# The Projector: An Interactive Annotation Projection Visualization Tool

**Alan Akbik** and **Roland Vollgraf**
Zalando Research
Charlottenstraße 4, 10969 Berlin
firstname.lastname@zalando.de

## Abstract

Previous works proposed *annotation projection* in parallel corpora to inexpensively generate treebanks or propbanks for new languages. In this approach, linguistic annotation is automatically transferred from a resource-rich source language (SL) to translations in a target language (TL). However, annotation projection may be adversely affected by *translational divergences* between specific language pairs. For this reason, previous work often required careful qualitative analysis of projectability of specific annotation in order to define strategies to address quality and coverage issues. In this demonstration, we present THE PROJECTOR, an interactive GUI designed to assist researchers in such analysis: it allows users to execute and visually inspect annotation projection in a range of different settings. We give an overview of the GUI, discuss use cases and illustrate how the tool can facilitate discussions with the research community.

## 1 Introduction

Natural language processing research relies heavily on the availability of textual corpora annotated with various levels of syntactic and semantic information such as *treebanks* (Marcus et al., 1993) or *propbanks* (Palmer et al., 2005). However, the manual creation of such resources is known to be highly costly and therefore difficult to scale across languages and domains (Hovy et al., 2006).

**Annotation Projection.** As a cost-effective alternative, previous work suggested the use of *annotation projection* (Yarowsky et al., 2001) in parallel corpora to automatically create NLP resources for new languages. This approach requires only a parallel corpus consisting of sentences in a resource-rich source language (SL) and their translations in a target language (TL), as well as existing parsers for the SL. It leverages the hypothesis that translated sentences will share a degree of syntactic and, in particular, semantic parallelism (Padó and Lapata, 2009), thus allowing us to automatically transfer linguistic annotations from SL to TL.

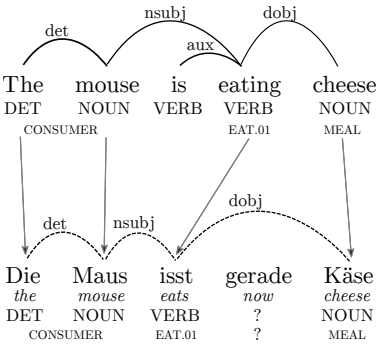


Figure 1: An English sentence with syntactic and semantic annotations predicted by a parser, and a German translation. These annotations are transferred onto aligned German words, thus automatically labeling the German sentence (with exception of the unaligned German word *gerade*).

For example, consider Figure 1 which shows an English SL sentence and its German TL translation, with word-level alignments indicated as lines between the sentences. State-of-the-art parsers and semantic role labelers (SRL) are used to predict labels for the English sentence. Following the word alignments, this annotation is then transferred onto the German sentence. The English word *cheese*, for instance, is aligned to the German *Käse* (cheese). We therefore learn that *Käse* is also a noun (NOUN), that in this sentence it is a direct object (dobj) and that it takes the semantic role of MEAL. Following this process, we can thus automatically annotate the German sentence with (partial) syntactic and semantic labels.

**A Need for Qualitative Analysis.** However,


*Proceedings of the 2017 EMNLP System Demonstrations*, pages 43–48
Copenhagen, Denmark, September 7–11, 2017. 

while annotation projection has been successfully employed to transfer various types of annotation (Yarowsky et al., 2001; Hwa et al., 2005; Padó and Lapata, 2009; Van der Plas et al., 2011; Akbik et al., 2015), it is not always clear how well specific annotation can be transferred to a specific TL. Previous work noted a range of issues including non-literal translations and general *translational divergences* (Dorr, 1994) between languages which cause incorrect annotation to be projected. For this reason, previous work often included qualitative analyses and carefully defined heuristics to address these problems.

**Contributions.** To facilitate such analysis and discussion, we present THE PROJECTOR, a web-based UI that visualizes the projection of syntactic and shallow semantic annotation in parallel sentences[1]. Our tool enables researchers to execute annotation projection for manually created examples or pre-loaded corpora, and allows researchers to visually inspect individual sentence pairs and types of linguistic annotation.

This paper is structured as follows: we first review relevant related work in annotation projection. We then give an overview of THE PROJECTOR, briefly sketch use cases for this tool and discuss directions for future research.

