# An Automatic Error Tagger for German

**Inga Kempfert and Christine Köhn**
Natural Lanuguage Systems Group
Department of Informatics
Universität Hamburg
{5kempfer,ckoehn}@informatik.uni-hamburg.de

## Abstract

Automatically classifying errors by language learners facilitates corpus analysis and tool development. We present a tag set and a rule-based classifier for automatically assigning error tags to edits in learner texts. In our manual evaluation, the tags assigned by the classifier are considered to be the best or close to best fitting tag by both raters in 91% of the cases.

## 1 Introduction

For a variety of tasks, it is useful to classify errors by language learners into error types. E. g. corpora which are annotated with error types can be used to extract examples for compiling teaching material or exercises. Errors can only be interpreted sensibly with respect to a reconstructed utterance, a so-called target hypothesis (TH) (Reznicek et al., 2013). An error type characterizes the divergence between the learner utterance and the corresponding TH.

Manually annotating error types is a time-consuming task and has to be repeated if an error tagging scheme changes. Therefore, automatic error tagging is desirable and in some use cases even inevitable when manual annotation is not feasible due to the amount of data (e. g. when selecting training data from Wikipedia edits for Grammatical Error Correction (GEC) systems (Boyd, 2018) or when evaluating the performance of GEC systems (Bryant et al., 2017)) or due to an interactive setting (automatic error tags could be used as an information source for student modeling and feedback generation if a reliable GEC system is available). In addition, automatic annotation has the advantage that it can be used to easily unify error annotations across different corpora as long as some form of correction is available[1].

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **orig** | Es | ist | **zeit** | für | | Abendessen |
| **TH2** | Es | ist | **Zeit** | für | **das** | Abendessen |
| | *It* | *is* | *time* | *for* | *the* | *dinner* |
| **tag** | | | **S:ORTH** | | **I:DET** | |

Table 1: Example for two edits and their classification. The original text **orig** is aligned with the extended target hypothesis **TH2**. The edit at position 3 corrects a case error (error tag: **S:ORTH**), the other at position 5 inserts a determiner (**I:DET**). (ComiGS corpus, text 2mVs_2)

Inspired by ERRANT (Bryant et al., 2017), a grammatical ERRor ANnotation Toolkit for extracting and classifying edits in English learner texts, we developed an error annotation tool for German: Gerrant. It classifies edits extracted from already aligned parallel learner corpora and assigns error tags using a rule-based approach. An example for two edits from the ComiGS corpus (Köhn and Köhn, 2018) and their error tags is shown in in Table 1.

We present the system, the error types and the design decisions that lead to this set. Although we have a rather large and diverse tag set, the assigned tags were regarded as best fitting in most of the cases in our manual evaluation.

## 2 Related Work

There have been several approaches to classifying edits in learner texts automatically in the past. The Falko corpus (Reznicek et al., 2012, 2013) which consists of essays written by learner of German was automatically annotated with simple tags

[1]The TH may be created automatically by a Grammatical Error Correction system. Grundkiewicz and Junczys-Dowmunt (2018) achieved a performance close to humans for English.



Inga Kempfert and Christine Köhn 2018. An automatic error tagger for German. *Proceedings of the 7th Workshop on NLP for Computer Assisted Language Learning at SLTC 2018 (NLP4CALL 2018)*. Linköping Electronic Conference Proceedings 152: 32–40.

| Tag | Description |
|---|---|
| **ADJ*** | Adjective error |
| **ADV*** | Adverb error |
| **CONJ*** | Conjunction error |
| **CONTR** | Contraction error |
| **DET*** | Determiner error |
| **MORPH** | Morphological error |
| **NOUN*** | Noun error |
| **OTHER*** | Default category |
| **ORTH** | Orthography error |
| **PREP*** | Preposition error |
| **PUNCT** | Punctuation error |
| **SPELL** | Spelling error |
| **VERB*** | Verb error |
| **WO** | Word order error |

Table 2: Main error categories. Every category can be prefixed with **S:** (substitution), categories marked with * can be combined with the prefixes **I:** (insertion) and **D:** (deletion). Word order errors have a special role (see text). Some categories have to be further specified to form a valid tag.

which classify the differences between the original and the target hypothesis based on the manual alignment into changes, insertions, deletions, merges, splits and movements.

