# FrameNet on the Way to Babel: Creating a Bilingual FrameNet Using Wiktionary as Interlingual Connection

**Silvana Hartmann**† **and Iryna Gurevych**†‡
† Ubiquitous Knowledge Processing Lap (UKP-TUDA)
Department of Computer Science, Technische Universität Darmstadt
‡ Ubiquitous Knowledge Processing Lap (UKP-DIPF)
German Institute for Educational Research and Educational Information
`www.ukp.tu-darmstadt.de`

## Abstract

We present a new bilingual FrameNet lexicon for English and German. It is created through a simple, but powerful approach to construct a FrameNet in any language using Wiktionary as an interlingual representation. Our approach is based on a sense alignment of FrameNet and Wiktionary, and subsequent translation disambiguation into the target language. We perform a detailed evaluation of the created resource and a discussion of Wiktionary as an interlingual connection for the cross-language transfer of lexical-semantic resources. The created resource is publicly available at `http://www.ukp.tu-darmstadt.de/fnwkde/`.

## 1 Introduction

FrameNet is a valuable resource for natural language processing (NLP): semantic role labeling (SRL) systems based on FrameNet provide semantic analysis for NLP applications, such as question answering (Narayanan and Harabagiu, 2004; Shi and Mihalcea, 2005) and information extraction (Mohit and Narayanan, 2003). However, their wide deployment has been prohibited by the poor coverage and limited availability of a similar resource in many languages.

Expert-built lexical-semantic resources are expensive to create. Previous cross-lingual transfer of FrameNet used corpus-based approaches, or resource alignment with multilingual expert-built resources, such as EuroWordNet. The latter indirectly also suffers from the high cost and constrained coverage of expert-built resources.

Recently, collaboratively created resources have been investigated for the multilingual extension of resources in NLP, beginning with Wikipedia (Navigli and Ponzetto, 2010). They rely on the so-called "Wisdom of the Crowds", contributions by a large number of volunteers, which results in a continuously updated high-quality resource available in hundreds of languages. Due to the encyclopedic nature of Wikipedia, previous work focused on encyclopedic information for Wikipedia entries, i.e., almost exclusively on nouns.

This is not enough for resources like FrameNet. Such resources need *lexical-semantic information* on *various POS*. For FrameNet, information on the predicates associated with a semantic frame – mostly verbs, nouns, and adjectives – is crucial, for instance gloss or syntactic subcategorization.

A solution for the problem of multilingual extension of lexical semantic resources is to use Wiktionary, a collaboratively created dictionary, as connection between languages. It provides high-quality lexical information on all POS, for instance glosses, sense relations, syntactic subcategorization, etc. Like Wikipedia, it is continuously extended and contains translations to hundreds of languages, including low-resource ones. To our knowledge, Wiktionary has not been evaluated as an interlingual index for the cross-lingual extension of lexical-semantic resources.

In this paper, we present a novel method for the creation of bilingual FrameNet lexicons based on an alignment to Wiktionary. We demonstrate our method on the language pair English-German and present the resulting resources, a lemma-based multilingual and a sense-disambiguated German-English FrameNet lexicon.

The understanding of lexical-semantic resources and their combinations, e.g., how alignment algorithms can be adapted to individual resource pairs and different POS, is essential for their effective use in NLP and a prerequisite for later in-task evaluation and application. To enhance this understanding for the presented resource pair, we perform a detailed analysis of the created resource and compare it to existing FrameNet resources for German.

The contributions of our work are the following: (1) We create a novel sense alignment between FrameNet and the English Wiktionary. It results in a multilingual FrameNet FNWKxx, which links FrameNet senses to lemmas in 280 languages. (2) We create a sense-disambiguated English-German FrameNet lexicon FNWKde based on FNWKxx and translation disambiguation on the German Wiktionary.[1] (3) We analyze the two resources and outline further steps for creating a multilingual FrameNet.

This is a major step towards the vision of this paper: a simple, but powerful approach to partially construct a FrameNet in any language using Wiktionary as an interlingual representation.

## 2 Resource Overview

**FrameNet** (Baker et al., 1998) is an expert-built lexical-semantic resource incorporating the theory of frame-semantics (Fillmore, 1976). It groups word senses in frames that represent particular situations. Thus, the verb *complete* and the noun *completion* belong to the *Activity finish* frame. The participants of these situations, typically realized as syntactic arguments, are the *semantic roles* of the frame, for instance the *Agent* performing an activity, or the *Activity* itself. FrameNet release 1.5 contains 1,015 frames, and 11,942 word senses. Corpus texts annotated with frames and their roles have been used to train automatic SRL systems.

**Wiktionary** is a collaboratively created dictionary available in over 500 language editions. It is continuously extended and revised by a community of volunteer editors. The English language edition contains over 500,000 word senses.[2]

Wiktionary is organized like a traditional dictionary in lexical entries and word senses. For the word senses, definitions and example sentences, as well as other lexical information, such as register (e.g., *colloquial*), phonetic transcription, inflection may be available, including language-specific types of information. Senses also provide translations to other languages. These are connected to lexical entries in the respective language editions via hyperlinks. This allows us to use Wiktionary as an interlingual connection between multiple languages.

---

[1]The xx in FNWKxx stands for all the languages in the resource. After translation disambiguation in a specific language, xx is replaced by the corresponding language code.

