The Impact of Attention Mechanism, Context and Genre Information when Classifying Semantic Clause Types with Recurrent Neural Networks

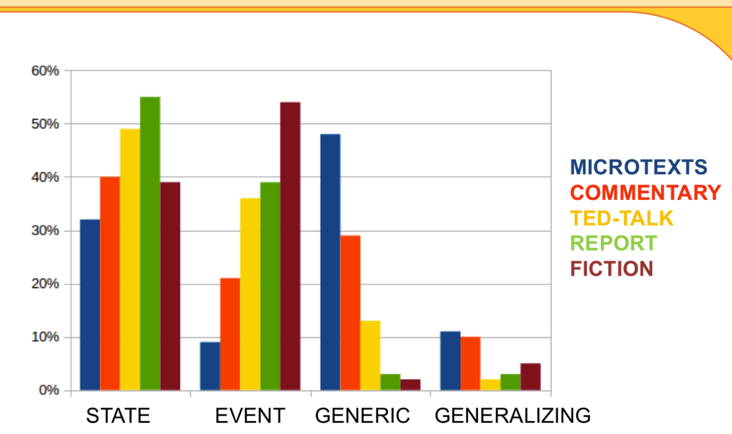
Maria Becker, Michael Staniek, Vivi Nastase, Alexis Palmer, Anette Frank

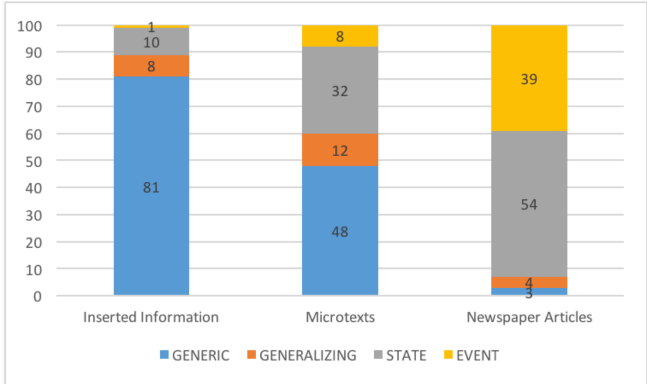
LIMO Empirical Linguistics and Computational Language Modeling

Motivation: Why are Semantic Clause Types Interesting?

The distribution of SCT in text passages correlates with **discourse modes** (Smith 2003) and plays a role in

- **Genre** characterization (Palmer and Friedrich, 2014)
- Detection of **generic and generalizing sentences** (Friedrich and Pinkal, 2015)
- **Argumentation** structure analysis (Becker et al., 2016)
- Characterization of **implicit knowledge** (Becker et al., 2017)





Related Work and Contribution

Automatic Classification of Semantic Clause Types

Prior work. **Feature-based classifiers** (Palmer et al. 2007, Friedrich et al. 2016)

- exploit language-specific and resource-intensive **features**
- **results**: with standard NLP Features – 69.8 accuracy
  with detailed features including external repositories – 71.4 accuracy
  with standard and detailed features used jointly – 74.7 accuracy

→ **Adaptation** to novel languages is **expensive**

Our Aim: Resource-lean **Recurrent Neural Network** model with **attention,** enhanced with **context** and **genre** information which is

- capable of modeling sequences
- capable of focusing on parts of the input
- easy to port to novel languages
- able to exploit context & genre

Data

- **English** Dataset: Friedrich et al. (2016): Wikipedia (10,607 clauses) and MASC (30,333 clauses), 13 genres *(Email, Essay, Letter, Newspaper, TED talk, Wikipedia...)*
- **German** Datasets: Mavridou et al. (2015) and Becker et al. (2016a,b) + self-annotated data (total: 18,194 clauses), 7 genres *(Fiction, Commentary, report...)*
- **Word embeddings**
  - English: 300-dim word2vec, trained on Google News (Mikolov et al. 2013)
  - German: 100-dim word2vec, trained on a web corpus (Reimers et al, 2014)

