LREC-COLING 2024

German SRL: Corpus Construction and Model Training

Maxim Konca, Andy Lücking, Alexander Mehler

Goethe University Frankfurt, Text Technology Lab

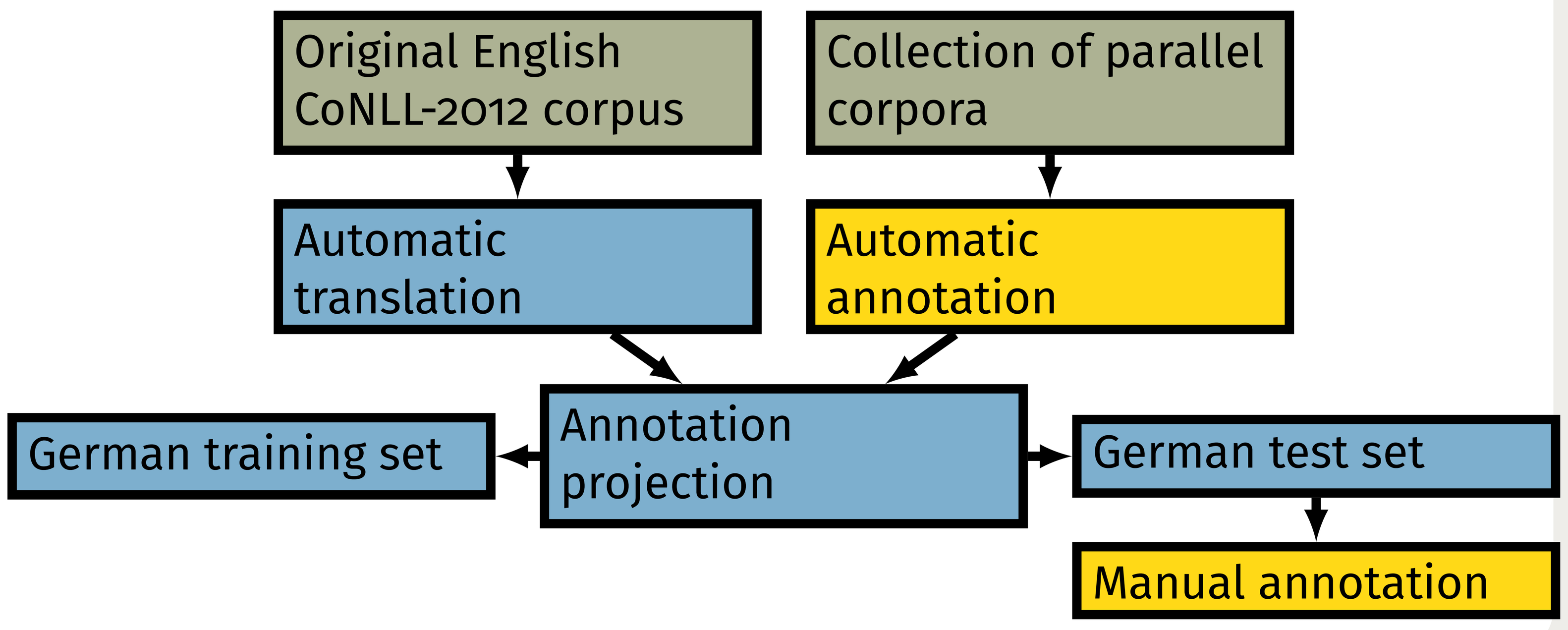
Abstract

We provide a semantic role-annotated resource for training semantic role models for the German language due to a combined translation and alignment process: The gold standard CoNLL-2012 semantic role annotations are translated into German. Semantic role labels are transferred due to alignment models. The resulting dataset is used to train a German semantic role model. With F1-scores around 0.7, the major roles achieve competitive evaluation scores, but avoid limitations of previous approaches.

Need for new resource

The most prominent corpus that is currently used for SRL in the German language, is CoNLL-2009 (Hajič et al., 2009). However, due to a couple of issues (e.g., incomplete mapping of SALSA to PropBank roles – see paper for details), a new resource for German SRL is needed.

Workflow



German SRL Model

We trained a semantic role labeling model for German utilizing the state-of-the-art crfsrl algorithm (Zhang et al., 2022) on the newly developed German SRL resource (training set).

Corpora

- The initial dataset is the English CoNLL-2012 corpus.
- In addition to the PropBank roles we use "PRED" to label the semantic-role licensing predicate).
- The original English CoNLL-2012 corpus was translated using the state-of-the-art machine translation system DeepL.
- Additionally, we used a variety of (pre-aligned) parallel corpora from the OPUS collection (Tiedemann, 2012).
- Automatic annotation of parallel corpora:
  1. Semantic roles with the model of Zhang et al. (2022) (F_1: 0.86).
  2. Token alignments provided by the parallel corpora were used to transfer the arguments to the corresponding German tokens.
- The merged dataset has been split into the German training set and the German test set.

