

Implicit Semantic Roles in a Multilingual Setting

Jennifer Sikos Yannick Versley Anette Frank

Department of Computational Linguistics

Heidelberg University, Germany

Leibniz Science Campus “Empirical Linguistics and Computational Language Modeling”

{sikos, versley, frank}@cl.uni-heidelberg.de

Abstract

Extending semantic role labeling (SRL) to detect and recover non-local arguments continues to be a challenge. Our work is the first to address the detection of implicit roles from a multilingual perspective. We map predicate-argument structures across English and German sentences, and we develop a classifier that distinguishes implicit arguments from other translation shifts. Using a combination of alignment statistics and linguistic features, we achieve a precision of 0.68 despite a limited training set, which is a significant gain over the majority baseline. Our approach does not rely on pre-existing knowledge bases and is extendible to any language pair with parallel data and dependency parses.

1 Introduction

Understanding events and their participants is a core NLP task, and SRL is the standard approach for identification and labeling of these events in text. SRL systems (Täckström et al., 2015; Roth and Woodsend, 2014) have benefited NLP applications, and many approaches have been proposed to transfer semantic roles from English to other languages without further reliance on manual annotation (Kozhevnikov and Titov, 2013; Padó and Lapata, 2009). However, event structures – both predicates and their arguments – are known to shift in the translation process, and this poor correspondence presents a bottleneck for the transference of semantic roles across languages. In some cases, the semantic content of an entire argument can be missing from the scope of its translated predicate.

Arguments that are omitted are often treated as noise in state-of-the-art projection models; however, our work views them as a valuable source of

data - such arguments serve as naturally occurring training data for *implicit* role detection. We target arguments that have been dislocated from their predicates, or are dropped entirely, in translated sentences. These non-isomorphic event structures can not only be leveraged as new training data for implicit role detection, but analyzing the shifts that trigger these implicit roles can guide improvements to systems that perform cross-lingual semantic role projection.

Implicit Roles If a predicate is known to have multiple semantic arguments, only a subset might be expressed within the local boundary of its clause or sentence. SRL models typically restrict their search for semantic arguments to this local domain and are not designed to recover arguments situated in the broader discourse context. Non-local role linking extends the SRL task by recovering the semantic arguments not instantiated in the local scope of the predicate. One complicating factor is that these implicit arguments can either be found in the context, and thereby are recoverable, or they could be existentially interpreted and might not correspond to any referent in the text at all. In the examples below, the argument for the predicate withdrawn in (1) is resolvable while the implicit argument for reading in (2) is not:

(1) El Salvador is now the only Latin American country which still has troops in [Iraq]¹. Nicaragua, Honduras, and the Dominican Republic have withdrawn their troops \emptyset .
Implicit role: Location

(2) I was sitting reading \emptyset in the chair.
Implicit role: Theme

Implicit role labeling systems consistently report low performance due to lack of training data. Combining the few existing resources improves

¹In this and other examples throughout the paper, the brackets [] indicate the antecedent of the implicit argument.

performance (Feizabadi and Padó, 2015) when they contribute diversity in predicate and argument types. Since much of the multilingual parallel corpora vary in domain and genre, mining these corpora for implicit roles should provide new training data that is sufficiently diverse to benefit the implicit role labeling task.

Predicate-Argument Structures across Languages Translational correspondences have been used in previous work to acquire resources for supervised monolingual tasks, such as word sense disambiguation (Diab and Resnik, 2002). Similarly, semantic role annotations can be transferred to new languages when predicate-argument structures are stable across language pairs (Padó and Lapata, 2009). In this work, we target predicate-argument structures that do not express such stability and have shifted in the translation process. In example (3), the role *farmers* is dropped entirely in the aligned German sentence:

- (3) The only change is that [farmers] are not required to produce.

Die einzige Neuerung ist, dass nicht gefordert
The only change is, that not required
wird zu produzieren.
are to produce.

The challenge in detecting implicit roles across languages is that these omissions represent only a fraction of the kinds of poor alignments that can occur. In fact, different types of translational shifts may occur that do not constitute cases of implicit role omission. Such factors include: change in part-of-speech from a verbal predicate to a noun or adjective, light verb constructions, single predicates that are expressed as both a verb and complement in the target language, and expressions with no direct translations (Samardžić et al., 2010).

