

# Prediction of Frame-to-Frame Relations in the FrameNet Hierarchy with Frame Embeddings

Teresa Botschen<sup>§†</sup>, Hatem Mousselly-Sergieh<sup>†</sup>, Iryna Gurevych<sup>§†</sup>

<sup>§</sup>Research Training Group AIPHES

<sup>†</sup>Ubiquitous Knowledge Processing (UKP) Lab

Department of Computer Science, Technische Universität Darmstadt

www.aiphes.tu-darmstadt.de, www.ukp.tu-darmstadt.de

## Abstract

Automatic completion of frame-to-frame (F2F) relations in the FrameNet (FN) hierarchy has received little attention, although they incorporate meta-level commonsense knowledge and are used in downstream approaches. We address the problem of sparsely annotated F2F relations. First, we examine whether the manually defined F2F relations emerge from text by learning text-based frame embeddings. Our analysis reveals insights about the difficulty of reconstructing F2F relations purely from text. Second, we present different systems for predicting F2F relations; our best-performing one uses the FN hierarchy to train on and to ground embeddings in. A comparison of systems and embeddings exposes the crucial influence of knowledge-based embeddings to a system’s performance in predicting F2F relations.

## 1 Introduction

FrameNet (FN) (Baker et al., 1998) is a lexical-semantic resource manually built by FN experts. It embodies the theory of frame semantics (Fillmore, 1976): the frames capture units of meaning corresponding to prototypical situations. Besides FN’s definitions of frame-specific roles and frame-evoking elements that are used for the task of Semantic Role Labeling, it also contains manual annotations for relations that connect pairs of frames. There are thirteen *frame-to-frame* (F2F) relations of which five are antonym relations (e.g. Precedes, Is\_Preceded\_by). To give an example, the frame “Waking\_up” is in relation Precedes to the frame “Being\_aware”. Figure 1 further elaborates this example by demon-

strating relationships to additional frames. Table 1 lists all F2F relation names with the number of frame pairs for each relation according to the FN hierarchy, and also restricted counts including only frame pairs that have lexical units (LU) in the FN hierarchy (e.g., the frame “Waking\_up” can be evoked by the LU “awake.v” of the verb “awake”). The FN hierarchy, a report version of FN, does not provide lexical units for 125 frames (e.g., the frame “Sleep\_wake\_cycle” has no LU). In fact, such frames are used as meta-frames for abstraction purposes, thus, they exist only to participate in F2F relations with other frames (Ruppenhofer et al., 2006). In general, each frame pair is connected via only one F2F relation with occasional exceptions and the F2F relations situate the frames in semantic space (Ruppenhofer et al., 2006). F2F relations are used in the context of other tasks, such as text understanding (Fillmore and Baker, 2001), paraphrase rule generation for the system LexPar (Coyne and Rambow, 2009) and recognition of textual entailment (Aharon et al., 2010). Furthermore, F2F can be used as a form of commonsense knowledge (Rastogi and Van Durme, 2014).

The incompleteness of the FN hierarchy is a known issue not only at the frame level (Rastogi and Van Durme, 2014; Pavlick et al., 2015; Hartmann and Gurevych, 2013) but also at the

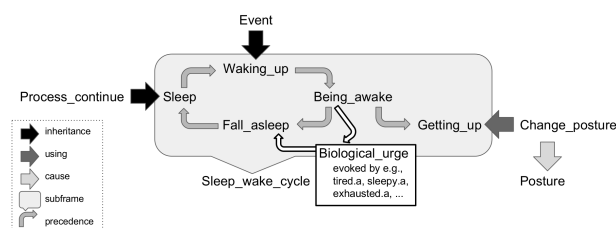


Figure 1: F2F relations example. Gray: FN hierarchy. White arrows: missing in FN hierarchy.

F2F Relation Name	Total	Restricted
Inherits_from	617	383
Is_inherited_by	617	383
Uses	491	430
Is_Used_by	490	430
Subframe_of	119	29
Has_Subframes	117	29
Perspective_on	99	15
Is_Perspectivized_in	99	15
Precedes	79	48
Is_Preceded_by	79	48
Is_Causative_of	48	47
See_also	41	40
Is_Inchoative_of	16	16
Sum	2,912	1,913

Table 1: F2F relation pair counts and restricted pair counts of frames with lexical units.

F2F relation level. Figure 1 exemplifies a missing precedence relation: “Fall\_asleep” is preceded by “Being\_aware” but inbetween yet another frame could be added, e.g. “Biological\_urge” (evoked by the predicate “tired”). Rastogi (2014) note a lack of research on automatically enriching the F2F relations, which would be beneficial given the large number of possible frame pairs for a relation and their use in other tasks. The automatic annotation of F2F relations involves three difficulties accounted for by the nature of FN. First, F2F relation annotations occur sparsely and for the majority of pairs in each relation there are few instances (see Table 1). Second, the relations themselves have no direct lexical correspondences in text and hence inferring them from text is not trivial. Third, if a relation involves a frame that does not have any lexical unit (see restricted counts in Table 1), this frame does not occur in text and hence inferring this relation from text is even more difficult.

To the best of our knowledge there is no work (a) addressing the emergence of F2F relations from text data or (b) enriching F2F relations automatically. We aim to address these problems with the following contributions:

### Contributions of the paper

- 1) We learn text-based frame embeddings and explore their limitations with respect to F2F relations to check whether the manually annotated F2F relations can naturally emerge from text. We find that, concerning the methods we explored, the embeddings have difficulties in showing structures directly corresponding to F2F relations.
- 2) We transfer the relation prediction task from research on Knowledge Graph Completion to the

case of FN and present our best-performing system for predicting F2F relations. The system involves training on the FN hierarchy and uses embeddings also trained on the FN hierarchy.