Annotation projection

- Both the automatically translated English CoNLL-2012 corpus and the automatically annotated parallel corpora were then merged through an annotation projection process.
- This step ensured that the German translations inherited the annotations from their corresponding English sentences.
- To align English and German tokens, we used SimAlign (Jalili Sabet et al., 2020) (F_1: 0.81).

Manual annotation

- Two expert annotators checked and corrected the test set annotations.
- This step provided a gold standard against which the performance of models could be benchmarked.
- Manual annotation has been carried out by making use of the PropAnnotator from the TextAnnotator collection of annotation tools (Abrami, Stoeckel, and Mehler, 2020).
- The annotator agreement measured as Krippendorff's α reached respectable 0.786.

Prediction results

development set

Argument	Precision	Recall	F1-Score	Support
ARG0	0.79	0.76	0.77	630
ARG1	0.74	0.67	0.70	1,050
ARG2	0.61	0.52	0.56	353
ARG3	0.62	0.22	0.33	36
ARG4	0.71	0.56	0.62	52
ARGM-ADV	0.53	0.28	0.36	181
ARGM-CAU	0.46	0.50	0.48	34
ARGM-COM	0.44	0.57	0.50	7
ARGM-DIR	0.46	0.38	0.41	16
ARGM-DIS	0.71	0.73	0.72	122
ARGM-EXT	0.50	0.17	0.25	12
ARGM-GOL	0.75	0.17	0.27	18
ARGM-LOC	0.53	0.56	0.54	102
ARGM-MNR	0.69	0.60	0.64	101
ARGM-MOD	0.89	0.80	0.84	142
ARGM-NEG	0.89	0.82	0.85	92
ARGM-PRD	0.60	0.23	0.33	13
ARGM-PRP	0.54	0.54	0.54	48
ARGM-REC	0.87	0.83	0.85	54
ARGM-TMP	0.65	0.77	0.70	220
C-ARG0	1.00	0.23	0.38	13
C-ARG1	0.33	0.19	0.24	79
C-ARG2	0.40	0.16	0.23	38
C-ARG4	1.00	0.14	0.25	7
R-ARG0	0.84	0.80	0.82	46
R-ARG1	0.85	0.55	0.67	51
R-ARGM-LOC	0.42	0.80	0.55	10
PRED	1.00	0.88	0.94	1,199
VG	0.91	0.93	0.92	68
micro	0.78	0.69	0.73	4,830
macro	0.45	0.35	0.37	4,830
weighted	0.77	0.69	0.73	4,830

Prediction results

test set

Argument	Precision	Recall	F1-Score	Support
ARG0	0.84	0.68	0.75	303
ARG1	0.71	0.68	0.70	600
ARG2	0.54	0.42	0.48	271
ARG3	0.00	0.00	0.00	30
ARG4	0.20	0.25	0.22	8
ARGM-ADV	0.30	0.17	0.21	206
ARGM-CAU	0.66	0.31	0.42	68
ARGM-COM	0.00	0.00	0.00	15
ARGM-CXN	0.00	0.00	0.00	22
ARGM-DIS	0.44	0.09	0.15	332
ARGM-EXT	0.50	0.11	0.18	27
ARGM-GOL	0.00	0.00	0.00	8
ARGM-LOC	0.18	0.18	0.18	136
ARGM-LVB	0.00	0.00	0.00	11
ARGM-MNR	0.35	0.28	0.31	79
ARGM-MOD	0.83	0.47	0.60	104
ARGM-NEG	0.69	0.84	0.76	37
ARGM-PRD	0.00	0.00	0.00	14
ARGM-PRP	0.47	0.33	0.39	21
ARGM-REC	0.00	0.00	0.00	47
ARGM-TMP	0.18	0.33	0.23	88
PRED	0.99	0.88	0.93	680
VG	0.84	0.83	0.84	71
micro	0.64	0.51	0.57	3,184
macro	0.26	0.20	0.22	3,184
weighted	0.63	0.51	0.55	3,184

Model ~ Annot. 1

Argument	Precision	Recall	F1-Score	Support
ARG0	0.88	0.65	0.75	588
ARG1	0.79	0.59	0.67	976
ARG2	0.72	0.56	0.63	326
ARG3	0.75	0.60	0.67	35
ARG4	0.83	0.60	0.70	50
ARGM-ADV	0.67	0.48	0.56	159
ARGM-CAU	0.83	0.65	0.73	31
ARGM-COM	1.00	0.71	0.83	7
ARGM-DIR	0.45	0.60	0.51	15
ARGM-DIS	0.76	0.65	0.70	110
ARGM-EXT	0.50	0.30	0.37	10
ARGM-GOL	0.71	0.83	0.77	18
ARGM-LOC	0.73	0.63	0.68	90
ARGM-MNR	0.68	0.62	0.65	95
ARGM-MOD	0.88	0.58	0.70	128
ARGM-NEG	0.77	0.55	0.64	75
ARGM-PRD	0.33	0.46	0.39	13
ARGM-PRP	0.52	0.51	0.52	47
ARGM-REC	0.96	0.46	0.62	50
ARGM-TMP	0.76	0.72	0.74	198
C-ARG0	0.50	0.62	0.55	13
C-ARG1	0.45	0.62	0.52	79
C-ARG2	0.31	0.54	0.40	35
C-ARG4	0.80	0.57	0.67	7
C-ARGM-MOD	1.00	0.50	0.67	2
C-ARGM-PRP	0.00	0.00	0.00	4
R-ARG0	0.83	0.42	0.56	45
R-ARG1	0.88	0.31	0.45	49
R-ARGM-LOC	0.57	0.50	0.53	8
PRED	1.00	0.76	0.86	1,111
VG	0.98	0.81	0.88	62
micro	0.81	0.64	0.71	4,462
macro	0.52	0.44	0.45	4,462
weighted	0.83	0.64	0.71	4,462