## 2 Previous Work

**Syntactic Annotation Projection.** Early work proposed the projection of shallow and deep syntactic information in parallel corpora, including part-of-speech tags (Yarowsky et al., 2001), syntactic chunks (Yarowsky and Ngai, 2001) and dependency trees (Hwa et al., 2005). However, these works also noted problems stemming from *translational divergences* (Dorr, 1994; Van Leuven-Zwart, 1989), i.e. systematic differences between languages on the structural and semantic realization levels. This may cause incorrect labels to be projected, or annotation gaps in the TL corpus. For instance, as Figure 1 shows, while a continuous process may be expressed in English using *gerunds* (the word *eating*), the continuous verb aspect generally does not exist in German which instead uses an adverbial construction (the word *gerade*, meaning *now*, which is unaligned and therefore remains unlabeled in Figure 1).

To filter out and correct such errors, previous work defined various heuristics such as filtering of infrequent alignments (Yarowsky et al., 2001), transformation rules that encode linguistic knowledge (Hwa et al., 2005) and the use of cross-lingual word clusters as constraints in projection (Täckström et al., 2012). More recently, Tiedemann (2014) argued that the ongoing harmonization of linguistic annotation across languages as pursued by the *universal dependencies* project (Nivre et al., 2016) has produced tagsets without language-specific syntax that can more easily be projected[2].

**Semantic Annotation Projection.** Previous work also investigated the applicability of annotation projection to shallow semantic annotation such as semantic role labels (SRL). Padó and Lapata (2009) first analyzed the viability of transferring SRL in the FRAMENET-formalism (Baker et al., 1998) and found a greater degree of parallelism for semantic than syntactic annotation. Van der Plas et al. (2011) applied this approach to the verb-centric PROPBANK-formalism of SRL (Palmer et al., 2005).

In our previous work, we defined a two-step process of filtering and semi-supervised learning to address problems caused by non-literal translations and coverage gaps (Akbik et al., 2015). We applied our approach to generate propbanks for 7 languages from 3 language groups (Akbik et al., 2015), and experimented with projecting both syntactic and semantic annotation to three low-resource languages (Akbik et al., 2016b). Qualitative analysis revealed propositions evoked by complex predication (Bonial et al., 2014) to be a major source of translational divergences of shallow semantics (Akbik et al., 2016a).

**Qualitative Analysis.** Previous works illustrated the need for qualitative analysis to identify error sources and define strategies to address translational divergences. However, to the best of our knowledge, no visualization tool is available that specifically addresses this task. While there exist tools that focus on inspecting and correcting word alignments (Gilmanov et al., 2014) as well as frameworks for visualization of various layers of linguistic annotation (Krause and Zeldes, 2016), THE PROJECTOR differs in that it specializes in the projection of various types of linguistic annotation in parallel corpora and interactive analysis.

[1] A screencast is available at https://vimeo.com/217035646

[2] The *gerund* verb type, for instance, is abstracted away from in universal PoS tags to a general VERB class.