3) We generically demonstrate the predictions of our best-performing system for unannotated frame pairs and suggest its application for automatic FN completion on the relation level.

**Structure of the paper** To start with, Section 2 reviews related research including algorithms and approaches that we will apply to our purposes. Next, Section 3 briefly presents the FN data that we will work with. Then, the paper is structured along our contributions: exploration of F2F relations in frame embedding space (Sec 4) and F2F Relation Prediction task (Sec 5), including a demonstration of predictions for unannotated frame pairs (Sec 5.3). Finally, Section 6 discusses our insights and traces options for future work.

## 2 Related Work

### 2.1 Frame Embeddings

**Frame embeddings for frame identification** To our knowledge, the only approach learning frame embeddings is a matrix factorization approach in the context of the task of Frame Identification (FrameId), which is the first step in FN Semantic Role Labeling. The state-of-the-art system (Hermann et al., 2014) for FrameId projects frames and predicates with their context words into the same latent space by using the *WSABIE* algorithm (Weston et al., 2011). Two projection matrices (one for frames and one for predicates) are learned using WARP loss and gradient-based updates such that the distance between the predicate’s latent representation and that of the correct frame are minimized. Consequently, latent representations of frames will end up close to each other if they are evoked by similar predicates and context words. As the focus of such systems is on the FrameId task, the latent representations of the frames are rather a sub-step contributing to FrameId but not studied further or applied to other tasks. We will extract these frame embeddings and explore them with respect to F2F relations.

**Word2Vec embeddings** The Neural Network (NN) architecture of the *Word2Vec* algorithm (Mikolov et al., 2013a) learns word embeddings by either predicting a target word given its context words (*CBOW model*) or by predicting context words given their target word (*skip-gram model*).

There are different tasks specifically designed for the evaluation of word embeddings (Mikolov et al., 2013b). They are formulated as analogy questions about syntax or semantics of the form “*a is to b as c is to \_\_\_*”. Mikolov (2013b) suggest a **vector offset method** based on cosine distance to solve these analogy tasks. This assumes that relationships are expressed by vector offsets: given two word pairs  $(a, b)$  and  $(c, d)$ , the question is to what extent the relations within the pairs are similar. We will apply this method to frame pairs that are connected via F2F relations in order to find out whether the frame embeddings incorporate F2F relations.

There is an interest in abstracting away from word embeddings towards embeddings for more coarse grained units: *Word2Vec* is used to learn embeddings for senses (Iacobacci et al., 2015) or for supersenses (Flekova and Gurevych, 2016). Iacobacci (2015) use the *CBOV model* on texts annotated with BabelNet senses (Navigli and Ponzetto, 2012). Flekova (2016) use the *skip-gram model* on texts with mapped WordNet supersenses (Miller, 1990; Fellbaum, 1990). For evaluation both works are oriented towards Mikolov’s (2013b) analogy tasks and perform qualitative analyses for the top  $k$  most similar embeddings for (super)senses or visualize the embeddings in vector space. To have text-based frame embeddings in line with related work, we will also use the *Word2Vec* algorithm to learn an additional version for frame embeddings.

## 2.2 Relation Prediction

This task stems from automatic Knowledge Graph Completion (KGC) and is known as “Link Prediction” (Bordes et al., 2011, 2012, 2013). We will transfer this task to F2F Relation Prediction for frame pairs. For this task, knowledge-based embeddings are well suited, which are not learned on text but on triples of a KG.

**TransE embeddings** We leverage an embedding learning approach from KGC to obtain embeddings for frames and for F2F relations that are grounded in the FN hierarchy. In translation models, all entities and relations of the triples of head-entity, relation and tail-entity  $(h, r, t)$  are projected into one latent vector space such that the relation-vector connects from the head-vector to the tail-vector as a translating vector operation. *TransE* (Bordes et al., 2013) introduced the idea of mod-

eling relations as translations that operate on the embeddings of the entities. The model is formulated to minimize  $|h + r - t|$  for a training set, with randomly initialized embeddings. The function to minimize resembles the idea of the vector offset by Mikolov (2013b).

**Answer selection model** Link Prediction is methodologically related to the key-task of Answer Selection from Question Answering (QA). The task is to rank a set of possible answer candidates with respect to a given question (Tan et al., 2015). State-of-the-art QA models are presented by (Feng et al., 2015) and by (Tan et al., 2015). They jointly learn vector representations for both the questions and the answers. Representations of the same dimensionality in the same space allow one to compute the cosine similarity between these vectors. We will orient ourselves by NN models for Answer Selection in order to adapt the ideas to F2F Relation Prediction. In our case, a question corresponds to a frame pair and an answer corresponds to a F2F relation. Optionally, pretrained frame embeddings can be used as initialization.

## 3 Data

**Textual data** In order to learn frame embeddings on textual data with *WSABIE* or *W2V*, we take the FrameNet 1.5 sentences provided by the Dependency-Parsed FrameNet Corpus (Bauer et al., 2012) which contains more than 170,000 sentences annotated manually with frame labels for 700 frames. We denote a frame as  $f$  where  $f \in F_t$  the set of frames in the textual data.