[2]as of May 2013, see http://en.wiktionary.org/wiki/Wiktionary:Statistics.
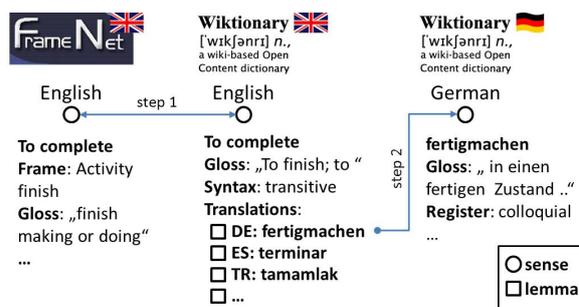


Figure 1: Method overview.

The quality of Wiktionary has been confirmed by Meyer and Gurevych (2012b) who also give an overview on the usage of Wiktionary in NLP applications such as speech synthesis.

## 3 Method Overview

Our method consists of two steps visualized in Fig. 1. In the first step, we create a novel sense alignment between FrameNet and the English Wiktionary following Niemann and Gurevych (2011). Thus, the FrameNet sense of *to complete* with frame *Activity finish* is assigned to the sense of *to complete* in Wiktionary meaning *to finish*.

This step establishes Wiktionary as an interlingual index between FrameNet senses and lemmas in many languages, and builds the foundation for the bilingual FrameNet extension.

It results in a basic multilingual FrameNet lexicon FNWKxx with translations to lemmas in 283 languages. An example: by aligning the FrameNet sense of the verb *complete* with gloss *to finish* with the corresponding English Wiktionary sense, we collect 39 translations to 22 languages, e.g., the German *fertigmachen* and the Spanish *terminar*.

The second step is the disambiguation of the translated lemmas with respect to the target language Wiktionary in order to retrieve the linguistic information of the corresponding word sense in the target language Wiktionary (Meyer and Gurevych, 2012a). We evaluate this step for English and German and create the bilingual FrameNet lexicon FNWKde. For the example sense of *complete*, we extract lexical information for the word sense of its German translation *fertigmachen*, for instance a German gloss, an example sentence, register information (*colloquial*), and synonyms, e.g., *beenden*. As a side-benefit of our method, we also extend the English FrameNet by the linguistic information in Wiktionary.

## 4 Related Work

### 4.1 Creating FrameNets in New Languages

There are two main lines of research in bootstrapping a FrameNet for languages other than English.

The first, corpus-based approach is to automatically extract word senses in the target language based on parallel corpora and frame annotations in the source language. In this vein, Padó and Lapata (2005) propose a cross-lingual FrameNet extension to German and French; Johansson and Nugues (2005) and Johansson and Nugues (2006) do this for Spanish and Swedish, and Basili et al. (2009) for Italian.

Padó and Lapata (2005) observe that their approach suffers from polysemy errors, because lemmas in the source language need to be disambiguated with respect to all the frames they evoke. To alleviate this problem, they use a disambiguation approach based on the most frequent frame; Basili et al. (2009) use distributional methods for frame disambiguation. Our approach is based on sense alignments and therefore explicitly aims to avoid such errors.

The second line of work is resource-based: FrameNet is aligned to multilingual resources in order to extract senses in the target language. Using monolingual resources, this approach has also been employed to extend FrameNet coverage for English (Shi and Mihalcea, 2005; Johansson and Nugues, 2007; Ferrandez et al., 2010).

De Cao et al. (2008) map FrameNet frames to WordNet synsets based on the embedding of FrameNet lemmas in WordNet. They use Multi-WordNet, an English-Italian wordnet, to induce an Italian FrameNet lexicon with 15,000 entries.

To create *MapNet*, Tonelli and Pianta (2009) align FrameNet senses with WordNet synsets by exploiting the textual similarity of their glosses. The similarity measure is based on stem overlap of the candidates' glosses expanded by WordNet domains, the WordNet synset, and the set of senses for a FrameNet frame. In Tonelli and Pighin (2009), they use these features to train an SVM-classifier to identify valid alignments and report an $F_1$-score of 0.66 on a manually annotated gold standard. They report 4,265 new English senses and 6,429 new Italian senses, which were derived via MultiWordNet.

*ExtendedWordFramenet* (Laparra and Rigau, 2009; Laparra and Rigau, 2010) is also based on the alignment of FrameNet senses to Word-Net synsets. The goal is the multilingual coverage extension of FrameNet, which is achieved by linking WordNet to wordnets in other languages (Spanish, Italian, Basque, and Catalan) in the Multilingual Central Repository. For each language, they add more then 10,000 senses to FrameNet. They rely on a knowledge-based word sense disambiguation algorithm to establish the alignment and report $F_1$=0.75 on a gold standard based on Tonelli and Pighin (2009).

Tonelli and Giuliano (2009) align FrameNet senses to Wikipedia entries with the goal to extract word senses and example sentences in Italian. Based on Wikipedia, this alignment is restricted to nouns. Subsequent work on Wikipedia and FrameNet follows a different path and tries to enhance the modeling of selectional preferences for FrameNet predicates (Tonelli et al., 2012).

Finally, there have been suggestions to combine the corpus-based and the resource-based approaches: Borin et al. (2012) do this for Finnish and Swedish. They prove the feasibility of their approach by creating a preliminary Finnish FrameNet with 2,694 senses.