German annotated dataset: www.cl.uni-heidelberg.de/english/research/downloads/resource_pages/GER_SET/GER_SET_data.shtml

Results

		English Testset		German Testset	
		Accuracy	F1-Score	Accuracy	F1-Score
Local Models	Local model (w/o attention)	66.55	59.14	74.94	67.12
	Local model with attention	69.18	68.31	74.51	74.02
	Local model with attention+genre	**71.12**	69.55	**75.56**	69.98
Context Model: Clauses	1 previous clause/genre	**71.67**	59.19	**74.51**	72.41
Local model with attention	2 previous clauses/genres	71.57	48.12	74.44	72.26
+ previous clauses (tokens, w/o attention)	3 previous clauses/genres	69.76	42.73	73.35	71.79
+ genre label	4 previous clauses/genres	69.29	41.55	73.11	71.12
	5 previous clauses/genres	68.99	30.78	72.89	70.61
Context Model: Labels	1 previous label/genre	69.55	60.21	71.78	52.88
Local model with attention	2 previous labels/genres	71.04	64.54	72.29	52.52
+ previous labels with attention	3 previous labels/genres	71.68	64.42	72.47	52.34
+ genre label	4 previous labels/genres	71.25	65.06	74.33	51.12
	5 previous labels/genres	**72.04**	64.74	**74.92**	50.76
Context Model: Labels + Clauses	1 previous label/clause/genre	**71.35**	70.82	**73.43**	59.51
Local model (w/o attention)	2 previous labels/clauses/genres	70.65	68.62	72.23	57.38
+ previous clauses (tokens, w/o attention)	3 previous labels/clauses/genres	69.90	68.83	71.69	57.99
+ previous labels (w/o attention)	4 previous labels/clauses/genres	69.26	67.47	71.11	56.48
+ genre label of previous labels	5 previous labels/clauses/genres	69.00	64.36	71.09	56.23

Conclusions

- Models that attend to **local clauses, context & genre jointly** perform best
- **Competitive performance** at the level of feature-based classifiers
- Model avoids reproducing linguistic features for **novel languages**

Leibniz ScienceCampus LiMo is funded by

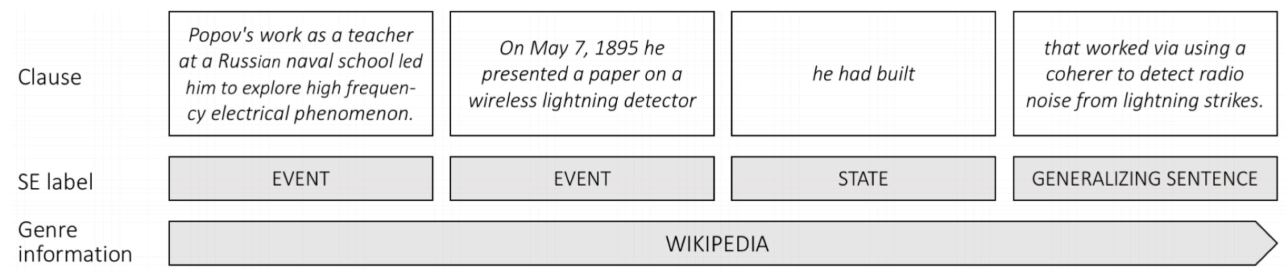
Inventory

Semantic Clause Types (SCT) (Smith 2003, Friedrich et al. 2016) characterize the **aspectual properties** of clauses and their **function** within a **text/discourse**:

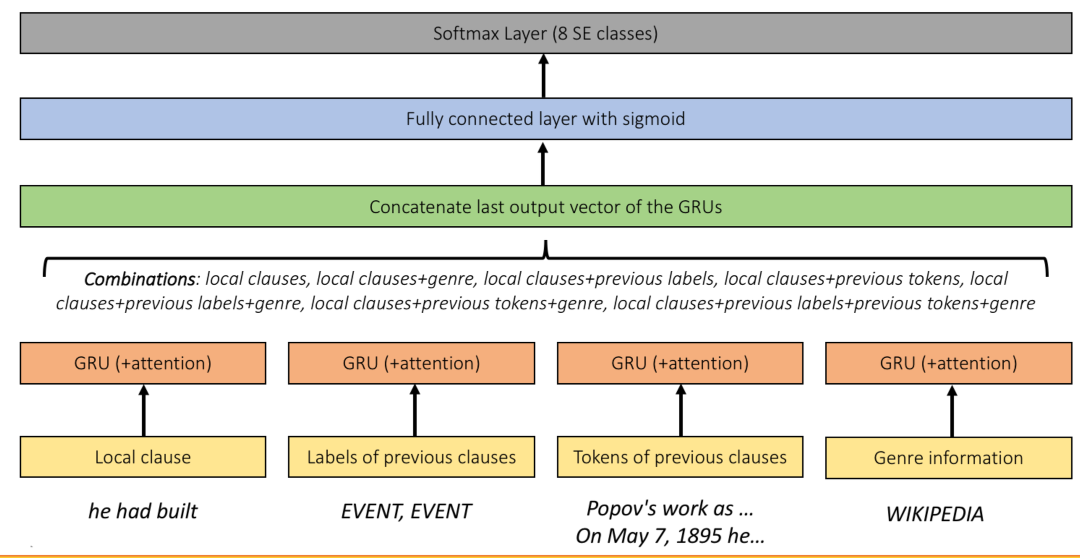
- STATES: John loves cake.
- EVENTS: Mike won the race.
- GENERALIZING SENTENCES: Mary often feeds my cat.
- GENERIC SENTENCES: Lions are carnivores.
- REPORT: John says that he loves cake.
- QUESTION: Why do you torment me so?
- IMPERATIVE: Listen to this.

