new word embeddings in one-shot, and use it as a proxy for assessing the structural organisation of Euclidean and Hyperbolic spaces. Our hypothesis is that a vector space organised according to the multi-relational structure induced by the definitions should allow for a more precise approximation of out-of-vocabulary word embeddings via relation-specific transformations.

In order to perform this experiment, we adopt the dev-set of SimVerb (Gerz et al., 2016) and MEN (Bruni et al., 2014), removing all the triples from our training set that contain a subject or an object word occurring in the benchmarks. Subsequently, we employ the pruned training set to re-train the models. After training, we derive the embeddings of the out-of-vocabulary words via geometric transformations applied to the in-vocabulary words. Specifically, given a target word (e.g., "dog") and its definition (e.g., "a domesticated carnivorous mammal that typically has a long snout") we jointly use the in-vocabulary definiens and their semantic relations (e.g., ["carnivorous",

Model	Dimension	Mean-Pooling		Multi-Relational		Differentia Quality		Supertype	
		SV	MEN	SV	MEN	SV	MEN	SV	MEN
`Euclidean`	40	**17.6**	**20.6**	23.7 (+6.1)	**31.7** (+11.1)	22.6	**26.0**	17.2	19.2
`Euclidean`	80	15.9	18.1	**24.6** (+8.7)	29.4 (+11.3)	23.4	23.3	18.4	18.8
`Euclidean`	200	14.5	18.4	23.7 **(+9.2)**	30.7 **(+12.3)**	**24.1**	22.2	18.7	19.1
`Euclidean`	300	15.1	18.8	24.3 **(+9.2)**	30.3 (+11.5)	23.8	22.7	**19.3**	**20.3**
`Hyperbolic`	40	15.9	22.8	25.4 (+9.5)	35.2 (+12.4)	22.7	25.5	14.0	20.2
`Hyperbolic`	80	17.9	25.1	27.7 **(+9.8)**	37.8 (+12.7)	25.9	**26.6**	15.4	20.1
`Hyperbolic`	200	19.2	24.9	28.4 (+9.2)	38.2 (+13.3)	27.9	25.5	17.3	**21.3**
`Hyperbolic`	300	**19.6**	**25.1**	**28.6** (+9.0)	**39.7 (+14.6)**	**28.5**	26.0	**18.1**	20.4

Table 5: Results on the one-shot approximation of out-of-vocabulary word embeddings. The numbers in the table represent the Spearman correlation computed over the out-of-vocabulary set after the approximation. (Left) impact of the multi-relational embeddings on the one-shot encoding of out-of-vocabulary words. (Right) ablations using the two most common semantic roles for one-shot approximation. The results demonstrate the superior capacity of the multi-relational Hyperbolic embeddings to capture the global semantic structure of definitions.

"supertype"], ["snout", "differentia-quality"]) to approximate a new word embedding e() for the definiendum via mean pooling and translation (i.e., e("dog") = mean(e("carnivorous"), ("snout")) + mean(e("supertype"), e("differentia-quality"))) and compare against a mean pooling baseline that does not have access to the semantic relations (i.e. e("dog") = mean(e("carnivorous"), ("snout"))).

The results reported in Table 5 demonstrate the impact of the multi-relational framework, also confirming the property of the Hyperbolic embeddings in better encoding the global semantic structure of natural language definitions.

5 Qualitative Analysis

In addition to the qualitative evaluation, we perform a qualitative analysis of the embeddings. This is performed in two different ways: traversal of the latent space and relation-adjusted transformations.

5.1 Latent Space Traversal

We perform traversal experiments to visualise the organisation of the latent space. This is done by sampling points at fixed intervals along the arc (i.e., geodesic) connecting the embeddings of a pair of predefined words (seeds), i.e., by interpolating along the shortest path between two embeddings. The choice of word pairs was done according to a group of semantic categories for which intermediate concepts can be understood to be semantically in between the pair. For example: $(car, bicycle) \to motorcycle$. Considering the latent space structure that should result from the proposed approach, we expect the traversal process to capture such intermediate concepts, while generalising the concepts towards the midpoint of the arc. In a latent space organised according to the semantic structure and concept hierarchy of definitions, in fact, we expect the midpoint to be close to concepts relating to both seed words.

The categories, sampled words and results for the midpoint of the arcs can be found in Table 6 (top). From the traversal analysis, we can observe that the intermediate concepts are indeed captured for all the categories, with a noticeable degree of generalisation in the Hyperbolic models. This indicates the consistent interpretable nature of the navigation for the latent space, and enables more robust semantic control, setting the desired embedded concept in terms of a symbolic conjunction of its vicinity. We can also observe that, the space between the pair of embeddings is populated mostly by concepts *related* to both entities of the pair in the Euclidean models, while being populated by concepts *relating* both entities in the Hyperbolic models.

