Universität Zürich UZH

Masterarbeit
zur Erlangung des akademischen Grades
Master of Arts
der Philosophischen Fakultät der Universität Zürich

(Titel)

Verfasserin/Verfasser: Jonathan Schaber
Matrikel-Nr: 11-771-359

Referentin/Referent: Dr. Simon Clematide

[Betreuerin/Betreuer: (Titel Vorname Name) [nur falls vom Ref. unterschiedlich]]

Institut für Computerlinguistik

Abgabedatum: (xx.xx.xxxx)

Abstract

This is the place to put the English version of the abstract.

Zusammenfassung

Und hier sollte die Zusammenfassung auf Deutsch erscheinen.

Acknowledgement

I want to thank X, Y and Z for their precious help. And many thanks to whoever for proofreading the present text.

Contents

List of Figures

List of Tables

List of Acronyms

USD	Universal Stanford Dependencies
BERT	Bidirectional Encoder Representations from Transformers
CPOSTAG	Coarse-grained Part-Of-Speech tag
LSTM	Long Short-Term Memory
ML	Machine Learning
NLP	Natural Language Processing
POS	Part-Of-Speech
POSTAG	Fine-grained Part-Of-Speech tag
RNN	Recurrent Neural Network
SRL	Semantic Role Labelling OR Semantic Role Labeller
STTS	Stuttgart-Tübingen-TagSet

1 Introduction

1.1 Motivation

Some words on your motivation would be nice.

1.2 Research Questions

The research questions that shall be answered in this thesis, are:

1. What do I do?
2. How do I do it?
3. And why?

1.3 Thesis Structure

In this first chapter ...
Chapter 2 introduces ...
Chapter 3 ...

2 Semantic Roles

2.1 Overview

"The main reason computational systems use semantic roles is to act as a shallow meaning representation that can let us make simple inferences that aren't possible from the pure surface string of words, or even from the parse tree." [Jurafsky and Martin, 2019, p. 375]

In the literature, often Gildea and Jurafsky [2002] is considered to have formally defined the task of automatic SRL.

3 Data Sets

3.1 Why create an own corpus?

3.2 Corpora

3.2.1 delSEAR

As Troiano et al. [2019] write in their

3.2.2 MLQA_V1

Lewis et al. [2019] compiled

3.2.3 PAWS-X

Yang et al. [2019]

3.2.4 SCARE

3.2.4.1 SCARE normal

The Sentiment Corpus of App Reviews with Fine-grained Annotations in German Sänger et al. [2016] is a hand-annotated corpus that asserts so sentiment to German mobile app reviews stemming from the Google Play Store. Since there are many users of In contrast to other data sets, e.g. [Socher et al., 2013; Go et al., 2009], that attributes one sentiment label to a whole text (may it be a review, a tweet, etc.), Sänger et al. [2016] annotated their data set on a lower textual level: Not each review gets labelled for a certain polarity — i.e. *positive*, *negative*, or *neutral* — but what the authors call *aspects* and *subjective (sub-)phrases*. An aspect is "part of an app or related to it", while a subjective (sub-)phrase "express opinions and statements of a personal evaluation regarding the app or a part of it, that are not based on (objective) facts but on individual opinions of the reviewers" [Sänger et al., 2016, p. 1116]. The authors therefore draw a distinction between objective facts regarding an app or parts of it and the sentiment connected to it ("functionality X is not working" $\rightarrow$ negative), and subjective user meanings concerning an app or parts of it ("I really like the color of X" $\rightarrow$ positive). This fine level of annotations leads often to several annotations per review, the sentiment of which may not always match. As illustration, consider the following review:

(3.1) guter wecker... || vom prinzip her echt gut...aber grade was die sprachausgabe betrifft noch etwas buggy....[0]

There are the following annotations for the several aspects and subjective (sub-)phrases are present in this example:

Aspects

- Wecker $\rightarrow$ neutral
- Prinzip $\rightarrow$ neutral
- Sprachausgabe $\rightarrow$ neutral

Subjectives

- guter $\rightarrow$ positive
- echt gut $\rightarrow$ positive
- etwas buggy $\rightarrow$ negative

3.2.4.2 SCARE reviews

Besides their carefully, hand-annotated corpus, the authours also provide a dataset comprising of XXX reviews along with the rating — one to five stars —, that were available in German on the Google App XXXXXXXXXXX.

3.2.4.3 Preprocessing

For integrating the SCARE corpus into my GerBLUE corpus, I need to prepare the data, so it can be handled by the model architecture. Following the original GLUE sentiment task, the model needs only to predict one sentiment label for each example. Since there exist mostly multiple annotations for each review in this data set, the data needs to be pre-processed in a way, so that there is one review-label per example.

