# Towards a Tool for Automatic Spelling Error Analysis and Feedback Generation for Freely Written German Texts Produced by Primary School Children

*Ronja Laarmann-Quante*

Ruhr-University Bochum, Germany
laarmann-quante@linguistics.rub.de

## Abstract

This paper proposes a tool for the automatic analysis of spelling errors in freely written German texts. It is based on automatic annotations of spelling errors that comprise various levels, such as linguistic properties of the target word (phonemes, syllables, morphemes) and error-related properties such as error categories which mark whether the misspelling changes the pronunciation of a word or whether the correct spelling can be derived from a related word form. These can be used to create an application that could, for example, help teachers analyze their students' orthographic skills and give feedback with little manual effort. For the future, it could also be implemented as an automatic tutoring system for children in which case the surface has to be child-oriented and should present error analysis as a kind of game. While the paper presents the capabilities of a first prototype, the concrete implementation for real-world use is open for discussion with experts on orthography instruction.
**Index Terms**: spelling errors, German, free text, primary school, individual diagnosis, automatic analysis, feedback

## 1. Introduction

When children learn to write, it is essential that feedback about spelling mistakes is relevant and helpful. A recent study with students of grades 3-5 in German schools has shown that an individual qualitative analysis of the spelling errors in freely written texts from children is a very good basis for providing feedback and fostering the children's spelling competence [1, 2]. However, even in primary schools, one teacher has to take care of 13-29 children in one class [3], which in reality makes it almost impossible to analyze every child's misspellings in detail. Furthermore, some teachers and teachers-to-be in Germany have been shown to have deficits in knowledge about the German writing system or in using this knowledge in didactic contexts [4, 5]. It is crucial, though, that feedback about a spelling is not misleading. For instance, in German, a careful pronunciation of a word is often the basis for the correct spelling but not always: the word <Kinder> 'children' is pronounced [kɪndɐ] in standard German, with the last phone sounding similar to [a]. If a child wrote *<Kinda> instead, it would be wrong to tell the child that it has to pronounce the word more carefully in order to 'hear' the correct spelling [4].

In order to support children's spelling acquisition more individually, automatic analyses of spelling errors with the help of natural language processing methods could reduce the time and cognitive effort that teachers need for a thorough assessment of a child's text. A close examination of the strengths and weaknesses of a child is then a good basis for further training. This paper proposes a tool which automatically categorizes spelling errors and provides additional information about them which can be used for preparing feedback or individual training material. We present a first prototype which could be adapted to the specific needs of teachers. For the future, if the automatic analysis becomes robust and reliable enough, it is also conceivable to design a child-oriented surface which presents the writing of a free text and a subsequent analysis and correction of spelling errors as a game.

An application with automatic error categorization for diagnostic purposes for German has been developed before [6, 7] but its focus is only on a small number of error categories and it does not provide feedback beyond statistics about possible and committed errors. Our goal is to facilitate a systematic analysis of all the errors in a text and to take different features beyond a one-dimensional error categorization into account. For instance, we want the application to automatically provide feedback about whether the pronunciation of the word changes with a misspelling or how one could otherwise arrive at the correct spelling using a previously developed annotation scheme for spelling errors and its automatic application [8, 9]. The aim of this paper is to show how these rather abstract annotations can be presented in a user-friendly way so that they could be used e.g. in a school setting.

The paper is structured as follows: Section 2 gives a short overview of word spelling in German and summarizes the annotation scheme which forms the basis of the spelling error analysis. Section 3 explains how the annotations are obtained automatically, before Section 4 presents a first prototype of how the annotations can be visualized in a user-friendly application. Section 5 concludes the paper with a summary and outlook.

## 2. Background: German Orthography and Spelling Error Annotation

The basis of German word spellings are grapheme-phoneme correspondences (GPCs).[1] For instance, the word *kalt* 'cold' can be spelled phonographically, which means that each phoneme in the pronunciation [kalt] corresponds to exactly one grapheme in the correct spelling <kalt>.[2] The correspondences can be read off from basic (context-independent) GPC-rules such as /k/ $\rightarrow$ <k> etc. However, there are phenomena which overwrite these basic correspondences, for example consonant doubling (<re**nn**en> [ʁɛ**n**ən]'(to) run') or that final devoicing is not reflected in the spelling ([hʊn**t**] is spelled <Hun**d**> 'dog'). These can be explained via the word's syllabic/prosodic structure (e.g. consonant doubling occurs in the spelling if in the pronunciation a single consonant occurs in a syllable joint

[1]The explanations of German word spelling follow [10, 11] in this paper.