5.2 Relation-Adjusted Transformations

We analyse the organisation of the latent space before and after the application of a translational operation. As discussed in Section 2.1, such operation should transform the embedding space according to the corresponding semantic role. For example, the operation $\phi_{\mathbb{B}}(dog, supertype, w_o)$ should cluster the space around the taxonomical branch related to *"dog"*. It is important to notice that this operation does not correspond to link prediction as we are not considering the scalar biases b_s, b_o. The goal here is to disentangle the impact of the semantic transformations on the latent space. We consider the `supertype` role for this analysis as it induces a global hierarchical structure that is eas-

Category	Word Pair	Euclidean	Hyperbolic
Concrete concepts	car - bycicle	bicycle, car, pedal_driven, motorcycle, banked, multiplying, swiveling, four_wheel, rented, no_parking	railcar, bicycle, car, pedal_driven, driving_axle, motorized_wheelchair, tricycle, bike, banked, live_axle
Gender, Role	man - woman	woman, man, procreation, men, non_jewish, three_cornered, middle_aged, bodice, boskop, soloensis	adulterer, boyfriend, ex-boyfriend, adult_female, manful, cuckold, virile, stateswoman, womanlike, wardress
Animal Hybrids	horse - donkey	donkey, horse, burro, hock_joint, neighing, dog_sized, tapirs, feathered_legged, racehorse, gasterophilidae	burro, cow_pony, unbridle, hackney, unbridled, equitation, sidesaddle, palfrey, roughrider, trotter
Process, Time	birth - death	death, birth, lifetime, childless, childhood, adityas, parturition, condemned, carta, liveborn	lifespan, life-time, firstborn, multiparous, full_term, teens, nonpregnant, childless, widowhood, gestational
Location	sea - land	land, sea, enderby, weddell, arafura, littoral, tyrrhenian, andaman, maud, toads	tellurian, litoral, seabed, high_sea, body_of_water, littoral_zone, international_waters, benthic_division, naval_forces, lake_michigan

w_s	**No Transformation:** $-d_{\mathbb{B}}(\mathbf{h}_s, \mathbf{h}_o)^2$	**Relation-Adjusted** (r = `supertype`)**:** $-d_{\mathbb{B}}(R \otimes \mathbf{h}_s, \mathbf{h}_o \oplus \mathbf{r})^2$
dog	dog, heavy_coated, smooth_coated, malamute, canidae, wolves, light_footed, long-established, whippet, greyhound	huntsman, hunting_dog, sledge_dog, coondog, sled_dog, working_dog, russian_wolfhound, guard_dog, tibetan_mastiff, housedog
car	car, railcar, telpherage, telferage, subcompact, cable_car, car_transporter, re_start, auto, railroad_car, driving_axle	railcar, marksman, subcompact, smoking_carriage, handcar, electric_automobile, limousine, taxicab, freight_car , slip_coach
star	star, armillary_sphere, charles's_wain, starlight, altair, drummer, northern_cross, photosphere, sterope, rigel	rigel, betelgeuse, film_star, movie_star, television_star, tv_star, starlight, supergiant, photosphere, starlet
king	louis_i, sultan, sir_gawain. uriah, camelot, dethrone, poitiers, excalibur, empress, divorcee	chessman, gustavus_vi, grandchild, alfred_the_great, jr, rajah, knights, louis_the_far, egbert, plantagenet, st._olav

Table 6: (Top) qualitative results for the latent space traversal, with midpoint nearest neighbours listed in descending order. (Bottom) nearest neighbours of seed words before and after applying a supertype-adjusted transformation.

ily inspectable. The results can be found in Table 6 (bottom). We observe that the transformation leads to a projection locus near all the closely defined terms (the types of dogs or stars), abstracting the subject words in terms of their conceptual extension (things that are dogs / stars). This displays a particular way of generalisation that is likely related to the arrangement of the roles and how they connect the concepts.

6 Related Work

Considering the basic characteristics of natural language definitions here discussed, efforts to leverage dictionary definitions for distributional models were proposed as a more efficient alternative to the large unlabeled corpora, following the rising popularity of the latter (Tsukagoshi et al., 2021; Hill et al., 2016; Tissier et al., 2017; Bosc and Vincent, 2018). Simultaneously, efforts to improve compositionality (Chen et al., 2015; Scheepers et al., 2018) and interpretability (de Carvalho and Le Nguyen, 2017; Silva et al., 2019) of word representations led to different approaches towards the incorporation of definition resources to language modelling, with the idea of modelling definitions becoming an established task (Noraset et al., 2017).

