FrameNet Semantic Role Classification by Analogy

Van-Duy Ngo[1,†], **Stergos Afantenos**[1,†], **Emiliano Lorini**[1], **Miguel Couceiro**[2]
[1]IRIT, CNRS, INP, Université de Toulouse
[2]INESC-ID, IST, Universidade de Lisboa
{van-duy.ngo, emiliano.lorini, stergos.afantenos}@irit.fr
miguel.j.couceiro@tecnico.ulisboa.pt

Abstract

In this paper, we adopt a relational view of analogies applied to Semantic Role Classification in FrameNet. We define analogies as formal relations over the Cartesian product of frame evoking lexical units (LUs) and frame element (FEs) pairs, which we use to construct a new dataset. Each element of this binary relation is labelled as a valid analogical instance if the frame elements share the same semantic role, or as invalid otherwise. This formulation allows us to transform Semantic Role Classification into binary classification and train a lightweight Artificial Neural Network (ANN) that exhibits rapid convergence with minimal parameters. Unconventionally, no Semantic Role information is introduced to the neural network during training. We recover semantic roles during inference by computing probability distributions over candidates of all semantic roles within a given frame through random sampling and analogical transfer. This approach allows us to surpass previous state-of-the-art results while maintaining computational efficiency and frugality.

Keywords: semantic role classification, FrameNet, analogy, analogical transfer

1. Introduction

Analogical reasoning is a human faculty that is "at the core of cognition" (Hofstadter, 2001) and although this might not be an opinion that is widely shared in the research community at large, it is undeniable that recent advances regarding Large Language Models (LLMs) have sparked a new interest in the research community for analogies, including the field of Natural Language Processing (NLP). More recently, Firt (2025) argues that reaching what has come to be known as Artificial General Intelligence (AGI) will only become possible if analogies serve as a fundamentally basic component of such systems. Despite the diachronic interest in analogies, it seems to be the case that by the term "analogies", different disciplines refer to varying viewpoints or even interpretations, since this term is used in fields ranging from cognitive science (Gentner, 1983; Hofstadter, 1984; Holyoak and Thagard, 1995) and computational linguistics (Mikolov et al., 2013b; Drozd et al., 2016) to formal logic (Prade and Richard, 2009) and machine learning (Bounhas and Prade, 2023; Cunha et al., 2026).

Most approaches in NLP adopt the proportional view of analogies (Aristotle, 2009) of form $a : b :: c : d$, which reads "a is to b as c is to d", with a, b, c, d being abstract entities (*e.g.* boolean variables, vectors, etc) and $a : b :: c : d$ representing a valid analogy if the same transformations (usually taking the form of a difference) that happen from a to b also hold from c to d (Barbot et al., 2019; Prade and Richard, 2017). From a cognitive science perspective, Gentner (1983) establishes the Structured Mapping Theory (SMT), which emphasises the rather fundamental role of structural relationships among components than their attributes in forming relational analogies. Hofstadter and FARG (1995); Mitchell (1993); Chalmers et al. (1992); Hofstadter and Sander (2013) argue that analogies is a ubiquitous and fluid mechanism involved in the formation of concepts. As such, it can be highly dependent on the context, in the sense that an object a can be considered analogous to object b only under a specific context.

In NLP, most researchers adopt the proportional point of view of analogies (Mikolov et al., 2013a,b; Drozd et al., 2016) and mainly focus on analogies at lexical levels, such as the lexicons in the well-cited $\texttt{king} : \texttt{man} :: \texttt{queen} : \texttt{woman}$ example of Mikolov et al. (2013b). Typically, words in the quadruplet are represented by either static or contextual vector embeddings, and the quadruplet's validity is realised by arithmetic or geometric differences. The relation that holds between pairs (a, b) and (c, d) is usually categorized into either inflectional, derivational, lexicographic, or encyclopedic relations (Drozd et al., 2016). Proportional views of analogies have also been applied for morphosyntactic patterns such as $\texttt{run} : \texttt{running} :: \texttt{drive} : \texttt{driving}$ (Lepage, 2005; Ulčar et al., 2020; Karpinska et al., 2018; Marquer and Couceiro, 2025). For languages that do not exhibit complex morphosyntactic patterns,[1] such a task is rather trivial (Ushio et al., 2021; Petersen and van der Plas, 2023).

Despite its usefulness for certain tasks, propor-

Authors with † are the main contributors.

[1]Some non-agglutinative languages, for example.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 3633–3644
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

tional approaches of analogies in Natural Language assume that the same common latent relation always holds between the two pairs of words that constitute a valid analogy, independently of anything else. This misses a crucial aspect regarding the pragmatics of Natural Language namely, the fact that such latent relations can dramatically shift depending on the *context* in which the lexical items of the quadruplet are found in.[2] Take for example the observation from Mickus et al. on the fact that the pairs `flower : petal` and `tonne : kilogram` both appear in the L06 meronym category of the BATS corpus (Drozd et al., 2016) thus forming an analogy, despite that the in-context validity of such analogy is very much debatable.

In this paper, we do not strictly subscribe to the proportional approach to analogies, but instead get inspired by the aforementioned theories in Cognitive Science. We propose that analogical validity in NLP—whether a quadruplet of textual elements[3] forms a valid analogy—depends vastly on the latent semantic relations established between those elements, which are themselves heavily influenced by their contextual environments. This highlights the fundamental principle that semantic interpretation is irrefutably shaped by pragmatic factors.

To demonstrate our point, consider the following quadruplet: `flower : petal :: tree : leaf`. On a lexical level, the meronymous relations are apparent hence the positive validity. Let us now place those four words in a context:

(1) Decomposed leaves provide a fertilizer for the trees.

(2) Infected petals imperil the life-span of the flowers.