[2]The pronunciation of a word in IPA is presented in square brackets, the spelling in angle brackets and phonemes in slashes.

between a stressed lax vowel and an unstressed vowel) or with reference to a related word form: The reference form of the word *Hund* [hʊnt] is a disyllabic word form such as the plural *Hunde* [hʊndə] which is spelled <Hunde> phonographically. The spelling usually remains constant in all related word forms (principle of morpheme constancy).

Spelling errors can be analyzed on different levels. For German, the classic approach which is used by standardized spelling tests such as the HSP [12] and by other popular spelling error analysis schemes such as AFRA [13] and OLFA [14] for freely written texts, is to categorize errors one-dimensionally by phenomena such as consonant doubling or final devoicing. However, with regard to giving feedback about errors, it is also important to take other features into account as well.[3] For instance, some misspellings change the pronunciation of a word (in *<renen> the first vowel is pronounced long and in <rennen> short) while others do not (*<Hunt>/<Hund> are both pronounced with a voiceless obstruent because of final devoicing), some spellings can be derived from related word forms (<Hund> from plural <Hunde>) while others cannot. To capture all these differences, we developed a multi-level annotation scheme for misspellings. For the annotation, the target word, i.e. the orthographically correct version of the word that the child wanted to write, has to be known. The annotation scheme then codes the following information:[4] For each target word we annotate

- graphemes
- phonemes
- syllables
- morphemes
- whether it structurally belongs to the German core vocabulary or can be seen as a 'foreign word' (according to [15])

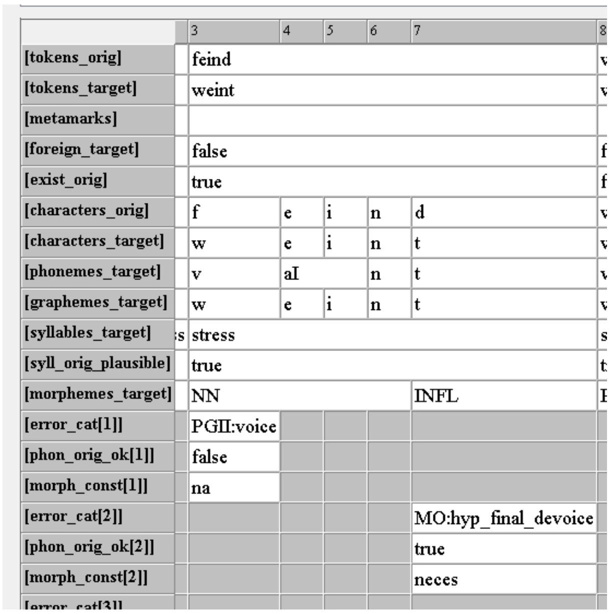
For each misspelled word, i.e. the original spelling of the child, we annotate whether the child's spelling

- resulted in another existing German word[5] (e.g. <feld> 'field' for <fällt> '(he/she) falls'): it might be interesting to know whether a child produces forms that he or she may have encountered before (e.g. in a book) vs. forms he or she has never seen before in other texts
- has a plausible syllable structure or violates graphotactic constraints: this can give a hint whether a child has a sense of the structure of German words

Finally, for each error (there can be more than one error in a word) we annotate

- error category
- whether the error influences the pronunciation of the word
- whether the correct spelling can be derived from a related word form

The annotations are stored in a custom-made XML format called *LearnerXML*[8] and can be visualized in *EXMARaLDA* [16, 17]. Figure 1 shows a screenshot from *EXMARaLDA* of all the annotations of the misspelling *<feind> for <weint> '(he/she) cries'.[6]



| | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| [tokens_orig] | feind | | | | | v |
| [tokens_target] | weint | | | | | v |
| [metamarks] | | | | | | |
| [foreign_target] | false | | | | | f |
| [exist_orig] | true | | | | | f |
| [characters_orig] | f | e | i | n | d | v |
| [characters_target] | w | e | i | n | t | v |
| [phonemes_target] | v | aI | | n | t | v |
| [graphemes_target] | w | e | i | n | t | v |
| [syllables_target] | s stress | | | | | s |
| [syll_orig_plausible] | true | | | | | t |
| [morphemes_target] | NN | | | | INFL | I |
| [error_cat[1]] | PGII:voice | | | | | |
| [phon_orig_ok[1]] | false | | | | | |
| [morph_const[1]] | na | | | | | |
| [error_cat[2]] | | | | | MO:hyp_final_devoice | |
| [phon_orig_ok[2]] | | | | | true | |
| [morph_const[2]] | | | | | neces | |
| [error_cat[3]] | | | | | | |

