We present a new multi-layered annotation scheme for orthographic errors in freely written German texts produced by primary school children. The scheme is closely linked to the German graphematic system and defines categories for both general structural word properties and error-related properties. It can also be used to investigate properties of correctly-spelled words.

In a pilot study, we achieved high inter-annotator agreement by the child (Fliss’ κ = .80) for our error categories (3 annotators, 295 annotations of 49 different categories in 951 tokens). Our aims are to get detailed and graphematically valid error profiles of learners, study the relationship between orthography acquisition and the graphematic system and to investigate properties of words that children are confronted with.

Annotation Layers I: General Properties

We code different general properties of the original spelling produced by the child (orig) and/or the intended spelling (target):

- **characters** (original/target)
- **phonemes** (target)
- **graphemes** (target)
- **syllables** (target): stressed, unstressed, reduced
- **morphemes** (target): morpheme class
- **foreign_target**: target word = foreign word? true/false
- **exist_orig**: original spelling = existing word? true/false
- **plausible_orig**: original syllable adheres to graphotactics? true/false

Annotation Layers II: Error-Related Categories

We distinguish four classes of error categories which refer to Eisenberg’s (2006) principles:

- phoneme-grapheme correspondence (PG, 19 tags), syllable (SL, 32 tags), morphology (MO, 6 tags), and phenomena beyond word spelling (e.g. syntax-based) (SN, 6 tags) → ‘other’ (4 tags)
- there is a systematic distinction between ignoring a principle, its hypercorrection (graphematically plausible) and its overuse (graphematically implausible)
- errors are annotated at the character level, which makes it possible to analyze the surrounding context in detail
- possible error categories code which orthographic properties a word possesses; thereby we can examine which orthographic phenomena a child already masters and which phenomena occur in texts that children are confronted with.

Two further properties are coded for each error:

- **phon_orig_ok**: word is pronounced the same way with correct spelling (necessary, hypercorrection, redundant, n.a.)

Examples

(1) *schönen → spielen* (‘to play’)

*phon_orig_ok = true, morph_const = na*

learner used GPC-compliant spelling, ignoring the exceptional spelling of a particular phoneme combination

(2) *kam → kommt* (‘he comes’)

*phon_orig_ok = true, morph_const = neces*

learner ignored consonant doubling before other consonants

(3) *Schn → Sand* (‘sand’)

*S1<NN>_final_devoice*

*phon_orig_ok = false, morph_const = na*

learner marked a long vowel for a phonetically short vowel

(4) *Hund → Hund’s dog* (‘dog’)

*M0<NN>_final_devoice*

*phon_orig_ok = true, morph_const = neces*

learner ignored that final devoicing is not reflected in the spelling

(5) *rät → rät* (‘he guesses’)

*M0<NN>_hyp_final_devoice*

*phon_orig_ok = true, morph_const = neces*

learner incorrectly assumed final devoicing

(6) *Hund → Hund’s dog* (‘dog’)

*SN<NN>_low_up*

learner ignored capitalization

Outlook

- large-scale annotation of the corpus from Frieg (2014): around 2000 texts written by primary school children
- automation of the annotation