Formalization of natural language processing tasks as optimization problems

Master thesis

Oleh Shkalikov (5102818)

Supervisor: Jannik Irmai

Reviewers: Prof. Dr. Bjoern Andres, Prof. Dr. Simon Razniewski

TU Dresden

January 24, 2025

Named Entity Recognition

What is NER?

Predict a class for every word and then aggregate entities, i.e. continuous sequences of words that belong to a single object.

B-PER	I-PER	O	O	O	B-LOC
Mark	Twain	was	born	in	Florida

Figure 1: Example of NER labeling in IOB2 format

Why NER is Important?

NER is a building block of various systems such as information retrieval, question answering, search engines, and more.

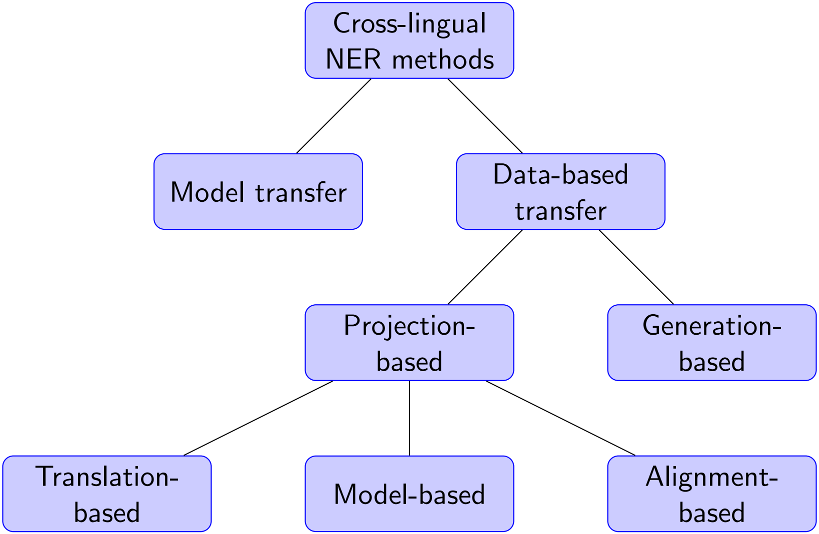
Issues of multilingual NER

The lack of (labelled) data is the main problem in multilingual NER!

Number of datasets for the token classification task on the HF Hub:

- English: 462
- German: 91
- Ukrainian: 21
- Estonian: 22
- Thai: 11
- Upper Sorbian: 8

Classification of approaches to XLNER



Model transfer

General idea

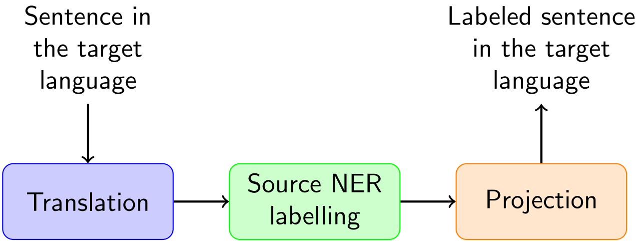
Transfer knoweledge from high-resource language to all languages

- Encoder-only: BERT-like
- Full Transformer: mT5, etc
- Decoder-only: LLMs
  - Prompting
  - Constrained Generation
  - Zero-/Few- shot

Language	Percent	Language	Percent
en	89.70	uk	0.07
unknown	8.38	ko	0.06
de	0.17	ca	0.04
fr	0.16	sr	0.04
sv	0.15	id	0.03
zh	0.13	cs	0.03
es	0.13	fi	0.03
ru	0.13	hu	0.03
nl	0.12	no	0.03
it	0.11	ro	0.03
ja	0.10	bg	0.02
pl	0.09	da	0.02
pt	0.09	sl	0.01
vi	0.08	hr	0.01

Table 1: Language distribution in pretraining data of the Llama2 with percentage above 0.005%

Projection-based pipeline



- Transformer (NLLB-200, M2M-100, MBart50)

- BERT-like
- LLM

- CROP, EasyProject, CODEC, CIAP
- TransFusion
- W2W alignment based

Figure 2: General scheme of projection-based pipelines

Translation-based projection: CROP and EasyProject

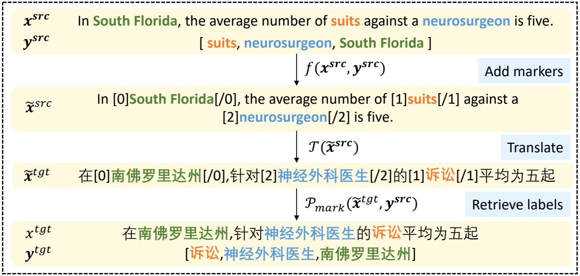


Figure 3: Marker-based methods. Illustration is from the CLAP paper.