ERRANT (Bryant et al., 2017) uses a more sophisticated approach and a broader tag set of 25 main error types for classifying edits in learner English. Most error types are based on the part of speech of the involved words. Since most of the types can be prefixed with "M:" (Missing), "R:" (Replacement) or "U:" (Unnecessary edit), there are 55 error categories in total. ERRANT uses the "linguistically-enhanced alignment algorithm" by Felice et al. (2016) for extracting the edits from a parallel corpus, which are then classified using a rule-based approach. ERRANT classifies edits based on automatically-obtained features such as PoS tag and dependency parse.

Recently, Boyd (2018) extended ERRANT to German and used it for enriching the training data for a GEC system by selecting edits from the German Wikipedia only for certain error types. This increased the performance of the GEC system over using all edits.

## 3 Error Types

Inspired by ERRANT and different manual error annotation schemes for German learner texts (Rogers, 1984; Boyd, 2010), we developed our set of error categories and error tags. Every tag is prefixed by either **S:** (Substitution), **D:** (Deletion) or **I:** (Insertion). Table 2 lists the main error categories. Most categories are based on the PoS of the involved words. We call the combination of prefix and main error category a *coarse tag*. Nearly all PoS-based coarse tags have to be further specified to form a precise tag. This is done by appending subcategories to the coarse tag, e. g. the coarse tag **S:DET** can be extended to **S:DET:NUM** to form the precise tag for determiner error in number. The complete list of precise error tags is shown in Appendix A.

Insertions and deletions are either punctuation errors or certain PoS that have been inserted or deleted. Table 1 shows an example for inserting a determiner in the extended target hypothesis (TH2) from the ComiGS corpus (Köhn and Köhn, 2018).

Often an error involves more than one property of a word, e. g. a determiner might differ in case and gender. Therefore, we allow combinations of certain parts (see Appendix A) within the same coarse error tag with "_" (*and*), e. g. **S:DET:CASE_GEN** for determiner error in case and number. Some errors cannot be narrowed down to one error tag and we allow the combination of alternatives: Combinations are build with "**:**" between different error parts, e. g. **S:DET:CASE:GEN** means that the error is either a **S:DET:CASE** *or* a **S:DET:GEN** error, meaning Gerrant is unable to narrow down the error further[2]. Combinations of alternatives and conjunctions are also possible as in **S:DET:CASE_GEN:NUM** (a determiner error in case and gender or a determiner error in number).

Although the error tags are token-based, the verb error **S:VERB:SVA** (subject-verb agreement) includes syntactic errors but on the token level. Lexical confusions or semantic replacements are recognized either by the respective PoS-based category such as **S:VERB:-** if a verb was replaced with a semantically better fitting one or by **S:MORPH** if the tokens have the same stem, but different PoS.

If words are rearranged and changed at the same time, ERRANT classifies this only as a word order error or cannot recognize the word order error

[2]Note that even humans cannot always narrow the error down completely due to ambiguities

at all. In contrast, Gerrant treats word order errors as token-based, i. e. instead of rearranging a span of tokens, individual tokens are moved which allows for an additional error tagging of the moved tokens. Because of this, the tag for word order errors **S:WO** has a special role: It is an error tag on its own if the moved token was not changed but it can also be a prefix for another error type, e. g. if the word moved was change from lower to upper case this would be a tagged as **S:WO:ORTH**.

Currently, Gerrant does not automatically align the input texts and since it relies on a manual alignment being available, it has only been used on the Falko corpus and the ComiGS corpus. The detailed classification of word order errors only works on the ComiGS corpus because tokens in that corpus are aligned via a so-called tokmovid (tmid) if they have been moved (Köhn and Köhn, 2018).

Also contrary to ERRANT, Gerrant is able to assign an error tag to discontinuous word errors e. g. if the original text is *ist [. . . ] liegend* ("is lying") and the TH *liegt* ("lies") and the tokens are annotated with a tokmovid, the error is tagged as **S:WO:VERB:FORM**, a combination of word order and verb form error. This is also important for classifying errors with separable verb prefixes because the verb and its prefix are often far apart (see **VERB:AVZ** in Table 5 in Appendix A).