More recently, research focus has shifted towards the fine-tuning of large language models and contextual embeddings for definition generation and classification (Gadetsky et al., 2018; Bosc and Vincent, 2018; Loureiro and Jorge, 2019; Mickus et al., 2022), with interest in the structural properties of definitions also gaining attention (Shu et al., 2020; Wang and Zaki, 2022).

Finally, research on Hyperbolic representation spaces has provided evidence of improvements in capturing hierarchical linguistic features, over traditional (Euclidean) ones (Balazevic et al., 2019; Nickel and Kiela, 2017; Tifrea et al., 2018; Zhao et al., 2020). This work builds upon the aforementioned developments, and proposes a novel approach to the incorporation of structural information extracted from natural language definitions by means of a translational objective guided by explicit semantic roles (Silva et al., 2016), combined with a Hyperbolic representation able to embed multi-relational structures.

7 Conclusion

This paper explored the semantic structure of *definitions* as a means to support novel learning paradigms able to preserve semantic interpretability and control. We proposed a *multi-relational* framework that can explicitly map terms and their corresponding *semantic relations* into a vector space. By automatically extracting the relations from external dictionaries, and specialising the framework in *Hyperbolic space*, we demonstrated that it is possible to capture the hierarchical and multi-relational structure induced by dictionary definitions while preserving, at the same time, the explicit mapping required for controllable semantic navigation.

8 Limitations

While the study here presented supports its findings with all the evidence compiled to the best of our knowledge, there are factors that limit the scope of the current state of the work, from which we understand as the most important:

1. The automatic semantic role labeling process is not 100% accurate, and thus is a limiting factor in analysing the impact of this information on the models. While we do not explore DSRLs with varying accuracy, future work can explicitly investigate the impact of the automatic annotation on the robustness of the multi-relational embeddings.

2. The embeddings obtained in this work are contextualizable (by means of a relation-adjusted transformation), but are not *contextualized*, i.e., they are not dependent on surrounding text. Therefore, they are not comparable on tasks dependant on contextualised embeddings.

3. The current version of the embeddings coalesces all senses of a definiendum into a single representation. This is a general limitation of models learning embeddings from dictionaries. Fixing this limitation is possible in future work, but it will require the non-trivial ability to disambiguate the terms appearing in the definitions (i.e., definiens).

4. The multi-relational embeddings presented in the paper were initialised from scratch in order to test their efficiency in capturing the semantic structure of dictionary definitions. Therefore, there is an open question regarding the possible benefits of initialising the models with pre-trained distributional embeddings such as `Word2Vec` and `Glove`.

Acknowledgements

This work was partially funded by the Swiss National Science Foundation (SNSF) project NeuMath (200021_204617), by the EPSRC grant EP/T026995/1 entitled "EnnCore: End-to-End Conceptual Guarding of Neural Architectures" under Security for all in an AI enabled society, by the CRUK National Biomarker Centre, and supported by the Manchester Experimental Cancer Medicine Centre and the NIHR Manchester Biomedical Research Centre.

References

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. Multi-relational poincaré graph embeddings. *Advances in Neural Information Processing Systems*, 32.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Tom Bosc and Pascal Vincent. 2018. Auto-encoding dictionary definitions into consistent word embeddings. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1522–1532.

Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *Journal of artificial intelligence research*, 49:1–47.

Danilo S Carvalho, Giangiacomo Mercatali, Yingji Zhang, and Andre Freitas. 2023. Learning disentangled representations for natural language definitions. *Findings of the European chapter of Association for Computational Linguistics (Findings of EACL)*.

Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2015. Improving distributed representation of word sense via wordnet gloss composition and context clustering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 15–20.

Danilo Silva de Carvalho and Minh Le Nguyen. 2017. Building lexical vector representations from concept definitions. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 905–915.

Jacob Devlin, 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)*, pages 4171–4186.

Manaal Faruqui, Yulia Tsvetkov, Pushpendre Rastogi, and Chris Dyer. 2016. Problems with evaluation of word embeddings using word similarity tasks. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, pages 30–35.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Jun Feng, Minlie Huang, Mingdong Wang, Mantong Zhou, Yu Hao, and Xiaoyan Zhu. 2016. Knowledge graph embedding by flexible translation. In *Fifteenth*

International Conference on the Principles of Knowledge Representation and Reasoning.

Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2001. Placing search in context: The concept revisited. In *Proceedings of the 10th international conference on World Wide Web*, pages 406–414.