Mouton et al. (2010) directly exploit the translations in the English and French Wiktionary editions to extend the French FrameNet. They match the FrameNet senses to Wiktionary lexical entries, thus encountering the problem of polysemy in the target language. To solve this, they define a set of filters that control how target lemmas are distributed over frames, increasing precision at the expense of recall (P=0.74, R=0.3, $F_1$=0.42). While their approach is in theory applicable to other languages, our approach goes beyond this by laying the ground for simultaneous FrameNet extension in multiple languages via FNWKxx.

### 4.2 Wiktionary Sense Alignments

Collaboratively created resources have become popular for sense alignments for NLP, starting with the alignment between WordNet and Wikipedia (Ruiz-Casado et al., 2005; Ponzetto and Navigli, 2009). Wiktionary has been subject to few alignment efforts: de Melo and Weikum (2009) integrate information from Wiktionary into Universal WordNet. Meyer and Gurevych (2011) map WordNet synsets to Wiktionary senses and show their complementary domain coverage.

## 5 FrameNet – Wiktionary Alignment

### 5.1 Alignment Technique

We follow the state-of-the-art sense alignment technique introduced by Niemann and Gurevych (2011). They align senses in WordNet to Wikipedia entries in a supervised setting using semantic similarity measures.

One reason to use their method was that it allows zero alignments or one-to-many alignments. This is crucial for obtaining a high-quality alignment of heterogeneous resources, such as the presented one, because their sense granularity and coverage can diverge a lot.

The alignment algorithm consists of two steps. In the *candidate extraction* step, we iterate over all FrameNet senses and match them with all senses from Wiktionary which have the same lemma and thus are likely to describe the same sense.

This step yields a set of candidate sense pairs $C_{all}$. In the *classification* step, a similarity score between the textual information associated with the senses in a candidate pair (e.g., their gloss) is computed and a threshold-based classifier decides for each pair on valid alignments.

Niemann and Gurevych (2011) combine two different types of similarity (i) cosine similarity on bag-of-words vectors (COS) and (ii) a personalized PageRank-based similarity measure (PPR). The PPR measure (Agirre and Soroa, 2009) maps the glosses of the two senses to a semantic vector space spanned up by WordNet synsets and then compares them using the chi-square measure.

The semantic vectors **ppr** are computed using the personalized PageRank algorithm on the WordNet graph. They determine the important nodes in the graph as the nodes that a random walker following the edges visits most frequently:

$$\mathbf{ppr} = cM\mathbf{ppr} + (1-c)\mathbf{v_{ppr}}, \quad (1)$$

where $M$ is a transition probability matrix between the n WordNet synsets, $c$ is a damping factor, and $\mathbf{v_{ppr}}$ is a vector of size n representing the probability of jumping to the node $i$ associated with each $\mathbf{v_i}$. For personalized PageRank, $\mathbf{v_{ppr}}$ is initialized in a particular way: the initial weight is distributed equally over the m vector components (i.e., synsets) associated with a word in the sense gloss, other components receive a 0 value.

For each similarity measure, Niemann and Gurevych (2011) determine a threshold ($t_{ppr}$ and $t_{cos}$) independently on a manually annotated gold standard. The final alignment decision is the conjunction of two decision functions:

$$a(s_s, s_t) = \\ PPR(s_s, s_t) > t_{ppr} \& COS(s_s, s_t) > t_{cos}. \quad (2)$$

We differ from Niemann and Gurevych (2011) in that we use a joint training setup which determines $t_{ppr}$ and $t_{cos}$ to optimize classification performance directly (as proposed in Gurevych et al. (2012)):

$$(t_{ppr}, t_{cos}) = \text{argmax}_{(t_{ppr}, t_{cos})} F_1(a), \quad (3)$$

where $F_1$ is the maximized evaluation score and $a$ is the decision function in equation (2).

### 5.2 Candidate Extraction

To compile the candidate set, we paired senses from both resources with identical lemma-POS combinations. FrameNet senses are defined by a lemma, a gloss, and a frame. Wiktionary senses are defined by a lemma and a gloss. For the FrameNet sense *Activity finish* of the verb *complete*, we find two candidate senses in Wiktionary (*to finish* and *to make whole*). There are on average 3.7 candidates per FrameNet sense. The full candidate set $C_{all}$ contains over 44,000 sense pairs and covers 97% of the 11,942 FrameNet senses.

### 5.3 Gold Standard Creation

For the gold standard, we sampled 2,900 candidate pairs from $C_{all}$. The properties of the gold standard mirror the properties of $C_{all}$: the sampling preserved the distribution of POS in $C_{all}$ (around 40% verbs and nouns, and 12% adjectives) and the average numbers of candidates per FrameNet sense. This ensures that highly polysemous words as well as words with few senses are selected.

Two human raters annotated the sense pairs based on their glosses. The annotation task consisted in a two-class annotation: *Do the presented senses have same meaning - (YES/NO)*. The raters received detailed guidelines and were trained on around 100 sense pairs drawn from the sample.

We computed Cohen's $\kappa$ to measure the inter-rater agreement between the two raters. It is $\kappa$=0.72 on the full set, which is considered acceptable according to Artstein and Poesio (2008). An additional expert annotator disambiguated ties.

For comparison: Meyer and Gurevych (2011) report $\kappa$=0.74 for their WordNet – Wiktionary gold standard, and Niemann and Gurevych (2011)

|  | adj | noun | verb | all |
|---|---|---|---|---|
| $\kappa$ | **.8** | **.77** | **.65** | **.72** |

Table 1: Inter-rater agreement.