Figure 1: *Screenshot of EXMARaLDA with the annotations for the misspelling *<feind> for <weint>*

One can see that the target word is not a foreign word (*foreign_target = false*) and that the misspelling resulted in another existing word (*exist_orig = true*; *Feind* = 'enemy'). Moreover, one can see that there are two errors in the word. The first one, <f> for <w>, has to do to with voicing (*error_cat[1] = PGII:voice*) and changes the pronunciation of the word (*phon_orig_ok = false*). Morpheme constancy does not explain the correct spelling here (*morph_const = na*). The second mistake, in contrast, concerns a substitution of <t> with <d>, which is a hypercorrection of final devoicing (*error_cat[2] = MO:hyp_final_devoice*). This does not affect the pronunciation of the word and morpheme constancy does play a role in that the child could have known that the word was spelled with a final <t> if it had recognized that this was an inflectional suffix that is always spelled <t>.

A more detailed description of the annotation scheme and the XML format can be found in [8]. In a current research project, we analyze a corpus of freely written texts (descriptions of picture stories) produced by primary school children of grades 2-4 that was collected by [18]. We want to investigate the relationship between spelling errors and a word's properties, including the information coded in the multi-layered annotation scheme presented above. In order to be able to annotate a large number of texts (in total we have over 1800 texts that were manually transcribed and corrected for spelling errors), we implemented an automation of the annotations in the scheme, which will be described in the next section. Thereby, we found that the analysis of orthographic errors can be carried out well with a rule-based approach. This has the advantage that the analyses do not depend on training material and can therefore be applied to any written text. The annotations of a text form then the backbone of the application described in Section 4, which presents

[3]AFRA sometimes allows one error to fall into multiple categories, which highlight different aspects, but this is not possible for all errors.

[4]The complete annotation scheme with all categories can be found under `https://www.linguistics.ruhr-uni-bochum.de/litkey/Scientific/Corpusanalysis/Resources.html`.

[5]Capitalization is not taken into account here.

[6]Phonemes are represented in SAMPA notation here, see `http://www.phon.ucl.ac.uk/home/sampa/german.htm`.

an analysis of all the spelling errors in a text in a way that is suitable for e.g. a teacher or other language trainer rather than a corpus linguist.

## 3. Automatic Spelling Error Analysis

The automatic annotation of a word's properties according to our annotation scheme described in the previous section is carried out with the help of the web service *G2P* of the Bavarian Archive of Speech Signals (BAS)[7] [19, 20] and our own processes. These are described in [9] but to convey how the spelling analysis of our proposed application works, the basic procedures will be sketched here again.

The *G2P* web service provides the phonemes, syllables and morphemes for any given word. For the word <fröhlich> [fʁøːlɪç] 'happy', for example, it returns

| | |
|---|---|
| phonemes (in SAMPA) | `fr2:lIC` |
| syllable boundaries (`.`) + stress marks (`'`) | `fr'2:.lIC` |
| morpheme boundaries | `fröh lich` |
| morpheme tags | `ADJ SFX` |

In German, units of more than one letter can correspond to a single phoneme (e.g. in *fröhlich*, <öh> corresponds to `[2:]`/[øː], <ch> corresponds to `[C]`/[ç]). To obtain these character units which we call *phoneme-corresponding units* or *PCUs*[8], we firstly use string alignment via weighted Levenshtein distance.[9] The resulting 1:1 (or 1:0, 0:1) correspondences between characters and phonemes are then rearranged via rules to get the correct PCUs.

The automatically obtained information about graphemes, phonemes, syllables and morphemes are then used to determine, for a given target word, which of the systematic spelling errors that our scheme defines, *could* be committed in this word. Currently, there are 65 error categories in total (+ 4 'other' categories), of which 62 are implemented. For instance, the category *final_devoice* (final devoicing), which describes an error such as *<Hunt> for <Hund> 'dog' (see Section 2) can apply to every <b, d, g, w, s> in a syllable coda. Every error category is connected with an error 'candidate' i.e. it is stored what the word would look like if the error was in fact committed. In this case, it would be that <b, d, g, w, s> are replaced by <p, t, k, f, ß>, respectively. If a misspelling is encountered, all the error candidates of the corresponding target word are checked if they equal the misspelling. Error candidates unambiguously point to a single error category, so for the example above, *<Hunt> is found among the error candidates of <Hund> and the error category *final_devoice* can be read off from the matching error candidate. Furthermore, it is stored in which PCU the error occurred. If there are multiple errors and/or unsystematic errors in a word, all combinations of all error candidates of a word are computed and compared with the original spelling, see [9] for details.