Aims and Contributions To find implicit (non-local) semantic roles in translation, we distinguish role omissions from other types of translation shifts. We test linguistic features to automatically detect such role omissions in parallel corpora. We divide our work into alignment (Section 3.1) and classification (Section 3.2), with an annotation task for data construction (Section 4).

Our contributions are (i) a novel method for automatically identifying implicit roles in discourse, (ii) a classifier that is able to distinguish general translational divergences from true cases of implicit roles, (iii) an annotated, multilingual dataset

of manually tagged implicit arguments, and (iv) a classifier that achieves precision of 0.68 despite a small training set size, which is a significant improvement over a majority class baseline. Finally, we perform detailed analysis of our annotation and automatic classification results.

2 Related Work

2.1 Implicit Semantic Role Labeling

Previous resources for implicit SRL were developed over diverging schemas, texts, and predicate types. An initial dataset was constructed in the SemEval-2010 Shared Task “Linking Events and Their Participants in Discourse”, under the FrameNet paradigm; authors annotated short stories with implicit arguments and their antecedents, resulting in approx. 500 resolvable and 700 non-resolvable implicit roles out of roughly 3,000 frame instances (Ruppenhofer et al., 2010). Gerber and Chai (2010) focused on the implicit arguments of a constrained set of 10 nominal predicates in the NomBank scheme, annotating 966 implicit role instances for these specific predicates.

Numerous studies on the recovery of implicit roles have concluded that a lack of training data has been the stopping point towards improvements on the implicit role labeling task (Gorinski et al., 2013; Laparra and Rigau, 2013). To address this problem, Silberer and Frank (2012) generated artificial training data by removing arguments from coreference chains and showed that adding such instances yields performance gains. However, their quality was low and later work (Roth and Frank, 2015) has shown that smaller numbers of naturally occurring training data performed better. Roth and Frank (2015) applied a graph-based method for automatically acquiring high-quality data for non-local SRL using comparable monolingual corpora. They detect implicit semantic roles across documents and their antecedents from the prior context, again following cross-document links. In contrast, our work does not rely on semantic resources (SRL and lexical ontologies), but builds on parallel corpora enriched with dependencies and word alignments. Finally, Stern and Dagan (2014) generate training data for implicit SRL from textual entailment data sets. However, this type of resource needs to be manually curated.

2.2 Cross-lingual Annotation Projection

Aside from English, resources for SRL only exist for a select number of languages. For the languages that have such resources, annotated data still tends to vastly underrepresent the variability and breadth of coverage that exists for English. To extend SRL to new languages without reliance on manual annotation, models for role transference have been developed under both the supervised (Padó and Lapata, 2009; Akbik et al., 2015) and unsupervised (Kozhevnikov and Titov, 2013) setting. Most relevant to our work are previous studies that address the problem of projecting semantic role annotations across parallel corpora.

To transfer semantic annotations across languages, Padó and Lapata (2009) score the constituents of word-aligned parallel sentences and project role labels for the arguments that achieve highest constituent alignment scores. Akbik et al (2015) use filtered projection by constraining alignments through lexical and syntactic filters to ensure accuracy in predicate and argument role projection. Complete predicate-argument mappings are then used to bootstrap a classifier to recover further unaligned predicates and arguments.

3 Detecting Implicit Roles across Languages

We hypothesize that implicit semantic roles can be found in translated sentences, even in corpora where sentences are typically close translations. Our goal is to distinguish implicit roles from other translation shifts that cause poor alignment in SRL projection. A model is constructed based on lexical, syntactic, and alignment properties of parallel predicate-argument structures, and this classifier is, to the best of our knowledge, the first to detect a wide range of omitted roles in multilingual, parallel corpora. Our implicit role detection applies to both core and non-core arguments and is not dependent on large-scale SRL resources.

3.1 Identifying Poorly Aligned Arguments

Our first goal is to find candidates for implicit arguments by aligning predicate-argument structures across parallel English and German sentences.

Predicate and Argument Identification We target all non-auxiliary verbs as predicates, and detect their dependents through grammatical relations in dependency parses. We extract subjects,

direct objects, indirect objects, prepositional objects, adverbial or nominal modifiers as well as embedded clauses. These recover both the core and non-core arguments (adjuncts) of the predicate.² Arguments are attached to their nearest predicate and cannot be attached to more than one, as might occur in cases of embedded clauses.