CROP

Markers: __SLOT1__, ..., __SLOTn__
Idea: save label for every slot
Projection: string matching between backtranslated and original enclosed in slots substrings

EasyProject

Markers: [,]
Idea: translate entities independently
Projection: fuzzy string matching between translated entities and enclosed in braces substrings

Translation-based projection: T-Projection

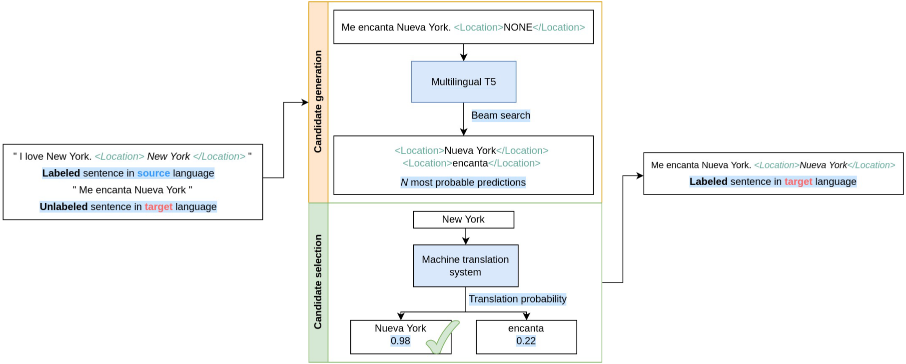


Figure 4: T-Projection. Illustration is from the original paper.

The closest related work

The idea is structurally similar to our method, but the projection step has not been formalized explicitly.

Translation-based projection: CLAP and CODEC

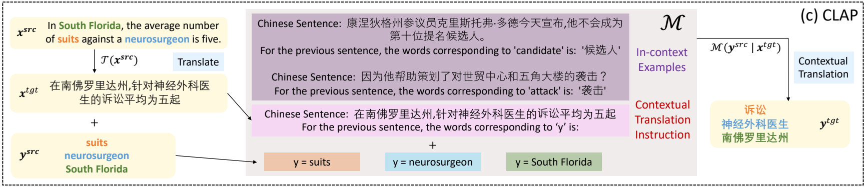


Figure 5: Contextual Label Projection for Cross-Lingual Structured Prediction. Illustration is from the original paper.

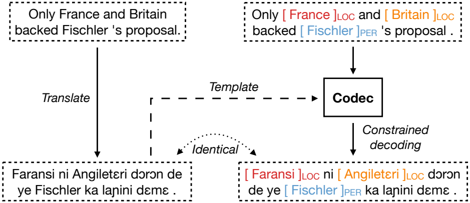


Figure 6: Constrained Decoding for Cross-lingual Label Projection. Illustration is from the original paper.

Model-based projection

Idea

Use model to perform projection

TransFusion Decoder-only input/output format:

$$[\text{GoLLIE Guidelines}, \text{src. sent}, \text{tgt. sent}, \text{Instructions}] \rightarrow [\text{src. ent.}, \text{tgt. ent.}]$$

TransFusion Encoder-only input format:

$$w_1^{src}, w_2^{src}, \langle \text{PER} \rangle, w_3^{src}, \langle /\text{PER} \rangle, w_4^{src}, \langle \text{sep} \rangle, w_1^{tgt}, w_2^{tgt}, w_3^{tgt}$$

Fundamental problem

TransFusion requires training, but if there are labelled parellel texts one can train a model in the target language directly

Word-to-word alignment-based projection

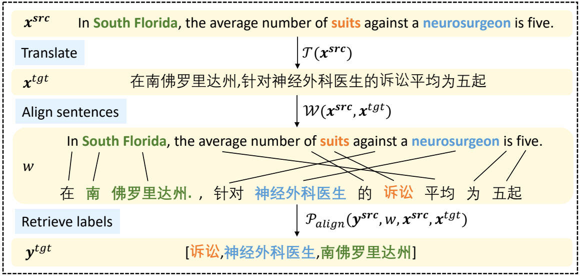


Figure 7: W2W alignment-based method. Illustration is from the CLAP paper.

Issues and heuristics to handle them:

Split annotations $\implies$ merge if only *d* non-aligned words are between

Annotation collisions $\implies$ project to the longest

Incorrect alignments $\implies$ project only to the top-k by length ranges, threshold by word length ratio

Idea and requirements to the formalization

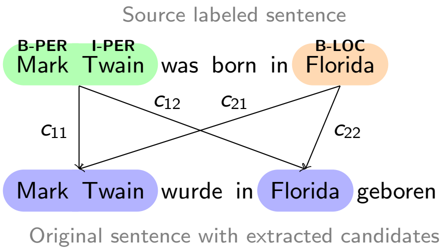


Figure 8: Matching of extracted target candidates with source entities

Requirements:

- Projection onto overlapping candidates should not be possible.
- There is a limit on the number of projections from any source entity.