**Hierarchy data** The FN hierarchy lists for each frame of the overall 1,019 frames the F2F relations to other frames. We denote with  $G$  the collection of triples  $(f_1, r, f_2)$  (standing for frame “ $f_1$  is in relation  $r$  to frame  $f_2$ ”), where  $f_1$  and  $f_2 \in F_h$  the set of frames in the FN hierarchy and  $r \in R$  the set of F2F relations. As listed in Table 2, there are 2,912 triples in the FN hierarchy with 1,913 triples remaining if considering only those where both frames have lexical units and

Corpus	Frames	F2F Relations
FN Hierarchy	1,019	2,912
FN Hierarchy restricted to frames with LU	894	1,913
Textual data FN 1.5 sentences	700	1,447

Table 2: Counts for frames and F2F relations.

with 1,447 triples remaining if considering only those where both frames occur in the textual data. We split the obtained triples whose frames have lexical units into a training and a test set such that the training set contains the first 70% of all the triples for each relation.

Table 2 summarizes frame counts per data source together with counts of F2F relations where both frames occur in the underlying source.

## 4 Exploration of Frame Embeddings

We aim at empirically analyzing whether F2F relations from the FN hierarchy are mirrored in frame embeddings learned on frame-labeled text in the context of other tasks. Thus, we want to identify whether a statistical analysis of text-based frame embeddings naturally yields the FN hierarchy. Indeed, the F2F relations are manually annotated by expert linguists but there is no guarantee that F2F relations can be observed in text. If these relations could emerge from raw text it would be reassuring for the definitions of the F2F relations that led to annotations of frame pairs and furthermore the annotations could be generated automatically. We hypothesize that distances and directions between frame embeddings learned on textual data can correspond to F2F relations. Figure 2 exemplifies this as known from word embeddings by Mikolov (Mikolov et al., 2013a): it highlights two frame pairs that are in the same relation: “Attempt” is in relation `precedes` with “Success\_or\_failure” and so is “Existence” in relation `precedes` with “Ceasing\_to\_be” and the connecting vectors are about the same direction and length.

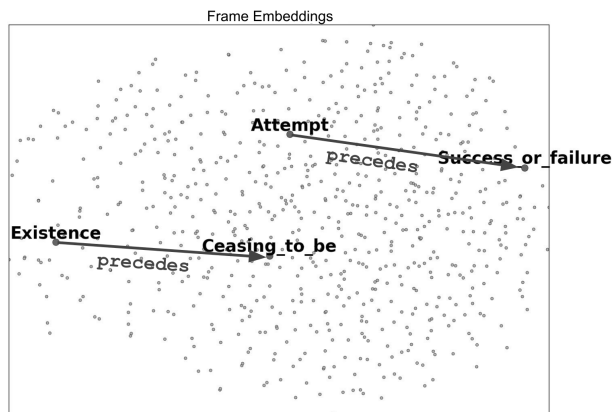


Figure 2: Intuition for frame embeddings incorporating F2F relations in vector space.

## 4.1 Methods

**WSABIE frame embeddings** Concerning the matrix factorization approach for learning text-based frame embeddings, we use the code provided by (Hartmann et al., 2017) as it is publically available. It is leaned on Hermann’s (2014) description of their state-of-the-art system and achieves comparable results on FrameId. Our hyperparameter choices are oriented towards (Hartmann et al., 2017): embedding dimension 100, maximum number of negative samples: 100, epochs: 1000 and initial representation of predicate and context: concatenation of pretrained dependency-based word embeddings (Levy and Goldberg, 2014).

**Word2Vec frame embeddings** Concerning the NN approach for learning text-based frame embeddings, we use the *Word2Vec* implementation in the python library gensim (Řehůřek and Sojka, 2010). To obtain frame embeddings we follow the same steps as if we would learn word embeddings on FN sentences plus we replace all predicates with their frames. For instance, in the sequence “Officials claim that Iran has produced bombs” the predicates “claim” and “bombs” are replaced by “STATEMENT” and “WEAPON” respectively. This procedure corresponds to Flekova’s (2016) setup for learning supersense embeddings and our hyperparameter choices are oriented towards their best performing ones: training algorithm: *skip-gram model*, embedding dimension: 300, minimal word frequency: 10, negative sampling of noise words: 5, window size 2, initial learning rate: 0.025 and iterations: 10.

**Prototypical relation embeddings** We denote learned embeddings with  $\vec{e}_1$  (for frame  $f_1$ ). We use the frame embeddings to infer prototypical F2F relation embeddings  $\vec{e}_r$  with the vector offset method in the following way: we denote with  $I_r$  the relation-specific subset of  $G$  with all the instances  $(f_1, r, f_2)$  for this relation (see frame pair counts in Table 1). The vector offset  $\vec{o}_{e_1, e_2}$  for two frames  $(f_1, f_2)$  is the difference of their embeddings, see Equation 1.

$$\text{offset}\{f_1, f_2\} = \vec{e}_2 - \vec{e}_1 \quad (1)$$

We denote with  $O_r$  the relation-specific set of vector offsets of all  $(f_1, f_2) \in I_r$ . We define the prototypical embedding  $\vec{e}_r$  for a relation  $r$  as the mean over all  $\vec{o}_{e_1, e_2} \in O_r$ . For visualizations in vector space we use t-SNE-

plots (t-distributed Stochastic Neighbor Embedding (Maaten and Hinton, 2008) algorithm).