This procedure allows a precise alignment of the original and target spelling which entails a correct allocation and categorization of the errors. This is not always trivial. For instance, if you consider the misspelling *<vermiest>[10] for <vermisst> '(he/she) misses', an alignment based on Levenshtein distance would be (1):

(1)
<table>
<tr><td>v</td><td>e</td><td>r</td><td>m</td><td>i</td><td>e</td><td>s</td><td>t</td></tr>
<tr><td>v</td><td>e</td><td>r</td><td>m</td><td>i</td><td>s</td><td>s</td><td>t</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PGIII:repl_C</td><td></td><td></td></tr>
</table>

This would mean that the child replaced an <s> with an <e> (in our annotation scheme this would be category *PGIII:repl_C*, i.e. an unsystematic incorrect character for a consonant), which does not capture the error correctly. Instead, what the child did was using a grapheme for a long vowel (<ie>) for a short vowel (category *SL:rem_Vlong_short*, here shortened as *SL:rVs*) and omitting consonant doubling in a position before another consonant (category *SL:Cdouble_beforeC*, here shortened as *SL:CbC*).[11] Hence, the alignment which characterizes the error(s) correctly and which our procedure produces looks like (2):

(2)
<table>
<tr><td>v</td><td>e</td><td>r</td><td>m</td><td>i</td><td>e</td><td colspan="2">s</td><td>t</td></tr>
<tr><td>v</td><td>e</td><td>r</td><td>m</td><td colspan="2">i</td><td>s</td><td>s</td><td>t</td></tr>
<tr><td></td><td></td><td></td><td></td><td colspan="2">SL:rVs</td><td colspan="2">SL:CbC</td><td></td></tr>
</table>

The two further annotations that are connected to an error, namely *phon_orig_ok* (whether the misspelling changes the pronunciation of the word) and *morph_const* (whether the correct spelling can be derived from a related word form) can be read off from the error category, sometimes connected with further restrictions. For example, for *final_devoice* (see *<Hunt>/<Hund> above), *phon_orig_ok* is always *true*, i.e. by definition, errors of this category never change the word's pronunciation. For *morph_const*, the syllables and morphemes play a role: For this error category, deriving the correct spelling from a related word form is said to be always possible/necessary if the devoiced letter is in final position of the syllable coda of an inflecting morpheme. This captures that words that do not inflect (such as <und> 'and') do not have related word forms and that in words where the devoiced consonant is not in final position of the syllable coda (such as <O**b**st> 'fruit'), the consonant is devoiced in *all* related word forms, hence they are of no help. For further information on these features, see [9].

In order to annotate *exist_orig* (does the misspelling result in another existing German word form?), one has to approximate what 'existing word forms' for children are, i.e. which words they are likely to have encountered before. For this purpose, we use an extract from *childLex*, the German Children's Book Corpus [22]. In order to exclude words which are specific to only particular books or book series, we extracted those words which occurred in at least 10 of the 500 books that were used for the compilation of the corpus. Besides these extracted words themselves, we also included all their related word forms, i.e. those with the same lemma. Furthermore, we did some corpus cleaning in that words in uppercase letters as well as lowercased words which were tagged as a noun and capitalized words which were not tagged as a noun were removed. The final list includes just over 39,000 word forms. For all misspellings which resulted in one of these word forms, *exist_orig* is tagged as *true*.

[7] `https://clarin.phonetik.uni-muenchen.de/BASWebServices/#/services/Grapheme2Phoneme`

[8] These units are sometimes referred to as *graphemes* [21] but we follow the grapheme definition in [10] where <ch>, <sch>, <ie> and <qu> are defined as the only multi-letter graphemes in German.

[9] For the alignment we are using a script created by Marcel Bollmann, `https://github.com/mbollmann/`.

[10] The misspelling resulted in another German word meaning '(he/she) spoiled (sth. for sb.)'.

[11] The abbreviations *PGIII* and *SL* before the colon mean that the errors refer to the level of grapheme-phoneme correspondence and the syllabic level, respectively. For more information see [8].