In this context, the holonym-meronym relations between `flower : petal` and `tree : leaf` become much less relevant. Instead, the semantics of the two sentences play a more prominent role in judging whether an analogy holds for this quadruplet. In this case, we are inclined to conclude that the lexical analogy no longer holds, since in (1) the decomposed leaves are beneficial for the trees while in (2) infected petals are harmful for the flowers, completely eclipsing the holonym-meronym relations. This is well reflected on the frame semantics of the two sentences, using FrameNet (Baker et al., 1998; Baker, 2017) annotations:

(3) *Frame: Supply*

Decomposed leaves	provide	a fertilizer	for	the trees
SUPPLIER	predicate	THEME		RECIPIENT

(4) *Frame: Endangering*

Infected petals	imperil	the life-span of the flowers
CAUSE	predicate	VALUED ENTITY

Suppose now that we have the following sentence annotated with FrameNet annotations:

(5) *Frame: Supply*

Blossoming petals	supply	precious pollinators	to	flowers
SUPPLIER	predicate	THEME		RECIPIENT

Although no analogy can be formed between the elements of sentences (3) and (4), we can discern an analogy on the semantic level between (3) and (5) in the sense that both `leaves` and `petals` are meant to *supply* something to `trees` and `flowers` respectively. This is reinforced by the fact that the focused predicates trigger the same frame *Supply* and the frame-role relations within the two predicate-frame element pairs remain identical.

Let us now consider the following quadruplet: `leaves : trees :: lentils : humans`. No obvious common relation, of meronymy or otherwise, can be discerned between the two pairs. Without further context, we cannot meaningfully claim that this quadruplet forms a valid analogy since we cannot say that "*leaves* are to *trees* as *lentils* are to *humans*". However, when we consider the following sentence:

(6) *Frame: Supply*

Lentils	furnish	proteins	to	humans
SUPPLIER	predicate	THEME		RECIPIENT

there is obviously a semantic analogy between (3) and (6).

Inspired by cognitive theories of analogy, particularly the approach put forth by Hofstadter and Sander (2013) as well as the Structure Mapping Theory (Gentner, 1983; Gentner and Hoyos, 2017), we formulate analogies based on the semantic roles that textual elements hold in relation to predicate words that trigger their containing frames. For instance, in sentences (3) and (6), we identify the following quadruplets $a : b :: c : d$ as valid analogies:

`provide : decomposed leaves :: furnish : lentils`
`provide : a fertilizer :: furnish : proteins`
`provide : the trees :: furnish : humans`

As Petersen et al. (2025) observe in their review of analogical reasoning at the intersection of NLP and cognitive science, this directly aligns with the *mapping* process in SMT: elements b and d playing the same semantic role within the common semantic frame triggered by *anchor* elements a and c constitute a valid analogy, while all other combinations with semantic role mismatch do not.

[2]This remark concerns primarily the construction of datasets, prior to any computational attempt in recognising analogies. It has to be noted that most extant approaches use one form or another of distributional semantics (Firth, 1968; Harris, 1954) and thus implicit analogy recognition relies on the context.

[3]Expressed as a couple of pairs, in our case.

Constructing pairs of predicate-arguments as described above allows us to build a novel dataset of positive and negative instances based on FrameNet annotations, from which we can train a model to learn the validity of such analogies. We subsequently use the model to identify pairs whose frame element's semantic role is analogous to the input, which corresponds to the *retrieval* process in SMT. In our case, the method is as follows: given a sentence with spans representing frame elements to be classified, we sample random annotated sentences corresponding to the same semantic frame and build quadruplets that we submit to a trained analogical model, assigning to each span probability distributions over all samples of semantic roles of the frame, then grant the frame element the semantic role with the highest probability.

The structure of the paper is as follows: in §2 we briefly introduce FrameNet and Semantic Role Classification, while in §3 we present in more detail how we construct analogical quadruplets from FrameNet annotations, which corresponds to the *mapping* phase of SMT, in order to create a training set with positive and negative instances for training neural network models. In the same section, we describe how this model can be used during inference for Semantic Role Classification. §4 presents the neural network architecture , the experimental settings, and the results obtained. Related work is briefed in §5. After discussing various aspects of this work and the prospect of applying analogy to diverse tasks in §6, we conclude.

2. FrameNet and Semantic Role Classification

The Berkeley FrameNet project (Baker et al., 1998; Baker, 2017; Ruppenhofer et al., 2006) is based on the theory of Frame Semantics established by Fillmore (1976). The main idea behind FrameNet is that the realisation of word's meaning requires a semantic frame that is evoked by a trigger. A semantic frame consists of a predicate—usually a verb, noun, or adjective—that triggers the frame, along with several semantic roles that serve as arguments to the frame. Semantic roles are either *core* that are almost always present in a sentence[4], *peripheral* whose instantiation is not mandatory, or *extra-thematic* that capture information related to time, location, manner, means, degree, etc. While the latter are shared among many frames, core and peripheral roles are frame-specific. Semantic roles share a common ontological structure across all frame types. Figure 1 shows an example of an annotated sentence with frame triggered by the predicate `decline`.

[Speaker Fenn] *declined*$_{\text{Trigger}}$ [Proposed_action the offer to buy] [Co-timed_event with a bemused wave of his hand]. [Interlocutor DNI]

Figure 1: An example of an annotated sentence for the frame type triggered by `decline.v`.

Frame Semantic Parsing (Gildea and Jurafsky, 2002; Baker et al., 2007) is the process that aims to extract full frame semantic structures starting from a sentence. The full process can be divided into three sub-tasks (Zheng et al., 2022). **Frame Identification**: given a sentence, identify all the frames and their respective evoking predicates. **Argument Identification:** given a sentence, a frame, and the evoking predicate, identify all the elements in the sentence that represent arguments for that predicate. **Semantic Role Classification:** This subtask takes as input the results of the previous two to assign the elements with corresponding semantic roles. On FrameNet, the task is termed FrameNet Parsing (Palmer and Sporleder, 2010). Evaluation is empirically performed either for each separate module or collectively on the entire FrameNet Parsing pipeline. The prevalent former usually presents the results of a single sub-task on gold input rather than the output of the preceding module, posing compatibility problems for integration. Integral pipelines for FrameNet Parsing—with modular components for each sub-task—emerged as the optimal solution for these problems (Lin et al., 2021; Chanin, 2023; Swayamdipta et al., 2017).