**Difficulty of associating frame pairs with prototypical relations** The association of the embedding of a frame pair  $\overrightarrow{o_{e_1, e_2}} \in O_r$  with the correct prototypical relation embedding  $\overrightarrow{e_r}$  is easier if the intra-relation variation (i.e. the deviation of frame pair embeddings from their prototypical embedding) is smaller than the inter-relation variation (i.e. the distances between prototypical embeddings). This means, the association is easier if two frame pairs which are members of the same F2F relation, on average, differ less from each other as they would differ from a member of another relation. As a way to capture this difficulty of association we compare the mean cosine distance between all prototypical relations embeddings  $\overrightarrow{e_r}$  of all  $r \in R$  to the relation-specific mean cosine distance between the frame pair embeddings in  $O_r$  and the prototypical embedding  $\overrightarrow{e_r}$ .

## 4.2 Experiments and Results

**Frame embeddings** Once the frame embeddings are learned, we perform a sanity check for frames and most similar frame embeddings by cosine similarity. Checking the top 10 most similar frame embeddings confirms that known properties from word or sense embeddings also apply to frame embeddings: their top 10 most similar frames are semantically related, both for frame embeddings learned with *WSABIE* and with *Word2Vec*. This is exemplified in Table 3 for the two most frequently occurring frames in the text data evoked by nouns (“Weapon”) and by verbs (“Statement”). For both *WSABIE* and *Word2Vec*, in many cases the most similar frames are obviously semantically related (which we marked in bold), with some exceptions where it is hard to judge or related via an asso-

frame	Top 10 most similar frames	
	<i>WSABIE</i>	<i>Word2Vec</i>
Weapon	Substance, <b>Shoot projectiles</b> , Manufacturing, <b>Bearing arms</b> , <b>Toxic substance</b> , <b>Hostile encounter</b> , Ingredients, Information, <b>Smuggling</b> , Active_substance	<b>Military</b> , Substance, Operational_testing, Store, Electricity, Process_completed_state, Active_substance, Range, Estimated_value, Cause_to_make_progress
Statement	Evidence, Causation, <b>Topic</b> , <b>Chatting</b> , <b>Point_of_dispute</b> , <b>Request</b> , <b>Text_creation</b> , Cognitive_connection, <b>Make_agreement_on_action</b> , <b>Communication</b>	<b>Reveal_secret</b> , <b>Telling</b> , <b>Complaining</b> , <b>Reasoning</b> , <b>Communication_response</b> , Awareness, <b>Reassuring</b> , <b>Bragging</b> , <b>Questioning</b> , Cogitation

Table 3: Top 10 most similar frames to two exemplary most frequent frames.

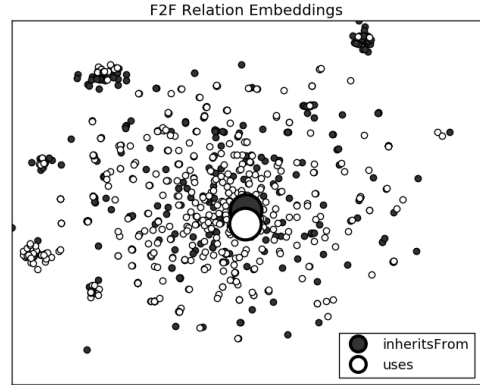


Figure 3: t-SNE plot of embeddings for two most frequent relations. Small: frame pair embeddings (offset). Large: prototypical embeddings (mean).

ciation chain. For the frame “Weapon”, the most similar frames by *Word2Vec* are weaker compared to *WSABIE*, however this does not allow a general conclusion over all frames learned with *Word2Vec* or *WSABIE*.

**F2F relations** To check whether the frame embeddings directly mirror F2F relations, we measure the difficulty of associating frame pairs with the correct prototypical relation embedding.

First, we visualize the frame pair embeddings in the training set and the inferred prototypical relation embeddings in vector space with t-SNE-plots. Figure 3 depicts examples of *WSABIE* embeddings for the most frequently occurring F2F relations *inherits\_from* and *uses*, and shows that the prototypical embeddings are very close to each other, whilst there are no separate relation-specific clusters for frame pairs. Vector space visualizations of embeddings stemming from both, *Word2Vec* and *WSABIE*, hint that the embeddings have difficulties in mirroring the F2F relations.

Second, we quantify the insights from the plots by comparing the distances between all prototypical embeddings to the mean over all mean distances between frame pair embeddings and their prototypical embeddings. Table 4 lists these vector space (cosine) distances. It shows that the distance between the prototypical embeddings (inter-relation) is smaller than that between frame pair embeddings and corresponding prototypical embeddings (intra-relation). In other words, two frame pairs which are members of the same relation, on average, differ as much from each other as

Mean distances between	<i>WSABIE</i>	<i>Word2Vec</i>
inter-relation variation (between prototypes)	$0.73 \pm 0.28$	$0.76 \pm 0.28$
intra-relation variation (between frame pairs and their prototypes)	$0.75 \pm 0.04$	$0.78 \pm 0.05$

Table 4: Cosine distances between the F2F relation embeddings.

they would differ from a member of another relation.

To sum up, we find that embeddings of frame pairs that are in the same relation do not have a similar vector offset which corresponds to the F2F relation. The FN hierarchy could not be reconstructed by the statistical analysis of text-based embeddings because there is as much intra-relation variation as inter-relation variation. We conclude that, concerning the methods we explored, the frame embeddings learned with *WSABIE* and *Word2Vec* have difficulties in showing structures in vector space corresponding to F2F relations and that F2F relations might not emerge purely from textual data. Hence, these text-based frame embeddings cannot be used as such to reliably infer the correct relation for a frame pair but might need some advanced learning. In the next section, we address the prediction of F2F relations with algorithms involving learning from the knowledge contained in the FN hierarchy.