Definitions

Let $s^{src} = \left(s_1^{src}, \ldots, s_n^{tgt}\right)$ and $s^{tgt} = \left(s_1^{tgt}, \ldots, s_m^{tgt}\right)$ be a source and target sentences respectively, L – set of classes.

Source entity

The source entity $p^{src} = (i_{p^{src}}, j_{p^{src}}, l_{p^{src}})$ is a continuous subrange of source words, where $i_{p^{src}} \in \{1, \ldots, n\}$ and $j_{p^{src}} \in \{1, \ldots, n\}$ are indices of the first and last word and $l_{p^{src}} \in L$ is the class of the entity.

Target candidate

The target candidate $p^{tgt} = (i_{p^{tgt}}, j_{p^{tgt}})$ is defined as a continuous subrange of target words.

Relation of overlapping

$p_1^{tgt} \cap p_2^{tgt} \neq \emptyset \Leftrightarrow (i_{p_1^{tgt}} \leq j_{p_2^{tgt}}) \wedge (i_{p_2^{tgt}} \leq j_{p_1^{tgt}})$, i.e. the sets of words are not disjoint.

Projection ILP problem

The projection problem is the following:

$$\max_x \sum_{(p^{src},p^{tgt})\in S\times T} c_{p^{src},p^{tgt}} x_{p^{src},p^{tgt}}$$

subject to

$$\sum_{p^{tgt}\in T} x_{p^{src},p^{tgt}} \lesseqgtr n_{proj} \qquad \forall p^{src} \in S$$

$$x_{p_1^{src},p_1^{tgt}} + x_{p_2^{src},p_2^{tgt}} \leq 1 \qquad \forall (p_1^{src}, p_2^{src}, p_1^{tgt}, p_2^{tgt}) \in \hat{\Pi}(S, T)$$

$$x_{p^{src},p^{tgt}} \in \{0, 1\} \qquad \forall (p^{src}, p^{tgt}) \in S \times T$$

where

$$\hat{\Pi}(S,T) = \Big\{ (p_1^{src}, p_2^{src}, p_1^{tgt}, p_2^{tgt}) \Big| p_1^{src}, p_2^{src} \in S, p_1^{tgt}, p_2^{tgt} \in T, \\ p_1^{tgt} \cap p_2^{tgt} \neq \emptyset, (p_1^{src} \neq p_2^{src}) \vee (p_1^{tgt} \neq p_2^{tgt}) \Big\}$$

Reduction of the non-overlapping constraints

If it is impossible to project one or any two source entities onto overlapping candidates, then it is also impossible to project any number of source entities onto these candidates

Theorem

The set of non-overlapping constraints is satisfied if and only if the following sets of reduced constraints are satisfied:

$$\sum_{p^{src} \in S} (x_{p^{src}, p_1^{tgt}} + x_{p^{src}, p_2^{tgt}}) \leq 1 \qquad \forall (p_1^{tgt}, p_2^{tgt}) \in \Pi(T)$$

$$\sum_{p^{src} \in S} x_{p^{src}, p^{tgt}} \leq 1 \qquad \forall p^{tgt} \in T \,\middle|\, \nexists p_2^{tgt} \in T : p^{tgt} \neq p_2^{tgt}, p^{tgt} \cap p_2^{tgt} \neq \emptyset$$

where

$$\Pi(T) = \left\{ (p_1^{tgt}, p_2^{tgt}) \,\middle|\, p_1^{tgt}, p_2^{tgt} \in T, \quad p_1^{tgt} \cap p_2^{tgt} \neq \emptyset, p_1^{tgt} \neq p_2^{tgt} \right\}$$

ILP problem with reduced number of constraints

The final form of the projection problem is

$$\max_{x} \sum_{(p^{src},p^{tgt})\in S\times T} c_{p^{src},p^{tgt}} x_{p^{src},p^{tgt}}$$

subject to

$$\sum_{p^{tgt}\in T} x_{p^{src},p^{tgt}} \lesseqgtr n_{proj} \qquad \forall p^{src} \in S$$

$$\sum_{p^{src}\in S} (x_{p^{src},p_1^{tgt}} + x_{p^{src},p_2^{tgt}}) \leq 1 \qquad \forall (p_1^{tgt}, p_2^{tgt}) \in \Pi(T)$$

$$\sum_{p^{src}\in S} x_{p^{src},p^{tgt}} \leq 1 \qquad \forall p^{tgt} \in T \Big| \nexists p_2^{tgt} \in T : p^{tgt} \neq p_2^{tgt}, p^{tgt} \cap p_2^{tgt} \neq \emptyset$$

$$x_{p^{src},p^{tgt}} \in \{0, 1\} \qquad \forall (p^{src}, p^{tgt}) \in S \times T$$

Candidate extraction

Let's consider the set of target candidates as all n-grams (composed of words from the target sentence).