## 4 Implementation and Rules

Gerrant uses several sources of information to classify an edit. It uses SpaCy[3] for dependency parsing, PoS tagging and lemmatization, Cistem[4] (Weissweiler and Fraser, 2018) for stemming and DEMorphy[5] (Altinok, 2018) for morphological analysis. We trained our own SpaCy model on the Hamburg Dependency Treebank (Foth et al., 2014) which uses the dependency scheme by Foth (2006) and the STTS tag set for PoS (Schiller et al., 1999).

Cistem is a state-of-the-art stemmer and segmenter for German and is available for several programming languages, including Python in which Gerrant is written. We chose Cistem over the Snowball stemmer provided by the python library nltk because it achieves better overall results.

We use DEMorphy's analyses for recognizing morphological errors such as case or gender errors. DEMorphy is an off-the-shelf FST-based German morphological analyzer implemented in native Python. For reducing the set of possible analyses for one token, we use PoS tags of the original and the corrected tokens and the case information of the corrected tokens obtained from the dependency tree. The dependency tree is also used for identifying subject-verb agreement errors.

In Gerrant, an edit is checked for the different error types one after the other. First, the prefix is assigned, then the error type in accordance with the prefix. Insertion and deletion errors can only be classified as either a PoS error or a punctuation error. Edits with the prefix **S:** (Substitution) can be further classified by comparing not only the PoS but also morphological properties of the words on each side. Additionally, the edit has to be checked for spelling, orthographic, morphological and punctuation errors. Punctuation and orthographic errors are checked before PoS errors, spelling and morphological errors are checked for afterwards. The checks are all capsuled in different functions, which makes it easy to adjust the checks if need be.

For some error tags, it is sufficient to check if certain properties hold, e. g. for an orthography error **S:ORTH**, we only need to check whether case and/or whitespace is different between the words. For categories such as **DET**, there can be different readings for a word due to ambiguities: When processing a substitution error, we take all readings of the original token and all readings of the correction, try to narrow them down e. g. by case information from the dependency parse, and compare them pair-wise. For each pair, we combine all the differences with "_" (*and*) (e. g. **CASE_NUM**) and collect the differences for all pairs in one set. Then, we take the minimal subsets[6] of this set and combine them with "**:**" (*or*). This way, we end up with minimal diagnoses of the difference between the two tokens. The complete rule set can be found on Gerrant's website[7].

At this point, Gerrant only works on the ComiGS Corpus and the Falko corpus. The original text and the target hypotheses were already aligned in both corpora. In the Falko data, edits were already labeled with CHA (change), INS

[3] https://spacy.io/

[4] https://github.com/LeonieWeissweiler/CISTEM

[5] https://github.com/DuyguA/DEMorphy

[6] A minimal subset of a set $S$ is a subset for which no other subset of $S$ is also a subset.

[7] https://nats.gitlab.io/gerrant

| | rater 1 | | rater 2 | | overall | |
|---|---|---|---|---|---|---|
| | coarse tag | precise tag | coarse tag | precise tag | coarse tag | precise tag |
| strongly agree | 96.0 | 81.5 | 93.0 | 83.5 | 94.5 | 82.5 |
| agree | 0.5 | 11.0 | 1.5 | 9.5 | 1.0 | 10.25 |
| disagree | 0.0 | 1.0 | 1.0 | 1.5 | 0.5 | 1.25 |
| strongly disagree | 3.5 | 6.5 | 4.5 | 5.5 | 4.0 | 6.0 |

Table 3: Results of evaluation showing how much the human raters agree with the tags assigned by the system (in percent).

(insertion), DEL (deletion), MERGE, SPLIT and MOVS/MOVT (move source and move target). In the ComiGS corpus, the tokens are aligned and tokens which have been moved are labeled with a tokmovid.

For both corpora, we implemented individual readers converting them to the same edit format, which is passed to the error classifier. To make Gerrant accessible for other corpora, new readers can be added, that convert input data to an edit format that is processable by Gerrant. The edit format contains the original token, its absolute position in the text (optional), its position in the sentence, the error category, the corrected token, its absolute position in the text (optional), its position in the sentence and edit type.

## 5 Evaluation and Discussion

To evaluate Gerrant, we (the authors) manually rated the tags for 200 randomly chosen edits independently. One half was from the ComiGS corpus, the other from the FalkoEssayL2v2.4 corpus. For each of these sets, one half was from the minimal target hypothesis and one was from the extended target hypothesis.