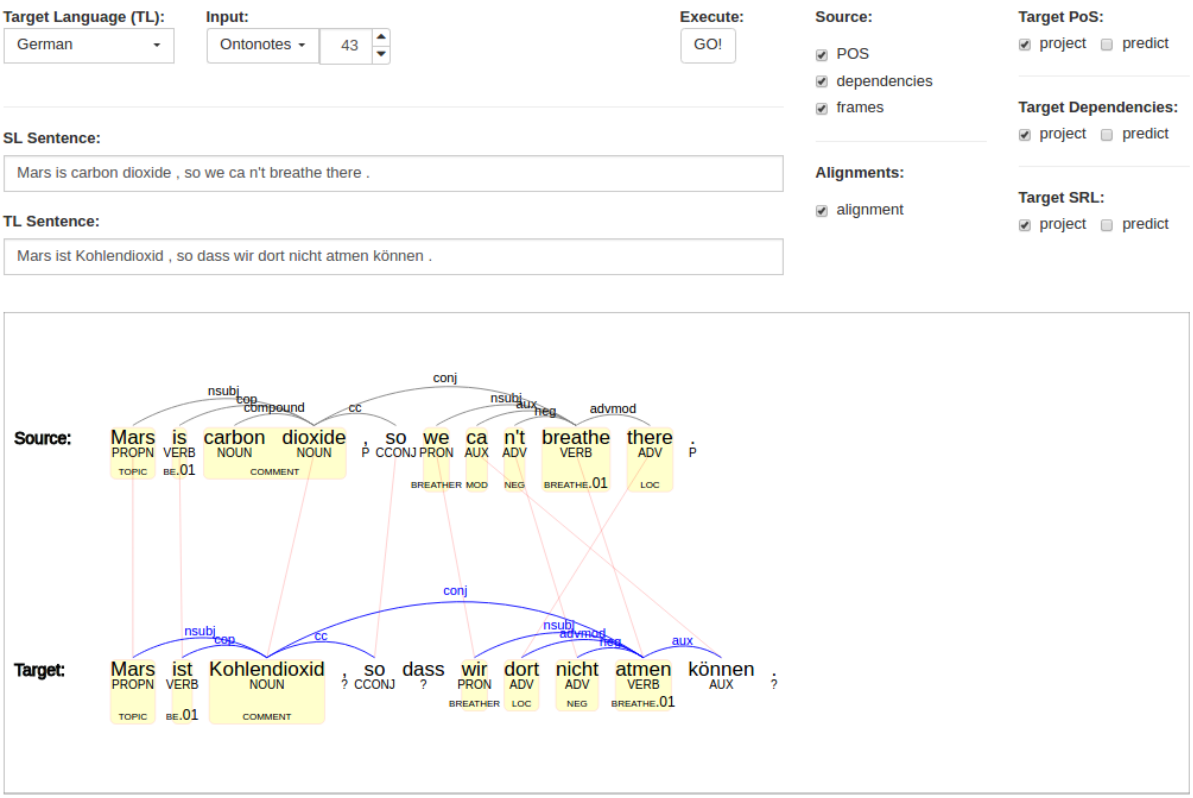


Figure 2: THE PROJECTOR's main view showing a gold-labeled English sentence from ONTONOTES and a word-aligned German translation. The German sentence is labeled using annotation projection.

## 3 THE PROJECTOR User Interface

THE PROJECTOR is a Web-based GUI that allows users to inspect alignments and projected annotation. We give an overview of the layout, input fields and visualization options (sec. 3.1) discuss the two main usage modes (sec. 3.2), and illustrate two example usage scenarios (sec. 3.3).

### 3.1 Layout

Figure 2 illustrates THE PROJECTOR's main view. It is divided into input fields (top right), visualization options (top left) and the visualization pane (bottom half).

#### 3.1.1 Input Fields

The input fields are grouped to the top left of the main screen. Two selection options are mandatory: the first is the **target language** dropdown option to indicate the TL of the annotation projection approach[3]. The second is the **input** option that can either be set to *manual* (users manually supply a sentence or sentence pair) or used to select a pre-loaded monolingual corpus in CoNLL-U format. At time of writing, we pre-loaded both the ONTONOTES (Hovy et al., 2006) and the *universal dependencies* corpora (Nivre et al., 2016).

There are two textual inputs, namely the **SL sentence** and the **TL sentence** fields. Users populate these fields with a translated sentence pair. At least one of the two fields must be populated - in this case, a translation is automatically generated using the Google Translate API (Wu et al., 2016). The fields can either be populated by manually entering a sentence, or - if a corpus is selected as input option - be populated by selecting a sentence using the **corpus navigator**. Once the selection is complete and at least one sentence field populated, users hit the **go** button to execute and visualize annotation projection.

#### 3.1.2 Visualization Options

Visualization options are divided into options that pertain to the source or target sentence. On the source side, users can check which layers of visualization should be displayed. Options include PoS tags, dependency trees and semantic frames and roles.

On the target side, users choose between several options for each layer. PoS tags and dependency trees can be either *predicted* using a TL parser

[3] At time of writing, we tested setups with the following TLs: Chinese, French, German and Japanese - through there is no principal limitation on the scope of TLs

or *projected* using annotation projection. By toggling between these options, users compare between predicted and projected annotation. Since SRL information is projected onto entire TL constituents, users additionally specify whether they are identified using predicted or projected dependency tree information. In addition, users choose whether or not to show the word alignments.

#### 3.1.3 Visualization Pane

The visualization pane displays annotation projection for a sentence pair according to the current settings. If activated, dependency trees are displayed above a sentence. PoS tags are placed directly beneath each word. SRL labels are displayed as boxes around constituents, where each layer corresponds to one semantic frame.

### 3.2 Modes

The tool supports two general modes of interaction: (1) an *interactive mode* in which users supply an example sentence (or sentence pair) and execute alignment, parsing and projection on-the-fly, and (2) a *corpus mode* in which a gold-labeled corpus is loaded that can be browsed and employed in annotation projection.