Problem: Number of candidates is too high

Limit the n-gram length to a fixed number.

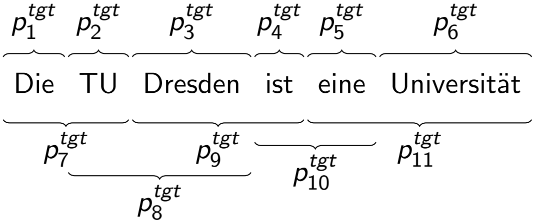


Figure 9: An example of a target candidate extraction algorithm with an n-gram length limit of 2 words

Word-to-word alignment-based matching score

Let a_{kl} be binary variable denoted if k-th source word aligned to l target word, then the word-to-word alignment-based mathcing score is the following:

$$c^{align}_{p^{src},p^{tgt}} = \frac{\sum\limits_{k=i_{p^{src}}}^{j_{p^{src}}} \sum\limits_{l=i_{p^{tgt}}}^{j_{p^{tgt}}} a_{kl}}{(j_{p^{src}} - i_{p^{src}}) + (j_{p^{tgt}} - i_{p^{tgt}})}$$

Pros:

- Alignments $\rightarrow$ quanitive score
- Allow to reduce set of target candidates

Cons:

- "Merging" heuristic is not always preserved
- Scores are not bounded by a shared constant

For $<$ and $\leq$ type of constraints, the set of candidates can be reduced by removing all candidates that lie outside the range between leftmost and rightmost word aligned with any entity words.

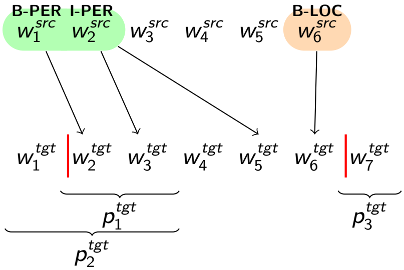


Figure 10: All candidates that start or end outside the red lines can be removed

Note, that $\forall p^{src} \in S$:

$$c_{p^{src}, p_3^{src}} = 0 \qquad c_{p^{src}, p_1^{src}} \geq c_{p^{src}, p_2^{src}}$$

NER model-based matching score

Let $p_{k,l}$ be the probability predicted by a NER model that the k-th target word has a label $l \in L$. Let $\mathcal{B}$ and $\mathcal{I}$ be mappings that return the indices of B- and I- labels in the model output matrix. Then the NER-based score is defined as follows:

$$c^{ner}_{p^{src},p^{tgt}} = \alpha^{(j_{p^{tgt}} - i_{p^{tgt}})-1} \frac{p_{i_{p^{tgt}},\mathcal{B}[l_{p^{src}}]} + \sum_{k=i_{p^{tgt}}+1}^{j_{p^{tgt}}} p_{k,\mathcal{I}[l_{p^{src}}]}}{j_{p^{tgt}} - i_{p^{tgt}}}$$

Pros:

- Projection $\implies$ can correct missed/wrongly predicted by a NER model entities

Cons:

- Projection $\implies$ depends on quality of source NER labelling
- Scores are identical for all source entities sharing the same class
- Overconfidence of NER models

NER model-based matching score: alpha parameter

Let s be a source entity with label PER. And the model output is the following:

$\underset{0.9}{\text{Mark}}$ $\underset{0.8}{\text{Twain}}$ wurde in $\underset{0.95}{\text{Florida}}$ geboren

If $\alpha = 1$, then the output of the projection step will be misaligned with the output of the NER model:

$$c^{ner}_{s,\text{"}Mark\text{"}} = 0.9 \qquad c^{ner}_{s,\text{"}MarkTwain\text{"}} = 0.85$$

Estimation of α

Let $M = \frac{1}{2|L|+1}$. Then, if $\alpha > 1 + \frac{1-M}{1+M}$, the predictions of the projection step and the NER model are aligned.