[0]The "||" denotes that the text left of it is the user given "title" of the review, and the part on the right is the actual review.

To generate the review-label, I simply carry out an majority class decision: The label that is most often annotated for a given review, regardless if it is an aspect or a subjective, is then also the review-label. If there is no majority label, the review-label is set to “neutral”. This is also the chosen strategy for 51 reviews that had no labels at all (e.g. “Ich bin die erfuinderin — — Ich bin die erfunden!!!!!!!!!!!!!!!!!!!!!!!!!!”).

3.2.5 XNLI

Conneau et al. [2018]

3.2.6 XQuAD

Artetxe et al. [2019]

4 Architecture

4.1 Overview

4.2 Semantic Role Labeller

A Semantic Role Labeller (SRL) is a system, that assigns automatically semantic roles to a given input text.[0]

State-of-the-art semantic role labellers (SRLs) are end-to-end models, nowadays often implementing deep learning techniques, like RNNs or attention, that render tedious feature engineering unnecessary. For my system, I implement the DAMESRL, a model presented by Do et al. [2018]. I use their pre-trained German Character-Attention model which, according to the authors, achieved an F1 score of 73.5% on the CoNLL'09 task [Hajič et al., 2009]. However, their SRL needs as input not only the sentence, but also "its predicate w_p as input" [Do et al., 2018].

"A major advantage of dependency grammars is their ability to deal with languages that are morphologically rich and have a relatively free word order." [Jurafsky and Martin, 2019, p. 274] For extracting predicates, I rely on the dependency tree the ParZu parser Sennrich et al. [2013] generates for a given sentence. Since one sentence can have multiple predicate-argument structures, I need to device an algorithm to extract the relevant predicates in a sentence. This is not as straight forward as it seems on the first look.

4.2.1 Finding Predicates

It is a known problem in the analysis of semantic roles that a proper procedure for predicate identification is a hard to tackle problem, consider e.g. the discussion concerning so called light verbs: Wittenberg [2016].

[0]This may be one or multiple sentences.

"First, the predicates which assign semantic roles to the constituents are identified prior to semantic role labelling proper. They are usually identified as the main verbs which head clauses." [Samardzic, 2013, p. 74] In a dependency framework like USD [De Marneffe et al., 2014], which explicitly sets the content verb as root, identification of the relevant predicate is straight-forward: One has simply to look at the dependency parse tree of a given sentence and select the heads — i.e. roots — of the clauses. However, the ParZu parser models not content words as heads but function words.[0]

(interestingly, this stands in contrast to the Pro3Gres parser [Schneider, 2008] which

"In a constituency parse, the finite verb is the head of a verb phrase or rather sentence. A dependency parse, on the other hand, does not consider auxiliaries as heads and therefore finite verbs are usually not the head of the sentence. Hence, the head of a sentence typically is the verb containing the meaning. In that sense, dependency structures are closer to the semantics of a sentence." [Aepli, 2018, p. 6f.]

According to the USD, function words are subordinated to content words, which means that in a sentence "He was hit by a ball.", *hit* would be analysed as root, not the finitely inflected *was*. This is an accordance with the view that XXXXXXXXXX However, there is a "substantial amount of evidence [that] delivers a strong argument for the [...] approach, which subordinates full verbs to auxiliaries" Groß and Osborne [2015].

"The parsing scheme that USD advocates takes the division between function word and content word as its guiding principle. One major difficulty with doing this is that the dividing line between function word and content word is often not clear." Groß and Osborne [2015]

Following Foth [2006]

(4.1) Die Keita-Dynastie regierte das vorkaiserliche und kaiserliche Mali vom 12. Jahrhundert bis Anfang des 17. Jahrhunderts.

(4.2) Im tibetischen Buddhismus werden die Dharma-Lehrer/innen gewöhnlich als Lama bezeichnet.

(4.3) Die Klage wurde abgewiesen, was als Sieg beschrieben werden kann.

whose dependency parse tree is shown in Figure 1: This sentence has five verbs in it, *wurde*, *abgewiesen*, *beschrieben*, *werden*, and *kann* (POS-tag "V" in the second

[0]This follows general dependency frameworks proposed for German, e.g. Gerdes and Kahane [2001]; Groß and Osborne [2015].

row), but only two of them are relevant predicates, i.e. predicates that carry “true” semantics.

