Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 976–986
Marseille, 20-25 June 2022

Frame Shift Prediction

Zheng-Xin Yong[1,†], Patrick D. Watson[2], Tiago Timponi Torrent[3], Oliver Czulo[4], Collin F. Baker[5]
[1]Brown University [2]Minerva University [3]Federal University of Juiz de Fora
[4]University of Leipzig [5]International Computer Science Institute
zheng_xin_yong@brown.edu, pwatson@uni.minerva.edu, tiago.torrent@ufjf.edu.br,
czulo@uni-leipzig.de, collinb@icsi.berkeley.edu

Abstract

Frame shift is a cross-linguistic phenomenon in translation which results in corresponding pairs of linguistic material evoking different frames. The ability to predict frame shifts would enable (semi-)automatic creation of multilingual frame annotations and thus speeding up FrameNet creation through annotation projection. Here, we first characterize how frame shifts result from other linguistic divergences such as translational divergences and construal differences. Our analysis also shows that many pairs of frames in frame shifts are multi-hop away from each other in Berkeley FrameNet's net-like configuration. Then, we propose the Frame Shift Prediction task and demonstrate that our graph attention networks, combined with auxiliary training, can learn cross-linguistic frame-to-frame correspondence and predict frame shifts.

Keywords: frame semantics, frame shifts, FrameNet, Multilinguality, Graph Neural Networks

1. Introduction

Frame Semantics is an approach to meaning that characterizes the background knowledge against which linguistic entities are understood via systems of interrelated concepts called *frames* (Fillmore, 1982). Frames are evoked by *lexical units* (LUs) and involve participants and props characterized as *frame elements* (see Figure 1 for example annotations in English and Brazilian Portuguese). The Berkeley FrameNet (BFN) project establishes a general-purpose resource for frame semantic descriptions of English (Ruppenhofer et al., 2016) and is widely adopted for many NLP applications. BFN has successfully been adapted for other languages such as Brazilian Portuguese (Torrent and Ellsworth, 2013; Torrent et al., 2018b) and German (Burchardt et al., 2006; Boas and Ziem, 2018). The success of these works implies that some frames are applicable across different languages, while others can be adapted to fit language specificities (Gilardi and Baker, 2018; Baker and Lorenzi, 2020).

Nonetheless, researchers working with non-English FrameNets find that differences in lexical-constructional patterns between languages result in *frame shifts* (Subirats and Sato, 2003; Litkowski, 2009; Padó and Lapata, 2009; Czulo, 2017; Lindén et al., 2019; Giouli et al., 2020; Ohara, 2020; Ellsworth et al., 2021). Because FrameNet frames are sensitive to variations e.g. in valency or syntactic configuration, when a given sentence is translated from one language to another, it is frequently the case that the frames evoked will be significantly different for orginals and translations. In a test set of 30 sentences, Torrent et al. (2018a) find a direct frame correspondence between sentences in English and their translation to Brazilian Portuguese of only 0.51. Here, as shown in Figure 1, we focus on a type of frame shift where, in a pair of parallel sentences, the corresponding LUs evoke two different frames. The issue of frame shift is of major importance for projecting annotations from one language to another and for bootstrapping FrameNets from parallel corpora, and we address this issue by using graph neural networks to predict frame shifts.

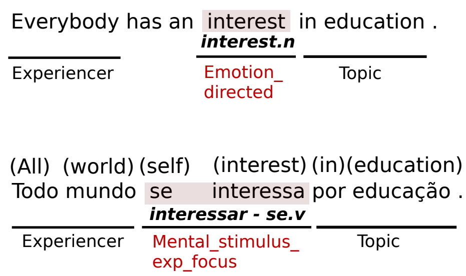

Figure 1: An example of frame shift in a pair of English and Brazilian Portuguese parallel sentences. The frame-evoking lexical units are bold and italicized, the frames are annotated under them and colored in red, and the frame elements are annotated in black.

In summary, this paper makes the following contributions: (a) we present a case study of frame shifts and identify their possible causes; (b) we propose a new task, Frame Shift Prediction (FSP), for predicting frame shifts in parallel texts; (c) our empirical results show that representing frames with graph attention networks (GAT) outperforms existing multilingual frame embedding methods in FSP.

2. Related Work

Multilingual FrameNet. Currently, there are more than 20 FrameNet (FN) resources in various lan-

† Work done when studying at Minerva University as an undergraduate.

guages.[1] One way to automatically expand FN to a new language is by projecting frame annotations (Johansson and Nugues, 2006). This is motivated by the fact that many frames constitute appropriate characterizations of events and situations that can be applied across difference languages, especially those related to basic human experiences (like eating and sleeping) and common cultural or social practices (like commercial transactions and employment). Gilardi and Baker (2018) show that frames from FN databases of different languages, developed independently, can, in principle, be aligned.

While FNs in each language use similar or even identical frames for annotation, recent work reports variations in how the translators "frame" the sentences on the conceptual level (Subirats and Sato, 2003; Litkowski, 2009; Padó and Lapata, 2009; Čulo, 2013; Lindén et al., 2019; Giouli et al., 2020; Ohara, 2020). Ellsworth et al. (2021) report that even when there exists an available equivalent frame for annotating a parallel sentence, the translated phrase may evoke a different target frame. Our work sheds light on the relationship between frame shifts and linguistic variation in translations.

Graph Neural Networks. Graph neural networks (GNNs) have been proven successful in encoding relational data and graphs in NLP tasks, such as knowledge graph completion (Shang et al., 2019; Zhang et al., 2020), syntactic and semantic parsing (Marcheggiani and Titov, 2017; Bogin et al., 2019; Gu et al., 2021). A graph consists of a set of nodes connected to one another by a set of relations, and the basic idea behind GNNs is to learn node embeddings that reflect the structure of the graph (Hamilton et al., 2017). In practice, GNNs learn the embedding for a node by aggregating the embeddings of its neighbors iteratively.