Model

Modeling Context and Genre Information

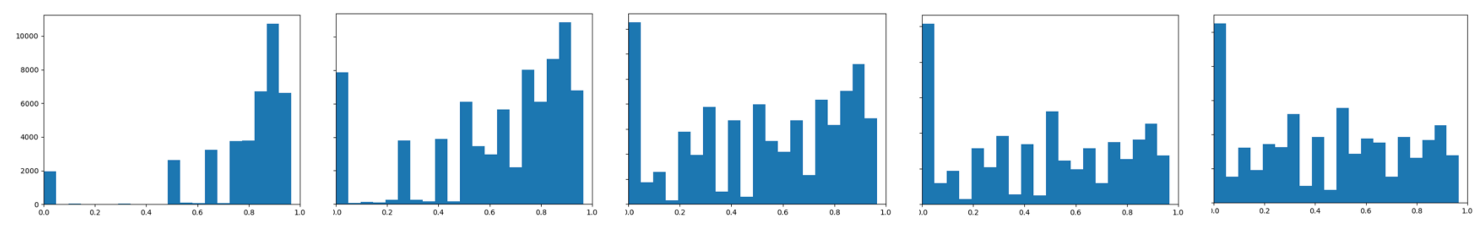


Model Architecture



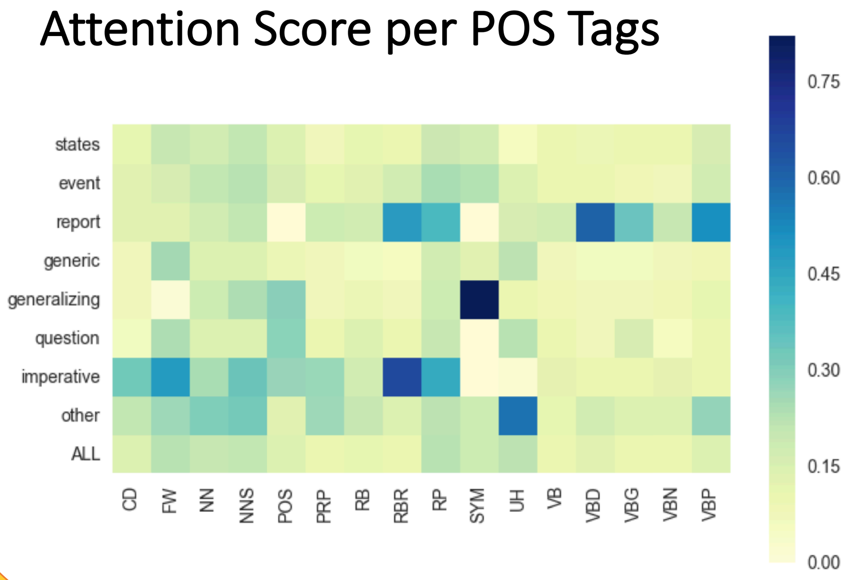
Analysis

Position of Words with High Attention Scores



1 word with highest attention | 2 words with highest attention | 3 words with highest attention | 4 words with highest attention | 5 words with highest attention

Attention Score per POS Tags

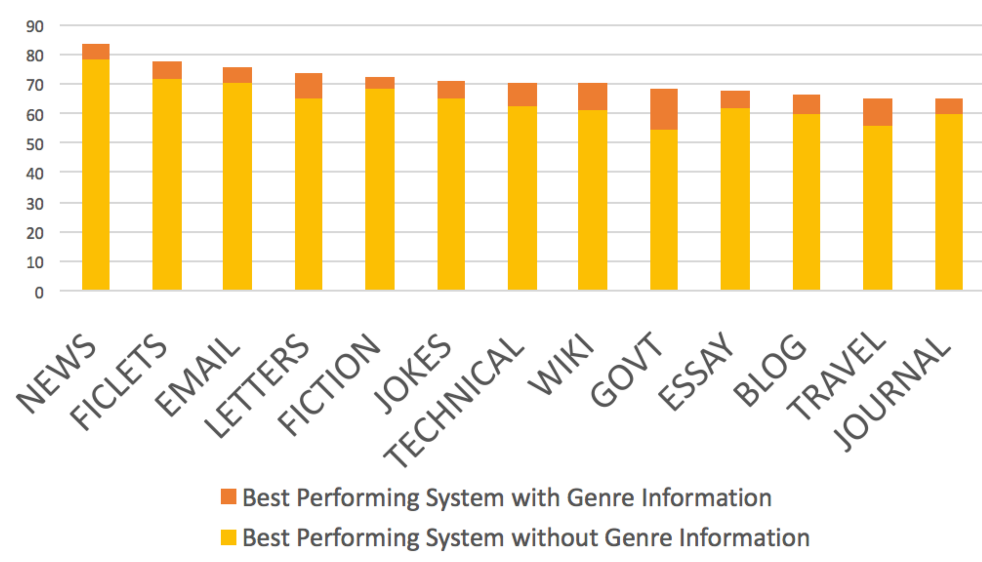


Words With High Attention Scores

- STATE: nouns, personal pronouns, pred. auxiliaries *(editors, I, am)*
- EVENT: gerunds *(thinking, writing)*
- GENERIC: adjectives, adverbs, modal verbs, indef. determiner *(awake, can, an)*
- GENERALIZING: names of official places/people *(York, States...)*

Impact of Genre

- Which genres are **easier** to classify?
- Which genre **helped** classifying correctly?



Similarity of Genres

- (sequences of) **SCT** differ among **genres**: most freq. n-grams per genre:
  - GENERIC → arg. texts, EVENTS → reports
  - STATE-STATE → Journals, GENERIC-GENERIC → Wikipedia
  - EVENT-EVENT-EVENT → Jokes, EVENT-STATE-STATE → gov. documents
- Distributions of SCT and their n-grams measured by **symmetric Kullback-Leibler divergence**