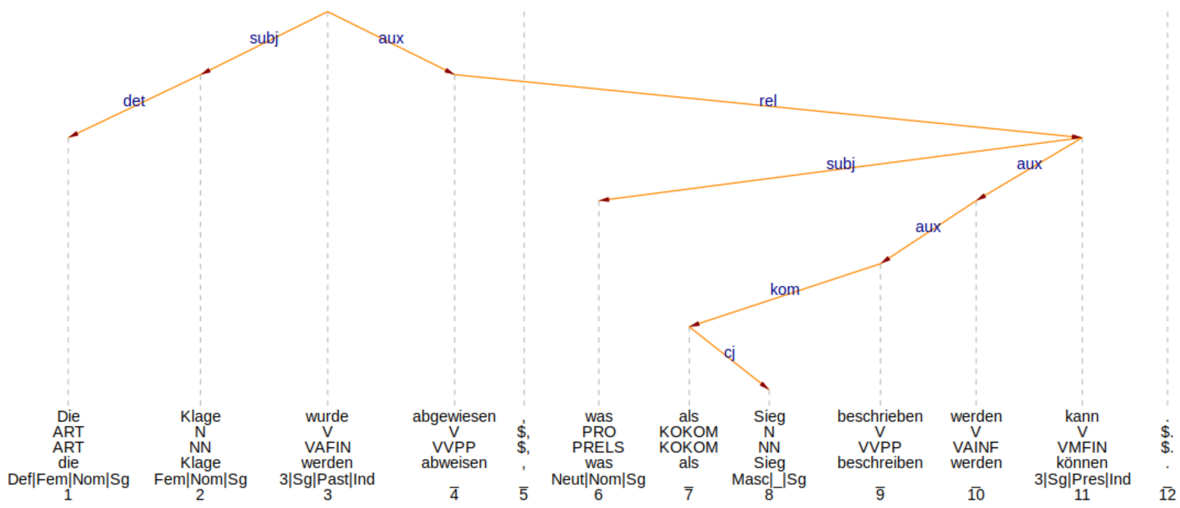

Figure 1: Example dependency parse tree for a sentence with multiple predicates.

I propose the following algorithm 1 deciding whether a verb in a sentence is or isn’t a predicate using a heuristic, relying on the token’s POS tag that the parser predicts. The ParZu parser’s default output follows the CoNLL scheme [Buchholz and Marsi, 2006] which means that there are two levels of POS tagging: coarse-grained (CPOSTAG) and fine-grained (POSTAG), where the POSTAG corresponds to the token’s STTS tag [Schiller et al., 1999].

The condition on line 9, that only tokens in the respective subclause are considered, is ensured by making sure that if a token u’s POS is “V” and it points to its head t, that it is not itself the head of a subclause — i.e. its dependency relation is e.g. “relative clause”. If that is the case the token u is considered to belong to another subclause and therefore not preventing token t from getting labelled as a predicate. Consider again the example 4.2.1: Let’s say we are in the for-loop at the token *weitergeleitet*. Because it is a verb but not a finite full-verb, we enter the else-clause on line 7. If we were now to loop through all token of sentence 4.2.1 we would find that token *führt* is a verb that points to our primary token. Without the above outlined constraint that only verbs in the same subclause pointing to our original verb are preventing it from being labelled a predicate, *weitergeleitet* would be labelled as non-predicate. This is obviously false. Taking into account the above considerations, we see that although *führt* points to *weitergeleitet*, its edge label is *rel* — which means that it’s the head of a relative subclause — therefore it is not anymore in the same subclause and *weitergeleitet* gets labelled as predicate.

Algorithm 1 Predicate finding algorithm

```
for all token t ∈ sentence do
    if CPOSTAG t ≠ 'V' then
        t ← NOT_PRED
    else
        if POSTAG t = 'VVFIN' then
            t ← PRED
        else
            FLAG ← True
            for all token u ≠ t ∈ subclause where t ∈ subclause do
                if CPOSTAG u = 'V' ∧ u dependent on t then
                    t ← NOT_PRED
                    FLAG ← False
                    break
                end if
            end for
            if FLAG = True then
                t ← PRED
            end if
        end if
    end if
end for
```

4.2.2 DAMESRL

4.3 German BERT

Since its publishing two years ago, BERT [Devlin et al., 2018] has often been called a “turning-point” in ML in NLP.

I use the `bert-base-german-cased` model from deepset which is available in pyTorch through the hugging face library[0].

[0] `https://huggingface.co/bert-base-german-cased`, accessed: 22.07.2020.