Aligning Arguments for Detection of Unaligned Roles We use word alignments between parallel source (sl) and target (tl) language sentences as input. A predicate in the source language p^{sl} is mapped to a predicate in the target language p^{tl} if there exists a word alignment link between them, and their arguments are then aligned using the scoring function $ArgAL_p$ (Eq 3). $ArgAL_p$ uses word alignment links between the source and target arguments a^{sl} , a^{tl} of the aligned predicate pair to produce an optimal mapping between corresponding predicate-argument structures.

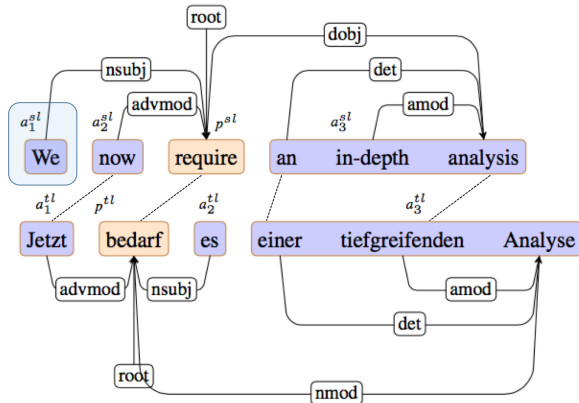
For scoring, we adapt Padó and Lapata (2009)’s constituent alignment-based overlap measure (Eq 1) to dependencies, where $yield(a)$ denotes the set of words in the yield (headword and dependents) of an argument a , and $align(a)$ the set of words in the target language that are aligned to the yield of a . Because the automatic word alignment tool gives predictions for links in both directions, we apply this asymmetric measure from the English-German and German-English links and average their results (Eq 2). The $ascore$ is computed for the Cartesian product $A^{sl} \times A^{tl}$ over all source and target arguments of the aligned predicates p^{sl} and p^{tl} . We select the argument alignments $A^{sl} \times A^{tl} \subseteq A^{sl} \times A^{tl}$ that return the maximal sum of scores for all arguments across the aligned argument structure (Eq 3).

Anticipating noise in the word alignments, we set a threshold to enforce accurate mappings between arguments. From the obtained mappings, we consider any argument whose alignment score does not exceed a threshold Θ as *unaligned* and thus as a candidate for an implicit role. The selection of threshold Θ is discussed in Section 5.

$$ovlp(a^{sl}, a^{tl}) = \frac{|align(a^{sl}) \cap yield(a^{tl})|}{|align(a^{sl}) \cup yield(a^{tl})|} \quad (1)$$

$$ascore(a^{sl}, a^{tl}) = \frac{ovlp(a^{sl}, a^{tl}) + ovlp(a^{tl}, a^{sl})}{2} \quad (2)$$

²Since we are treating arguments and adjuncts alike, in the following we loosely refer to both types of dependents as ‘arguments’.



Alignment Type	Aligned Arguments	Alignment Score
<i>headword</i>	$a_1^{sl}, -$	0
	a_2^{sl}, a_1^{tl}	1
	$a_3^{sl}, -$	0
<i>ascore</i>	$a_1^{sl}, -^{tl}$	0
	a_2^{sl}, a_1^{tl}	$(1 + 1)/2 = 1.0$
	a_3^{sl}, a_3^{tl}	$(2/3 + 2/3)/2 = 0.67$

Figure 1: Predicate-argument structures with noisy word alignments (left), and alignment scores for the arguments (right). *Headword* scoring aligns only headwords of the source (a^{sl}) and target (a^{tl}) arguments, while *ascore* uses headwords and dependents of an entire argument span for alignment.

$$ArgAL_p = \arg \max_{A^{sl} \times A^{tl} \subseteq A^{sl} \times A^{tl}} \sum_{A^{sl} \times A^{tl}} ascore(a^{sl}, a^{tl})$$

where

$$A^{sl} \times A^{tl} = \{ \langle a^{sl}, a^{tl} \rangle \mid a^{sl} \in \langle p^{sl}, a^{sl} \rangle, a^{tl} \in \langle p^{tl}, a^{tl} \rangle \} \quad (3)$$

An example of the alignment scoring is given in Figure 1, where predicates and arguments are detected over parallel English-German sentences, and word alignments are automatically generated. The argument ‘an in-depth analysis’ consists of a headword and two dependents, with two noisy word alignments that link the arguments across languages. Given these word alignment links, the *ascore* (Eq 2) is computed by taking the number of alignments and the yield of the arguments for both English and German, and these scores are then averaged for a final alignment score of 0.67. In this case, the scoring function still produces correct mappings across the predicate-argument structures despite imperfect word alignments, and an implicit role, *We*, is correctly unaligned to the German sentence.