Frames in FN are connected by different types of frame-to-frame relations such as *Inheritance* or *Precedence* (in a temporal sense). Therefore, GNNs are ideally suited to modeling frames because of their ability to capture those relational dependencies. Li et al. (2017) and Suhail and Sigal (2019) applied GNNs to learn the dependencies between verb and frame-semantic roles for situation recognition. Here, we use graph attention network (Veličković et al., 2018), which is a type of GNNs, to model frame relations and predict frame shifts.

3. Linguistic and Quantitative Analysis of Frame Shifts

3.1. Frame Shifts Dataset

We create the dataset for frame shifts from the Global FrameNet Shared Annotation Task (Torrent et al., 2018a), which has been devised to assess whether frames in BFN 1.7 suit the semantics of LUs in different languages. At the time of the study, the English, Brazilian and German datasets were among the most comparable in terms of annotation coverage and could be assessed by the research team involved in the study. Given an English sentence and its translation in German (DE) and Brazilian Portuguese (PT), we first construct the word-to-word correspondence with Fast Align (Dyer et al., 2013) and then extract the corresponding LUs. We also use the Open Multilingual WordNet (Bond and Paik, 2012) to filter out false positives, that is, LUs which are not translation equivalents. In the end, we extract 95 EN-DE and 316 EN-PT annotation pairs for FSP. Frame shifts are found in 36% of the EN-DE and 22.4% of the EN-PT pairs.

3.2. Translational divergences

We rely on the model for description or translational divergences as developed by Dorr (1994). Next, we present examples of frame shifts in our dataset for each of the divergence categories.

Categorial Categorial divergence happens when two languages use words of different parts-of-speech to express the same meaning. Figure 1 illustrates the example of frame shift caused by categorial divergence: the English noun *interest* in the phrase *has an interest* corresponds to the Portuguese verb phrase *se interessa ("to interest oneself")*. They evoke different frames as the former refers to the feeling of interest, whereas the latter refers to the evocation of an emotional response in the EXPERIENCER to the TOPIC.

Conflational/Inflational Conflational divergence occurs when two or more words in one language are translated into one word in another language, whereas inflational divergence is the opposite. In Figure 2, the inflational divergence splits the English word *everywhere* into two Portuguese words and causes the translated frame-evoking counterpart *lugar (place)* to lose the conceptual relativity to other locations.

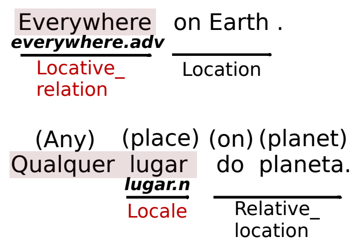

Figure 2: Frame shift due to inflational divergence.

Lexical The unavailability of an exact translation for a construction in a language leads to lexical divergence. In frame semantics, divergence of LUs often causes divergence in meaning. For instance, as shown in Figure 3, the German translation for the English phrase *play out* is *enden (end)*, which differs on the dimension of the realization of the terminal state.

[1] See a potentially non-exhaustive list under www.globalframenet.org.

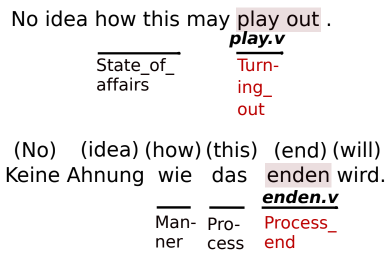

Figure 3: Frame shift due to lexical divergence.

Structural This divergence happens when verb arguments result in different syntactic configurations. In Figure 4, the English verb *move* does not take a reflexive direct object so it evokes the `Motion` frame, in which the subject of motion is the THEME (entity that changes location). On the other hand, the Portuguese verb *mexer (move)* is adjacent to the reflexive particle *se (self)*, which is interpreted as the AGENT moving his/her body; therefore, the verb evokes the `Body_movement` frame.

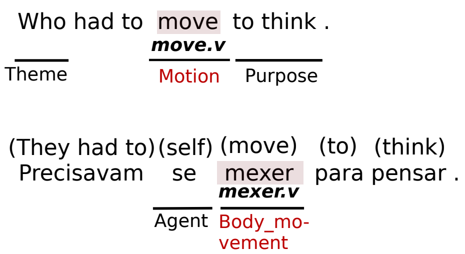

Figure 4: Frame shift due to structural divergence.

Thematic and Head Swapping We did not observe frame shifts caused by thematic divergence (inversion of semantic roles) and head swapping (inverted direction of the dependency relations) in our dataset.

3.3. Construal Differences

We find that translational divergences alone (Section 3.2) are insufficient to account for all instances of frame shift. The reason is that the linguistic expressions can be almost identical in the semantic content but differ in the decoding of their meanings along certain dimensions of construal (Verhagen and others, 2007; Trott et al., 2020). The following analysis of the effects of construal operations is by no means exhaustive.

Resolution As lexical categories form taxonomic hierarchies consisting of various levels of specificity (e.g., cat < mammal < animal < organism), language users can express a concept with different degrees of granularity. Therefore, the expressions can evoke different frames. In Figure 5, the lexical item *said* is more schematic compared to the word *perguntei (ask)* where the former denotes the generic action of communicating a message, whereas the latter provides additional information about the nature of the message.