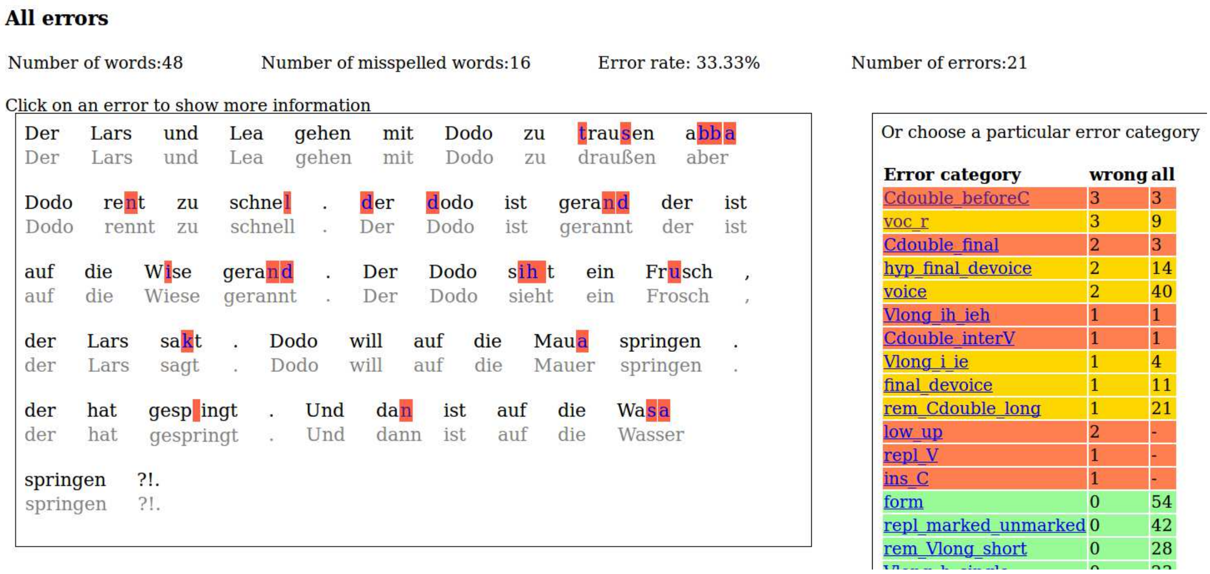



Figure 2: *Main page showing the learner's text with marked errors and a color-coded table of error categories*

The two remaining layers of our annotation scheme, namely *foreign_target* (is the target word a foreign word?) and *syll_orig_plausible* (does the syllable in the original spelling adhere to graphotactics?), are currently being implemented.

A big issue when talking about automatic analyses or annotations is of course the reliability. So far, we have studied the agreement between human and automatic annotations for the error category, *phon_orig_ok* and *morph_const* (see [9]). Three human annotators (students) and the automatic system annotated 11 texts (866 target tokens) taken from our corpus of freely written children's texts addressed in Section 2. Table 1 (extracted from the result table presented in [9]) shows the average agreement of human and automatic annotations and among human annotations only.

| | Level | Size | Auto/Human perc. | Auto/Human $\kappa$ | Human perc. | Human $\kappa$ |
|---|---|---|---|---|---|---|
| *complete* | error cat. | 261 | 78.67% | .77 | 77.39% | .82 |
| | phon_orig_ok | 227 | 80.35% | .63 | 78.03% | .72 |
| | morph_const | 227 | 79.54% | .57 | 70.99% | .55 |
| *w/o SN* | error cat. | 170 | 72.16% | .71 | 72.94% | .79 |
| | phon_orig_ok | 145 | 80.26% | .63 | 78.03% | .72 |
| | morph_const | 145 | 79.46% | .57 | 70.99% | .55 |

Table 1: *Average inter-annotator agreement (raw percent and Fleiss' $\kappa$) between human and automatic annotations and among human annotations only.*

Since errors of category *SN* (syntactically determined errors, i.e. capitalization and writing words together or separately) are trivial to detect when the target word is known, the table also presents agreement figures disregarding errors from this category to get a sense of how well the more challenging cases are handled. The features *morph_const* and *phon_orig_ok* were only evaluated if the error category was the same for all human annotators (and the system, respectively). As one can see, agreement with the automatic annotations is quite comparable to the agreement among the human annotations. We are currently working on a gold standard for further evaluation results and our automatic procedures are still being improved by considering the sources of disagreements that were discussed in [9]. We are also currently evaluating the information about phonemes, syllables and morphemes obtained from the BAS web service, which is also not 100% correct as one can see in Figure 1, where the stem of the verb *weinen* is analyzed as a noun. As they form the basis of our error analysis, the correctness of these information in part determines the correctness of our error annotation, and hence the feedback that is generated about a spelling, as well.