Translation-based matching score

The idea is to compute how likely a source entity is translated into a target candidate. To accomplish this, we use the NMTScore (denoted as the *sim* function):

$$c^{nmt}_{p^{src},p^{tgt}} = sim\left(w^{src}[i_{p^{src}} : j_{p^{src}}], w^{tgt}[i_{p^{tgt}} : j_{p^{tgt}}]\right)$$

Pros:

- Interpretability

Cons:

- Translation without context
- Slower then other scores

Fused matching score

Since all matching score are just real numbers we can compute a weighted sum of them, i.e. combine different projection strategies together:

$$c^{fused}_{p^{src},p^{tgt}} = \lambda_{align} c^{align}_{p^{src},p^{tgt}} + \lambda_{ner} c^{ner}_{p^{src},p^{tgt}} + \lambda_{nmt} c^{nmt}_{p^{src},p^{tgt}}$$

where $\lambda_{align} \geq 0, \lambda_{ner} \geq 0, \lambda_{nmt} \geq 0 \in \mathbb{R}$.

Then, NER-based score can be used to evaluate only spans without labels:

$$c^{ner}_{p^{src},p^{tgt}} = \alpha^{(j_{p^{tgt}} - i_{p^{tgt}})-1} \max_{l \in L} \frac{p_{i_{p^{tgt}},B[l]} + \sum\limits_{k=i_{p^{tgt}}+1}^{j_{p^{tgt}}} p_{k,I[l]}}{j_{p^{tgt}} - i_{p^{tgt}}}$$

Complexity

Current results:

- MaxIS is reducible to the generalized problem with target candidates beyond intervals $\implies$ NP-Hard
- The ILP problem without constraint on number of projections reduces to MaxWIS on **interval graphs** $\implies$ can be solved in polynomial time

Complexity of the original projection problem remains an open question.

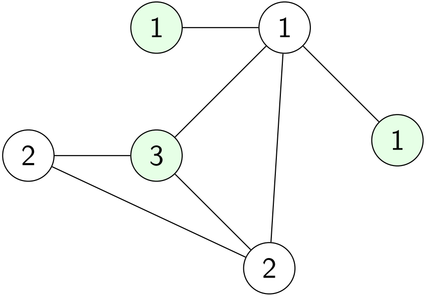


Figure 11: Maximum weight independent set problem (MaxWIS). Vertices that form an optimal solution are colored in green

Simplified projection problem

$$\max_{x} \sum_{(p^{src}, p^{tgt}) \in S \times T} c_{p^{src}, p^{tgt}} x_{p^{src}, p^{tgt}}$$

subject to

$$x_{p_1^{src}, p_1^{tgt}} + x_{p_2^{src}, p_2^{tgt}} \leq 1$$

$$x_{p^{src}, p^{tgt}} \in \{0, 1\}$$

MaxWIS problem

$$\max \sum_{v \in V} w_v x_v$$

subject to

$$x_u + x_v \leq 1 \qquad \forall \{u, v\} \in E$$

$$x_v \in \{0, 1\}$$

Reduction

$$w_v = \max_{s \in S} c_{s, t_v} \qquad s_v^* = \arg\max_{s \in S} x_{s, t_v}$$

$$x_v = 1 \implies \begin{cases} x_{s^*, t_v} = 1 \\ x_{s, t_v} = 0 \quad \forall s \in S \setminus \{s^*\} \end{cases}$$

$$\{v, u\} \in E \Leftrightarrow t_u \cap t_v \neq \emptyset$$

Reduction to the MaxWIS problem

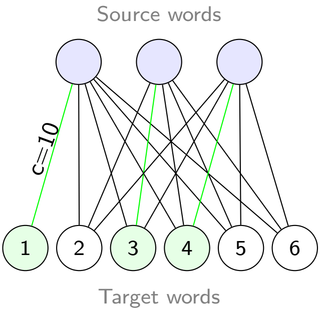


Figure 12: The simplified projeciton problem, maximum weight edges are colored in green