The raters were given the original sentence, the corrected sentence, the edit and the tag assigned by the system. The raters were asked to judge on a 4-point Likert scale how appropriate the error tag is. Since there can be multiple tags for one coarse tag (combined with "**:**") and multiple parts combined in one tag (combined with "_") and we wanted to give partial credit for partially correct tags, the rating should be given as follows:

**Strongly agree** When the error in the text matches the error type in the description of the error tag exactly and no other tag fits better. If there are multiple tags combined with "**:**", every one of them fits exactly. Example 1: If **S:DET:NUM_CASE** is the best fitting tag and Gerrant assigns exactly **S:DET:NUM_CASE**. Example 2: If Gerrant assigns **S:DET:CASE:GEN** and both **S:DET:CASE** and **S:DET:GEN** fit exactly.

**Agree** When Gerrant assigns one error type (without combinations of parts with "**:**") and the error matches the type but another error type fits better. Or: When Gerrant assigns a combination of error types (combinations of parts with "**:**") and the error matches one of the assigned error types in the description of the error tag, which include the best fitting label. Example: If **S:DET:NUM_CASE** is the best fitting tag and Gerrant assigns **S:DET:NUM_CASE:GEN**.

**Disagree** When the error matches the error type in the description of the error tag without the context. Considering the sentence context, the tag is incorrect. Or: If more than one tag was assigned, no label fits perfectly, but parts of the label are correct (e. g. if the assigned tag is **S:NOUN:CASE_NUM:-**, but it is only a **S:NOUN:NUM**).

**Strongly disagree** When the error does not match the error type described in the error tag description. If more than one error tag is assigned, not even partial tags fit.

If none of the above cases apply, the most appropriate rating should be chosen.

In addition to the precise error tags, the raters also evaluated the coarse error tags for the same edits. The coarse error tag consists of the prefix and the first part of the error tag, e. g. **S:NOUN** or **S:MORPH**. The coarse tag for all word order errors is **S:WO** even if the word error's precise tag classifies the error further as in **S:WO:NOUN:CASE**.

The evaluation results for both raters are shown in Table 3. When averaging over both annotators, Gerrant assigns the best or close to best fitting pre-

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **orig** | Er | **hat** | seinen | Mund | mit | die | Hand | **anzuhalten** | und | nur | gucken | | |
| **TH2** | Er | **hält** | seinen | Mund | mit | der | Hand | zu | und | | guckt | nur | **zu** |
| | *He* | *shuts-1* | *his* | *mouth* | *with* | *his* | *Hand* | *shuts-2* | *and* | | *watches-1* | *only* | *watches-2* |
| **tmid** | | 1 | | | | | | 1 | | 2 | | 2 | |

Table 4: Sentence which contains a complex verb error (positions 2 and 8, marked with tokmovid **tmid** 1) where two verb forms are jointly replaced by two other verb forms. (ComiGS Corpus, text 2mVs_1)

cise tag in $92.75\%$ of the cases (coarse tag: $95.5\%$, see ). While there is only a small difference between coarse and precise tags if "strongly agree" and "agree" are considered in sum, there is a considerable drop in "strongly agree" ($-12$ percentage points on average) and a considerable increase in "agree" ($+9.25$ percentage points on average). This shows that Gerrant most often assigns the best fitting coarse tag but not as often also the best fitting precise tag but only the close to best. In only $3\%$ of the cases on average, the precise error tag was considered as not fitting (*disagree* or *strongly disagree*), although the coarse tag was considered fitting (*strongly agree* or *agree*).

Both raters give the same rating for the precise tags in $91.5\%$ of the cases (coarse tag: $95.5\%$) and $91\%$ of the precise tags are rated as *strongly agree* or *agree* by both annotators.

There are are a number of errors which Gerrant can improve on. Some error types do not behave as expected because Gerrant only extracts differences between the original and the correction, e. g. if the first word of a sentence is moved and the case is changed, this would be classified as an **S:WO:ORTH**, although technically it is not an orthographic error if the case was correct in the original text. For other error types, the rules can be further refined to match the tags more precisely: E. g. if the verb is changed by inserting the particle *zu* ("to") into the word as in *wegfahren* → *wegzufahren* ("to drive off"), Gerrant classifies this as a **S:VERB:AVZ**, although the separable verb prefix (*weg*) has not been changed. Currently insertions or deletions of the particle *zu* as a token on its own when it is not used as a separable verb prefix are classified as **OTHER**. It might be sensible to introduce an error category **PART** to cover all cases where the particle *zu* is deleted or inserted.