3.2 Classification of Poor Alignments as Implicit Roles

Our objective is to build a classifier that automatically detects implicit roles across parallel corpora. To achieve this goal, we construct a classifier that takes as input an unaligned argument in English and, based on linguistic features in the aligned English and German sentences, determines whether this unaligned argument is an implicit role in German. Our dataset, described in Section 4.2, consists of instances of poorly aligned roles that have been annotated as either *implicit*, *not implicit*,

or *not a role of the predicate*. In classification, we reduce the annotation classes (*implicit/not implicit/not a role of the predicate*) to a binary decision where the positive class represents the implicit roles, and the negative class is any unaligned argument that annotators determined as either *not implicit* or *not a semantic role*. We reduced the task to a binary decision to avoid sparsity in the classification.

Features We hypothesize that we can predict the existence of an implicit role through features of the predicate-argument structures in the source and target languages. These features include monolingual predicate-argument structures, as well as cross-lingual features that represent the quality of the alignments across the parallel sentences. Monolingual features encode the syntactic properties of the arguments and predicates for source and target sentences, as well as sentential-level features that include the presence of modal and auxiliary verbs and conjunctions. To incorporate cross-lingual information, the alignment scores described in Section 3.1 are kept as features to the classifier, based on our assumption that the overall alignment between source and target predicate-argument structures should impact the classification of an implicit role. Both monolingual and cross-lingual features apply to surrounding predicate/arguments, where arguments can either be aligned or unaligned, and predicates that have fully aligned structures are considered *complete*. A complete list of features is shown in Table 1.

Classifiers We experimented with three classifiers, a Support Vector Machine (SVM) with a lin-

TYPE	FEATURE	XLING
Argument	Lemma	
	POS	
	Grammatical relation to predicate	
	Distance to predicate	
	Number of dependents	
	Syntactic path to predicate	
	Alignment type of neighboring argument (<i>aligned, unaligned</i>)	+
Predicate	Lemma	
	POS	
	Total number of arguments	
	Number of aligned arguments	+
	Number of unaligned arguments	+
	$ArgAL_p$ score	+
	Alignment type of neighboring predicate (<i>complete, incomplete</i>)	+
Sentential	Presence of a modal or auxiliary	
	Sentence-final punctuation marks before end	
	Conjunction between $p^{sl}, p^{sl}-I$	
	Sum of $ArgAL_p$ scores	
	Total number of arguments	
	Total number of predicates	
	Sum of $ArgAL_p$ scores over all predicates	+

Table 1: Features investigated for classification, where Xling are cross-lingual features.

ear kernel, Decision Tree, and Gradient Boosting, under the framework of the Scikit-learn library (Pedregosa et al., 2011).

4 Constructing a Dataset for Classifying Implicit Arguments

This section presents the construction of our experimental dataset for implicit role detection.

4.1 Corpora and Tools

We conduct our experiments over the Europarl corpus (Koehn, 2005), which contains over 1.9 million aligned sentences in our target languages. Anticipating noise in the automatic word alignments, we first take sentences from manually word-aligned German-English Europarl data (Padó and Lapata, 2005) to conduct our initial experiments. These sentences give us an upper bound for the number of implicit roles we should expect to obtain. Automatic word alignments are generated with GIZA++ (Och and Ney, 2003).

Predicates and their arguments are first detected

through dependency parses on English and German parallel corpora. Parses are generated for English with ClearNLP (Choi and McCallum, 2013). German sentences are run through the MarMot morphological analyzer (Mueller et al., 2013), and dependency parses for German are then generated using the RBG Parser (Lei et al., 2014). The Universal Dependencies project facilitates cross-lingual consistency in parsing and provides better compatibility amongst multiple languages. We trained the RBG Parser with the Universal Dependencies tagset (Rosa et al., 2014), and thus our argument detection can be applied to other languages in the Universal Dependencies project.