Generally, one can see that it is not trivial to interpret spelling errors. As indicated in the introduction, also teachers, whose assessment of spelling errors can have a direct influence on the feedback given to the child and the evaluation of his/her spelling competence, do not always analyze spelling errors correctly. Hence, a "joint effort" of human and automatic analyses should be even more reliable than analyses made by teachers "from scratch" in a school context. With a precise allocation of the error(s) in a misspelled word and various information about the target word and the error, it is possible to create an application which uses these features to create individual diagnoses and feedback. The main task is to find a suitable visualization for the target users and to present the relevant information accordingly. The next section presents a very first prototype of how the rather abstract annotations can be visualized in an application that can be used for example by teachers in order to assess a child's spellings in a freely written text.

## 4. Application Prototype

While a manual analysis of spelling errors is very time-consuming, the automatic tool analyzes the spelling errors in the input text within seconds.[12] Figure 2 shows the main page of the application with an example text from our corpus addressed

[12] Currently, both the original spelling and the target spelling of a word have to be provided as input for the tool. However, we are working on a robust automatic spell checking procedure for children's writings so that at one point the manual effort of correcting the texts will be reduced to counter-checking the automatic corrections.

in Section 2. Since in our project we are studying orthography errors only, grammatical errors are not corrected and addressed in the texts, and are not part of our annotation scheme. We are planning to deal with grammatical errors in the future.

At the top of the page, one can see some general statistics regarding the number of words and errors in the text. The box on the left hand side displays the input text with each target word printed in gray under the original spelling. Each error is marked in red in the original word. This is achieved with the procedure described in Section 3 that allows a precise allocation of an error. On the right hand side of the page one can see a table of error categories. The column "wrong" shows how many times an error of a particular category was committed and "all" shows how many times this error could have been committed ('base rate' [13, 23]). In that our automatic error analysis first determines which errors could theoretically be committed in a given word before committed errors are interpreted, we automatically obtain the base rate for each category. Errors which can occur in every word (such as capitalization, here *low_up*) or every letter (like unsystematic replacement or deletion of a letter, here *repl_V* and *ins_C*) are not given a base rate. What you can see in this table, for instance, is that there are nine occurrences of *voc_r*, i.e. vocalized <r> (as in [kɪndɐ]/<Kinder> 'children') in the text and three of them were misspelled. As [6] emphasizes, it is important to have a "dual view" on the absolute and relative error rate of a category. Therefore, two different visual markers are used to be able to see the prominence of an error category without effort: The categories are ranked by the absolute number of errors with the category with most errors put on top of the table (errors without base rate are currently displayed under those with a base rate). The relative error rate of a category is indicated with colors: red means the ratio of committed and possible errors is $\geq .5$, yellow means $< .5$ and green means that no error was produced in this category.

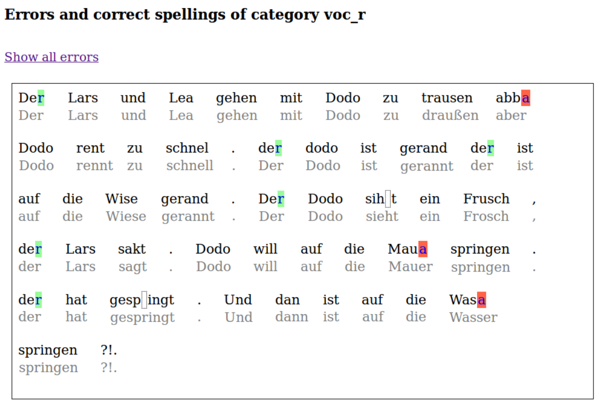



Figure 3: *Page showing all errors (red) and correct applications (green) of r-vocalization (voc_r) in the learner's text*

While the main page gives a quick overview of the main problem areas in the child's spellings, you can click on an error category in the table to get more detailed information. Clicking on *voc_r* yields the page in Figure 3: Here you can see all occurrences of a vocalized <r> in the text with correct spellings in green and mistakes in red. Looking at the text itself gives insights which are hidden if you only look at the abstract statistics. While an error rate of 3/9 for *voc_r* does not look too alarming, seeing the occurrences in the text quickly reveals that the child only spelled the vocalized <r> correctly in the definite article <der>, which is probably memorized holistically, and made a mistake in all other occurrences, including two nouns (*<Mau**a**>/<Mau**er**> 'wall', *<Was**a**>/<Wass**er**> 'water') and one conjunction (*<abb**a**>/<ab**er**> 'but').[13] From seeing this, you can quickly conclude that the phenomenon of r-vocalization has not yet been successfully acquired by the child. In general, it is important to not only analyze errors but also to see where a child dealt with a particular phenomenon correctly (see also [23]). While it is tedious to search a text manually for all occurrences of a phenomenon, the proposed application does exactly this for all of the error categories, which should make it easier for a teacher to evaluate a child's spelling competence adequately.