3. Learning Analogies for Semantic Role Classification

Our methodology to solving FrameNet Semantic Role Classification is inspired by cognitive science approaches to analogies, such as per Hofstadter and Sander (2013), in which the context dictates the validity of an analogy. Our analysis roots for the use of predicates to encompass the context of the sentences, relying on (i) the theoretically sound activation of the frame by the predicate, and (ii) the frame-specific relations that come alongside the evoked frame. This is advantageous because the same encoder can generate latent representations for both the context and the target frame element, thereby ensuring consistency and preventing representation mismatches.

By this principle, we build a dataset of analogical instances from FrameNet annotations and train a model for binary classification. During training, the model receives no semantic role label information.

[4] Niche grammatical constructions do not allow instantiation of some core elements. In such cases a NI (Not Instantiate) label is provided. Non-instantiations are divided into Definite (DNI) Indefinite (DNI) and Constructional (CNI).

For inference, we construct a reference dataset of predicate-argument pairs by randomly selecting annotated instances that meet the frame constraint. Specifically, for every *target* frame element $\mathtt{e}^{\mathtt{t}}$ to classify with its known predicate $\mathtt{p}^{\mathtt{t}}$, we randomly sample n_π pairs of *source* frame elements and associated predicates $\mathtt{p}_\mathtt{i}^\mathtt{s} : \mathtt{e}_\mathtt{i}^\mathtt{s}$ of the same frame type to build n_π analogical instances of $\mathtt{p}_\mathtt{i}^\mathtt{s} : \mathtt{e}_\mathtt{i}^\mathtt{s} :: \mathtt{p}^\mathtt{t} : \mathtt{e}^\mathtt{t}, 0 < i \leq n_\pi$, then submit them to the trained model for classification. Instances classified as positive are kept and the semantic role type of the source frame element is transferred to the target's.

To summarise, we divide our approach into three phases: **a)** creating a dataset of positive and negative instances representing analogies, **b)** training a neural network model for binary classification, and **c)** decoding of this model's output for multi-class classification via analogical transfer, based on obtained probability distributions over all semantic role types within a given frame.

3.1. Dataset Creation and Learning

In order to train our binary analogical model, it is necessary to construct a dataset that contains positive and negative instances of analogies. Given a set S of sentences annotated with Frame types and Semantic Roles according to FrameNet, we consider each sentence $s \in S$ as a sequence of tokens $s = \{\mathtt{t}_i\}_{i=1}^{l}$ with l being the length of the sentence. For each sentence, along with the frame and the evoking predicate $\mathtt{p}$, we extract the frame elements $E = \{\mathtt{e}_\mathtt{j}\}_{j=1}^{k}$, with k representing the total number of frame elements. Each frame element $\mathtt{e}_\mathtt{j}$ should be classified into its semantic role. For each frame ϕ in the set of all frames Φ, we construct the set of all predicate-argument pairs $\mathtt{P}_\phi$. Note that a sentence s may contain multiple triggers that by definition evoke multiple frames. Semantic roles for each frame element are obtained using function $sr : \mathcal{E} \times \Phi \mapsto \mathcal{SR}_\phi$ with $\mathcal{E}$ representing the set of all frame elements and $\mathcal{SR}_\phi$ the set of all semantic role types for a given semantic frame $\phi \in \Phi$.

We then define two formal relations:[5]

$$\mathcal{A}_{\mathbb{1}}, \mathcal{A}_{\mathbb{0}} \subseteq \bigcup_{\phi \in \Phi} \mathtt{P}_\phi \times \mathtt{P}_\phi$$

that respectively denote the positive/negative validity of an instance—which is a pair of predicate-argument pairs. This allows for direct construction of a dataset containing positive and negative analogy instances. More precisely, given two pairs $(\mathtt{p}_1, \mathtt{e}_1) \in \mathtt{P}_\phi, (\mathtt{p}_2, \mathtt{e}_2) \in \mathtt{P}_\phi$ for any given $\phi \in \Phi$, the relation containing valid instances of analogies $\mathcal{A}_{\mathbb{1}}$, is defined as follows:

$$((\mathtt{p}_1, \mathtt{e}_1), (\mathtt{p}_2, \mathtt{e}_2)) \in \mathcal{A}_{\mathbb{1}} \Leftrightarrow sr(\phi, \mathtt{e}_1) = sr(\phi, \mathtt{e}_2)$$

Similarly, the set of quadruplets representing non-valid analogies is defined as:

$$((\mathtt{p}_1, \mathtt{e}_1), (\mathtt{p}_2, \mathtt{e}_2)) \in \mathcal{A}_{\mathbb{0}} \Leftrightarrow sr(\phi, \mathtt{e}_1) \neq sr(\phi, \mathtt{e}_2)$$

$\mathcal{A}_{\mathbb{1}}$ is an equivalence relation with equivalence classes representing semantic role types. $\mathcal{A}_{\mathbb{0}}$ retains symmetry but not reflexivity or transitivity.

The definition of these two relations allows us to construct a set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ of positive and negative instances. More precisely:

$$\begin{aligned} &\forall\ ((\mathtt{p}_1, \mathtt{e}_1), (\mathtt{p}_2, \mathtt{e}_2)) \in \mathcal{A}_{\mathbb{1}} \cup \mathcal{A}_{\mathbb{0}} : \\ &\mathbf{x} = [\mathtt{emb}(\mathtt{p}_1) \oplus \mathtt{emb}(\mathtt{e}_1) \oplus \mathtt{emb}(\mathtt{p}_2) \oplus \mathtt{emb}(\mathtt{e}_2)], \\ &y = \begin{cases} 1, & \text{if } sr(\phi, \mathtt{e}_1) = sr(\phi, \mathtt{e}_2) \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where $\mathtt{emb}(\lambda)$ denotes an embedding function that provides a fixed-size vector for any text λ, and $\oplus$ denotes concatenation. Quadruplets that belong to $\mathcal{A}_{\mathbb{1}}$ provide positive instances while those falling under $\mathcal{A}_{\mathbb{0}}$ provide negative ones. These instances are intended for the binary classifier regardless of their analogical validity. Crucially, we never *explicitly* provide semantic role information to the model during training. This binary classifier $\mathcal{A} : \mathcal{A}_{\mathbb{1}} \cup \mathcal{A}_{\mathbb{0}} \mapsto \{0, 1\}$—following our methodology with the analogical transfer mechanism—enables the transformation of a multi-class classification problem involving approximately $1,200$ distinct semantic roles in FrameNet into binary classification.