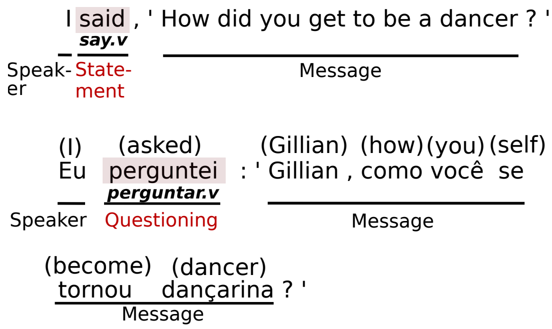

Figure 5: Frame shift due to differences in resolution.

Prominence Prominence refers to the relative focus of attention on elements against the rest in a scene. In Figure 6, while both sentences characterize the style of thinking with adverbial phrases, the linguistic expression *in sound* makes explicit the auditory sensation. In contrast, the Portuguese adverb *auditivamente (aurally)* foregrounds the thinking action *Pensamos (We think)*, which is labeled with the frame element COMPARISON_ACTIVITY that indicates the activity characterized by the `Manner` frame.

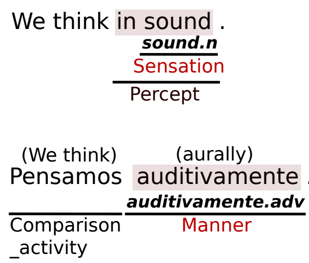

Figure 6: Frame shift due to differences in prominence.

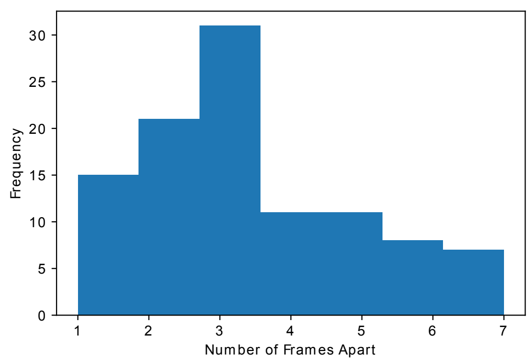

Figure 7: Distribution of the number of nodes apart for frame shifts.

3.4. Quantitative Analysis of Frame Shifts

Figure 7 demonstrates a unimodal distribution of the distance in the FN network of diverging frames. Diverging frames do not necessarily exhibit *first-order* frame-to-frame relations; many are more than one hop away from each other (see Figure 8). Most of the frame pairs are connected to each other. Even though not the full potential of connections FN is exploited to date, only two out of the 104 pairs of diverging frames do not have a path connecting them. In other words, frame shifts can be accounted for by the net-like configuration of FN, which is similar to the conclusion drawn by Torrent et al. (2018a).

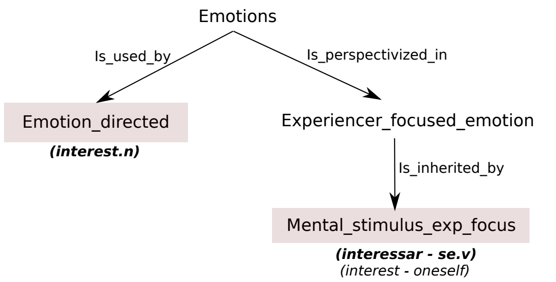

Figure 8: Visualization of the path connecting the pair of frames (`Emotion_directed` and `Mental_stimulus_exp_focus`) in frame shift in Figure 1.

4. Frame Shift Prediction (FSP)

4.1. Task Description

FSP is a multi-class classification task. Given a labeled frame $f_{src} \in \mathcal{F}$, where $\mathcal{F}$ denotes the set of all 1224 frames in BFN 1.7 (Ruppenhofer et al., 2016), for a lexical unit LU_{src} in the source English sentence, our goal is to predict the frame $f_{tgt} \in \mathcal{F}$ for the corresponding lexical unit LU_{tgt} in the target German and Brazilian Portuguese sentences. Following the Global FrameNet Shared Annotation Task (Torrent et al., 2018a), we use the same $\mathcal{F}$ for all three languages during training and inference.

4.2. Proposed Model

Figure 9 shows our proposed approach to FSP[2]. We propose using graph attention networks (GATs) (Veličković et al., 2018) to represent frames to capture the relational structure of FN. The attention mechanism in GATs learns the weights of neighboring nodes according to node similarity and improves semantic clustering of frames (Wang et al., 2019).

We propose using GAT because it models the interactions between frames through learnable scalar coffecients. In other words, it allows us to model non-structure-driven interactions among frames in frame shifts. In FSP, the structural components are the frame-to-frame relationships, whereas the non-structural counterparts, which include frame-evoking LUs, translational divergences and construal differences, are absent from our FN network.

[2]https://github.com/yongzx/Semantic-Frame-Shift

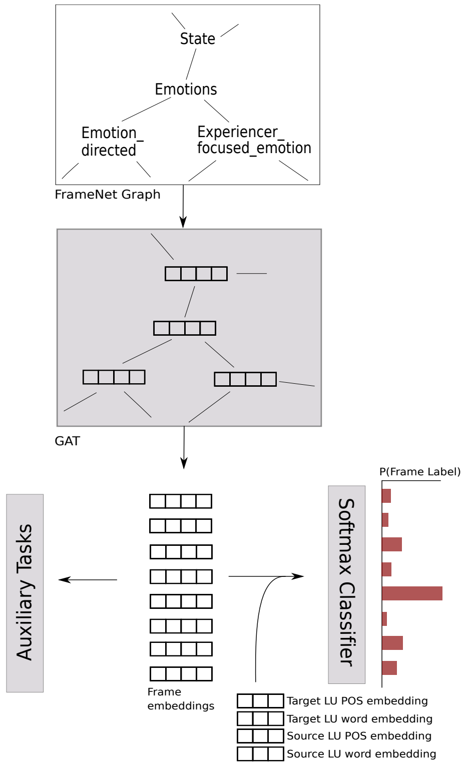

Figure 9: Proposed approach for FSP using graph attention networks (GAT) and auxiliary training to represent frames.