$\kappa$=0.87 for their WordNet – Wikipedia gold standard. These gold standards only consist of nouns, which appear to be an easier annotation task than verb senses. This is supported by our analysis of the agreement by POS (see Table 1): the agreement on nouns and adjectives lies between the two agreement scores previously reported on nouns. Thus our annotation is of similar quality. Only the agreement on verbs is slightly below the acceptability threshold of 0.67 (Artstein and Poesio, 2008). The verb senses are very fine-grained and thus present a difficult alignment task. Therefore, we had an expert annotator correct the verbal part of the gold standard set. After removing the training set for the raters, the final gold standard contains 2,789 sense pairs. 28% of these are aligned.

## 5.4 Alignment Experiments

We determined the best setting for the alignment of FrameNet and Wiktionary in a ten-fold cross-validation on the gold standard.

Besides the parameters for the computation of the PPR vectors (we used the publicly available UKB tool by Agirre and Soroa (2009)), the main parameter in the experiments is the textual information that is used to represent the senses. For FrameNet senses, we used the *lemma-pos*, *sense gloss*, *example sentences*, *frame name* and *frame definition* as textual features; for Wiktionary senses, we considered *lemma-pos*, *sense gloss*, *example sentences*, *hyponyms* and *synonyms*.

We computed the similarity scores on tokenized, lemmatized and stop-word-filtered texts.

First, we evaluated models for COS and PPR independently based on various combinations of the textual features listed above. We then used the parameter setting of the best-performing single models to train the model that jointly optimizes the thresholds for PPR and COS (see eqn. (5)). In Table 2, we report on the results of the best single models and the best joint model.

For the evaluation, we compute precision P, recall R and $F_1$ on the positive class (aligned=true), e.g., precision P is the number of pairs correctly aligned divided by all aligned pairs.

We achieved the highest precision and $F_1$-score

|  | Evaluation | verb | noun | adj | all |
|---|---|---|---|---|---|
| P | Random-1 BL | 0.503 | 0.559 | 0.661 | 0.557 |
|  | WKT-1 BL | 0.620 | 0.664 | 0.725 | 0.66 |
|  | BEST COS | 0.639 | **0.778** | 0.706 | 0.703 |
|  | BEST PPR | 0.66 | 0.754 | 0.729 | 0.713 |
|  | BEST JOINT | **0.677** | 0.766 | **0.742** | **0.728** |
| R | Random-1 BL | 0.471 | 0.546 | 0.683 | 0.540 |
|  | WKT-1 BL | 0.581 | 0.65 | 0.75 | 0.64 |
|  | BEST COS | 0.658 | 0.758 | 0.754 | 0.715 |
|  | BEST PPR | 0.666 | 0.724 | 0.754 | 0.699 |
|  | BEST JOINT | **0.683** | **0.783** | **0.83** | **0.75** |
| $F_1$ | Random-1 BL | 0.487 | 0.552 | 0.672 | 0.549 |
|  | WKT-1 BL | 0.60 | 0.657 | 0.737 | 0.65 |
|  | BEST COS | 0.648 | 0.768 | 0.729 | 0.709 |
|  | BEST PPR | 0.663 | 0.739 | 0.741 | 0.706 |
|  | BEST JOINT | **0.68** | **0.775** | **0.784** | **0.739** |
|  | UBound | 0.735 | 0.834 | 0.864 | 0.797 |

Table 2: Alignment performance by POS.

for COS using all available features, but excluding FrameNet *example sentences* because they introduce too much noise. Adding the *frame name* and *frame definition* to the often short glosses provides a richer sense representation for the COS measure.

The best-performing PPR configuration uses *sense gloss* and *lemma-pos*. For the joint model, we employed the best single PPR configuration, and a COS configuration that uses *sense gloss* extended by Wiktionary *hypernyms*, *synonyms* and FrameNet *frame name* and *frame definition*, to achieve the highest score, an $F_1$-score of 0.739.

## 5.5 Gold Standard Evaluation

We compared the performance of our alignment on the gold standard to a baseline which randomly selects one target sense from the candidate set of each source sense (Random-1). We also consider the more competitive Wiktionary first sense baseline (WKT-1). It is guided by the heuristic that more frequent senses are listed first in Wiktionary (Meyer and Gurevych, 2010). It is a stronger baseline with an $F_1$-score of 0.65 (see Table 2).

To derive the upper bound for the alignment performance (UBound), we computed the $F_1$ score from the average pairwise $F_1$-score of the annotators according to Hripcsak and Rothschild (2005).

As the evaluation set mirrors the POS distribution in FrameNet and is sufficiently large, unlike earlier alignments, we can analyze the performance by POS. The BEST JOINT model performs well on nouns, slightly better on adjectives, and worse on verbs, see Table 2. For the baselines and the UBound the same applies, with the difference that adjectives receive even better results

in comparison. This fits in with the perceived degree of difficulty according to the observed polysemy for the POS: for verbs we have many candidate sets with two or more candidates, i.e., we observe higher polysemy, while for nouns and even stronger for adjectives, many small candidate sets occur, which stand for an easier alignment decision. This is in line with the reported higher complexity of lexical resources with respect to verbs and greater difficulty in alignments and word sense disambiguation (Laparra and Rigau, 2010).