3.2. Decoding Analogies

Inspired by the notion of decoding termed in structured prediction approaches (Smith, 2011), we use the trained analogical model to obtain probability distributions over all semantic role candidates of a given frame. More precisely, given a target sentence of which the predicate $\mathtt{p}^\mathtt{t}$ evokes frame ϕ, as well as k frame elements $\{\mathtt{e}_\mathtt{i}^\mathtt{t}\}_{\mathtt{i}=1}^{\mathtt{k}}$ identified, we view semantic role classification of each $\mathtt{e}_\mathtt{i}^\mathtt{t}$ as an independent event for which we compute probability distributions over all possible semantic roles types of ϕ by sampling a statistically significant n_e number of source predicate-argument pairs $\{(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s})\}_{j=1}^{n_e}$ from source instances under the same frame type ϕ. The number n_e remains identical to each and every semantic role that ϕ covers. For each pair $(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s})$ and target $(\mathtt{p}^\mathtt{t}, \mathtt{e}_\mathtt{i}^\mathtt{t})$ we apply our trained model $\mathcal{A}$. If $\mathcal{A}(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s}, \mathtt{p}^\mathtt{t}, \mathtt{e}_\mathtt{i}^\mathtt{t})$ yields 1, we apply analogical transfer, shortlisting the semantic role of $\mathtt{e}_\mathtt{j}^\mathtt{s}$ for $\mathtt{e}_\mathtt{i}^\mathtt{t}$. To obtain the final semantic role for target argument

[5] Verb diathesis and multiple valences of a verb could lead us to think that we should "encode" each verb as an n-ary relation. This is not what we try to do here. Instead we focus on a single argument for which we want to identify its semantic role.

$\mathtt{e}_\mathtt{i}^\mathtt{t}$, we calculate the following score for each semantic role $\rho_\kappa \in \phi$ with κ representing the number of semantic role types in ϕ:

$$\sigma(\rho_\kappa) = \sum_{j=1}^{n_e} \mathcal{A}(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s}, \mathtt{p}^\mathtt{t}, \mathtt{e}_\mathtt{i}^\mathtt{t}) \text{ s.t. } sr(\phi, \mathtt{e}_\mathtt{j}^\mathtt{s}) = \rho_\kappa$$

Finally, we attribute to the target argument $\mathtt{e}_\mathtt{i}^\mathtt{t}$ the semantic role with the highest score:

$$sr(\mathtt{e}_\mathtt{i}^\mathtt{t}) = \arg\max_\kappa (\sigma(\rho_\kappa))$$

4. Experiments and Results

In this section, we outline the experimental configuration and discuss the principal results.

4.1. FrameNet Dataset

For our experiments, we employ the latest FrameNet 1.7 dataset following the standard train-development-test partitions established in prior work of Das et al. (2014); Swayamdipta et al. (2017); Lin et al. (2021). This partitioning scheme has been consistently adopted across the literature to ensure comparability with existing frame semantic parsing approaches. We extract all sentences from each partitioned document along with their associated annotations, organising them into corresponding training, development, and test splits—whose statistics are presented in Table 1—for our experimental setup.

Dataset	N. sents	N. frames	N. docs
All	27,228	796	101
train	18,772	754	70
dev	2,192	368	8
test	6,264	563	23

Table 1: The number of sentences, frames, and documents in the datasets

4.2. Analogy Dataset

For analogical instance construction, we iterate within the scope of each semantic frame $\phi \in \Phi$ to extract all predicate-argument pairs present in the annotated sentences, yielding the frame-specific set $\mathtt{P}_\phi$. We subsequently generate the set $\mathtt{AP}_\phi$ of analogy instances, subsuming both negative and positive instances, through the Cartesian product $\mathtt{AP}_\phi = \mathtt{P}_\phi \times \mathtt{P}_\phi$. The complete set $\mathtt{AP}_\Phi$ of analogical instances is constructed as the union of all frame-specific sets $\mathtt{AP}_\Phi = \bigcup_{\phi\in\Phi} \mathtt{AP}_\phi$.

To mitigate class imbalance, we implemented dataset balancing techniques to create a balanced training partition `train_balanced`, which serves as the training set for our experiments. Table 2 presents the detailed distribution of instances across all data partitions.

Dataset	all	positive	negative
train	7,638,128	3,036,470	4,601,658
train_balanced	6,072,940	3,036,470	3,036,470
dev	154,802	72,290	82,512
test	984,724	397,880	586,844

Table 2: Number of analogical instances

4.3. Analogical Model

We design and train a simple feed-forward model, aiming to minimise the number of parameters required to solve the classification problem. The model architecture consists of a number of intermediate blocks followed by a final dense layer, where each intermediate block contains a dense, a dropout, and an activation layer. Throughout the feed-forward network, the model learns weight matrices from concatenated input vectors and outputs a 2-dimensional vector representing class scores for NEG (invalid) and POS (valid). We train this model of feed-forward architecture for binary classification of analogical relations. Hyperparameters are tuned to maximise accuracy on the development set. The hyperparameters tuned include: number of intermediate blocks, feature reduction rate function, dropout rate, and activation function for introducing non-linearity. Experimental results demonstrate that a feed-forward model with two intermediate blocks, a geometric function for feature reduction, a dropout rate of 0.3, and ReLU activation functions achieves optimal training accuracy with only $\approx 800,000$ parameters. The model architecture is pictured in Figure 2.

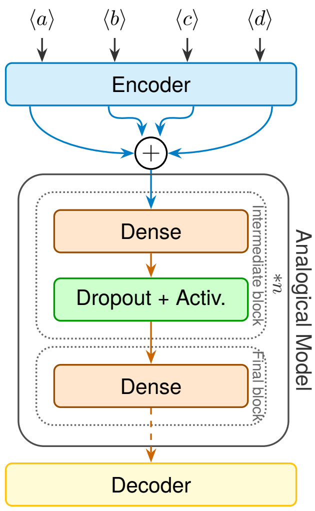

Figure 2: Architecture of our pipeline

We train the model on embeddings of analogy in-

Label	0 (negative)	1(positive)
Precision	91.80	86.03
Recall	90.18	88.23
F1 score	90.98	87.12
Accuracy	89.39	

Table 3: Scores of the binary analogy model (in %)

stances obtained from the `bert-base-uncased` encoder (Devlin et al., 2019). Due to the substantial size of the dataset, we randomly partition the `train_balanced` set into 20 disjoint data segments of similar sizes and employ incremental training, processing each segment sequentially until completion. Following each segment, we save the model to the corresponding checkpoint, resulting in a total of 20 checkpoints throughout the training process. The model optimises Cross-Entropy loss and converges rapidly to optimal performance within the initial segments, subsequently stabilising by the end of the approximately 7-minute training procedure on a single Tesla V100 GPU. Training accuracy and loss across checkpoints are shown in Figure 3.