4.2.1. Graph Initialization

We define a graph $\mathcal{G} = (V, E, H)$ composed of a set of graph nodes V, node representations $H = (h_1, ..., h_{|V|})$ and a set of directed edges $E = (E_1, ..., E_K)$ where K is the number of edges. Each of the 1224 nodes corresponds to a frame in $\mathcal{F}$. Each directed edge represents a frame-to-frame relation, and we do not distinguish between their types since, for the purposes of this study, it does not matter whether an edge captures a generic-specific relation (such as $Inheritance$) or causative-stative relation (such as $Causative\_of$), as both can be due to differences on how source and target language encode meaning. We initialize the nodes with multilingual LASER sentence representations of the frame definitions, where sentences from different languages are mapped into the same embedding space through a BiLSTM encoder (Artetxe and Schwenk, 2019).

4.2.2. Graph Attention Network (GAT)

GAT consists of a stack of graph attentional layers (Veličković et al., 2018). Each layer applies multi-headed self-attention mechanism to transform its inputs, which is a set of node representations $H = (h_1, ..., h_{|V|}), h_i \in \mathbb{R}^D$ (where $|V|$ is the number of nodes, and D is the dimension of the node representation), to a new set of node representations, $H' = (h'_1, ..., h'_{|V|}), h'_i \in \mathbb{R}^{D'}$, of potentially different dimension D'. For m-th attention head, the output fea-

Data	# lus	# frames	# sents	langs	tasks
Semantic Frame Shift Prediction Dataset (Section 3.1)	952	179	788	de, en, pt	Frame Shift Prediction
Berkeley FrameNet 1.7 (Ruppenhofer et al., 2016)	8404	1224	174527	en	Link Prediction Path Length Prediction Binary Frame Prediction Frame Label Reconstruction
Multilingual frame-annotated corpus (Johannsen et al., 2015)	7558	729	18442	bg, da, de, el, en, es, fr, it, sv	Binary Frame Prediction Frame Label Reconstruction

Table 1: Statistics of the all datasets in training GAT for FSP.

ture for a node i is the linear combination of the input features of the node's first-order neighbors j (including itself i), weighted by the normalized attention coefficients $\alpha_{ij}^{(m)}$. Then, the output features (which may undergo non-linear transformation σ) from all attention heads are concatenated to produce h_i'. The transformation from h_i to h_i' is as follows.

$$e_{ij}^{(m)} = \text{LeakyRELU}(\mathbf{a}^{(m)T}[\mathbf{W}^{(m)}h_i \| \mathbf{W}^{(m)}h_j])$$

$$\alpha_{ij}^{(m)} = \frac{\exp(e_{ij}^{(m)})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik}^{(m)})}$$

$$h_i' = \big\Vert_{m=1}^{M} \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(m)} \mathbf{W}^{(m)} h_j \right)$$

We follow the basic architecture of Veličković et al. (2018) and implement a two-layer GAT model. Differing from Veličković et al. (2018), the output layer is a linear transformation layer followed by the softmax layer to include the LUs when the model calculates the probability for each frame $f_{tgt} \in \mathcal{F}$. The probability distribution $P(f_{tgt}|f_{src}, LU_{src}, LU_{tgt})$ is computed as follows:

$$\text{softmax}(W([h_{f_{src}}; w_{src}; w_{tgt}; p_{src}; p_{tgt}]) + b)$$

where $h_{f_{src}}$ is the node representation of f_{src}; w_{src} and w_{tgt} are the pretrained mBERT embeddings of LU_{src} and LU_{tgt} respectively; p_{src} and p_{tgt} are randomly initialized embeddings for parts-of-speech tags; W and b are the trainable parameters of the output layer. The whole system is trained with the cross-entropy loss function.

4.2.3. Graph Regularization

To improve the generalization ability of GAT and prevent it from overfitting, we further introduce two graph regularization techniques: NodeNorm (Zhou et al., 2020) and DropEdge (Rong et al., 2020). NodeNorm normalizes the embedding of every node in each layer by its own mean and standard deviation. It increases the smoothness of the model w.r.t. node features, so similar frames share similar representations. On the other hand, DropEdge randomly drops out a certain rate of edges of the input graph for each training time. As an unbiased data augmentation technique (Rong et al., 2020), it enables a random subset aggregation instead of the full aggregation during GAT training, thus better capable of preventing overfitting.

4.3. Auxiliary Training

We propose several auxiliary training tasks for GAT for two purposes. First, tasks 1 and 2 train our model to explicitly learn the connections among frames since we observe that 102 out of 104 pairs of diverging frames in our dataset are connected in FN. On the other hand, tasks 3 and 4 help our GAT associate frames with LUs since the relational structure of FN does not inform GAT how LUs evoke frames. These two tasks enable GAT model to implicitly learn the information about frame-evoking LUs (instead of explicitly representing LUs as nodes in the FrameNet graph).

1. **Link Prediction.** A binary classification problem where the model predicts if there is a frame-to-frame relation between two semantic frames, $(f_1, f_2) \in E$.

2. **Path Length Prediction.** A regression task where the model predicts the number of edges between two frames, $(f_1, f_2) \in E$.

3. **Binary Frame Prediction.** A binary classification task where, given a pair of randomly chosen frame f and an LU LU_x, the model predicts if LU_x evokes f.