The performance of BEST JOINT on all POS is $F_1$=0.73, which is significantly higher than the WKT-1 baseline ($p<0.05$ according to McNemar's test). The performance on nouns ($F_1$=0.775) is on par with the results reported by Niemann and Gurevych (2011) for nouns ($F_1$=0.78).

### 5.6 Error Analysis

The confusion matrix from the evaluation of BEST JOINT on the gold standard shows 214 false positives and 191 false negatives. The false negatives suffer from low overlap between the glosses, which are often quite short (*contend - assert*), sometimes circular (*sinful - relating to sin*). Aligning senses with such glosses is difficult for a system based on semantic similarity. In about 50% of the analyzed pairs, highly similar words are used in the gloss, that we should be able to detect with second-order representations, for instance by expanding short definitions with the definitions of the contained words, or via derivational similarity.

A number of false positives occur because the gold standard was developed in a very fine-grained manner: distinctions such as causative vs. inchoative (*enlarge: become large* vs. *enlarge: make large*) were explicitly stressed in the definitions and thus annotated as different senses by the annotators. This was motivated by the fact that this distinction is relevant for many frames in FrameNet. The first meaning of *enlarge* belongs to the frame *Expansion*, the second to *Cause expansion*. Our similarity based approach cannot capture such differences well.

## 6 Intermediate Resource FNWKxx

### 6.1 Statistics

We applied the best system setup to the full candidate set of over 44,000 candidates to create the intermediate resource FNWKxx. The alignment consists of 12,094 sense pairs. It covers 82% of

|  | fine-grained P | | | coarse-grained P | | |
|---|---|---|---|---|---|---|
| All POS | 0.67 | | | 0.78 | | |
| By POS | verb | noun | adj | verb | noun | adj |
|  | 0.53 | 0.73 | 0.80 | 0.73 | 0.82 | 0.85 |

Table 3: Post-hoc evaluation (precision P).

the senses in FrameNet and 86% of the frames. It connects more than 9,800 unique FrameNet senses with more than 10,000 unique Wiktionary senses, which shows that both non-alignments and multiple alignments occur for some source senses.

### 6.2 Post-hoc Evaluation

Our cross-validation approach entails the danger of over-fitting. In order to verify the quality of the alignment, we performed a detailed post-hoc analysis on a sample of 270 aligned sense pairs randomly drawn from the set of aligned senses.

Because sense granularity was an issue in the error analysis, we considered two alignment decisions: (a) fine-grained alignment: the two glosses describe the same sense; (b) coarse-grained alignment. The causative/inchoative distinction is, among others, ignored.

The evaluation results are listed in Table 3. The precision for the fine-grained (a) is lower than the allover precision on the gold standard. The evaluation by POS shows that the result for nouns and adjectives is equal or superior to the evaluation result on the gold standard, while it is worse for verbs. This shows that over-fitting, if at all, is only a risk for the verb senses.

The allover precision for (b) exceeds the precision on the gold standard. Particularly verbs receive much better results. This shows that a coarse-grained alignment may suffice for the FrameNet extension.

This evaluation confirms the quality of the sense alignment, in particular with respect to the FrameNet extension. But it also elicits the question whether a coarse-grained alignment would suffice. We will discuss this question below.

### 6.3 Resource Analysis

For each of the aligned senses in the 12,094 aligned sense pairs, we extracted **glosses** from Wiktionary. Because FrameNet glosses are often very brief, the additional glosses will benefit algorithms such as frame detection for SRL. We also added 4,352 new **example sentences** from Wik-

tionary to FrameNet.

We can extract 2,151 new **lemma-POS** for FrameNet frames from the synonyms of the aligned senses in Wiktionary. We also extract other related lemma-POS, for instance 487 antonyms, 126 hyponyms, and 19 hypernyms.

This step establishes Wiktionary as an interlingual connection between FrameNet and a large number of languages, including low-resource ones: via Wiktionary, we connect FrameNet senses to **translations** in 283 languages, e.g., we translate the sense of the verb *complete* associated with the frame *Activity Finish* to the German colloquial *fertigmachen*, the Spanish *terminar*, the Turkish *tamamlamak*, and 19 other languages.

For 36 languages, we can extract more than 1,000 translations each, among them low-resource languages such as Telugu, Swahili, or Kurdish. The languages with most translations are: Finnish (9,333), Russian (7,790), and German (6,871). The number of Finnish translations is more than three times larger than the preliminary Finnish FrameNet by Borin et al. (2012). Likewise, we get three times the number of German lemma-POS than provided by the SALSA corpus.

## 7 Translation Disambiguation

### 7.1 Disambiguation Method

FNWKxx initially does not provide lexical-semantic information for the German translations: the translations link to a lemma in the German Wiktionary, not a target sense. In order to integrate the information attached to a German Wiktionary sense, e.g., the gloss, into our resource, the lemmas need to be disambiguated.

We use the sense-disambiguated Wiktionary resulting from a recently published approach for the disambiguation of relations and translations in Wiktionary (Meyer and Gurevych, 2012a) to create our new bilingual (German-English) FrameNet lexicon FNWKde.

Their approach combines information on the source sense and all potential target senses in order to determine the best target sense in a rule-based disambiguation strategy. The information is encoded as binary features, which are ordered in a back-off hierarchy: if the first feature applies, the target sense is selected, otherwise the second feature is considered, and so forth.