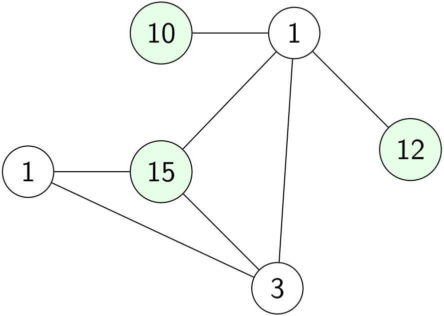
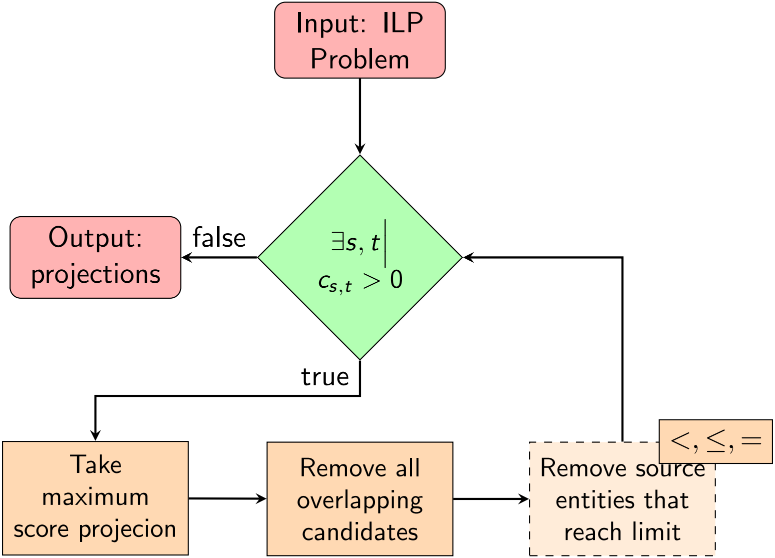


Figure 13: Induced maximum weight independent set problem on an interval graph

Approximate algorithm



Example of approximate algorithm failure

Consider the ILP problem with $n_{proj} = 2$ and the $\leq$ type of constraints. Let $T = \{p_1^{tgt} = (1,2), p_2^{tgt} = (2,3), p_3^{tgt} = (3,5)\}$ and $S = \{p^{src}\}$, and let the matching scores be as follows:

	p_1^{tgt}	p_2^{tgt}	p_3^{tgt}
p^{src}	0.2	0.3	0.2

Note, that $p_1^{tgt} \cap p_2^{tgt} \neq \emptyset$ and $p_2^{tgt} \cap p_3^{tgt} \neq \emptyset$, but $p_2^{tgt} \cap p_3^{tgt} = \emptyset$

The output of the Algorithm:

$x_{p^{src},p_1^{tgt}} = 0$
$x_{p^{src},p_2^{tgt}} = 1$
$x_{p^{src},p_3^{tgt}} = 0$

Objective value: 0.3

Optimal solution:

$x_{p^{src},p_1^{tgt}} = 1$
$x_{p^{src},p_2^{tgt}} = 0$
$x_{p^{src},p_3^{tgt}} = 1$

Objective value: 0.4

Experiments setup

- Intrinsic evaluation
- Translation and NER models are shared between experiments
- Translation model: NLLB-200
- NER model: MDeBERTa-v3-base finetuned on the English split of the CONLL-2003
- Aligner: non-finetuned AWeSOME with default parameters
- Maximum candidate length: 10 words
- Exact ILP solver: GUROBI

			tgt_lang	de	es	it
d	k	only_i	thr			
0	1	False	0.8	0.814	0.873	0.838
			-	0.896	0.819	0.789
		True	0.8	0.814	0.873	0.838
			-	0.896	0.819	0.789
	-	False	0.8	0.814	0.872	0.840
			-	0.875	0.763	0.735
		True	0.8	0.814	0.872	0.840
			-	0.875	0.763	0.735
1	1	False	0.8	0.819	**0.903**	0.871
			-	**0.916**	0.875	0.846
		True	0.8	0.815	0.886	0.848
			-	0.899	0.847	0.813
	-	False	0.8	0.819	**0.903**	**0.873**
			-	0.912	0.858	0.832
		True	0.8	0.815	0.885	0.850
			-	0.886	0.816	0.782

Table 2: Overall F1 scores for word-to-word alignments-based heuristic algorithm with different hyperparameter on the Europarl NER dataset

Europarl-NER: ILP and model transfer

pipeline	constr. type	tgt_lang / solver / n_{proj}	de GREEDY	de GUROBI	es GREEDY	es GUROBI	it GREEDY	it GUROBI
align	$\leq$	1	0.920	**0.921**	0.883	0.883	0.866	0.864
		2	0.920	0.650	0.883	0.488	0.866	0.471
	$=$	1	0.920	0.918	0.883	0.883	0.866	0.864
	$\geq$	0	0.908	0.569	0.863	0.417	0.853	0.400
		1	0.908	0.569	0.863	0.417	0.853	0.400
ner	$\leq$	1	0.713	0.714	0.739	0.734	0.713	0.705
		2	0.713	0.669	0.739	0.665	0.713	0.660
	$=$	1	0.713	0.670	0.739	0.706	0.713	0.665
	$\geq$	0	0.690	0.641	0.716	0.629	0.685	0.653
		1	0.690	0.598	0.716	0.607	0.685	0.613
nmt	$\leq$	1	0.876	0.886	0.906	**0.916**	0.872	**0.879**
		2	0.876	0.602	0.906	0.635	0.872	0.601
	$=$	1	0.876	0.886	0.906	**0.916**	0.872	**0.879**
	$\geq$	0	0.233	0.180	0.315	0.220	0.271	0.203
		1	0.233	0.181	0.315	0.220	0.271	0.203
Model transfer	-	-	0.621		0.653		0.657	

Table 3: Overall F1 scores for the model transfer and ILP based projection pipelines on the Europarl NER dataset.