4. **Frame Label Reconstruction.** A multi-class classification task where some of the frame labels f for annotated sentences are randomly "perturbed" into incorrect frame labels f_x with a probability p, and the model is trained to recover the correct frame f.

Task 2 uses the mean squared error as the objective function whereas the rest uses cross entropy loss. The combined loss for training GAT is the sum of losses from the auxiliary tasks and the primary FSP task,

Tasks	# Layers	Layer Parameters	Objective Functions
Frame Shift Prediction	1	$\mathbb{R}^{D_f+2\times(D_w+D_{pos})}$	Cross-Entropy Loss
Link Prediction	1	$\mathbb{R}^{2\times D_f} \to \mathbb{R}^2$	Cross-Entropy Loss
Path Length Prediction	2	$\mathbb{R}^{2\times D_f} \to \mathbb{R}^{1024} \to \mathbb{R}$	Mean-Squared Error
Binary Frame Prediction	1	$\mathbb{R}^{D_f+D_w+D_{pos}} \to \mathbb{R}^2$	Cross-Entropy Loss
Frame Label Reconstruction	1	$\mathbb{R}^{D_f+2\times(D_w+D_{pos})} \to \mathbb{R}^{1224}$	Cross-Entropy Loss

Table 2: Parameters of output layers for frame shift prediction and auxiliary tasks.

weighted by the homoscedastic uncertainty of each task (Kendall et al., 2018).

4.4. Datasets

Table 1 shows the statistics of the datasets used for FSP (primary task) and the auxiliary tasks. FSP experiments use the frame shifts dataset described in Section 3.1. On the other hand, auxiliary tasks 1 and 2 use the frame-to-frame relationships information in BFN 1.7 (Ruppenhofer et al., 2016) for training, whereas tasks 3 and 4 use the lexicographic annotations for the LUs in BFN 1.7 and the multilingual frame-annotated corpus (Johannsen et al., 2015).

4.5. Experimental Setup

We selected hyperparameters for GAT via Bayesian optimization on the Frame Label Reconstruction task and used the same hyperparameters for FSP. The resulting first layer of GAT consists of 9 attention heads computing 109 features each, and the second layer 10 attention heads 256 features each. The final softmax classifier only has a single linear transformation layer that receives 1824 input features from GAT and lexical units and outputs 1224 features (as frame classes).

Table 2 shows the details of the final output layers for FSP and auxiliary tasks. We optimize their hyperparameters, namely the number of layers and the hidden features' dimension, on the respective auxiliary tasks before reusing hyperparameters for FSP. Here, we use $D_f = 256$ to denote the dimension of frame representations, $D_{pos} = 16$ the dimension of POS tag embeddings, and $D_w = 768$ the dimension of mBERT embeddings of LUs.

We represent the LUs with mBERT embeddings (Devlin et al., 2019). Parts-of-speech tags are represented with randomly initialized embeddings of dimension 16. We train the model using the Adam optimizer with a batch size of 512, learning rate of 0.005, and weight decay of $\lambda = 0.0005$. For each setting, we perform five runs of nested five-fold cross-validation on a Nvidia Tesla P100 GPU and report their average F1 scores as well as their standard deviations. The inner cross-validation is used to find the suitable number of training epochs. Training and evaluation model take approximately three hours.

4.6. Baselines

Since our paper is the first attempt to predict frame shifts, we do not have other classifiers to directly compare with. As we are proposing a novel frame representation method for the multilingual task, we use other recent multilingual frame representation methods as baselines. The frame representations obtained from the baselines are concatenated with the word embeddings and part-of-speech tag embeddings of the LUs. Subsequently, the tensors are passed through a single linear transformation layer and a softmax final layer for classification.

Direct Transfer. This method assumes that frame shifts are absent and projects the frame labels without changes. In other words, $f_{tgt} = f_{src}$.

Randomized Frame Embeddings. This method represents each frame with a trainable, randomized embedding of dimension 256. The embeddings are cross-lingual because they are trained with the FSP dataset.

Sikos and Padó (2018). The authors embedded English and German frames from BFN 1.5 and SALSA corpus in the same vector space. In our setup, the frame embeddings are only used for FSP between English and German. We directly transfer the frames that are not embedded by the authors.

Sikos and Padó (2019). The authors embedded frames with the pretrained BERT model with and without finetuning. Without finetuning, frame embeddings are the unweighted centroid of the contextualized embeddings of the corresponding LUs. Otherwise, the frame embeddings are finetuned to predict frame labels for each word token in the full-text annotations. In our setup, we use the multilingual BERT (mBERT) model to represent frames and finetune them on all the datasets in Table 1 following the authors' instructions.

Baker and Lorenzi (2020) The authors created frame embeddings using the unweighted centroid of the FastText embeddings of the LUs. In our experiments, we embed the LUs in the source and target sentences with FastText representations.

4.7. Evaluation

To ensure a robust evaluation of models on a small FSP dataset, we evaluate each model with the five-fold nested cross-validation (CV) method. It separates the CV fold used for model development (including feature selection and parameter tuning) from the one used for model evaluation; therefore, the performance estimates are unaffected by and unbiased to the sample sizes (Vabalas et al., 2019).