#### 3.2.1 Interactive Mode

The first mode is intended for analysis of specific linguistic constructs in the source or target language. Researchers select "manual" as input option and create an example sentence for the construct of interest. To analyze how a SL construct transfers to a TL, users enter the example sentence in the source field. Similarly, to investigate a specific TL construct, they enter the example sentence in the target field. Users may supply the corresponding SL or TL translation themselves or simply leave the other field blank - if only one sentence is provided, our tool uses the Google Translate API to automatically retrieve a translation and fill in the missing field.

**Parsing pipelines.** Upon clicking **go**, the SL sentence is sent to a pipeline of NLP tools, namely the StanfordNLP tools (Manning et al., 2014) to tokenize, lemmatize, PoS tag and dependency parse the sentence and MatePlus (Roth and Woodsend, 2014) to predict SRL annotation. We used the standard models provided for StanfordNLP and trained MatePlus over the version 3 release of propbank annotations (Bonial et al., 2014) for the OntoNotes corpus. The target language sentence is parsed using the transition-based MATE parser (Bohnet and Nivre, 2012) which we trained for each TL over version 1.4 release of the universal dependency treebank.

**Alignment and projection.** Word alignments are heuristically detected on-the-fly using word co-occurrence weights as determined by the BerkeleyAligner (DeNero and Liang, 2007) over the 2016 release of the OpenSubtitles parallel corpus for all supported language pairs (Tiedemann, 2012). Using these alignments, we execute annotation projection of all syntactic and semantic information. Word-level PoS tags and semantic frames are simply transferred to aligned words in the target language, while dependencies are transferred to corresponding word pairs. For semantic roles (which label entire constituents), we identify the best matching TL constituent using the Jaccard distance as described in (Padó and Lapata, 2009).

**Results.** The GUI displays the sentence pair with all predicted and projected annotations. Users may change visualization options, and experiment with modifications to the sentence pair (for instance, chose a different translation).

#### 3.2.2 Corpus Mode

The second mode is to enable qualitative analysis for cases in which a gold labeled corpus already exists either for the source or target language. This setting allows us to inspect annotation projection without interference from potentially incorrect parses[4]. Users select a corpus in the input field, and then browse sentences using the navigation field. Depending on whether the gold-labeled corpus is loaded for the SL or TL side, a different pipeline of tools is executed:

**Gold-labeled SL corpus.** If users select an English gold corpus (OntoNotes in the current setup), no SL parsing pipeline is executed. Instead, the sentence is automatically translated into the selected TL and the alignment, TL parsing and projection pipeline executed. If the translation is incorrect or lacking, users may manually enter a better TL translation and re-execute the approach.

**Gold-labeled TL corpus.** If users select a TL gold corpus (the universal dependency treebanks in the current setup), the currently selected sentence is first translated into English and then parsed using the default English parsing pipeline

[4] In previous work, we showed that many TL annotation errors were caused by parsing errors on the SL side that were propagated during projection (Akbik et al., 2015).

and word-aligned. The annotation is then projected back onto the TL sentence where it can be compared to the original gold labels.

### 3.3 Example Scenarios

We now present two example usage scenarios.

#### 3.3.1 Scenario 1: Study of Translational Divergences

In the first scenario, a user may be interested to study the effects of specific items of SL or TL syntax known to be divergent between languages. For instance, as previously discussed, one might study how continuous aspects in English verbs are transferred to a language that has no such verb aspect. German, for instance, expresses this information either implicitly or through a variety of more complex constructions.

To investigate this, the user may type in a number of sentences in both English and German that convey continuous information and investigate annotation projection. One example may be the sentence pair in Figure 1: Here, the user finds that (1) the continuous aspect does not introduce errors since universal PoS tags and dependencies do not reflect such information, but that (2) the adverbial construction in German remains unlabeled. Other examples (see the accompanying screencast) show that syntactic projection is sometimes affected, while semantic projection is more robust. Based on these investigations, the user may conclude that either heuristic rules or a semi-supervised learning approach are appropriate to close the quality and coverage gap.

#### 3.3.2 Scenario 2: Adding A Layer of Annotation

A second example scenario is to investigate adding a layer of annotation to an already existing TL treebank. For instance, the universal dependencies treebanks are annotated with gold-standard PoS and dependency information for over 40 languages. However, there is as-yet no semantic layer of annotation. Previous work proposed to re-use English propositions as a layer of annotations for the universal treebanks (Akbik et al., 2015; Haverinen et al., 2015), but the applicability of these labels is a matter of ongoing discussion.