The most important features are: definition overlap between source and automatically trans-

|  | SALSA2 | P&L05 | FNWKde |
|---|---|---|---|
| Type Creation | Corpus Manual | Corpus Automatic | Lexicon Automatic |
| Frames(+p) | 266(907) | 468 | 755 |
| Senses | 1,813 | 9,851 | 5,897 |
| Examples | 24,184 | 1,672,551 | 6,933 |
| Glosses | - | - | 5,897 |

Table 4: Frame-semantic resources for German.

lated target definition; occurrence of the source lemma in the target definition; shared linguistic information (e.g., same register); inverse translation relations (i.e., the source lemma occurs on the translation list of the target sense); relation overlap; Lesk measure between original and translated glosses in source and target language; and finally, backing off to the first target sense.

For the gold standard evaluation of the approach we refer to Meyer and Gurevych (2012a): their system obtained an $F_1$-score of 0.67 for the task of disambiguating translations from English to German, and an $F_1$-score of 0.79 for the disambiguation of English sense relations. We use the latter to identify target senses of synonyms in FNWKxx.

## 8 Resource FNWKde

### 8.1 Statistics

Table 4 gives an overview of FNWKde. It contains 5,897 pairs of German Wiktionary senses and FrameNet senses, i.e., 86% of the translations could be disambiguated. Each sense has a gloss, and there are 6,933 example sentences.

Based on the relation disambiguation and inference of new relations by Meyer and Gurevych (2012a), we can also disambiguate synonyms in the English Wiktionary. This leads to a further extension of the English FrameNet summarized in Table 5. The number of Wiktionary senses aligned to FrameNet senses is increased by 50%.

We also provide results for other sense relations, e.g., antonyms. We will discuss whether and how they can be integrated as FrameNet senses in our resource below.

### 8.2 Post-hoc Evaluation

Because the errors of two subsequently applied automatic methods can multiply, we provide a post-hoc evaluation of the results.

To evaluate the quality of the German FrameNet lexicon, we collected the FrameNet senses for a list of 15 frames that were sampled by Padó and

| Relation | # English senses per FrameNet sense | # English senses per frame |
|---|---|---|
| SYNONYM | 17,713 | 13,288 |
| HYPONYM | 4,818 | 3,347 |
| HYPERNYM | 6,369 | 3,961 |
| ANTONYM | 9,626 | 6,737 |

Table 5: Statistics after relation disambiguation.

| | Resource r | % of r | % of FNWKde |
|---|---|---|---|
| Frame | SALSA 2 | 89% | 31% |
| | P&L05 | 90% | 55% |
| Sense | SALSA 2 | 15% | 5% |
| | P&L05 | 10% | 19% |

Table 6: Overlap of FNWKde with resource r.

Lapata (2005) according to three frequency bands on a large corpus. There are 115 senses associated with these frames in our resource. In a manual evaluation of these 115 senses, we find that 67% were assigned correctly to their frames. This is higher than expected, considering the errors from the applied methods add up.

Further analysis revealed that both resource creation steps contribute equally to the 39 errors. For 17 of the evaluated sense pairs, redundancy confirms their quality: they were obtained independently by two or three alignment-and-translation paths and do not contain alignment errors.

### 8.3 Comparison

We compare FNWKde to two German frame-semantic resources, the manually annotated SALSA corpus (Burchardt et al., 2006) and a resource from Padó and Lapata (2005), henceforth P&L05. Note that both resources are frame-annotated corpora, while FNWKde is a FrameNet-like *lexicon* and contains information complementary to the corpora. The different properties of the resources are contrasted in Table 4.

The automatically developed resources, including FNWKde, provide a larger number of senses than SALSA. The annotated corpora contain a large number of examples, but they do not provide any glosses, which are useful for frame detection in SRL, nor do they contain any other lexical-semantic information.

FNWKde covers a larger number of FrameNet frames than the other two resources. 266 of the 907 frames in SALSA are connected to original FrameNet frames, the others are newly-developed proto-frames p (shown in parentheses in Table 4).

Table 6 describes the proportion of the overlapping frames and senses[3] to the respective resources. The numbers on frame overlap show that our resource covers the frames in the other

resources well (89% and 90% coverage respectively), and that it adds frames not covered in the other resources: P&L05 only covers 55% of the frames in FNWKde. The sense overlap shows that the resources have senses in common, which confirms the quality of the automatically developed resources, but they also complement each other. FNWKde, for instance, adds 3,041 senses to P&L05.

## 9 Discussion: a Multilingual FrameNet

FNWKxx builds an excellent starting point to create FrameNet lexicons in various languages: the translation counts, for instance 6,871 for German, compare favorably to FrameNet 1.5, which contains 9,700 English lemma-POS.

To create those FrameNet lexicons, the translation disambiguation approach used for FNWKde (step 2 in Fig. 1) needs to be adapted to other languages. The approach is in theory applicable to any language, but there are some obstacles: first, it relies on the availability of the target sense in the target language Wiktionary. For many of the top 30 languages in FNWKxx, the Wiktionary editions seem sufficiently large to provide targets for translation disambiguation,[4] and they are continuously extended. Second, our approach requires access to the target language Wiktionary, but the data format across Wiktionary language editions is not standardized. Third, the approach requires machine translation into the target language. For languages, where such a tool is not available, we could default to the first-sense-heuristic, or encourage the Wiktionary community to link the translations to their target Wiktionary senses inspired by Sajous et al. (2010).