Europarl-NER: Heuristic vs ILP-based projection

Crucial difference

The heuristic algorithm is capable of merging nearby candidates, whereas the ILP-based approach can only do so in specific cases.

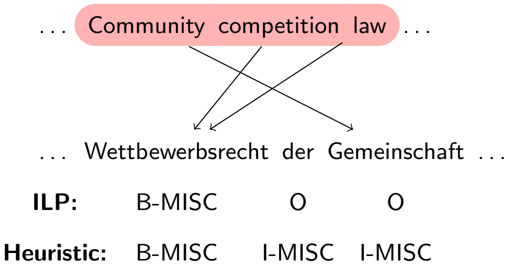


Figure 14: Example where ILP-based approach fails. Alignments are denoted with edges

Europarl-NER: Model transfer vs NER-based score

Source labeled sentence

Delays in the CEN 's work are now making it difficult to apply this very directive .

Model transfer

Aufgetretene Verzögerungen bei den Arbeiten des CEN führen nun zu Problemen bei der Anwendung eben dieser Richtlinie

ILP with NERScore

Aufgetretene Verzögerungen bei den Arbeiten des CEN führen nun zu Problemen bei der Anwendung eben dieser Richtlinie

Figure 15: Example of ignoring a incorrectly predicted entity

Source labeled sentence

Parliament rejected the request ...

Model transfer

Das Parlament lehnt den Antrag ...

ILP with NERScore

Das Parlament lehnt den Antrag ...

Figure 16: Example of restoring the target entity that was missed by the NER model

Europarl-NER: Comparison of matching scores

Source labeled sentence

It is irresponsible of EU Member States to refuse to renew the embargo

ILP with AlignScore

Die Ablehnung einer Verlängerung des Embargos seitens der EU-Mitgliedstaaten ist unverantwortlich

ILP with NERScore

Die Ablehnung einer Verlängerung des Embargos seitens der EU-Mitgliedstaaten ist unverantwortlich

Figure 17: Example where the alignment-based score outperforms NER model-based

Source labeled sentence

We shall now hear Mr Wurtz speaking against this request

ILP with AlignScore

Wir kommen nun zu Herrn Wurtz , der gegen den Antrag spricht

ILP with NERScore

Wir kommen nun zu Herrn Wurtz , der gegen den Antrag spricht

Figure 18: Example where the NER model-based score outperforms alignment-based

MasakhaNER2: Part 1

tgt_lang pipeline	bam	ewe	fon	hau	ibo	kin	lug	luo	mos	nya
Model transfer	0.295	0.776	0.542	0.721	0.618	0.673	0.755	0.542	0.522	**0.802**
Heuristic	0.492	0.756	0.624	0.709	0.714	0.658	0.790	**0.742**	0.428	0.761
align	0.471	0.733	0.616	0.695	**<u>0.736</u>**	0.652	**<u>0.808</u>**	<u>0.727</u>	0.436	0.757
ner	0.386	0.782	0.683	**<u>0.727</u>**	0.635	0.687	0.781	0.610	0.518	0.759
nmt	0.245	0.760	0.612	0.713	0.623	0.665	0.778	0.641	0.401	0.729
align_ner	0.477	**<u>0.791</u>**	0.651	0.719	0.668	**<u>0.689</u>**	0.806	0.723	0.502	<u>0.761</u>
align_ner_spans	**<u>0.499</u>**	0.784	0.652	0.716	0.641	0.681	0.800	0.712	**<u>0.523</u>**	0.754
align_nmt	0.310	0.771	0.644	0.718	0.657	0.674	0.788	0.672	0.446	0.739
ner_spans_nmt	0.319	0.780	0.668	0.713	0.660	0.676	0.797	0.686	0.504	0.749
all_fusion	0.370	0.786	**<u>0.684</u>**	0.712	0.697	0.685	0.798	0.693	0.508	0.755

Table 4: Overall F1 scores for XLNER pipelines with different projection steps on the MasakhaNER2 dataset