## 5 Frame-to-Frame Relation Prediction

We aim at developing a system for finding the correct F2F relation given two frames, which can potentially be used for automatic completion of the F2F relation annotations in the FN hierarchy. This task transfers the principles of Link Prediction from KGC to the case of FN. As the previous experiment suggested that text-based frame embeddings do not mirror the F2F relations, we develop a system that learns from the knowledge contained in the FN hierarchy and that uses pretrained frame embeddings as input representations. Related work in KGC also demonstrates the strengths of representations trained directly on the KG for this task. For our systems involving learning, we experiment with different embeddings as input representations: in addition to the text-based frame embeddings, we also learn knowledge-based embeddings for frames and for F2F relations on the structure of the FN hierarchy with *TransE*, an approach well-known from KGC.

We want to quantify which combination of pre-trained embeddings and system is most promising for the F2F Relation Prediction task.

### 5.1 Methods

***TransE* embeddings** In addition to the text-based frame embeddings, we also learn embeddings for frames as well as for F2F relations by applying the well-known translation model *TransE*. *TransE* leverages the structure of the knowledge base, which is in our case the FN hierarchy with the collection of the  $(frame, relation, frame)$  triples, and learns low dimensional vector representations for frames and for F2F relations in the same space. These embeddings will have the property of being learned explicitly for incorporating the annotations from the FN hierarchy. Concerning this knowledge-based approach for learning frame and F2F relation embeddings, we use an implementation of *TransE* provided by (Lin et al., 2015) yielding embeddings of dimension 50.

**Neural network for relation selection** We propose a nonlinear model based on NNs to identify the best F2F relation  $r$  between a frame pair  $(f_1, f_2)$ . Figure 4 demonstrates the proposed NN architecture. Given a training instance, i.e. a triple  $(f_1, r, f_2)$ , we feed a vector representation for each element into the NN. By default the input vector representations are initialized randomly but they can also stem from a pretraining step (more details in Sec 5.2). Within the NN, the initial vector representations of the two frames are combined into an internal dense layer  $c$ , followed by the calculation of the cosine similarity between this combina-

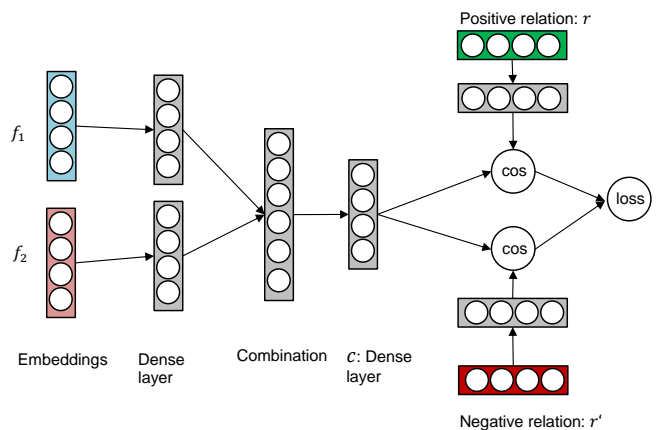


Figure 4: NN architecture for training on the F2F Relation Prediction task.

tion and the representation for the F2F relation  $r$ . Meanwhile, a negative relation  $r'$  is sampled randomly (by selecting a F2F relation which does not hold between the two frames) and its vector representation is also fed into the NN. The negative relation is processed as the correct one, yielding a second cosine similarity. Finally, the NN minimizes the following ranking loss:

$$loss = \max\{0, m - \cos(c, r) + \cos(c, r')\} \quad (2)$$

$m$  is a margin and  $\cos$  is the cosine similarity function. This means, the internal representations are trained to maximize the similarity between frame pair and correct relation and to minimize it for the negative relation. Our hyperparameter choices are: epochs: 550, size of dense layers: 128, dropout: 0.2, margin: 0.1, activation function: hyperbolic tangent, batch size: 2, learning rate 0.001.

## 5.2 Experiments and Results

Given a triple  $(f_3, r, f_4)$  from the test set, we want to predict the correct relation  $r$  for  $(f_3, f_4)$ . As described in Sec 3, 70% of the triples in the FN hierarchy are used for training. Our systems are:

- **0a): rand.bsl** A random guessing baseline that chooses a relation randomly out of  $R$ .
- **0b): maj.bsl** Informed majority baseline that leverages the skewed distribution in the training set and predicts most frequent relation.
- **1: off** A test of the pretrained frame embeddings (*WSABIE* and *Word2Vec*) as introduced in Section 4. It computes the vector offset  $\vec{o}_{e_3, e_4}$  (therefore “off”) between the test frame embeddings, measures the similarity with the prototypical mean relation embeddings  $\vec{e}_r$  of the training set and ranks the relations with respect to similarity (cosine) to output the closest one. No further training with respect to the FN hierarchy.
- **2: reg** A test of the pretrained frame embeddings (*WSABIE* and *Word2Vec*) as introduced in Section 4 involving training with respect to the FN hierarchy. It is a multinomial logistic regression model (therefore “reg”) that trains the weights and biases on the training triples. It takes the test frame embeddings  $e_3, e_4$  as input and ranks the prediction for a relation via the softmax function.
- **3: NN** NN architecture as described in Section 5.1 for training with respect to the FN hierarchy in the training triples. By de-

fault, it uses randomly initialized input representations, but it can also take pretrained representations as input: (a) the pretrained frame embeddings (*WSABIE* and *Word2Vec*) and inferred prototypical mean relation embeddings as introduced in Section 4 and (b) the *TransE* frame and relation embeddings trained on the training triples from the FN hierarchy as introduced in Section 5.1.