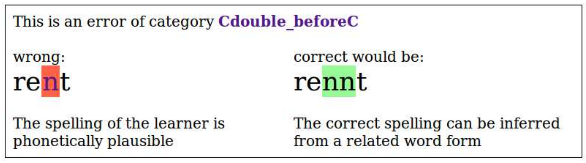



Figure 4: *Additional generated information concerning the misspelling *<rent> for <rennt> '(he/she) runs'*

Clicking on a specific mistake in the text opens a box with additional information about certain particularities of a misspelling. This is supposed to help giving relevant feedback about a particular spelling. In principle, all the annotations from our annotation scheme could be used here. For instance, it could be interesting to know which morpheme class was affected by the error: errors in frequent function words may point at more severe deficits in spelling competence. If, in contrast, a foreign word with a non-native structure was misspelled (as could be read-off from the feature *foreign_target*), these are not very indicative of the child's spelling abilities per se. So far, the features *phon_orig_ok* and *morph_const* have been implemented into the application. Depending on the feature value, a sentence is generated which states whether the error has influenced the pronunciation of the word and whether the spelling can be derived from a related word form, respectively. Figure 4 gives an example for the misspelling *<rent> for <rennt> '(he/she) runs'. The application shows here that the error does not influence the pronunciation of the word (so feedback should not refer to the child's articulation here) but that the correct spelling can be derived from a related word form. For future versions, a desideratum is to also show the related word forms that the spelling can be derived from, in the case of <rent> it would for example be <rennen>, or to pick words from a database which show a similar spelling pattern that can be used for training. What and how to display information here is worth discussing with experts on orthography instruction.

## 5. Conclusion and Outlook

This paper presented a proposal for the automatic spelling error analysis in freely written German texts. The aim of the paper was to show how previously developed techniques for analyzing spelling errors on various levels could be used to develop a user-friendly application. In the focus of the analyses are misspellings produced by German primary school children.

[13] It could also play a role that in <der> the <er> is pronounced [eːɐ], i.e. you can hear an [e], while in the misspelled words <er> is pronounced [ɐ].

Such an application can for example assist teachers in analyzing children's spelling errors in more detail without a big manual effort and is also capable of giving hints for useful feedback about spellings. The prototype presented here is supposed to give an idea of what our proposed application would be able to do. The final implementation, however, is highly dependent on the needs of the target users. So far, the application was presented as an aid for teachers or other language trainers who want to examine the main problem areas of a child's spellings and be assisted in how to give feedback. Errors are marked in the text itself just like a teacher would do but at the same time one can see where the same phenomenon was handled correctly by the child, which would be tedious to annotate manually. For a smooth integration into the daily routine in schools, one obstacle is that children usually produce handwritten texts while the tool needs digital input. As was also noted by [7], it would be desirable to cooperate with the field of automatic handwriting recognition technology in this respect. On the other hand, with digitization being on the rise in all areas, it is not unlikely that in the future, children will produce a lot of typewritten texts as well. It is also conceivable to implement the automatic spelling error analysis as an application that can directly be used by children. In this case, the whole surface has to look differently of course and one should think about implementing the error analysis as a game.

Anyway, the further development of the tool would gain a lot from the expertise of professionals in orthography instruction. What has to be discussed, for instance, is what kinds of error categories to use (our categorization system is very fine-grained and can easily adapt to more specific needs), what information about an error to display and how to organize the whole application. A goal would then be to have the application tested and evaluated in real-world settings, i.e. by teachers or even children.

Furthermore, it would be interesting to see to what extent L2 learners of German could benefit from such an application as well. To investigate how their orthography errors fit into our annotation scheme is part of our future work.

# 6. Acknowledgements

This research is part of the project *Literacy as the key to social participation: Psycholinguistic perspectives on orthography instruction and literacy acquisition* funded by the Volkswagen Foundation as part of the research initiative “Key Issues for Research and Society”.

# 7. References

[1] K. Siekmann, “Kompetenzorientierte orthographische Individualförderung: Ergebnisse einer Förderstudie,” in *Theorie, Empirie und Praxis effektiver Rechtschreibdiagnostik*, K. Siekmann, Ed. Tübingen: Stauffenburg Verlag, 2014, pp. 191–205.

[2] ——, “Evidenzbasierte Förderung des Orthographieerwerbs auf der Grundlage individueller Fehleranalysen,” in *Lernprozessbegleitung und adaptives Lernen in der Grundschule*, K. Liebers, B. Landwehr, A. Marquardt, and K. Schlotter, Eds. Wiesbaden: Springer VS, 2015, pp. 199–204.