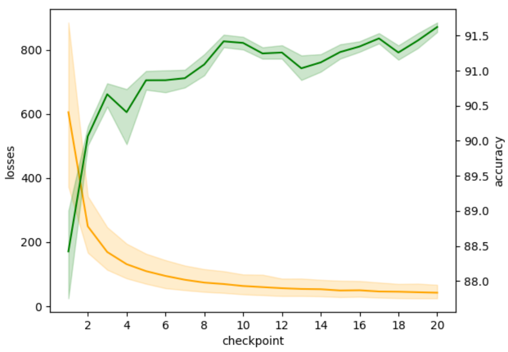

Figure 3: Training statistics wrt checkpoints

We evaluate the binary analogical model on instances derived from the test partition and report per-class performance in Table 3. The model achieves strong results across both NEG and POS classes. A slight bias toward negative predictions is observed, which may be attributable to class imbalance in the development set used for validation.

4.4. Analogical Transfer

For every *target* frame element $\mathtt{e}_\mathtt{i}^\mathtt{t}$ that needs classification in the test set, we extract the corresponding predicate $\mathtt{p}^\mathtt{t}$ for the unique predicate-argument pair $(\mathtt{p}^\mathtt{t}, \mathtt{e}_\mathtt{i}^\mathtt{t})$ of the *target* set $\mathtt{P}^\mathtt{t}$. We subsequently build a source set $\mathtt{P}_\rho^\mathtt{s}$ such that for each semantic role ρ of the target frame type ϕ, a fixed number n_e of exemplary frame elements are retrieved from the *source* dataset along with their corresponding predicates. The statistics of our source dataset suggest $n_e = 7$ following the minimum occurrence rate of 90% of all semantic roles in this dataset. We retrieve source predicate-argument pairs of form $(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s})$ such that:

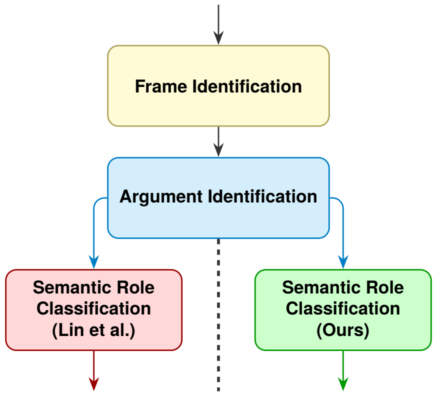

Figure 4: Semantic parsing workflows with two distinct semantic role classifiers

$$\forall \rho \in \phi, \mathtt{P}_\rho^\mathtt{s} = \bigcup_{j=1}^{n_e} \{(\mathtt{p}_\mathtt{j}^\mathtt{s}, \mathtt{e}_\mathtt{j}^\mathtt{s})\} \text{ with } sr(\phi, e_j^s) = \rho$$

We perform classification using the trained model on the analogical instances in the set $\mathtt{AB}_\rho = \mathtt{P}_\mathtt{t} \times \mathtt{P}_\rho^\mathtt{s}$ and aggregate the scores of predicted positive instances to obtain the score for the corresponding role ρ. The group with the highest score transfers its semantic role label to $\mathtt{e}_\mathtt{i}^\mathtt{t}$. Let us repeat that at no point during training nor inference was information on semantic roles passed to the analogical model.

We evaluate the analogical transfer pipeline on the test set using the train set as the source of annotated sentences. This experimental design ensures that the model has never encountered the complete analogical instances—with target pairs from the test set—before evaluation, thereby maintaining the integrity of our experimental setup.

To compare our results with state-of-the-art approaches, we integrate our pipeline as a module for semantic role classification into the framework of Lin et al. (2021), creating a complete end-to-end FrameNet Parsing system. This integration facilitates a direct comparison with existing methods by ensuring a level playing field. Figure 4 illustrates the experimental workflow and system architecture.

Table 4 presents the comparative performance of both end-to-end semantic parsers. Our module as a semantic role classifier demonstrates improvements over the reported results, advancing the state-of-the-art on the FrameNet Parsing task. The improvements are attributable to our approach given the identical modules for Frame Identification and Argument Identification, proving the superior performance of our analogical transfer approach on the Semantic Role Classification task.

Frame id.	Arg id.	SR cl.	Accuracy
Lin et al.	Lin et al.	Lin et al.	48.95
Lin et al.	Lin et al.	Ours	**49.81**

Table 4: Overall performance of state-of-the-art end-to-end parser vs ours on FrameNet 1.7

To further assess our analogical semantic role classifier, we integrated it into Open-SESAME (Swayamdipta et al., 2017) to perform classification on their predicted spans. The experimental configuration, while similar to the framework presented in Figure 4, employs gold-standard frame types and predicates as input. We train their models on FrameNet 1.7 following their standard configuration. Performance results for both approaches on the test set are presented in Table 5, where we can observe again a clear performance gain of our approach.

Furthermore, to comprehensively evaluate our approach to semantic role classification, we assess its performance on instances derived from the test set using gold frame types and frame elements. The results are presented in Table 6. We include comparisons with two baselines, both employing similar MLP architecture for multi-class semantic role classification using BERT embeddings of the input. The first baseline performs classification on the sole frame element $\mathtt{e}^{\mathtt{t}}$, while the second baseline combines predicates with frame elements $(\mathtt{p}^{\mathtt{t}}, \mathtt{e}^{\mathtt{t}})$. The results demonstrate that our analogy-based approach on predicate-frame element pairs $(\mathtt{p}^{\mathtt{s}}, \mathtt{e}^{\mathtt{s}}, \mathtt{p}^{\mathtt{t}}, \mathtt{e}^{\mathtt{t}})$ substantially outperforms both baselines.