4.2 Annotation of Poorly Aligned Arguments

Annotation Instances Our goal is to find any argument that is either missing or dislocated from its predicate in translation. With this objective in mind, we focused our annotation on incomplete predicate structures whose argument(s) remained unaligned. Any argument with scores below the alignment threshold (see Section 3.1) was a candidate for annotation.

Annotation Task and Guidelines Three annotators worked on this task. Each annotator was a native German speaker with high fluency in English, and had taken at least one undergraduate course in linguistics. Annotators were given guidelines that define predicates as *events* or *scenarios*, and semantic roles as an element that has a semantic dependence on the predicate, including the *who*, *what*, *where*, *when*, and *why* type of information. Implicit roles were defined as “any role that is missing from the scope, or clausal boundary, of the predicate”. Each annotator was trained on a test set of 10 example sentences.

Annotators were given pairs of sentences with aligned predicates in English and German, where the English predicate had a poorly aligned argument. Annotation instances were presented as: two preceding English sentences, the English sentence with both the argument and predicate highlighted, the German sentence with the aligned predicate highlighted, and two preceding German sentences. An example of the annotation task is shown in Figure 2.

The annotation task was broken into two sub-tasks. First, annotators were asked to judge whether the marked argument is a correct semantic role for the English predicate. The second sub-

Context - 2 preceding English sentences
—
The only change is that [farmers] are not –required– to produce . Die einzige Neuerung ist , dass nicht –gefordert– wird zu produzieren .
—
Context - 2 preceding German sentences
—————
[farmers] Can ‘farmers’ be considered a role of the English predicate ‘required’?
If ‘no’: please choose: not a role of English predicate
Can ‘farmers’ be considered an implicit role for the German predicate ‘gefordert’?
If ‘no’: please choose: not an implicit role of German predicate If ‘yes’: please indicate the location of the German translation of ‘farmers’ by marking it in (**)

Figure 2: Example annotation task. Aligned predicates are marked in dashes (–) and implicit role candidates are surrounded by squared brackets [].

task asked annotators to judge whether a translation for the argument was available in the scope of the highlighted German predicate. If it was not available in the scope, they were asked to annotate the example as *implicit*.

Difficult Annotation Cases The annotations were adjudicated by one of the authors, and the annotator with the highest agreement with the adjudicator was asked to complete the entire dataset.

Cases that resulted in higher annotator disagreement included arguments of nominal predicates that were themselves the argument of the aligned predicate. In Example 4 below, *30 August* is a role for the nominal predicate *participation* but not *continue*:

- (4) The massive participation [from 30 August] must continue.

Other difficult annotation cases included roles that were partially, or entirely, encoded in the translated predicate. These included temporal adjuncts that could either be interpreted as present tense or implicit in the translated sentence:

- (5) I will [now] give the floor to the President
Ich gebe dem Präsidenten das Wort
I give the President the floor

After a review of these difficult cases, annotation guidelines were modified and annotators were re-trained.

Annotation Quality Inter-annotator agreement was measured by Cohen’s Kappa scores over 114 instances, and the entire 700 candidates were then completed by Annotator 1. One of the authors adjudicated for agreement. Results are given in Table 2 where “Role + Implicit” reports Kappa scores over all three categories - *not a role*, *implicit*, and *not implicit*, while “Implicit” reports agreement over binary *implicit vs non-implicit* decisions.

ANNOTATOR vs ADJUDICATOR	ROLE + IMPLICIT	IMPLICIT
ANNOTATOR 1	0.76	0.96
ANNOTATOR 2	0.48	0.92
ANNOTATOR 3	0.29	0.69

Table 2: Kappa agreements

Annotation Results In total, we took 700 poorly aligned arguments whose scores were below the alignment threshold (Section 3.1), where 500 were selected from manual word alignments and 200 from GIZA++ alignments. The 500 candidate arguments were sampled from 987 gold-aligned Europarl sentences, in which over 3,000 arguments fell below the threshold. The 200 candidates were sampled from 500 automatically aligned Europarl sentence pairs (excluding the sentences from the manually aligned dataset), with nearly 3,000 arguments below the threshold, to estimate the difference in implicit roles between manual and automatic word alignments.