We evaluate FSP with the top-5 F1 score. As long

Models	EN → PT	EN → DE	EN → (PT + DE)
Direct Transfer	77.5	63.9	74.3
Randomized Embeddings	44.2 (± 3.1)	37.7 (± 2.4)	39.8 (± 2.9)
Sikos and Padó (2018)	-	55.2 (± 3.5)	-
mBERT (w/o finetuning) (Sikos and Padó, 2019)	53.4 (± 1.7)	47.8 (± 4.3)	49.3 (± 2.4)
mBERT (with finetuning) (Sikos and Padó, 2019)	71.5 (± 2.3)	65.6 (± 0.9)	68.7 (± 2.2)
FastText (Baker and Lorenzi, 2020)	54.8 (± 1.1)	43.7 (± 2.6)	50.1 (± 1.8)
GAT (w/o auxiliary training)	57.1 (± 1.3)	40.2 (± 1.8)	55.9 (± 4.1)
GAT (with auxiliary training)	**83.1** (± 1.5)	**68.0** (± 1.9)	**79.7** (± 2.0)

Table 3: 5-Fold nested cross-validation with top-5 F1 scores (± standard deviation) for each model in predicting frame shifts. $X \rightarrow Y$ denotes that projecting frames from language X to language Y (EN: English, PT: Portuguese, DE: German).

as the correct frame label is among the top-5 most probable predicted frame shifts——hence the term "top-5"——we consider the model to have successfully predicted the frame shift. The reason for this metric choice is that the size of our FSP dataset is much smaller than the number of classes (1224 frames with varying granularity) in this experiment. As a result, the models have to perform FSP on frames they have not seen before in FSP training. Furthermore, the frame labels vary with respect to granularity, which can cause the model to suffer from class ambiguity. Therefore, we argue that the top-5 F1 score gives a more realistic performance evaluation.

5. Discussion

5.1. Frame Shift Prediction

Table 3 illustrates the performance of different models in FSP. Embedding-based baseline models generally perform worse than the Direct Tansfer approach. Hence, we argue that simply aggregating embeddings of LUs from pretrained models to represent frames cannot capture the fine-grained semantic distinctions between frames. One solution is to finetune the embeddings of LUs by learning to map them to the frames they evoke. The reason is that, after finetuning, the embeddings of LUs are more similar if the LUs evoke the same frame, and less similar if they evoke different frames. In other words, finetuned embeddings better encode the similarities and variations between frames. As seen in Table 3, the finetuned embeddings of mBERT (Sikos and Padó, 2019) demonstrate the best FSP performance among the embedding-based baselines.

Our use of GAT and auxiliary training to represent frames achieves the best performance. We want to highlight the differences in our approach: our model treats frames and LUs as independent units and learns their relations through the auxiliary tasks 3 and 4, as opposed to representing frames as a combination of LU representations (Peng et al., 2018; Sikos and Padó, 2018; Sikos and Padó, 2019; Popov and Sikos, 2019; Alhoshan et al., 2019; Baker and Lorenzi, 2020). Furthermore, our approach can represent so-called non-lexical frames – i.e. frames that are assumed to be present in the conceptual system but are not linguistically realized in a language – because of the message propagation from their surrounding nodes. The main takeaway here is that learning the relational structure of FN enables FSP.

We see a significant decrease in performance when auxiliary training is absent. Auxiliary training boosts performance for the FSP in Brazilian Portuguese (EN → PT) even though the auxiliary datasets (see Table 1) do not contain Brazilian Portuguese sentences. This shows that the auxiliary learning successfully encodes cross-linguistic information about frames — it helps the GAT model to learn the frame-to-frame and LU-to-frame relationships, thus creating more generalized and meaningful frame representations.

The study limitation is the small sample size. Global FrameNet (Torrent et al., 2018a) is an initiative involving several languages, and the shared annotation task requires fine grained annotation of the parallel corpora. Nonetheless, it is the dataset the FN community currently uses for studying frame adequacy across languages.

5.2. Visualization of Frame Representations

Figure 10 illustrates the semantic frame representations with UMAP dimensionality reduction (McInnes et al., 2018). To obtain the frame representations, we average the node representations from five different GAT models trained in the five-fold nested cross-validation. We foreground two clear clusters of frames, where one is related to commerce and the other related to occupation, to show that GAT learns the underlying relationships between frames.

There are two main findings. First, the frames in the clusters are connected to one another in FN; for instance, `Member_of_military` inherits from `People_by_vocation`, and `Being_employed` is a subframe of `Employee_scenario`. Second, we obtain similar representations for frames that share a domain but are not connected. `Price_per_unit` and `Commercial_goods-transfer` are both conceptually related to commerce, and there is no path connecting them in FN, but they are still clustered together

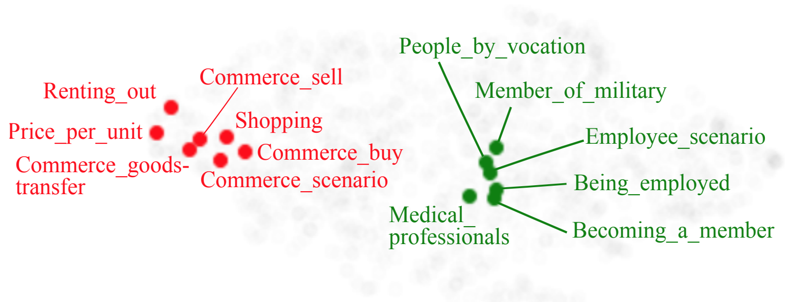

Figure 10: UMAP visualization of semantic frame vectors learned by our proposed graph attention networks model. We color-code the frames related to commerce (in red) and occupation (in green) to show the clustering of frames.

because of their semantic associations. The clustering is thus a sign that the GAT model has successfully learned the relationships between frames also beyond the explicit connections made in FrameNet.