5. Related Work

Turney (2008) introduced Latent Relational Analysis (LRA) for identifying analogies, which he tested in 20 scientific and metaphorical examples. Later, Mikolov et al. (2013a,b) initialised the trend of using analogies for word embedding quality evaluation, with Word2vec being the baseline encoder. Gladkova et al. (2016) demonstrated that poorly balanced datasets, such as the Google analogy test set (Mikolov et al., 2013b,a), do not guarantee that static embeddings combined with a vector offset approach are able to capture analogies. This led the authors to introduce the Bigger Analogy Test Set (BATS), showing that even with balanced data, derivational and lexicographic relations remain a challenge. Rogers et al. (2017) showed that the vector offset approach as well as 3CosAdd (Levy and Goldberg, 2014) suffer from dependence on vector similarity, thus arguing against the use of such datasets to evaluate the intrinsic quality of word embeddings.

Frame id.	Arg id.	SR cl.	Accuracy
gold	Swayamdipta et al.	Swayamdipta et al.	58.46
gold	Swayamdipta et al.	Ours	**59.35**

Table 5: Overall performance of former state-of-the-art SRL approach vs ours on FrameNet 1.7

	Precision	Recall	F1 score	Accuracy
Ours	80.77	79.26	79.17	79.30
Baseline 1 $(\mathtt{e}^{\mathtt{t}})$	-	-	-	9.14
Baseline 2 $(\mathtt{p}^{\mathtt{t}}, \mathtt{e}^{\mathtt{t}})$	-	-	-	20.64

Table 6: Baseline evaluation results on FrameNet 1.7 test set with gold frame and spans, using the exact same architecture as our approach. Precision, recall, and F1 scores are weighted.

Sultan and Shahaf (2022) adapted Structure Mapping Theory (Gentner, 1983) to procedural texts, extracting entities and their relationships, and finding suitable mappings between two different domains based on relational similarity. Relations are sets of ordered verbs between entities that are extracted on the basis of question/answer pairs. Here, higher similarity suggests that sets share more relations. Mappings are identified heuristically based on the cosine similarity of BERT vectors representing the questions that provided the entities. Their approach successfully extracts mappings in two different datasets.

On Semantic Role Classification, Gildea and Jurafsky (2002) introduced a statistical classifier to classify spans into corresponding semantic roles on FrameNet. Their approach requires a deal of rigorous processing and laborious manual annotation of the sentences that engineer the features for classification. The authors reported 80.4% accuracy on gold spans. Comparing that with our results of 80.62% accuracy on gold spans is not straightforward due to the fact that (i) we do not work on the same version of FrameNet—we use FrameNet 1.7 while they FrameNet 1.5—as well as (ii) the two datasets exhibit statistical differences in the train–dev–test splits such as the number of spans; we classified 11,206 spans while they did 8,167. Similar feature engineering approaches were also adopted by Cohn and Blunsom (2005) to tackle the SRL problem on the CoNLL shared task and dataset (Carreras and Màrquez, 2005). Syntactic parsing – as a technique to obtain syntactic features from structured texts – is deemed one of the key factors in their success, consistent with the finding of Wang et al. (2019).

To the best of our knowledge, no existing work has ever employed analogies to solve similar problems in FrameNet Parsing and its sub-tasks. Swayamdipta et al. (2017) presented a Softmax-margin semi-Markov model. Their `Open-SESAME` framework initially combines a bidirectional RNN with `SegRNN` (Gimpel and Smith, 2010) then employs syntactic scaffolding to further its initial results to state-of-the-art. More recently, Lin et al. (2021) use Graph Neural Networks based on `BERT` embeddings and `BiHLSTM` architecture (Srivastava et al., 2015) for full FrameNet Parsing, achieving new highs. Finally, Chanin (2023) implemented a framework that fine-tunes a `T5` model (Raffel et al., 2020) on FrameNet and PropBank (Kingsbury and Palmer, 2002) data, rivalling the above parsers.

6. Discussion and Perspectives

Following concerns over coverage gaps in semantic role labelling addressed by Palmer and Sporleder (2010), we examined our results in order to gain valuable insights and identify potential improvements to our methodology. Among the five gaps discussed by Palmer and Sporleder (2010), we hypothesise that our model may be susceptible to no-training-data (NOTR) gaps, attributable to the differences between frame types and semantic roles covered across the train, development, and test partitions. When evaluated on approximately $1{,}800$ instances across 238 out of 563 frame types present in the test set but entirely absent from the development set, which makes up more than 42% of total frame types evaluated, the model suffered a $\approx 4\%$ accuracy loss. On approximately $2{,}000$ test instances containing semantic roles absent from the development set, performance loss increased to $\approx 7\%$. While these results present losses compared to the overall accuracy of 79.30% on test set, the comparably acceptable rates demonstrate the model's generalisation capability across unseen properties in line with the findings of Afantenos et al. (2026). Though not strictly impactful, these limitations may be mitigated by enriching the training and development sets with additional resources such as CoNLL-05 (Carreras and Màrquez, 2005) and CoNLL-09 (Hajic et al., 2009) datasets as suggested in FitzGerald et al. (2015).

On the results of recent approaches, namely `Framenet-Parser` (Lin et al., 2021) and `Open-SESAME` (Swayamdipta et al., 2017), we encountered discrepancies between the reported and reproduced scores. On `Framenet-Parser`, though we used the reported score of 48.95% for comparison, the reproduced accuracy of 47.85% reflected original inaccuracies in both argument identification and classification tasks, consequently having a negative effect on our classifier that relies on their argument identification module. Their SRL module also remains inaccessible, hindering the possibilities for separate module evaluation. For `Open-SESAME`, the inconsistency seems consistent across FrameNet 1.5 and 1.7 datasets. The authors acknowledged the problem on their repository[6]. Since their pre-trained `argid` model on FrameNet 1.7 cannot be loaded directly due to configuration mismatch, we had to reproduce their standalone model from scratch and reported on the collected results. That being said, even the improvement by our pipeline remains consistent given the best effort we invested in reproducibility across experiments, further investigation is encouraged to discover any overlooked variables.