Over the completed dataset, results for the annotation types are given in Table 3. Out of the manually aligned Europarl sentences, annotations produced 45 positive implicit role instances (9% of the annotated candidates). The automatic alignments, with 200 examples, contained 6 instances (3% of the annotated candidates) of implicit roles. Over the total 700 instances, 24.5% were classified as ‘not a predicate role’, 68.3% as ‘not implicit’, and 7.2% as ‘implicit’.

	INSTANCES	NOT A ROLE	NOT IMPLICIT	IMPLICIT
Manual	500	154	301	45
GIZA++	200	18	176	6

Table 3: Final annotation dataset.

Classifier	P	R	F1
Majority Baseline	0	0	0
SVM-ablated	0.6805	0.4444	0.5128
SVM-all	0.1555	0.2238	0.18333
Decision Tree-ablated	0.6682	0.4155	0.4934
Decision Tree-all	0.4134	0.2222	0.2748
Gradient Boosting-ablated	0.6688	0.3777	0.4631
Gradient Boosting-all	0.6466	0.2222	0.3268

Table 4: Precision, Recall and F_1 for the positive class (*implicit role*), with stratified 5-fold CV.

5 Classification Experiments and Results

5.1 Argument Alignment and Scoring

With the scoring function described in Section 3.1, perfectly aligned arguments should produce a score of 1.0. We experimentally set the threshold Θ for the minimum alignment score at 0.2 for arguments such that arguments with imperfect word alignments will still be aligned.

5.2 Classification of Implicit Arguments

The data set constructed in Section 4 resulted in 51 manually validated implicit roles and 649 negative instances that were input for classification.

We measure precision, recall, and F_1 scores, and for the SVM and Gradient Boosting classifiers we experimented with parameters to optimize precision. The SVM classifier with a linear kernel produced the highest scores, but results were closely followed by Decision Tree and Gradient Boosting classifiers. For the SVM classifier, we experimented with different regularization $\{0.5, 1, 10, 20\}$ and class weight increments $\{\text{None}, 1:2, 1:10\}$ and found the highest precision scores were achieved with $C=0.5$ and class weight 1:2. In Gradient Boosting, we experimented with max depth $\{1, 2, 3\}$ and found the highest precision scores were obtained with a max depth of 2. Since the data set is heavily biased towards the negative class, we divided training and test sets with a stratified 5-fold cross-validation (CV). We later experimented with upsampling for the positive class but found no significant improvement.

Feature Ablation To determine the optimal feature set, we performed ablation tests by incrementally removing a feature and performing training/testing over the reduced feature set. Ablation was performed individually for each classifier. After these tests, we eliminated features that caused

Type	Feature
a^{sl}	lemma, POS, path to predicate
$a^{sl}+1$	POS, path to predicate
$a^{sl}-1$	alignment type, number of dependents
p^{tl}	POS
$p^{tl}-1$	sum of $ArgAL_p$ scores
$p^{tl}+1$	POS, number of arguments, number of unaligned arguments, sum of $ArgAL_p$ scores, alignment type

Table 5: Final feature set used in classification. Notation is defined in Section 3.1, where ± 1 are the arguments/predicates preceding (-1) and following (+1) the candidate.

a drop in performance and used only the best performing features in the final classification. The final feature set is shown in Table 5.

The SVM model obtains the best results of 0.68 precision and F_1 -score of 0.51 with the ablated feature set, closely followed by the other classifier models and outperforming the majority baseline, which always predicts the negative class (see Table 4 for both ablated and full feature results).

Feature Analysis The final feature set used in the classification experiment included both cross-lingual features of the predicate and arguments on source/target sentences, as well as monolingual predicate and argument features. The ablation results support our initial hypothesis that the surrounding predicate/argument structures and alignment scores are relevant to the detection of an omitted role.

5.3 Analysis of Results

Translation Shifts that Trigger Implicit Roles

Through observation of the positive instances, we determined a number of syntactic environments that trigger omission of semantic roles from English to German. Shift in voice, finite to infinite verb forms, and coordination could all motivate the deletion of a role across translated sentences. While these syntactically licensed implicit roles composed 57% of our positive instances, a large number (43%) were not found to have an explanation on syntactic grounds alone. In these cases, the arguments seem to have been omitted by pragmatic or semantic factors. The distribution of these shift types over our dataset is given in Table 6.