When a substitution error has more than one token on any side and the spans are not contiguous, Gerrant makes the simplifying assumption that this is always a word order error and uses **S:WO** as a prefix, although this might not be a word order error.

Gerrant can classify verb errors which contain more than one verb form on one side or both sides, e. g. for identifying tense errors. However, there are cases which Gerrant does not yet handle well: In the example in Table 5, the edit containing tokens 2 and 8 *hat anzuhalten* → *hält zu* ("has to stop" → "shuts") is tagged as a **S:WO:VERB:AVZ** error due to the differences in verb prefixes, although this should rather be modeled as a semantic and form error because *anzuhalten* ("to stop", an infinitive with the particle *zu*) was confused with *zuhalten* ("shut", a verb with the separable verb prefix *zu*).

Gerrant classifies verb errors based on the PoS of the original and the correction. Both sides must contain a verb form in order to check for verb errors. Because of this, some errors are not classified as verb errors due to the assigned PoS tags (an incorrect participle might be tagged as adjective and therefore is not treated as a verb).

Some improvements can also be made for recognizing **ADJ:FORM** and **ADV:FORM**, e. g. check if the adverb is accompanied with a particle (STTS tag: PTKA) or certain words such as *mehr* ("more").

Moreover, Gerrant could narrow down the assigned error tags further by taking more of the sentence context into account when disambiguating tokens.

## 6 Conclusions and Outlook

We presented Gerrant, an error annotation tool for German, which assigns error tags to given edits. Our evaluation shows that Gerrant chooses the most appropriate tag in the majority of cases. While the coarse tag is mostly correct, the precise tag is more often not the best fitting tag.

In future work, we plan to include more disambiguating information to further narrow down the possible error tags, currently the dependency tree is often used for disambiguating the corrected tokens but only rarely for the original tokens. Such

information might also be useful for reducing the set of analyses of the original tokens.

In addition, word order errors are assigned in certain rare cases in the ComiGS corpus (due to a simplifying assumption) where no reordering has taken place. Also, word order errors are currently only treated token-based which allows for a straightforward further classification of the error. However, groups of moved or rearranged tokens should be combined into one error, which would require that error spans for different errors can overlap.

Until now Gerrant has only been used on manually aligned corpora. It should be extended to be able to automatically align input.

Gerrant can be downloaded from `https://nats.gitlab.io/gerrant`.

## References


D. Altinok. 2018. DEMorphy, German Language Morphological Analyzer. *ArXiv e-prints*.

Adriane Boyd. 2010. EAGLE: an Error-Annotated Corpus of Beginning Learner German. In *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta. European Language Resources Association (ELRA).

Adriane Boyd. 2018. Using Wikipedia Edits in Low Resource Grammatical Error Correction. In *Proceedings of the 4th Workshop on Noisy User-generated Text*.

Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 793–805, Vancouver, Canada. Association for Computational Linguistics.

Mariano Felice, Christopher Bryant, and Ted Briscoe. 2016. Automatic Extraction of Learner Errors in ESL Sentences Using Linguistically Enhanced Alignments. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 825–835, Osaka, Japan. The COLING 2016 Organizing Committee.

Kilian A. Foth. 2006. *Eine umfassende Constraint-Dependenz-Grammatik des Deutschen*. Fachbereich Informatik, Universität Hamburg. URN: urn:nbn:de:gbv:18-228-7-2048.

Kilian A. Foth, Arne Köhn, Niels Beuck, and Wolfgang Menzel. 2014. Because size does matter: The Hamburg Dependency Treebank. In *Proceedings of the Language Resources and Evaluation Conference 2014*, Reykjavik, Iceland. LREC, European Language Resources Association (ELRA).

Roman Grundkiewicz and Marcin Junczys-Dowmunt. 2018. Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 284–290. Association for Computational Linguistics.

Christine Köhn and Arne Köhn. 2018. An Annotated Corpus of Picture Stories Retold by Language Learners. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 121–132. Association for Computational Linguistics.