Another issue that applies to all automatic (and also manual) approaches of cross-lingual FrameNet extension is the restricted cross-language applicability of frames. Boas (2005) reports that, while many frames are largely

---

[3]Note that the senses in SALSA and P&L05 are defined by frame, lemma, and POS. In Table 6, FNWKde senses with identical frame, lemma, and POS, but different gloss are therefore conflated to one sense.

language-independent, other frames receive culture-specific or language-specific interpretations, for example calendars or holidays. Also, fine-grained sense and frame distinctions may be more relevant in one language than in another language. Such granularity differences also led to the addition of proto-frames in SALSA 2 (Rehbein et al., 2012). Therefore, manual correction or extension of a multilingual FrameNet based on FNWKde may be desired for specific applications. In this case, the automatically created FrameNets in other languages are good starting points that can be quickly and efficiently compiled.

The quality of the multilingual FNWKxx depends on i) the translations in the interlingual connection Wiktionary, which are manually created, controlled by the community, and therefore reliable, and ii) on the FrameNet–Wiktionary alignment. Therefore, we evaluated our sense alignment method in detail. The alignment reaches state-of-the-art results, and the analysis shows that the method is particularly fit for a coarse-grained alignment. We however find lower performance for verbs in a fine-grained setting. We argue that an improved alignment algorithm, for instance taking subcategorization information into account, can identify the fine-grained distinctions.

The post-hoc analysis raised the question of FrameNet frame granularity. Do separate frames exist for causative/inchoative alternations (as *Being dry* and *Cause to be dry* for *to dry*), or do they belong to the same frame (*Make noise* for *to creak* and *to creak something*)? For the coarse-grained frames, fine-grained decisions can be merged in a second classification step. Alternatively, we could map Wiktionary senses directly to frames, and include features that cover the granularity distinctions, e.g., whether the existing senses of a frame show the semantic alternation.

We could use the same approach to assign senses to a frame which are derived via sense relations other than synonymy, i.e., for linking antonyms or hyponyms to a frame. Some frames do cover antonymous predicates, others do not.

Based on Wiktionary, our approach suffers less from the disadvantages of previous resource-based work, i.e., the constraints of expert-built resources and the lack of lexical information in Wikipedia. Unlike corpus-based approaches for cross-lingual FrameNet extension, our approach does not provide frame-semantic annotations for the example sentences. Our advantage is that we create a FrameNet *lexicon* with lexical-semantic information in the target language. Example annotations can be additionally obtained via cross-lingual annotation projection (Padó and Lapata, 2009), and the lexical information in FNWKde can be used to guide this process.

## 10 Conclusion

The resource-coverage bottleneck for frame-semantic resources is particularly severe for less well-resourced languages. We present a simple, but effective approach to solve this problem using the English Wiktionary as an interlingual representation and subsequent translation disambiguation in the target language. We validate our approach on the language pair English-German and discuss the options and requirements for creating FrameNets in further languages.

As part of this work, we created the first sense alignment between FrameNet and the English Wiktionary. The resulting resource FNWKxx connects FrameNet senses to over 280 languages. The bilingual English-German FrameNet lexicon FNWKde competes with manually created resources, as shown by a comparison to the SALSA corpus.

We make both resources publicly available in the standardized format UBY-LMF (Eckle-Kohler et al., 2012), which supports automatic processing of the resources via the UBY Java API, see `http://www.ukp.tu-darmstadt.de/fnwkde/`.

We also extended FrameNet by several thousand new English senses from Wiktionary which are provided as part of FNWKde. In our future work, we will evaluate the benefits of the extracted information to SRL.

# References

Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 33–41, Athens, Greece.

Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.

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 (COLING-ACL'98)*, pages 86–90, Montreal, Canada.

Roberto Basili, Diego Cao, Danilo Croce, Bonaventura Coppola, and Alessandro Moschitti. 2009. Cross-language frame semantics transfer in bilingual corpora. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 5449 of *Lecture Notes in Computer Science*, pages 332–345. Springer Berlin Heidelberg.

Hans C. Boas. 2005. Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography*, 18(4):445–478.

Lars Borin, Markus Forsberg, Richard Johansson, Kristiina Muhonen, Tanja Purtonen, and Kaarlo Voionmaa. 2012. Transferring frames: Utilization of linked lexical resources. In *Proceedings of the NAACL-HLT Workshop on the Induction of Linguistic Structure*, pages 8–15, Montréal, Canada.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, and Manfred Pinkal. 2006. The SALSA corpus: a German corpus resource for lexical semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 969–974, Genoa, Italy.

Diego De Cao, Danilo Croce, Marco Pennacchiotti, and Roberto Basili. 2008. Combining word sense and usage for modeling frame semantics. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, STEP '08, pages 85–101, Stroudsburg, PA, USA.

Gerard de Melo and Gerhard Weikum. 2009. Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, pages 513–522, New York, NY, USA.

Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek, and Christian M. Meyer. 2012. UBY-LMF - A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pages 275–282, Istanbul, Turkey.

Oscar Ferrandez, Michael Ellsworth, Rafael Munoz, and Collin F. Baker. 2010. Aligning FrameNet and WordNet based on Semantic Neighborhoods. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 310–314, Valletta, Malta.

Charles J. Fillmore. 1976. Frame Semantics and the Nature of Language. In *Annuals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32. New York Academy of Sciences, New York, NY, USA.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. Uby - A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 580–590, Avignon, France.