5.3. Ablation Study

Models	F1	Δ
GAT (all Auxiliary Tasks)	79.7	-
– Link Prediction	76.1	-3.6
– Path Length Prediction	75.5	-4.2
– Binary Frame Prediction	69.4	-10.3
– Frame Label Reconstruction	63.7	-16.0
– All Auxiliary Tasks	55.9	-23.8

Table 4: Ablation study on auxiliary tasks.

Table 4 shows the result of an ablation study of auxiliary tasks. We conclude that the auxiliary tasks are suitable for learning frame shifts as ablation of any task hurts FSP. Out of the four tasks, Frame Label Reconstruction is the most helpful for learning FSP. This could be due to shared task structure and classifier parameters between the auxiliary task and the FSP task, as both tasks compute the posterior probability of frame labels for a LU given a prior (source) frame. In contrast, Link Prediction contributes the least to FSP. This is possibly due the presence of frames that are not immediate neighbors in FrameNet, making supervised learning of frame-to-frame relations less informative.

6. Limitations and Future Work

Since the main limitation of our study is the small amount of FSP data, we encourage future work to explore methods to augment the training data to better model frames' relationship and methods to incorporate relational information about frames into zero-shot or few-shot learning techniques such as prompt learning (Lin et al., 2021; Sanh et al., 2022; Wei et al., 2022), where we perform a task with zero or few demonstration examples and task instructions (i.e., prompts).
Our work focuses on two langauge pairs: EN-DE and EN-PT. Therefore, further research will be necessary into the extent to which the methodology can be transferred to other language pairs. Our study involves typology closely related languages, and we expect our methods to be transferable to other closely related languages. The divergence categories we borrowed from Dorr (1994) present a formal way of describing some types of shifts observed, at least some of which also cover phenomena seen in the interaction between typologically more distant languages such as the head-switch in translations between English and Japanese as, e.g., described by Ohara (2020).[3] In (Czulo, 2017; Ohara, 2020), i.a., the authors discuss potential factors of frame shifts which range from formal factors to such factors as register conventions or differences in framing preferences between languages; this is an open list of factors which may not be as easily formalized by means of above-referenced categories. Further studies should include both testing for yet undescribed frame shift factors as well as divergence categories.

7. Conclusion

Our research analyzes frame shifts that co-occur with variation on the morpho-syntactic level or may come down to differences in construal. We also pioneer a new task, Frame Shift Prediction (FSP), and show that Graph Attention Networks (GATs) can predict frame shifts by learning FrameNet's relational structure and the lexical units.

8. Acknowledgements

This material is based in part upon work supported by the U.S. National Science Foundation under Grant No. (1629989). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The annotation of the TED Talk for Brazilian Portuguese and German were made possible by the CAPES/PROBRAL Grant No. 88881.144044/2017-00.

[3]Note that head switches in translation also occur, e.g., between English and German.

9. Bibliographical References

Alhoshan, W., Batista-Navarro, R., and Zhao, L. (2019). Semantic frame embeddings for detecting relations between software requirements. In *Proceedings of the 13th International Conference on Computational Semantics - Student Papers*, pages 44–51, Gothenburg, Sweden, May. Association for Computational Linguistics.

Artetxe, M. and Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

Baker, C. F. and Lorenzi, A. (2020). Exploring crosslinguistic frame alignment. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 77–84, Marseille, France, May. European Language Resources Association.

Boas, H. C. and Ziem, A. (2018). Constructing a constructicon for german. *Constructicography: Constructicon development across languages*, 22:183.

Bogin, B., Berant, J., and Gardner, M. (2019). Representing schema structure with graph neural networks for text-to-SQL parsing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4560–4565, Florence, Italy, July. Association for Computational Linguistics.

Bond, F. and Paik, K. (2012). A Survey of WordNets and their Licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue. 64–71.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The SALSA corpus: a German corpus resource for lexical semantics. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May. European Language Resources Association (ELRA).

Čulo, O. (2013). Constructions-and-frames analysis of translations: The interplay of syntax and semantics in translations between English and German. *Constructions and Frames*, 5(2):143–167.

Czulo, O. (2017). Aspects of a primacy of frame model of translation. In S. Hansen-Schirra, et al., editors, *Empirical modelling of translation and interpreting*, number 6 in Translation and Multilingual Natural Language Processing, pages 465–490. Language Science Press, Berlin.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Dorr, B. J. (1994). Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20(4):597–633.

Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia, June. Association for Computational Linguistics.

Ellsworth, M., Baker, C., and Petruck, M. R. L. (2021). FrameNet and typology. In *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP*, pages 61–66, Online, June. Association for Computational Linguistics.

Fillmore, C. J., (1982). *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Gilardi, L. and Baker, C. (2018). Learning to align across languages: Toward multilingual framenet. In Tiago Timponi Torrent, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, may. European Language Resources Association (ELRA).

Giouli, V., Pilitsidou, V., and Christopoulos, H. (2020). Greek within the global FrameNet initiative: Challenges and conclusions so far. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 48–55, Marseille, France, May. European Language Resources Association.

Gu, M., Gu, Y., Luo, W., Xu, G., Yang, Z., Zhou, J., and Qu, W. (2021). From text to graph: a general transition-based amr parsing using neural network. *Neural Computing and Applications*, 33(11):6009–6025.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, 40(3):52–74.

Johannsen, A., Martínez Alonso, H., and Søgaard, A. (2015). Any-language frame-semantic parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2062–2066, Lisbon, Portugal, September. Association for Computational Linguistics.

Johansson, R. and Nugues, P. (2006). A FrameNet-based semantic role labeler for Swedish. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 436–443, Sydney, Australia, July. Association for Computational Linguistics.

Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

Li, R., Tapaswi, M., Liao, R., Jia, J., Urtasun, R., and Fidler, S. (2017). Situation recognition with graph neural networks. In *Proceedings of the IEEE In-*

ternational Conference on Computer Vision, pages 4173–4182.

Lin, X. V., Mihaylov, T., Artetxe, M., Wang, T., Chen, S., Simig, D., Ott, M., Goyal, N., Bhosale, S., Du, J., Pasunuru, R., Shleifer, S., Koura, P. S., Chaudhary, V., O'Horo, B., Wang, J., Zettlemoyer, L., Kozareva, Z., Diab, M. T., Stoyanov, V., and Li, X. (2021). Few-shot learning with multilingual language models. *CoRR*, abs/2112.10668.

Lindén, K., Haltia, H., Laine, A., Luukkonen, J., Piitulainen, J., and Väisänen, N. (2019). Finntransframe: translating frames in the finnframenet project. *Language Resources and Evaluation*, 53(1):141–171.

Litkowski, K. C. (2009). Multilingual Framenets in Computational Lexicography: Methods and Applications.

Marcheggiani, D. and Titov, I. (2017). Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark, September. Association for Computational Linguistics.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29):861.

Ohara, K. (2020). Finding corresponding constructions in English and Japanese in a TED talk parallel corpus using frames-and-constructions analysis. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 8–12, Marseille, France, May. European Language Resources Association.

Padó, S. and Lapata, M. (2009). Cross-Lingual Annotation Projection of Semantic roles. *Journal of Artificial Intelligence Research*, 36(1).

Peng, H., Thomson, S., Swayamdipta, S., and Smith, N. A. (2018). Learning joint semantic parsers from disjoint data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1492–1502, New Orleans, Louisiana, June. Association for Computational Linguistics.

Popov, A. and Sikos, J. (2019). Graph embeddings for frame identification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 939–948, Varna, Bulgaria, September. INCOMA Ltd.

Rong, Y., Huang, W., Xu, T., and Huang, J. (2020). DropEdge: Towards Deep Graph Convolutional Networks on Node Classification. In *International Conference on Learning Representations*.

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., Baker, C. F., and Scheffczyk, J. (2016). FrameNet II: Extended Theory and Practice. *FrameNet Project*.

Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Scao, T. L., Raja, A., Dey, M., Bari, M. S., Xu, C., Thakker, U., Sharma, S. S., Szczechla, E., Kim, T., Chhablani, G., Nayak, N., Datta, D., Chang, J., Jiang, M. T.-J., Wang, H., Manica, M., Shen, S., Yong, Z. X., Pandey, H., Bawden, R., Wang, T., Neeraj, T., Rozen, J., Sharma, A., Santilli, A., Fevry, T., Fries, J. A., Teehan, R., Biderman, S., Gao, L., Bers, T., Wolf, T., and Rush, A. M. (2022). Multitask prompted training enables zero-shot task generalization. International Conference on Learning Representations (ICLR).

Shang, C., Tang, Y., Huang, J., Bi, J., He, X., and Zhou, B. (2019). End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3060–3067.

Sikos, J. and Padó, S. (2018). Using embeddings to compare FrameNet frames across languages. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, pages 91–101, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Sikos, J. and Padó, S. (2019). Frame identification as categorization: Exemplars vs prototypes in embeddingland. In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 295–306, Gothenburg, Sweden, May. Association for Computational Linguistics.

Subirats, C. and Sato, H. (2003). Surprise! Spanish FrameNet. In *Proceedings of the Workshop on Frame Semantics at the XVII. International Congress of Linguists*. Citeseer.

Suhail, M. and Sigal, L. (2019). Mixture-kernel graph attention network for situation recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10363–10372.

Torrent, T. T. and Ellsworth, M. (2013). Behind the labels: criteria for defining analytical categories in framenet brasil. *Veredas-Revista de Estudos Linguisticos*, 17(1):44–66.

Torrent, T. T., Ellsworth, M., Baker, C., and Matos, E. E. d. S. (2018a). The Multilingual FrameNet Shared Annotation Task: a Preliminary Report. In Tiago Timponi Torrent, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, may. European Language Resources Association (ELRA).

Torrent, T. T., Matos, E. E. d. S., Lage, L., Laviola, A., Tavares, T., Almeida, V. G. d., and Sigiliano, N. (2018b). Towards continuity between the lexicon and the constructicon in framenet brasil. In Benjamin Lyngfelt, et al., editors, *Constructicography: constructicon development across languages*, pages 107–140. John Benjamins, Amsterdam, The Netherlands.

Trott, S., Torrent, T. T., Chang, N., and Schneider, N. (2020). (re)construing meaning in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5170–5184, Online, July. Association for Computational Linguistics.

Vabalas, A., Gowen, E., Poliakoff, E., and Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one*, 14(11):e0224365.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph Attention Networks. In *International Conference on Learning Representations*.

Verhagen, A. et al. (2007). Construal and perspectivization. *The Oxford handbook of cognitive linguistics*, 48:81.

Wang, C., Pan, S., Hu, R., Long, G., Jiang, J., and Zhang, C. (2019). Attributed graph clustering: A deep attentional embedding approach. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 3670–3676. International Joint Conferences on Artificial Intelligence Organization, 7.

Wei, J., Bosma, M., Zhao, V., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., and Le, Q. V. (2022). Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*.

Zhang, Z., Zhuang, F., Zhu, H., Shi, Z., Xiong, H., and He, Q. (2020). Relational graph neural network with hierarchical attention for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9612–9619.

Zhou, K., Dong, Y., Lee, W. S., Hooi, B., Xu, H., and Feng, J. (2020). Effective training strategies for deep graph neural networks. *arXiv preprint arXiv:2006.07107*.