Voice A change from active (source) to passive (target). Subjects are dropped in translation:

- (6) The more [we] refuse to democratize the institutions

Je mehr die Demokratisierung der
The more the democratization of the
Institutionen verweigert wird ...
institution refused are

Coordination An argument might be the repeated subject of two conjoined clauses, but expressed as a shared argument in the parallel sentence:

- (7) I was faced with this system *and* [I] do not know any parliament

Ich fand dieses System vor *und* kenne kein
I faced this system before and know no
Parlament
parliament

Extrapolation Complex clausal embeddings can cause roles to be extraposed from their predicates in the target language text:

- (8) ...but would also want to encourage both parties [to observe the spirit of this new agreement].

...er kann die beiden Parteien nur veranlassen
...it can that both parties only encourage
wollen, [den Geist dieses neuen Abkommens
want, the spirit of-this new agreement
zu achten].
to observe

Coordination and extraposition are borderline cases with regard to the non-locality of roles. PropBank does annotate coordinated arguments, and in these cases the syntactic parse tree can be leveraged for recovery of the non-local role. However, we still consider these implicit arguments since they are expressed outside of the local scope of the predicates.

Nonfinite Similar to change in voice, the subject of a finite verb can be dropped when the translated verb is nonfinite:

- (9) I would ask that [they] reconsider these decisions

Ich bitte, diese Entscheidung zu überdenken
I ask, these decisions to reconsider

Semantic/Pragmatic A role can be dropped in translation without a structural shift that licenses the omission. In these instances, the role could have been incorporated into the aligned sentence without a change to the syntactic environment.

- (10) ... I am asking you to do this directly, [in this House].

...wende ich mich hiermit direkt an Sie .
...turn I myself hereby directly to you

Since the directionality of our implicit role search focused on English to German, we do not account for syntactic shifts that could cause omissions in the opposite direction, i.e. German to English. There are imperative constructions in German that overtly encode the addressee of the command (“go outside” in English can be translated as “go *you* outside” in German) which can trigger implicit roles in translation from German to English.

Shift Type	Count	%
Voice	7	14%
Coordination	8	16%
Extrapolation	8	16%
Nonfinite	6	11%
Semantic/Pragmatic	22	43%

Table 6: Shift types that trigger implicit roles.

Semantic Role Types of Omitted Arguments We adopt the VerbNet roleset (Kipper et al., 2000) to manually label semantic role across all our implicit argument instances. A full analysis of the role types, shown in Table 7, found that a majority of implicit roles are *Agent* and *Theme*. This reflects the general distributions for role frequency (Merlo and Van Der Plas, 2009), but could also be due to the syntactic shifts that produce a higher omission of the subject, such as passivization and coordination, which are commonly filled by the *Agent* and *Theme* roles.

Core Role	Count	Non-Core Role	Count
Agent	15	Time	6
Theme	14	Topic	5
Recipient	3	Location	3
Experiencer	2	Manner	1
Cause	2		

Table 7: Thematic roles, both core and non-core, of the implicit cases.

Antecedents to the Implicit Role The analyses above described the shift types that trigger argument omission, but only two of these types, coordination and extraposition, would guarantee the missing argument to be recoverable from the non-local context. Cases where the annotators were

able to recover the antecedent roles, either from the previous clause or sentences, were less than the majority (21 out of the 51 cases), while many instances were not instantiated in the non-local context. Table 8 gives the proportion of recovered antecedents according to shift types. The fact that extraposition and coordination cases yield higher number of resolvable roles can be exploited in future work for antecedent linking.

Shift Type	Resolvable	Not resolvable
Voice	1	6
Coordination	8	0
Extraposition	8	0
Nonfinite	1	5
Semantic/Pragmatic	3	19

Table 8: Availability of the antecedent in the surrounding context.

6 Conclusion and Future Work

In this work, we investigated the hypothesis that implicit semantic roles can be identified in translation. Our method is knowledge-lean and achieves respectable performance despite a small training set. While the present work has focused on missing arguments of verbal predicates, implicit role detection in this multilingual framework can be easily extended to nominal predicates. Combining both predicate types is expected to improve the overall results, as some of the noise we are currently observing pertains to implicit roles occurring with nouns. Additional noise is produced by the automatic word alignments, which can be addressed by employing triangulation techniques using multiple language pairs. Further, with our current classifier we can predict role omissions across parallel sentences with better accuracy than reliance on noisy word alignments alone, and with these predictions we can generate better candidates for annotation and reduce the time and cost of future annotation effort.