George Hripcsak and Adam S. Rothschild. 2005. Agreement, the F-Measure, and Reliability in Information Retrieval. *Journal of the American Medical Informatics Association*, 12(3):296–298.

Richard Johansson and Pierre Nugues. 2005. Using Parallel Corpora for Automatic Transfer of FrameNet Annotation. In *Proceedings of the 1st ROMANCE FrameNet Workshop*, Cluj-Napoca, Romania.

Richard Johansson and Pierre Nugues. 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.

Richard Johansson and Pierre Nugues. 2007. Using WordNet to extend FrameNet coverage. In *Proceedings of the Workshop on Building Frame-semantic Resources for Scandinavian and Baltic Languages, at NODALIDA*, pages 27–30, Tartu, Estonia.

Egoitz Laparra and German Rigau. 2009. Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm. In *Proceedings of the International Conference RANLP-2009*, pages 208–213, Borovets, Bulgaria.

Egoitz Laparra and German Rigau. 2010. eXtended WordFrameNet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 1214–1419, Valletta, Malta.

Christian M. Meyer and Iryna Gurevych. 2010. How Web Communities Analyze Human Language: Word Senses in Wiktionary. In *Proceedings of the Second Web Science Conference*, Raleigh, NC, USA.

Christian M. Meyer and Iryna Gurevych. 2011. What Psycholinguists Know About Chemistry: Aligning Wiktionary and WordNet for Increased Domain Coverage. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 883–892, Chiang Mai, Thailand.

Christian M. Meyer and Iryna Gurevych. 2012a. To Exhibit is not to Loiter: A Multilingual, Sense-Disambiguated Wiktionary for Measuring Verb Similarity. In *Proceedings of COLING 2012*, pages 1763–1780, Mumbai, India.

Christian M. Meyer and Iryna Gurevych. 2012b. Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In Sylviane Granger and Magali Paquot, editors, *Electronic Lexicography*, pages 259–291. Oxford University Press, Oxford.

Behrang Mohit and Srini Narayanan. 2003. Semantic Extraction with Wide-Coverage Lexical Resources. In *Proceedings of HLT-NAACL 2003: Companion Volume*, pages 64–66, Edmonton, Canada.

Claire Mouton, Gaël de Chalendar, and Benoît Richert. 2010. FrameNet Translation Using Bilingual Dictionaries with Evaluation on the English-French Pair. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 20–27, Valletta, Malta.

Srini Narayanan and Sanda Harabagiu. 2004. Question Answering Based on Semantic Structures. In *Proceedings of the 20th international conference on Computational Linguistics - COLING '04*, pages 693–701, Geneva, Switzerland.

Roberto Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden.

Elisabeth (geb. Wolf) Niemann and Iryna Gurevych. 2011. The People's Web meets Linguistic Knowledge: Automatic Sense Alignment of Wikipedia and WordNet. In *Proceedings of the International Conference on Computational Semantics (IWCS)*, pages 205–214, Singapore.

Sebastian Padó and Mirella Lapata. 2005. Cross-lingual bootstrapping of semantic lexicons: the case of FrameNet. In *Proceedings of the 20th national conference on Artificial intelligence - Volume 3*, AAAI'05, pages 1087–1092, Pittsburgh, PA, USA.

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual Annotation Projection for Semantic Roles. *Journal of Artificial Intelligence Research*, 36:307–340.

Simone Paolo Ponzetto and Roberto Navigli. 2009. Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia. In *Proceedings of the 21st International Joint Conference on AI*, pages 2083–2088, Pasadena, CA, USA.

Ines Rehbein, Joseph Ruppenhofer, Caroline Sporleder, and Manfred Pinkal. 2012. Adding nominal spice to SALSA - frame-semantic annotation of German nouns and verbs. In *Proceedings of the 11th Conference on Natural Language Processing (KONVENS'12)*, pages 89–97, Vienna, Austria.

Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. Automatic Assignment of Wikipedia Encyclopedic Entries to WordNet Synsets. In *Advances in Web Intelligence*, volume 3528 of *Lecture Notes in Computer Science*, pages 380–386. Springer, Berlin Heidelberg.

Franck Sajous, Emmanuel Navarro, Bruno Gaume, Laurent Prévot, and Yannick Chudy. 2010. Semi-automatic endogenous enrichment of collaboratively constructed lexical resources: piggybacking onto wiktionary. In *Proceedings of the 7th international conference on Advances in natural language processing*, IceTAL'10, pages 332–344. Springer, Berlin, Heidelberg.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *Computational Linguistics and Intelligent Text Processing*, pages 100–111. Springer, Berlin Heidelberg.

Sara Tonelli and Claudio Giuliano. 2009. Wikipedia as frame information repository. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 276–285, Singapore.

Sara Tonelli and Emanuele Pianta. 2009. A novel approach to mapping FrameNet lexical units to WordNet synsets. In *IWCS-8 '09: Proceedings of the Eighth International Conference on Computational Semantics*, pages 342–345, Tilburg, The Netherlands.

Sara Tonelli and Daniele Pighin. 2009. New Features for FrameNet - WordNet Mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 219–227, Boulder, CO, USA.

Sara Tonelli, Volha Bryl, Claudio Giuliano, and Luciano Serafini. 2012. Investigating the semantics of frame elements. In *Knowledge Engineering and Knowledge Management*, volume 7603 of *Lecture Notes in Computer Science*, pages 130–143. Springer Berlin Heidelberg.