To evaluate the predictions of our systems for the F2F Relation Prediction task, we compare the measurements of accuracy, mean rank of the true relation and hits amongst the 5 first predictions, see Table 5.

**Accuracy** measures the proportion of correctly predicted relations amongst all predictions.

For the next two measures, not only the one predicted relation is of interest, but the ranked list of all relations with the predicted relation at rank 1.

**Mean rank** measures the mean of the rank of the true relation label over all predictions, aiming at low mean rank (best is  $mr = 1$ ).

**Hits@5** measures the proportion of true relation labels ranked in the top 5.

The random guessing baseline is a weak baseline that is outperformed by all approaches. The informed majority baseline, however, is a strong baseline given the skewed distribution of F2F relations in the FN hierarchy.

A comparison of this strong baseline with system 1 using the text-based frame embeddings (*WSABIE* and *Word2Vec*) and the similarity with prototypical relation embeddings, emphasizes the difficulties of these embeddings for reconstructing the F2F relations. Concerning accuracy scores, system 1 performs slightly better than the strong baseline but concerning the other two measures, mean rank and hits at 5, it is the other way round. An-

System	Embed.	acc $\uparrow$	mr $\downarrow$	hits@5 $\uparrow$
0: rand.bsl	-	7.69	6.5	38.46
0: maj.bsl	-	22.48	3.27	87.51
1: off	<i>WSABIE</i>	25.22	4.50	68.52
1: off	<i>Word2Vec</i>	30.61	4.53	66.96
1: off	<i>TransE</i>	51.13	2.99	83.30
2: reg	<i>WSABIE</i>	35.65	3.14	84.00
2: reg	<i>Word2Vec</i>	41.91	2.81	88.00
2: reg	<i>TransE</i>	<b>66.61</b>	<b>1.93</b>	<b>93.22</b>
3: NN	random	26.89	3.67	77.00
3: NN	<i>WSABIE</i>	27.46	3.59	79.98
3: NN	<i>Word2Vec</i>	30.55	3.27	82.61
3: NN	<i>TransE</i>	<b>67.73</b>	<b>1.83</b>	<b>94.39</b>

Table 5: Performances on relation prediction task.

other point made by system 1 is the fact that it does not involve training on the triples but is still competitive with the strong baseline that leverages the underlying distribution from the triples. This indicates that to some extent the textual frame embeddings still capture useful information for the F2F Prediction Task. In a further step, we also use the embeddings pretrained on the F2F relations of the FN hierarchy (*TransE*), even if in this setting we do not need to calculate prototypical relation embeddings as *TransE* provides embeddings for frames and relations. Thus, system 1 uses the *TransE* embeddings directly to calculate the similarity of the frame embeddings’ vector offset and the relation embeddings. The large improvement in all performance measures shows the strength of knowledge-based embeddings over text-based embeddings and confirms the difficulty of text-based embeddings in reconstructing the F2F relations.

Performance increases with system 2, the softmax regression model involving learning. This shows the effect of training with respect to the F2F relations. It indicates that training should be involved for leveraging the text-based frame embeddings in the F2F Prediction Task. Using embeddings pretrained on the F2F relations of the FN hierarchy (*TransE*) instead, again leads to a large improvement in all performance measures. This confirms that embeddings designed to incorporate the knowledge from the FN hierarchy are better suited for the F2F relation prediction task and it emphasizes the large improvement over the textual embeddings.

Overall, we achieve best results in all performance measures with system 3, the NN approach, in combination with the knowledge-based *TransE* embeddings as input representations. Interestingly, the difference between NN and the regression model is only marginal when using the *TransE* embeddings, indicating the crucial influence of the knowledge-based embeddings and not necessarily the system. Moreover, when using the text-based *WSABIE* and *Word2Vec* the softmax regression model is stronger than the NN, which might be due to little training data. Furthermore, the randomly initialized embeddings for system 3 could be seen as another baseline which is not only beaten by the knowledge-based *TransE* embeddings but also by the text-based *WSABIE* and *Word2Vec* embeddings in systems 2 and 3. This again indicates the capability of the textual frame

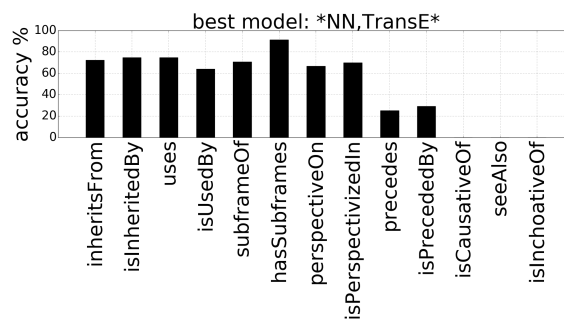


Figure 5: Relation-specific analysis of the best-performing model with respect to accuracy.

embeddings of capturing useful information for the F2F Prediction Task to at least some extent.

The systems could reach higher scores if the split of the data into training and test triples would be done random per relation such that the train and test set have some (random) relation-specific overlap in frames on the position  $f_1$  in the triple. But in this case, it would not clear whether the systems would just perform “lexical memorization” as pointed out by (Levy et al., 2015) when the test set contains partial instances that were in the training set. We leave it for future work to contrast and explore different splits, e.g., random split, zero-overlap by relation or by all relations.