Model ~ Annot. 2

Argument	Precision	Recall	F1-Score	Support
ARG0	0.92	0.65	0.76	613
ARG1	0.80	0.62	0.70	952
ARG2	0.68	0.57	0.62	339
ARG3	0.59	0.73	0.66	30
ARG4	0.86	0.54	0.67	57
ARGM-ADV	0.61	0.65	0.62	93
ARGM-CAU	0.81	0.50	0.62	34
ARGM-COM	0.80	0.89	0.84	9
ARGM-DIR	0.39	1.00	0.56	7
ARGM-DIS	0.75	0.80	0.78	115
ARGM-EXT	0.86	0.86	0.86	7
ARGM-GOL	0.47	1.00	0.64	9
ARGM-LOC	0.73	0.66	0.69	91
ARGM-MNR	0.74	0.62	0.67	78
ARGM-MOD	0.85	0.60	0.70	121
ARGM-NEG	0.81	0.63	0.71	70
ARGM-PNC	0.57	1.00	0.73	8
ARGM-PRD	0.60	0.90	0.72	10
ARGM-PRP	0.36	0.41	0.38	32
ARGM-REC	0.97	0.58	0.72	52
ARGM-TMP	0.79	0.71	0.75	199
C-ARG0	0.42	0.62	0.50	13
C-ARG1	0.39	0.64	0.48	76
C-ARG2	0.39	0.61	0.48	46
C-ARG4	0.40	0.40	0.40	5
R-ARG0	0.88	0.56	0.69	39
R-ARG1	0.76	0.64	0.70	45
PRED	1.00	0.80	0.89	1,097
VG	0.98	0.78	0.87	51
micro	0.81	0.67	0.74	4,322
macro	0.52	0.53	0.50	4,322
weighted	0.83	0.67	0.74	4,322

Summary

- Core argument roles such as ARG0 achieve an F1-score of 0.75 (support: 303)
- ARG1 exhibits an F1-score of 0.70 (support: 600)
- ARG2 reaches an F1-score of 0.51 (support: 201)
- The performance varies across the modifier roles.
- The identification of PRED reaches an F1-score of 0.93 (support: 680)
- VG abbreviates "verb group", i.e. components of discontinuous verb phrases
- Core roles such as ARG0, ARG1 and PRED show commendable performance

Conclusions

- For comparisons with a model trained on CoNLL-2009 and with X-SRL (Daza and Frank, 2020) see the paper
- In particular modifiers seem to be difficult.
- The German SRL resource will be published via LDC.

References

Abrami, Giuseppe, Manuel Stoeckel, and Alexander Mehler (May 2020). "TextAnnotator: A UIMA Based Tool for the Simultaneous and Collaborative Annotation of Texts". In: *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 891–900. URL: https://www.aclweb.org/anthology/2020.lrec-1.112.

Daza, Angel and Anette Frank (Nov. 2020). "X-SRL: A Parallel Cross-Lingual Semantic Role Labeling Dataset". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, pp. 3904–3914. DOI: 10.18653/v1/2020.emnlp-main.321. URL: https://aclanthology.org/2020.emnlp-main.321.

Hajič, Jan et al. (2009). "The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages". In: *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*. CoNLL '09. Boulder, Colorado: Association for Computational Linguistics, pp. 1–18.

Jalili Sabet, Masoud et al. (Nov. 2020). "SimAlign: High Quality Word Alignments without Parallel Training Data using Static and Contextualized Embeddings". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. Online: Association for Computational Linguistics, pp. 1627–1643. URL: https://www.aclweb.org/anthology/2020.findings-emnlp.147.

Tiedemann, Jörg (May 2012). "Parallel Data, Tools and Interfaces in OPUS". In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Istanbul, Turkey: European Language Resources Association (ELRA).

Zhang, Yu et al. (2022). "Semantic Role Labeling as Dependency Parsing: Exploring Latent Tree Structures inside Arguments". In: *Proceedings of COLING*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 4212–4227. URL: https://aclanthology.org/2022.coling-1.370.

LaTeX TikZposter