Marc Reznicek, Anke Lüdeling, and Hagen Hirschmann. 2013. Competing target hypotheses in the Falko corpus. In Ana Ballier Díaz-Negrillo and Paul Nicolas Thompson, editors, *Automatic Treatment and Analysis of Learner Corpus Data*, pages 101–123. John Benjamins Publishing Company, Amsterdam, NLD.

Marc Reznicek, Anke Lüdeling, Cedric Krummes, Franziska Schwantuschke, Maik Walter, Karin Schmidt, Hagen Hirschmann, and Torsten Andreas. 2012. *Das Falko-Handbuch*.

Margaret Rogers. 1984. On major types of written error in advanced students of German. *International Review of Applied Linguistics in Language Teaching*, 22(1):1–39.

Anne Schiller, Simone Teufel, Christine Stöckert, and Christine Thielen. 1999. Guidelines für das Tagging deutscher Textcorpora mit STTS. Technical report, Universität Stuttgart / Universität Tübingen.

Leonie Weissweiler and Alexander Fraser. 2018. Developing a Stemmer for German Based on a Comparative Analysis of Publicly Available Stemmers. In *Proceedings of the 27th International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2017): Language Technologies for the Challenges of the Digital Age*, pages 81–94, Cham. Springer International Publishing.

## A Error Types

Categories which can be combined with **D:** (deletion) or **I:** (insertion) to form a precise error tag:

| Category | Description |
|---|---|
| | insertion or deletion of |
| **ADJ** | adjective |
| **ADV** | adverb |
| **CONJ:COORD** | coordinating conjunction |
| **CONJ:SUBORD** | subordinating conjunction |
| **DET** | determiner |
| **NOUN** | noun |
| **OTHER** | (default category) |
| **PREP** | preposition |
| **PRON** | pronoun |
| **PUNCT** | punctuation |
| **VERB** | verb |
| **VERB:AVZ** | separable verb prefix |

| Category | Description | Example |
|---|---|---|
| **ADJ:FORM** | Either the token in the original sentence is not a valid form or the degree is incorrect. | *Der **freundlichere** Mann → Der **freundliche** Mann* |
| **ADJ:INFL*** | The inflection degree (weak/strong) of the adjective in the original text is incorrect | *Ein **schlafende** Löwe → Ein **schlafender** Löwe* |
| **ADJ:NUM*** | The number of the adjective in the original text is incorrect. | ***Ungeduldiges** Pferde wiehern. → **Ungeduldige** Pferde wiehern.* |
| **ADJ:CASE*** | The case of the adjective in the original text is incorrect. | *Der **schlafendem** Löwe → Der **schlafende** Löwe* |
| **ADJ:GEN*** | The gender of the adjective in the original text is incorrect. | *Die **schöner** Frau geht spazieren. → Die **schöne** Frau geht spazieren.* |
| **ADJ:-*** | Any adjective error other than NUM, CASE, GEN, INFL and FORM e. g. the adjective was semantically replaced by a different one. | *Das **freundliche** Kind → Das **fröhliche** Kind* |
| **DET:NUM*** | The number of the determiner in the original text is incorrect. | ***Das** Pferde stehen auf der Weide. → **Die** Pferde stehen auf der Weide.* |
| **DET:CASE*** | The case of the determiner in the original text is incorrect. | *Ich gebe **den** Hund den Ball. → Ich gebe **dem** Hund den Ball.* |
| **DET:GEN*** | The gender of the determiner in the original text is incorrect. | ***Das** Hund bellt. → **Der** Hund bellt.* |
| **DET:DEF*** | The definiteness of the determiner in the original text is incorrect. | ***Ein** Hund bellt. → **Der** Hund bellt.* |
| **PRON:NUM*** | The number of the pronoun in the original text is incorrect. | ***Er** gingen nach Hause. → **Sie** gingen nach Hause.* |
| **PRON:CASE*** | The case of the pronoun in the original text is incorrect. | *Er gab mir **seiner** Jacke. → Er gab mir **seine** Jacke.* |
| **PRON:GEN*** | The gender of the pronoun in the original text is incorrect. | ***Er** läuft. → **Sie** läuft.* |