To sum up, on the one hand, the results confirm the conclusions from the exploration in Section 4: the frame embeddings learned on frame-labeled text in the context of other tasks are not able to reliably mirror the F2F relations, not even when used as input representations to a classifier. On the other hand, our results clearly emphasize the influence of the knowledge-based embeddings on the performance of our best-performing system. Thus, we propose this NN architecture in combination with the *TransE* embeddings as the first system for automatic F2F relation annotation for frame pairs in the FN hierarchy.

Figure 5 depicts a relation-specific analysis of the best-performing model showing good performances (above 60% accuracy) for frequent relations, a drop for the less frequent precedence relations and no capability at all in predicting infrequent relations, such as *is\_Causative\_of*, *see\_Also* and *is\_Inchoative\_of*.

### 5.3 Demonstration of Predictions

We generically demonstrate the best-performing system’s prediction for examples of frame pairs which are not annotated so far. Looking back at



$(f_1, f_2)$	top 3 F2F relation predictions
(Biological_urge, Sleep_wake_cycle)	<b>Subframe_of</b> , Inherits_from, Uses
(Biological_urge, Being_awake)	Is_Inherited_by, <b>Precedes, Is_Preceded_by</b>
(Biological_urge, Fall_asleep)	Is_Inherited_by, <b>Precedes, Is_Preceded_by</b>

Table 6: F2F relation predictions of best system.

the motivational example from the beginning, Figure 1 illustrated the incompleteness of the FN hierarchy at the F2F relation level with the example of a possibly missing `precedence` relation from “Being\_aware” to “Biological\_urge” (evoked by the predicate “tired”). Table 6 displays the top 3 F2F relation predictions for the frame pairs around “Biological\_urge” in the figure. The expected F2F relation (printed bold) is indeed amongst the top 3 predictions of the best performing system for this example, even for the `precedence` relation which is rather underrepresented in the data. If this system was used to make suggestions to human expert annotators, they should be informed about the system being biased against the infrequent relations. However, it is hard to do a proper manual evaluation as judging the suggested relations requires expert knowledge of the definitions and annotation best-practices for the F2F relations. We propose using the best-performing system for semi-automatic FN completion on the relation level in cooperation with FN annotation experts. The system can be used to make reasonable suggestions of relations for frame pairs and the final decision could be made by experienced FN annotators. This would be a first step towards improving the incompleteness of F2F relation annotations in FN, which in turn could improve the performance in other tasks that take these F2F relations as input.

## 6 Discussion and Future Work

As the F2F relations of the FN hierarchy did not emerge from frame embeddings learned on frame-labeled text, the F2F relations should be seen as meta-structures not having direct evidence in text. On the one hand, more advanced approaches might be needed to distill F2F relations for frames occurring in raw text, by learning about common-sense knowledge involving frames, and then inferring the implicit relations. Here, it could also be helpful to exploit inter-sentential clues e.g., event chains, to enrich the frame embeddings which so

far are built on sentence-level. On the other hand, the automatic completion of F2F relations can rely on knowledge-based embeddings trained on the hierarchy. To this end, an expert evaluation of the best-performing system’s predictions for frame pairs could give clues for further system improvements. It could also yield an expert upper bound and may pave the way for developing advanced systems using frame embeddings for the prediction of F2F relations. Finally, we plan to investigate the case of FN for embeddings learned on both, frame-labeled texts and F2F relation annotations. By having such a combination, the limitation of the text-based embeddings on frames that have LUs (and hence occur in text) can be overcome as the knowledge-based embeddings also have access to frames without LUs. Last but not least, for different tasks, different representations of frames and relations might be better suited: embeddings purely learned on text, or embeddings purely learned on the FN hierarchy, or a combination of both.

## 7 Conclusion

We raised the question whether text-based frame embeddings naturally mirror F2F relations in the FN hierarchy. We set up the F2F Relation Prediction task as an adaptation of the link prediction task from KGC to the case of FN. Through this task, we quantify the ability of systems and embeddings to predict F2F relations. The F2F Relation Prediction task addresses the need for automatically completing F2F relations that are used in down-stream tasks. Our best-performing system for predicting F2F relations is a NN trained on the FN hierarchy and uses knowledge-based embeddings that by design incorporate the F2F relation. It can be used to suggest more F2F relation annotations in the FN hierarchy. The comparison of our different systems and embeddings reveals insights about the difficulty of reconstructing F2F relations purely from text. We encourage the development of advanced systems and embeddings for the F2F Relation Prediction task.

## Acknowledgments

This work has been supported by the DFG-funded research training group “Adaptive Preparation of Information from Heterogeneous Sources” (AIPHES, GRK 1994/1). We also acknowledge the useful comments of the anonymous reviewers.