MasakhaNER2: Part 2

tgt_lang pipeline	sna	swa	tsn	twi	wol	xho	yor	zul	**avg**
Model transfer	0.365	**0.883**	0.646	0.482	0.442	0.244	0.394	0.437	0.563
Heuristic	0.678	0.792	**0.796**	**0.742**	0.583	0.526	0.489	0.641	0.662
align	0.673	0.785	<u>0.787</u>	0.706	0.574	0.517	0.506	0.638	0.656
ner	0.621	<u>0.826</u>	0.703	0.602	0.541	0.527	0.459	0.535	0.632
nmt	0.713	0.812	0.763	0.682	0.596	0.642	0.532	0.739	0.647
align_ner	0.704	0.817	0.778	0.683	0.617	0.585	0.536	0.652	0.675
align_ner_spans	0.698	0.812	0.763	0.667	0.638	0.585	0.519	0.648	0.672
align_nmt	0.722	0.811	0.772	0.692	0.631	0.646	0.565	0.741	0.667
ner_spans_nmt	0.722	0.817	0.771	0.707	0.651	<u>**0.654**</u>	0.552	0.740	0.676
all_fusion	<u>**0.745**</u>	0.817	0.776	<u>0.711</u>	<u>**0.661**</u>	0.654	<u>**0.566**</u>	<u>**0.745**</u>	<u>**0.687**</u>

Table 5: Overall F1 scores for XLNER pipelines with different projection steps on the MasakhaNER2 dataset

MasakhaNER2: Alternative methods

tgt_lang Method	bam	ewe	fon	hau	ibo	kin	lug	luo	mos	nya
EasyProject	0.458	0.785	0.614	0.722	0.656	0.710	0.767	0.502	0.531	0.753
CODEC	0.458	0.791	0.655	0.724	0.709	0.712	0.772	0.496	0.556	0.768
GPT-4	0.422	0.722	0.394	0.659	0.422	0.475	0.625	0.472	0.432	0.711
GoLLIE-TF	0.548	0.732	0.579	0.671	0.566	0.585	0.755	0.517	0.488	0.782
CLaP	-	-	-	0.699	0.605	-	-	-	-	0.587

tgt_lang Method	sna	swa	tsn	twi	wol	xho	yor	zul	**avg**
EasyProject	0.559	0.836	0.740	0.653	0.589	0.711	0.368	0.730	0.649
CODEC	0.724	0.831	0.747	0.646	0.631	0.704	0.414	0.748	0.671
GPT-4	0.395	0.792	0.563	0.442	0.526	0.498	0.547	0.369	0.526
GoLLIE-TF	0.574	0.735	0.710	0.742	0.619	0.499	0.544	0.528	0.621
CLaP	0.597	0.807	-	-	-	0.613	0.306	0.544	-

Table 6: Overall F1 scores for various XLNER methods evaluated under different settings compared to our experiments

Conclusions

- Formalization of the projection step has been proposed $\Longrightarrow$ better interpretability

Conclusions

- Formalization of the projection step has been proposed $\Longrightarrow$ better interpretability
- "Less or equal" type constraints on the number of projections from each source entity show the best results

Conclusions

- Formalization of the projection step has been proposed $\Longrightarrow$ better interpretability
- "Less or equal" type constraints on the number of projections from each source entity show the best results
- The proposed greedy algorithm performs on par or better than the exact solver $\Longrightarrow$ optimal solution to the problem is not always the best solution from the NER point of view

Conclusions

- Formalization of the projection step has been proposed $\Longrightarrow$ better interpretability
- "Less or equal" type constraints on the number of projections from each source entity show the best results
- The proposed greedy algorithm performs on par or better than the exact solver $\Longrightarrow$ optimal solution to the problem is not always the best solution from the NER point of view
- Projection strategies are encapsulated into matching scores $\Longrightarrow$ combination (fusion) is possible

Conclusions

- Formalization of the projection step has been proposed $\implies$ better interpretability
- "Less or equal" type constraints on the number of projections from each source entity show the best results
- The proposed greedy algorithm performs on par or better than the exact solver $\implies$ optimal solution to the problem is not always the best solution from the NER point of view
- Projection strategies are encapsulated into matching scores $\implies$ combination (fusion) is possible
- Experiments show that ILP-based projection with fused scores outperforms other methods

Conclusions

- Formalization of the projection step has been proposed $\implies$ better interpretability
- "Less or equal" type constraints on the number of projections from each source entity show the best results
- The proposed greedy algorithm performs on par or better than the exact solver $\implies$ optimal solution to the problem is not always the best solution from the NER point of view
- Projection strategies are encapsulated into matching scores $\implies$ combination (fusion) is possible
- Experiments show that ILP-based projection with fused scores outperforms other methods
- Complexity remains an open question

Directions for Further Work

- Complexity of the problem
- New matching scores, especially those that can take negative values
- New target candidate extraction methods
- Word-to-word alignment problem as an instance of the proposed ILP problem