

**Inga Kempfert** and Christine Köhn  
{5kempfer,ckoehn}@informatik.uni-hamburg.de

---

**AN AUTOMATIC ERROR TAGGER FOR GERMAN**

# Error Tagging for Language Learner Texts

- Learner data annotated with error tags is useful, e.g. for evaluating GEC systems (*Bryant et al., 2017*)
  - Manual annotation is time-consuming and in some use cases infeasible
- Develop a rule-based system to tag error types in edits
- ERRANT: a grammatical ERRor ANnotation Toolkit (*Bryant et al., 2017*)
  - **Gerrant**: Error annotation tool for German

## Example for Edits and their Classification

	1	2	3	4	5	6
<b>orig</b>	Es	ist	<b>zeit</b>	für		Abendessen
<b>TH</b>	Es	ist	<b>Zeit</b>	für	<b>das</b>	Abendessen
	<i>It</i>	<i>is</i>	<i>time</i>	<i>for</i>	<i>the</i>	<i>dinner</i>
<b>tag</b>			<b>S:ORTH</b>		<b>I:DET</b>	

(ComiGS corpus (*Köhn and Köhn, 2018*), text 2mVs\_2)

# Error Types

- largely based on ERRANT's error types but adapted and extended
- try to cover frequent phenomena in learner language of German
- token-based but:
  - edits can contain multiple tokens on each side

# Error Types



- **Prefixes:** **D** (deletion), **I** (insertion), **S** (substitution)
- **15 Categories:** largely based on PoS such as **DET** for determiner errors, but also **SPELL, MORPH, ORTH, WO**
- **Further classification:** for nouns, pronouns, determiners, adjectives, adverbs and verbs
- overall: 61 precise tags

# Combinations and Alternatives

More than one inflectional property can differ at the same time within same category

→ allow combinations of error types with “\_”

- E. g. determiner error in case **and** number: **S:DET:CASE\_NUM**

# Combinations and Alternatives

More than one inflectional property can differ at the same time within same category

→ allow combinations of error types with “\_”

- E. g. determiner error in case **and** number: **S:DET:CASE\_NUM**

Some errors cannot be narrowed down to one error tag within same category

→ allow alternatives with “:” for further classifications of certain categories

- E. g. determiner error in case **or** gender: **S:DET:CASE:GEN**

# Word order errors

- word order error **S:WO** is token-based and can be further classified
- further classification only possible with suitably annotated input data
- e. g. The house **Blue** → The **blue** house: orthographic and word order error  
**S:WO:ORTH**



## Discontinuous Errors

- Discontinuous error spans can be classified (if made available by alignment)
- interesting for e.g. separable verb prefix or verb tense errors in German

	1	2	3	4	5
<b>orig</b>	Er	<b>hat</b>	das	Buch	<b>gelesen</b>
	<i>He</i>	<i>has</i>	<i>the</i>	<i>book</i>	<i>read</i>
<b>TH</b>	Er	<b>liest</b>	das	Buch	
	<i>He</i>	<i>reads</i>	<i>the</i>	<i>book</i>	
<b>edit</b>		<b>x</b>			<b>x</b>
<b>tag</b>		<b>S:VERB:TENSE</b>			

# Rules

- select prefix, then check error types depending on the prefix
- insertions (**I**) and deletions (**D**): based on PoS or **OTHER**
  - PoS tagger
- check substitutions for category and further classification
  - categories mostly based on PoS tag
  - use information provided by stemmer (*Cistem*), lemmatizer (*spaCy*), PoS tagger (*spaCy*), dependency parser (*spaCy*) and morphological analyser (*DEMorphy*) to narrow down error types

# Examples for Rules

- **S:ORTH**: delete whitespace and lowercase tokens, then check whether tokens are the same (same as in ERRANT)
- **S:DET**: ambiguities! Possibly number, gender, case or definiteness error or combinations thereof
  - narrow down set of morphological analyses of corrected word by including case information (from dependency label)
  - compare all morphological analyses of original word and corrected word pairwise and find minimal set of differences

# Evaluation

- 200 random edits
- 50 each from
  - ComiGS corpus, minimal target hypothesis (TH1)
  - ComiGS corpus, extended target hypothesis (TH2)
  - Falko corpus (*Reznicek et al., 2012*), TH1
  - Falko corpus, TH2

# Evaluation

- **Strongly Agree:** Tag fits best or all alternatives fit best
  - case *and* number: **S:NOUN:CASE\_NUM** or
  - number *or* case: **S:NOUN:CASE:NUM**
- **Agree:** Tag fits, but another fits better or an alternatives fits best
  - e.g. if **S:DET:CASE\_NUM** is best fitting but **S:DET:CASE\_NUM:GEN** is assigned
- **Disagree:** Tag fits without context, or a part of a combination fits best
  - e.g. if **S:NOUN:NUM** is best-fitting but **S:NOUN:CASE\_NUM** is assigned
- **Strongly Disagree:** Error tag does not match error type, no partial tags fit

## Evaluation Results

	overall (%)	
	coarse tag	precise tag
strongly agree	94.50	82.50
agree	1.00	10.25
disagree	0.50	1.25
strongly disagree	4.00	6.00

91% of precise tags rated as “strongly agree” or “agree” by both raters

# Outlook

- refine rules for some error types (e. g. word order **S:WO** or **S:VERB:AVZ**)
  - currently, edits extracted under simplifying assumption from ComiGS corpus:
    - **S:WO** in rare cases also assigned to non-WO errors
- improve disambiguation of word forms
- currently, focus only on the differences between the tokens → counter-intuitive classification possible (e.g. Houses **blue** → **Blue** houses: **S:WO:ORTH**)
- introduce new error categories for frequent errors that now fall under **OTHER**

# Summary and Conclusions

- Gerrant assigns fitting error tags in the majority of cases
- can classify discontinuous error spans
- word order errors can be further classified
- some errors could be covered more precisely by the rules
- extend error categories



Thank you

[nats.gitlab.io/gerrant](https://nats.gitlab.io/gerrant)