To investigate, a user may load a TL universal dependency corpus and browse example sentences. Each sentence is automatically translated into English, labeled with semantic roles which are then projected back onto the TL. Users qualitatively analyze whether the propositions are fitting and identify sources of errors such as suboptimal automatic translations, source language parsing errors and translational divergences.

## 4 Demonstration and Outlook

We present THE PROJECTOR as a hands-on demo where users can enter sentences or sentence pairs and request parsing and on-the-fly annotation projection. In order to enable the research community to quickly set up annotation projection experiments and discuss crosslingual syntax and semantics, we plan to make the toolkit publicly available, either through an online demo or in form of open source code. Our current work on THE PROJECTOR focuses on extending the functionality of the demo and the projection framework. This includes adding additional layers of annotation, such as named entities and word senses, into the parsing and projection pipeline. We also aim to enable more flexibility in choosing SL parsers and heuristics to address translational divergences, as proposed in previous work.

## Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 732328 ("FashionBrain").

## References

Alan Akbik, Laura Chiticariu, Marina Danilevsky, Yunyao Li, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2015. Generating high quality proposition banks for multilingual semantic role labeling. In *ACL 2015, 53rd Annual Meeting of the Association for Computational Linguistics*, pages 397–407.

Alan Akbik, Xinyu Guan, and Yunyao Li. 2016a. Multilingual aliasing for auto-generating proposition banks. In *COLING 2016, 26th International Conference on Computational Linguistics*, pages 3466–3474.

Alan Akbik, Kumar Vishwajeet, and Yunyao Li. 2016b. Towards semi-automatic generation of proposition banks for low-resource languages. In *EMNLP 2016, Conference on Empirical Methods on Natural Language Processing*, pages 993–998.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *ACL 1998, 36th Annual Meeting of the Association for Computational Linguistics*, pages 86–90.

Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *EMNLP-CoNLL 2012: Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena D. Hwang, and Martha Palmer. 2014. Propbank: Semantics of new predicate types. In *LREC 2014, 9th International Conference on Language Resources and Evaluation*, pages 3013–3019.

John DeNero and Percy Liang. 2007. The berkeley aligner. http://code.google.com/p/berkeleyaligner/.

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

Timur Gilmanov, Olga Scrivner, and Sandra Kübler. 2014. Swift aligner, a multifunctional tool for parallel corpora: Visualization, word alignment, and (morpho)-syntactic cross-language transfer. In *LREC 2014, 9th International Conference on Language Resources and Evaluation*, pages 2913–2919.

Katri Haverinen, Jenna Kanerva, Samuel Kohonen, Anna Missilä, Stina Ojala, Timo Viljanen, Veronika Laippala, and Filip Ginter. 2015. The finnish proposition bank. *Language Resources and Evaluation*, 49(4):907–926.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. Ontonotes: the 90% solution. In *Human Language Technology Conference of the NAACL*, pages 57–60.

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11(03):311–325.

Thomas Krause and Amir Zeldes. 2016. Annis3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31(1):118–139.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *ACL 2014. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *LREC 2016, 10th International Conference on Language Resources and Evaluation*, pages 1659–1666.

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36(1):307–340.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Lonneke Van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In *ACL 2011, 49th Annual Meeting of the Association for Computational Linguistics*, pages 299–304.

Michael Roth and Kristian Woodsend. 2014. Composition of word representations improves semantic role labelling. In *EMNLP 2014, Conference on Empirical Methods in Natural Language Processing*, pages 407–413.

Oscar Täckström, Ryan McDonald, and Jakob Uszkoreit. 2012. Cross-lingual word clusters for direct transfer of linguistic structure. In *NAACL 2012, Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pages 477–487.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *LREC 2012, 8th International Conference on Language Resources and Evaluation*, pages 2214–2218.

Jörg Tiedemann. 2014. Rediscovering annotation projection for cross-lingual parser induction. In *COLING 2014, 25th International Conference on Computational Linguistics*, pages 1854–1864.

Kitty Van Leuven-Zwart. 1989. Translation and original: Similarities and dissimilarities, i. *International Journal of Translation Studies*, 1(2):151–181.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

David Yarowsky and Grace Ngai. 2001. Inducing multilingual pos taggers and np bracketers via robust projection across aligned corpora. In *NAACL 2001, 2nd Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, pages 1–8.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *HLT 2001, 1st International Conference on Human Language Technology Research*, pages 1–8.