5 Results

5.1 BLEU Scores

Table 1 shows how to use the predefined tab command to have it listed.

language pair	ABC	YYY
EN→DE	20.56	32.53
DE→EN	43.35	52.53

Table 1: BLEU scores of different MT systems

And we can reference the large table in the appendix as Table 2

5.2 Evaluation

We saw in section 5.1

We will see in subsection 5.2.1 some more evaluations.

5.2.1 More evaluation

5.3 Citations

Although BLEU scores should be taken with caution (see ?) or if you prefer to cite like this: [?] ...

to cite: [?, 30-31]
to cite within parentheses/brackets: [?], [?, 30-32]

to cite within the text: **?**, **?**, 37
only the author(s): **?**
only the year: **?**

5.4 Graphics

To include a graphic that appears in the list of figures, use the predefined fig command:

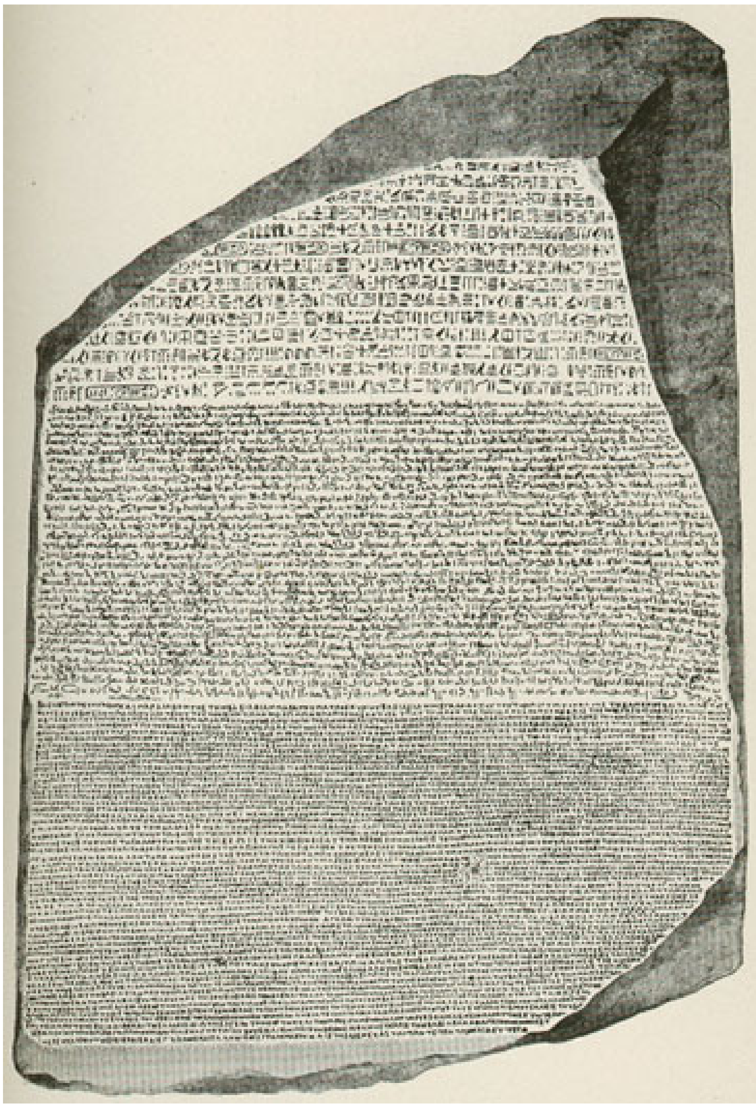

Figure 2: The Rosetta Stone

And then reference it as Figure 2 is easy.

5.5 Some Linguistics

(With the package 'covington')

Gloss:

(5.1)	*The*	*cat*	*sits*	*on*	*the*	*table.*
	die	Katze	sitzt	auf	dem	Tisch

'Die Katze sitzt auf dem Tisch.'

Gloss with morphology:

(5.2)	*La*	*gata*	*duerm*	*-e*	*en*	*la*	*cama.*
	Art.Fem.Sg	Katze	schlaf	-3.Sg	in	Art.Fem.Sg	Bett

'Die Katze schläft im Bett.'

6 Conclusion

In this project we have done so much.[1]

We could show that ...

Future research is needed.
The show must go on.

[1] Thanks to many people that helped me.

Glossary

Of course there are plenty of glossaries out there! One (not too serious) example is the online MT glossary of Kevin Knight [2] in which MT itself is defined as

> techniques for allowing construction workers and architects from all over the world to communicate better with each other so they can get back to work on that really tall tower.

accuracy A basic score for evaluating automatic **annotation tools** such as **parsers** or **part-of-speech taggers**. It is equal to the number of **tokens** correctly tagged, divided by the total number of tokens. [...]. (See **precision and recall**.)

clitic A morpheme that has the syntactic characteristics of a word, but is phonologically and lexically bound to another word, for example *n't* in the word *hasn't*. Possessive forms can also be clitics, e.g. The dog*'s* dinner. When **part-of-speech tagging** is carried out on a corpus, clitics are often separated from the word they are joined to.