Our experiments were designed and conducted around FrameNet with proof-of-concept objectives rather than providing a holistic approach to semantic parsing or semantic role labelling tasks. Beyond the performance gains, this work is designated to introduce a novel direction for knowledge and representation transfer that transforms multi-label classification into intuitive, human-interpretable, explainable, robust, and scalable binary classification. The reasoning process relies entirely on human-like mechanisms that enhance transparency and explainability. Unlike traditional classification approaches, our analogical model requires no retraining when new labels emerge, highlighting the competitive scalability of the analogical transfer approach.

Regarding computation and energy efficiency, the entire pipeline can be trained and evaluated on a single Tesla V100 GPU—including `BERT` embedding extraction—within 15 minutes. Consuming approximately 50Wh, our approach achieves state-of-the-art performance while remaining considerably more energy-efficient than existing methods.

In this study, we selected `bert-base-uncased` based on the model's computational efficiency and simplicity rather than optimal performance. Future work will examine top notch models from the MTEB leaderboard[7] (Muennighoff et al., 2022), as well as non-contextualised encoders such as GloVe (Pennington et al., 2014). This is necessary to address and assess the correlation between the representation quality and downstream model performance.

Our additional experiments also indicate that sentence embeddings can also serve as representations of context, albeit with a modest decrease in overall performance (approximately 4% in this FrameNet Parsing task). Nevertheless, this suggests promising directions, as it allows for more flexible constructions of analogy instances and may open new avenues for solving other problems, especially when encapsulating the context through its components is infeasible. That being said, we plan

[6] https://github.com/swabhs/open-sesame

[7] https://huggingface.co/spaces/mteb/leaderboard

to evaluate the analogical model on other tasks beyond Semantic Role Classification to explore the broader applicability of analogical transfer. We also foresee the expansion of the analogical transfer methodology to and across other datasets to learn whether comparable performance improvements can be achieved while preserving parameter efficiency and model frugality.

7. Conclusion

In this paper, we adopted a relational view of analogies that allows us to transform a multi-class classification task into binary classification, while recuperating original classes during inference by decoding model output into probability distributions with random sampling. We applied our approach to the Semantic Role Classification task on FrameNet 1.7. We define two binary relations, representing valid and invalid instances of analogies, using pairs of frame evoking predicates and frame elements covered by the same semantic frame ϕ as the domain and co-domain of both relations. Frame elements of valid instances share the same semantic role type, while invalid ones do not.

Using BERT embeddings to train a lightweight ANN with minimal parameters, we have achieved state-of-the-art performance on this task. Our approach exhibits strong energy efficiency, with rapid convergence suggesting potential for further optimisation. The absence of explicit class information during training enables analogical transfer to new classes without model retraining, requiring only annotated examples for sampling during inference, thereby enhancing both scalability and energy efficiency. At the same time, our method preserves interpretability through analogical transfer, providing transparent reasoning mechanisms that promote human understanding of model decisions. We publish the detailed implementation of this work on our repository[8].

Limitations

Optimal decoding of the binary model's output assumes a sufficiently large number of annotated instances for reference. Failure to obtain necessary reference instances may hinder accurate analogical transfer of under-represented classes at the decoding phase. Nonetheless, this not-enough-data gap is a rather general problem in Machine Learning (Ghosh et al., 2024).

[8]https://github.com/thebugcreator/franabelling

Acknowledgements

This work is supported by the AT2TA project[9] funded by the French National Research Agency ("Agence Nationale de la Recherche" – ANR) under grant ANR-22-CE23-0023.
This work has also been supported by the Occitanie region, the European Regional Development Fund (ERDF), and the French government through the France 2030 project managed by the ANR with reference number ANR-22-EXES-0015.

References

Stergos Afantenos, Miguel Couceiro, Emiliano Lorini, and Van-Duy Ngo. 2026. Learning proportional analogies: Lightweight neural network vs large language models. In *Proceedings of the 18th International Conference on Agents and Artificial Intelligence*, pages 3604–3614.

Aristotle. 2009. *Nicomachean Ethics*. Oxford World's Classics. Oxford University Press, Oxford.

Collin F. Baker. 2017. Framenet: Frame semantic annotation in practice. *Handbook of Linguistic Annotation*, pages 771–811.

Collin F. Baker, Michael Ellsworth, and Katrin Erk. 2007. SemEval-2007 task 19: Frame semantic structure extraction. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 99–104, Prague, Czech Republic. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *International Conference on Computational Linguistics*, volume 1, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Nelly Barbot, Laurent Miclet, and Henri Prade. 2019. Analogy between concepts. *Artificial Intelligence*, 275:487–539.

Myriam Bounhas and Henri Prade. 2023. Analogy-based classifiers: An improved algorithm exploiting competent data pairs. *International Journal of Approximate Reasoning*, 158:108923.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the ninth conference on computational natural language learning (CoNLL-2005)*, pages 152–164.

[9]https://at2ta.loria.fr/

David J. Chalmers, Robert M. French, and Douglas R. Hofstadter. 1992. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, 4(3):185–211.

David Chanin. 2023. Open-source frame semantic parsing. *arXiv preprint arXiv:2303.12788*.

Trevor Cohn and Phil Blunsom. 2005. Semantic role labelling with tree conditional random fields. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 169–172.

Francisco Cunha, Yves Lepage, Miguel Couceiro, and Zied Bouraoui. 2026. Generalizing analogical inference from boolean to continuous domains. In *AAAI-26, Sponsored by the Association for the Advancement of Artificial Intelligence, January 20 - January 27, 2026, Singapore*. AAAI Press. To appear.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186. Association for Computational Linguistics.

Aleksandr Drozd, Anna Gladkova, and Satoshi Matsuoka. 2016. Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *International Conference on Computational Linguistics, Technical Papers*, pages 3519–3530, Osaka, Japan. The COLING 2016 Organizing Committee.

Charles J Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

Erez Firt. 2025. Analogical reasoning as a core agi capability. *AI and Ethics*, 5(5):5501–5515.

John Rupert Firth. 1968. *Selected Papers of J. R. Firth, 1952–59*. Indiana University Studies in the History and Theory of Linguistics. Indiana University Press, Bloomington.

Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Semantic role labeling with neural network factors. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 960–970.

Dedre Gentner. 1983. Structure Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7:155–170.

Dedre Gentner and Christian Hoyos. 2017. Analogy and abstraction. *Topics in Cognitive Science*, 9(3):672–693.

Kushankur Ghosh, Colin Bellinger, Roberto Corizzo, Paula Branco, Bartosz Krawczyk, and Nathalie Japkowicz. 2024. The class imbalance problem in deep learning. *Machine Learning*, 113(7):4845–4901.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.

Kevin Gimpel and Noah A Smith. 2010. Softmax-margin crfs: Training log-linear models with cost functions. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 733–736.

Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. 2016. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *North American Chapter of the Association for Computational Linguistics, Student Research Workshop*, pages 8–15, San Diego, California. Association for Computational Linguistics.

Jan Hajic, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, M Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18.

Zellig S. Harris. 1954. Distributional structure. *Word*, 10(2–3):146–162.

Douglas R. Hofstadter. 1984. The copycat project: An experiment in nondeterminism and creative analogies.

Douglas R. Hofstadter. 2001. Analogy as the Core of Cognition. In Dedre Gentner, Keith J. Holyoak, and Boicho N. Kokinov, editors, *The Analogical Mind: Perspectives from Cognitive Science*, chapter 15, pages 499–538. The MIT Press, Cambridge, Massachusetts.

Douglas R. Hofstadter and FARG. 1995. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books. Publication Title: Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought.

Douglas R. Hofstadter and Emmanuel Sander. 2013. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books.

Keith J. Holyoak and Paul Thagard. 1995. *Mental Leaps: Analogy in Creative Thought*. A Bradford Book. MIT Press.

Marzena Karpinska, Bofang Li, Anna Rogers, and Aleksandr Drozd. 2018. Subcharacter information in Japanese embeddings: When is it worth it? In *Proceedings of the Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP*, pages 28–37, Melbourne, Australia. Association for Computational Linguistics.

Paul Kingsbury and Martha Palmer. 2002. From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Spain*.

Yves Lepage. 2005. Translation of sentences by analogy principle. In *Second Language and Technology Conference*.

Omer Levy and Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 171–180, Ann Arbor, Michigan. Association for Computational Linguistics.

ZhiChao Lin, Yueheng Sun, and Meishan Zhang. 2021. A graph-based neural model for end-to-end frame semantic parsing. In *Conference on Conference on Empirical Methods in Natural Language Processing*, pages 3864–3874, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Esteban Marquer and Miguel Couceiro. 2025. Solving morphological analogies: from retrieval to generation. *Ann. Math. Artif. Intell.*, 93(2):263–298.

Timothee Mickus, Eduardo Calò, Léo Jacqmin, Denis Paperno, and Mathieu Constant. 2023. „mann“ is to “donna” as “guówáng” is to «reine» adapting the analogy task for multilingual and contextual embeddings. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 270–283, Toronto, Canada. Association for Computational Linguistics.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations, Workshop*.

Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26, pages 3111–3119. Curran Associates Inc.

Melanie Mitchell. 1993. *Analogy Making as Perception: A Computer Model*. Neural Network Modeling and Connectionism. The MIT Press.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. 2022. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*.

Alexis Palmer and Caroline Sporleder. 2010. Evaluating framenet-style semantic parsing: the role of coverage gaps in framenet. In *Coling 2010: Posters*, pages 928–936.

Jeffrey Pennington, Richard Socher, and Christopher 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, Doha, Qatar. Association for Computational Linguistics.

Molly Petersen and Lonneke van der Plas. 2023. Can language models learn analogical reasoning? investigating training objectives and comparisons to human performance. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 16414–16425, Singapore. Association for Computational Linguistics.

Molly R. Petersen, Claire. E Stevenson, and Lonneke van der Plas. 2025. Modelling analogies and analogical reasoning: Connecting cognitive science theory and nlp research.

Henri Prade and Gilles Richard. 2009. Analogy, paralogy and reverse analogy: Postulates and inferences. In *Annual conference on artificial intelligence*, pages 306–314. Springer.

Henri Prade and Gilles Richard. 2017. Analogical proportions and analogical reasoning - an introduction. In *Case-Based Reasoning Research and Development - 25th International Conference, ICCBR 2017, Trondheim, Norway, June 26-28, 2017, Proceedings*, volume 10339 of *LNCS*, pages 16–32. Springer.

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. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Anna Rogers, Aleksandr Drozd, and Bofang Li. 2017. The (too many) problems of analogical reasoning with word vectors. In *Joint Conference on Lexical and Computational Semantics*, pages 135–148, Vancouver, Canada. Association for Computational Linguistics.

Josef Ruppenhofer, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

Noah A. Smith. 2011. *Linguistic Structure Prediction*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool.

Rupesh K. Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Training very deep networks. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Oren Sultan and Dafna Shahaf. 2022. Life is a circus and we are the clowns: Automatically finding analogies between situations and processes. In *Conference on Empirical Methods in Natural Language Processing*, pages 3547–3562, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold.

Peter D. Turney. 2008. The Latent Relation Mapping Engine: Algorithm and Experiments. *Journal of Artificial Intelligence Research*, 33:615–655.

Matej Ulčar, Kristiina Vaik, Jessica Lindström, Milda Dailidėnaitė, and Marko Robnik-Šikonja. 2020. Multilingual culture-independent word analogy datasets. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4074–4080, Marseille, France. European Language Resources Association.

Asahi Ushio, Luis Espinosa Anke, Steven Schockaert, and Jose Camacho-Collados. 2021. BERT is to NLP what AlexNet is to CV: Can pre-trained language models identify analogies? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3609–3624, Online. Association for Computational Linguistics.

Yufei Wang, Mark Johnson, Stephen Wan, Yifang Sun, and Wei Wang. 2019. How to best use syntax in semantic role labelling. *arXiv preprint arXiv:1906.00266*.

Ce Zheng, Xudong Chen, Runxin Xu, and Baobao Chang. 2022. A double-graph based framework for frame semantic parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4998–5011, Seattle, United States. Association for Computational Linguistics.