| Category | Description | Example |
|---|---|---|
| **PRON:-*** | Any pronoun error other than NUM, CASE or GEN. | ***Er*** *rennt.* → ***Wer*** *rennt?* |
| **NOUN:CASE*** | The case of the noun in the original text is incorrect. | *Ich sehe das Auto des* ***Mann****.* → *Ich sehe das Auto des* ***Mannes****.* |
| **NOUN:NUM*** | The number of the noun in the original text is incorrect. | *Die* ***Ball*** *rollen.* → *Die* ***Bälle*** *rollen.* |
| **NOUN:-*** | Any noun error other than CASE or NUM e. g. the noun was semantically replaced by a differnt one. | *Das* ***Kalb*** *schlief.* → *Das* ***Fohlen*** *schlief.* |
| **VERB:INFL** | The verb is not a valid form. | *Die Vögel* ***fliegten****.* → *Die Vögel* ***flogen****.* |
| **VERB:AVZ** | The separable verb affix is incorrect in the original sentence | *Er* ***beibringt*** *seinem Sohn etwas.* → *Er* ***bringt*** *seinem Sohn etwas* ***bei****.* |
| **VERB:FORM** | The infinitive form is incorrect or the use of infinitive forms or participles is incorrect | *Das Kind* ***ist lesend****.* → *Das Kind* ***liest****.* |
| **VERB:SVA*** | Number and/or person of the verb in the original text are incorrect. | *Das Mädchen* ***spielen*** *draußen.* → *Das Mädchen* ***spielt*** *draußen.* |
| **VERB:TENSE*** | The tense of the verb in the original text is incorrect. | *Das Mädchen* ***spielt*** *draußen.* → *Das Mädchen* ***spielte*** *draußen.* |
| **VERB:MODE*** | Passive or subjunctive error in the original text. | *Das Mädchen* ***hätte*** *gespielt.* → *Das Mädchen* ***hat*** *gespielt.* |
| **VERB:-*** | Any verb error other than INFL, AVZ, FORM, SVA, TENSE or MODE | *Das Kind* ***hat gehend*** *nach Hause.* → *Das Kind* ***rannte*** *nach Hause.* |
| **ADV:FORM** | Either the token in the original sentence is not a valid adverb form or the degree of the adverb is incorrect. | *Ich tanze* ***guter*** *als du.* → *Ich tanze* ***besser*** *als du.* |
| **ADV:-** | Any adverb error e. g. the adverb was semantically replaced by a different one. | *Ich lese* ***immer****.* → *Ich lese* ***gerne****.* |
| **CONJ:COORD** | Both tokens are conjunctions for a coordinate clause. | *und* → *aber* |
| **CONJ:SUBORD** | Both tokens are conjunctions for a subordinate clause | ***weil*** *das Kind lief* → ***während*** *das Kind lief* |
| **CONJ:-** | Any conjunction error which is neither CONJ:COORD nor CONJ:SUBORD | *weil* → *aber* |
| **CONTR** | A preposition and a determiner were contracted to a preposition or a preposition was split into a preposition and a determiner. | *Ich gehe* ***zu das*** *Haus.* → *Ich gehe* ***zum*** *Haus.* |
| **PREP** | All involved tokens are prepositions. | ***zu*** *dem Tisch* → ***auf*** *dem Tisch* |
| **PUNCT** | Any punctuation error. | . → , |
| **MORPH** | Morphology error: The word in the original text and the target hypothesis have the same stem but have different PoS tags. | *Er* ***Liebe*** *sie* → *Er* ***liebt*** *sie* |
| **OTHER** | Default category if none of the error tags are applicable | |
| **ORTH** | Orthography error: Whitespace or case error | *hunde Korb* → *Hundekorb* |

| Category | Description | Example |
|---|---|---|
| **SPELL** | Spelling error where the original lemma is unknown and has a certain similarity to the corrected token. | *Weinahtcen* → *Weihnachten* |
| **WO** | Word order error | *Das Haus* ***blaue*** → *Das* ***blaue*** *Haus* |

Table 5: Error categories which can be combined with the prefix **S:** to form a precise tag. * indicates that this tag can be combined with other tags in the same coarse category, e. g. case *or* number as in **S:ADJ:CASE:NUM** or case *and* number as in **S:ADJ:CASE_NUM**. Note that "**-**" cannot be combined with "_" (*and*). **WO** has a special role as it can be combined with any other category in this table (see Section 3).