[2] Machine Translation Glossary (Kevin Knight): `http://www.isi.edu/natural-language/people/dvl.html`

References

N. Aepli. *Parsing Approaches for Swiss German*. PhD thesis, University of Zurich, 2018.

M. Artetxe, S. Ruder, and D. Yogatama. On the cross-lingual transferability of monolingual representations. *arXiv preprint arXiv:1910.11856*, 2019.

S. Buchholz and E. Marsi. Conll-x shared task on multilingual dependency parsing. In *Proceedings of the tenth conference on computational natural language learning (CoNLL-X)*, pages 149–164, 2006.

A. Conneau, G. Lample, R. Rinott, A. Williams, S. R. Bowman, H. Schwenk, and V. Stoyanov. Xnli: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053*, 2018.

M.-C. De Marneffe, T. Dozat, N. Silveira, K. Haverinen, F. Ginter, J. Nivre, and C. D. Manning. Universal stanford dependencies: A cross-linguistic typology. In *LREC*, volume 14, pages 4585–4592, 2014.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Q. N. T. Do, A. Leeuwenberg, G. Heyman, and M. F. Moens. A flexible and easy-to-use semantic role labeling framework for different languages. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 161–165, 2018.

K. A. Foth. Eine umfassende constraint-dependenz-grammatik des deutschen. 2006.

K. Gerdes and S. Kahane. Word order in german: A formal dependency grammar using a topological hierarchy. In *Proceedings of the 39th annual meeting of the Association for Computational Linguistics*, pages 220–227, 2001.

D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288, 2002.

A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009, 2009.

T. Groß and T. Osborne. The dependency status of function words: Auxiliaries. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 111–120, 2015.

J. Hajič, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štepánek, et al. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. 2009.

D. Jurafsky and J. H. Martin. Speech and language processing (draft). october 2019. *URL https://web. stanford. edu/˜ jurafsky/slp3*, 2019.

P. Lewis, B. Oğuz, R. Rinott, S. Riedel, and H. Schwenk. Mlqa: Evaluating cross-lingual extractive question answering. *arXiv preprint arXiv:1910.07475*, 2019.

T. Samardzic. *Dynamics, causation, duration in the predicate-argument structure of verbs: a computational approach based on parallel corpora*. PhD thesis, University of Geneva, 2013.

M. Sänger, U. Leser, S. Kemmerer, P. Adolphs, and R. Klinger. Scare—the sentiment corpus of app reviews with fine-grained annotations in german. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1114–1121, 2016.

A. Schiller, S. Teufel, C. Stöckert, and C. Thielen. Guidelines für das tagging deutscher textcorpora. *University of Stuttgart/University of Tübingen*, 1999.

G. Schneider. *Hybrid long-distance functional dependency parsing*. PhD thesis, University of Zurich, 2008.

R. Sennrich, M. Volk, and G. Schneider. Exploiting synergies between open resources for german dependency parsing, pos-tagging, and morphological analysis. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 601–609, 2013.

R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

E. Troiano, S. Padó, and R. Klinger. Crowdsourcing and validating event-focused emotion corpora for german and english. *arXiv preprint arXiv:1905.13618*, 2019.

E. Wittenberg. *With light verb constructions from syntax to concepts*, volume 7. Universitätsverlag Potsdam, 2016.

Y. Yang, Y. Zhang, C. Tar, and J. Baldridge. Paws-x: A cross-lingual adversarial dataset for paraphrase identification. *arXiv preprint arXiv:1908.11828*, 2019.

Lebenslauf

Persönliche Angaben

Ich Persönlich
Meinestrasse Nr
PLZ Wohnort
ichpersoenlich@uzh.ch

Schulbildung

2012-2014	Bachelor-Studium Computerlinguistik und Sprachtechnologie an der Universität Zürich
seit 2014	Master

Berufliche und nebenberufliche Tätigkeiten

2012–2013	Tutorate PCL I+II

A Tables

Part of speech	POS type	POS	number of labels in my corpus
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	DET	**35**	280
14	Total	**35**	280

Table 2: Some very large table in the appendix

B List of something

This appendix contains a list of things I used for my work.

- apples
  - `export2someformat`
- bananas
- oranges
  - `bleu4orange`
  - `rouge2orange`