## References

- Roni Ben Aharon, Idan Szpektor, and Ido Dagan. 2010. [Generating Entailment Rules from FrameNet](#). In *Proceedings of the ACL 2010 Conference Short Papers*. Association for Computational Linguistics, pages 241–246. <http://www.aclweb.org/anthology/P10-2045>.
- Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. [The Berkeley FrameNet Project](#). In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, Montreal, Quebec, Canada, pages 86–90. <https://doi.org/10.3115/980451.980860>.
- Daniel Bauer, Hagen Fürstenau, and Owen Rambow. 2012. [The Dependency-Parsed FrameNet Corpus](#). In *Proceedings of the 8th Language Resources and Evaluation Conference (LREC 2012)*. Istanbul, Turkey, pages 3861–3867. <http://hdl.handle.net/10022/AC:P:21192>.
- Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. 2012. [Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing](#). In Neil D. Lawrence and Mark Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. PMLR, La Palma, Canary Islands, volume 22 of *Proceedings of Machine Learning Research*, pages 127–135. <http://proceedings.mlr.press/v22/bordes12.html>.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. [Translating Embeddings for Modeling Multi-relational Data](#). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., pages 2787–2795. <http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf>.
- Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. [Learning Structured Embeddings of Knowledge Bases](#). In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*. AAAI Press, AAAI'11, pages 301–306. <http://dl.acm.org/citation.cfm?id=2900423.2900470>.
- Bob Coyne and Owen Rambow. 2009. [LexPar: A Freely Available English Paraphrase Lexicon Automatically Extracted from FrameNet](#). In *Proceedings of the Third IEEE International Conference on Semantic Computing (ICSC 2009)*. pages 53–58. <https://doi.org/10.1109/ICSC.2009.56>.
- Christiane Fellbaum. 1990. [English Verbs as a Semantic Net](#). *International Journal of Lexicography* 3(4):278–301. <https://doi.org/10.1093/ijl/3.4.278>.
- Minwei Feng, Bing Xiang, Michael R Glass, Lidan Wang, and Bowen Zhou. 2015. [Applying Deep Learning to Answer Selection: A Study and An Open Task](#). In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, pages 813–820. <https://doi.org/10.1109/ASRU.2015.7404872>.
- Charles J Fillmore. 1976. [Frame Semantics and the Nature of Language](#). *Annals of the New York Academy of Sciences* 280(1):20–32. <https://doi.org/10.1111/j.1749-6632.1976.tb25467.x>.
- Charles J Fillmore and Collin F Baker. 2001. [Frame Semantics for Text Understanding](#). In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*.
- Lucie Flekova and Iryna Gurevych. 2016. [Supersense Embeddings: A Unified Model for Supersense Interpretation, Prediction, and Utilization](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*. Association for Computational Linguistics, Berlin, Germany, volume Volume 1: Long Papers, pages 2029–2041.
- Silvana Hartmann and Iryna Gurevych. 2013. [FrameNet on the Way to Babel: Creating a Bilingual FrameNet Using Wiktionary as Interlingual Connection](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*. Association for Computational Linguistics, Stroudsburg, PA, USA, volume 1, pages 1363–1373.
- Silvana Hartmann, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. [Out-of-domain FrameNet Semantic Role Labeling](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*. Association for Computational Linguistics, pages 471–482.
- Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. [Semantic Frame Identification with Distributed Word Representations](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 1448–1458. <http://www.aclweb.org/anthology/P14-1136>.
- Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. [SenseEmbed: Learning Sense Embeddings for Word and Relational Similarity](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 95–105. <http://www.aclweb.org/anthology/P15-1010>.

- Omer Levy and Yoav Goldberg. 2014. [Dependency-Based Word Embeddings](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*. The Association for Computer Linguistics, pages 302–308. <https://doi.org/10.3115/v1/P14-2050>.
- Omer Levy, Steffen Remus, Chris Biemann, Ido Dagan, and Israel Ramat-Gan. 2015. [Do Supervised Distributional Methods Really Learn Lexical Inference Relations?](#) In *HLT-NAACL*. pages 970–976. <https://doi.org/10.3115/v1/N15-1098>.
- Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *AAAI*. pages 2181–2187.
- Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9(Nov):2579–2605.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*.
- Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. [Linguistic Regularities in Continuous Space Word Representations](#). In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)*. Association for Computational Linguistics, volume 13, pages 746–751. <https://www.microsoft.com/en-us/research/publication/linguistic-regularities-in-continuous-space-word-representations/>.
- George A Miller. 1990. [Nouns in WordNet: A Lexical Inheritance System](#). *International journal of Lexicography* 3(4):245–264. <https://doi.org/10.1093/ijl/3.4.245>.
- Roberto Navigli and Simone Paolo Ponzetto. 2012. [BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network](#). *Artificial Intelligence* 193:217–250. <https://doi.org/10.1016/j.artint.2012.07.001>.
- Ellie Pavlick, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze, and Benjamin Van Durme. 2015. [FrameNet+: Fast Paraphrastic Tripling of FrameNet](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*. Beijing, China, pages 408–413.
- Pushpendre Rastogi and Benjamin Van Durme. 2014. [Augmenting FrameNet Via PPDB](#). In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. Association for Computational Linguistics, Baltimore, Maryland, USA, pages 1–5.
- Radim Řehůřek and Petr Sojka. 2010. [Software Framework for Topic Modelling with Large Corpora](#). In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, pages 45–50. <http://is.muni.cz/publication/884893/en>.
- Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Scheffczyk. 2006. [FrameNet II: Extended Theory and Practice](#). Distributed with the FrameNet data.
- Ming Tan, Bing Xiang, and Bowen Zhou. 2015. [LSTM-based Deep Learning Models for Non-factoid Answer Selection](#). *arXiv preprint arXiv:1511.04108*.
- Jason Weston, Samy Bengio, and Nicolas Usunier. 2011. [WSABIE: Scaling Up to Large Vocabulary Image Annotation](#). In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*. AAAI Press, Barcelona, Catalonia, Spain, IJCAI'11, pages 2764–2770. <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-460>.