A next step from the current work would be to automatically recover the antecedent of the implicit role in the target language when it is available. By doing so, we can construct new training data for monolingual implicit role labeling, improve transference of semantic roles across parallel corpora, and generate novel training data for implicit role labeling for new languages.

7 Acknowledgments

This research has been conducted within the Leibniz Science Campus “Empirical Linguistics and Computational Modeling”, funded by the Leibniz Association under grant no. SAS-2015-IDS-LWC and by the Ministry of Science, Research, and Art (MWK) of the state of Baden-Württemberg. We thank our annotators Leo Born, Max Müller-Eberstein and Julius Steen for their contribution.

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 *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 397–407, Beijing, China.
- Jinho D. Choi and Andrew McCallum. 2013. Transition-based Dependency Parsing with Selectional Branching. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1052–1062, Sofia, Bulgaria.
- Mona Diab and Philip Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 255–262. Association for Computational Linguistics.
- Parvin Sadat Feizabadi and Sebastian Padó. 2015. Combining Seemingly Incompatible Corpora for Implicit Semantic Role Labeling. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 40–50, Denver, Colorado, June.
- Matthew Gerber and Joyce Y Chai. 2010. Beyond NomBank: A study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1583–1592.
- Philip Gorinski, Josef Ruppenhofer, and Caroline Sporleder. 2013. Towards weakly supervised resolution of null instantiations. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 119–130.
- Karin Kipper, Hoa Trang Dang, Martha Palmer, et al. 2000. Class-based construction of a verb lexicon. In *AAAI/IAAI*, pages 691–696.
- Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86.

- Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual Transfer of Semantic Role Labeling Models. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1190–1200, Sofia, Bulgaria, August.
- Egoitz Laparra and German Rigau. 2013. ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling. In *ACL (1)*, pages 1180–1189.
- Tao Lei, Yu Xin, Yuan Zhang, Regina Barzilay, and Tommi Jaakkola. 2014. Low-Rank Tensors for Scoring Dependency Structures. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1381–1391, Baltimore, Maryland.
- Paola Merlo and Lonneke Van Der Plas. 2009. Abstraction and generalisation in semantic role labels: PropBank, VerbNet or both? In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 288–296. Association for Computational Linguistics.
- Thomas Mueller, Helmut Schmid, and Hinrich Schütze. 2013. Efficient Higher-Order CRFs for Morphological Tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 322–332, Seattle, Washington, USA.
- Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.
- Sebastian Padó and Mirella Lapata. 2005. Cross-lingual projection of role-semantic information. In *Proceedings of HLT/EMNLP 2005*, Vancouver, BC.
- Sebastian Padó and Mirella Lapata. 2009. Cross-lingual Annotation Projection for Semantic Roles. *Journal of Artificial Intelligence Research*, 36(1):307–340.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Rudolf Rosa, Jan Masek, David Marecek, Martin Popel, Daniel Zeman, and Zdenek Zabokrtský. 2014. HamleDT 2.0: Thirty Dependency Treebanks Stanfordized. In *LREC*, pages 2334–2341.
- Michael Roth and Anette Frank. 2015. Inducing Implicit Arguments from Comparable Texts: A Framework and its Applications. *Computational Linguistics*, 41(4):625–664.
- Michael Roth and Kristian Woodsend. 2014. Composition of Word Representations Improves Semantic Role Labelling. In *EMNLP*, pages 407–413.
- Josef Ruppenhofer, Caroline Sporleder, Roser Morante, Collin Baker, and Martha Palmer. 2010. Semeval-2010 task 10: Linking events and their participants in discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 45–50.
- Tanja Samardžić, Lonneke Van Der Plas, Goljihan Kashaeva, and Paola Merlo. 2010. The Scope and the Sources of Variation in Verbal Predicates in English and French. In *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories*, pages 199–211.
- Carina Silberer and Anette Frank. 2012. Casting Implicit Role Linking as an Anaphora Resolution Task. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 1–10.
- Asher Stern and Ido Dagan. 2014. Recognizing implied predicate-argument relationships in textual inference. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 739–744, Baltimore, Maryland.
- Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Efficient inference and structured learning for semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:29–41.