[3] Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, “Vorgaben für die Klassenbildung: Schuljahr 2016/2017.” [Online]. Available: https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Klassenbildung_2016.pdf

[4] I. Corvacho del Toro, *Fachwissen von Grundschullehrkräften: Effekt auf die Rechtschreibleistung von Grundschülern*. Bamberg: University of Bamberg Press, 2013.

[5] S. Jagemann, “Was wissen Studierende über die Regeln der deutschen Wortschreibung?: Eine explorative Studie zum graphematischen Wissen,” in *Lehrende im Blick*, C. Bräuer and D. Wieser, Eds. Wiesbaden: Springer Fachmedien Wiesbaden, 2015, pp. 255–279.

[6] K. Berkling, “A case study using data exploration of spelling errors towards designing automated interactive diagnostics,” in *WOCCI*, 2012, pp. 97–103.

[7] ——, “A non-experts user interface for obtaining automatic diagnostic spelling evaluations for learners of the German writing system,” in *INTERSPEECH*, 2013, pp. 1879–1881.

[8] R. Laarmann-Quante, L. Knichel, S. Dipper, and C. Betken, “Annotating spelling errors in German texts produced by primary school children,” in *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, A. Friedrich and K. Tomanek, Eds., 2016, pp. 32–42.

[9] R. Laarmann-Quante, “Automating multi-level annotations of orthographic properties of German words and children's spelling errors,” in *Language Teaching, Learning and Technology*. ISCA, 2016, pp. 14–22.

[10] P. Eisenberg, *Das Wort*, 3rd ed., ser. Grundriss der deutschen Grammatik. Stuttgart: J.B. Metzler, 2006, vol. 1.

[11] ——, *Das Wort*, 4th ed., ser. Grundriss der deutschen Grammatik. Stuttgart: Metzler, 2013, vol. 1.

[12] P. May, *Hamburger Schreib-Probe zur Erfassung der grundlegenden Rechtschreibstrategien: Manual/Handbuch Diagnose orthografischer Kompetenz*. Stuttgart: vpm, 2013.

[13] K.-L. Herné and C. L. Naumann, *Aachener Förderdiagnostische Rechtschreibfehler-Analyse*, 4th ed. Aachen: Alfa Zentaurus, 2002.

[14] G. Thomé and D. Thomé, *Oldenburger Fehleranalyse OLFA: Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz aus freien Texten ab Klasse 3 und zur Qualitätssicherung von Fördermaßnahmen*. Oldenburg: isb Verlag, 2004.

[15] P. Eisenberg, *Das Fremdwort im Deutschen*, 2nd ed. Berlin: de Gruyter, 2012.

[16] T. Schmidt and K. Wörner, “EXMARaLDA: Creating, analysing and sharing spoken language corpora for pragmatic research,” *Pragmatics*, vol. 19, no. 4, pp. 565–582, 2009.

[17] T. Schmidt, K. Wörner, H. Hedeland, and T. Lehmberg, “New and future developments in EXMARaLDA,” in *Multilingual Resources and Multilingual Applications. Proceedings of GSCL Conference 2011 Hamburg*, T. Schmidt and K. Wörner, Eds., 2011.

[18] H. Frieg, “Sprachförderung im Regelunterricht der Grundschule: Eine Evaluation der Generativen Textproduktion,” Dissertation, Ruhr-Universität Bochum, 2014. [Online]. Available: http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/FriegHendrike/diss.pdf

[19] U. D. Reichel, “PermA and Balloon: Tools for string alignment and text processing,” in *INTERSPEECH*, Portland, Oregon, 2012.

[20] U. D. Reichel and T. Kisler, “Language-independent grapheme-phoneme conversion and word stress assignment as a web service,” in *Elektronische Sprachverarbeitung: Studientexte zur Sprachkommunikation 71*, R. Hoffmann, Ed. TUDpress, 2014, pp. 42–49.

[21] G. Thomé, *Orthographieerwerb: Qualitative Fehleranalysen zum Aufbau der orthographischen Kompetenz*. Frankfurt am Main: Peter Lang, 1999.

[22] S. Schroeder, K.-M. Würzner, J. Heister, A. Geyken, and R. Kliegl, “childLex: A lexical database of German read by children,” *Behavior Research Methods*, vol. 47, no. 4, pp. 1085–1094, 2015.

[23] J. Fay, *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben: Eine empirische Untersuchung in Klasse 1 bis 4*. Frankfurt a. M.: Peter Lang, 2010.