Artyom Gadetsky, Ilya Yakubovskiy, and Dmitry Vetrov. 2018. Conditional generators of words definitions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 266–271.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic neural networks. *Advances in neural information processing systems*, 31.

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen. 2016. Simverb-3500: A large-scale evaluation set of verb similarity. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2173–2182.

Felix Hill, Kyunghyun Cho, Anna Korhonen, and Yoshua Bengio. 2016. Learning to understand phrases by embedding the dictionary. *Transactions of the Association for Computational Linguistics*, 4:17–30.

Felix Hill, Roi Reichart, and Anna Korhonen. 2015. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.

Eric H Huang, Richard Socher, Christopher D Manning, and Andrew Y Ng. 2012. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 873–882.

Matthias Leimeister and Benjamin J Wilson. 2018. Skip-gram word embeddings in hyperbolic space. *arXiv preprint arXiv:1809.01498*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Daniel Loureiro and Alipio Jorge. 2019. Language modelling makes sense: Propagating representations through wordnet for full-coverage word sense disambiguation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5682–5691.

Timothee Mickus, Kees van Deemter, Mathieu Constant, and Denis Paperno. 2022. Semeval-2022 task 1: Codwoe–comparing dictionaries and word embeddings. *arXiv preprint arXiv:2205.13858*.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. 2022. Sentence-t5: Scalable sentence encoders from pretrained text-to-text models. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1864–1874.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30.

Thanapon Noraset, Chen Liang, Larry Birnbaum, and Doug Downey. 2017. Definition modeling: Learning to define word embeddings in natural language. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

Herbert Rubenstein and John B Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Thijs Scheepers, Evangelos Kanoulas, and Efstratios Gavves. 2018. Improving word embedding compositionality using lexicographic definitions. In *Proceedings of the 2018 World Wide Web Conference*, pages 1083–1093.

Xiaobo Shu, Bowen Yu, Zhenyu Zhang, and Tingwen Liu. 2020. Drg2vec: Learning word representations from definition relational graph. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE.

Vivian Silva, André Freitas, and Siegfried Handschuh. 2018a. Building a knowledge graph from natural language definitions for interpretable text entailment recognition. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Vivian Silva, Siegfried Handschuh, and André Freitas. 2016. Categorization of semantic roles for dictionary definitions. In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pages 176–184.

Vivian S Silva, André Freitas, and Siegfried Handschuh. 2019. Exploring knowledge graphs in an interpretable composite approach for text entailment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7023–7030.

Vivian S Silva, Siegfried Handschuh, and André Freitas. 2018b. Recognizing and justifying text entailment through distributional navigation on definition graphs. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33:16857–16867.

Alexandru Tifrea, Gary Becigneul, and Octavian-Eugen Ganea. 2018. Poincare glove: Hyperbolic word embeddings. In *International Conference on Learning Representations*.

Julien Tissier, Christophe Gravier, and Amaury Habrard. 2017. Dict2vec: Learning word embeddings using lexical dictionaries. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 254–263.

Hayato Tsukagoshi, Ryohei Sasano, and Koichi Takeda. 2021. Defsent: Sentence embeddings using definition sentences. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 411–418.

Abraham A Ungar. 2001. Hyperbolic trigonometry and its application in the poincaré ball model of hyperbolic geometry. *Computers & Mathematics with Applications*, 41(1-2):135–147.

Qitong Wang and Mohammed J Zaki. 2022. Hg2vec: Improved word embeddings from dictionary and thesaurus based heterogeneous graph. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3154–3163.

Torsten Zesch, Christof Müller, and Iryna Gurevych. 2008. Using wiktionary for computing semantic relatedness. In *AAAI*, volume 8, pages 861–866.

Wenyu Zhao, Dong Zhou, Lin Li, and Jinjun Chen. 2020. Manifold learning-based word representation refinement incorporating global and local information. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3401–3412.

A Multi-Relational Embeddings

The multi-relational embeddings are trained on a total of $\approx 400k$ definitions from which we are able to extract ≈ 2 million triples. We experiment with varying dimensions for both Euclidean and Hyperbolic embeddings (i.e., 40, 80, 200, and 300), training the models for a total of 300 iterations with batch size 128 and learning rate 50 on 16GB Nvidia Tesla V100 GPU. The multi-relational models are optimised via a Bernoulli negative log-likelihood loss:

$$\begin{aligned} L(y,p) = -1\frac{1}{N}\sum_{i=1}^{N}(y^{(i)}log(p^{(i)}) \\ +(1-y^{(i)})log(1-p^{(i)})) \end{aligned} \tag{6}$$

where $p^{(i)}$ represents the predictions made by the model and $y^{(i)}$ represents the actual label.