

Ronja Laarmann-Quante, Lukas Knichel, Stefanie Dipper, Carina Betken

laarmann-quante@linguistics.rub.de, lukas.knichel@rub.de, dipper@linguistics.rub.de, carina.betken@rub.de

Ruhr-University Bochum

Overview

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 (Fleiss' 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*

Theoretic Background: Eisenberg (2006)

The spelling of isolated words in German can be explained with three kinds of principles:

1. **grapheme-phoneme correspondences** *phonographic spellings*, e.g.
/b/ → , /v/ → <u>, /n/ → <n>, /t/ → <t> ⇒ <bunt> 'colorful'
2. **syllabic principles**, e.g.
consonant doubling <kommen> '(to) come', syllable-separating *h* <Ruhe> 'quietness'
3. **morphological principles** *morpheme constancy*, e.g.
[hʊnt] → <Hund> 'dog' because of <Hunde> 'dogs',
[kʊmst] → <kommst> '(you) come' because of <kommen>

Annotation Example in EXMARALDA¹

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
[tokens_orig]	fä	l			d		e	i	n	b	i	e	s		j	e	n	
[tokens_target]	fä	l			t		e	i	n	b	i	s	s		c	h	e	n
[foreign_target]	false																	
[exist_orig]	false																	
[characters_orig]	f	ä	l		d		e	i	n	b	i	e	s		j	e	n	
[characters_target]	f	ä	l		t		e	i	n	b	i	s	s		c	h	e	n
[phonemes_target]	f	E	l		t		?	a	I	n	b	I	s		C		@	n
[graphemes_target]	f	ä	l		t			e	i	n	b	i	s		s	ch	e	n
[syllables_target]																		
[syll_orig_plausible]	true																	
[morphemes_target]	NN				INFL					PTKVZ								
[error_cat[1]]					SL:Cdouble_beforeC									SL:rem_Vlong_short				
[phon_orig_ok[1]]					true									false				
[morph_const[1]]					neces									na				
[error_cat[2]]										MO:hyp_final_devoice								
[phon_orig_ok[2]]										true								
[morph_const[2]]										neces								
[error_cat[3]]																		
[phon_orig_ok[3]]																		
[morph_const[3]]																		

Annotation Layers II: Error-Related Categories

Well-established instruments of assessing spelling abilities in German such as the HSP (May, 2013), OLFA (Thomé and Thomé, 2004) or AFRA (Herné and Naumann, 2002) only partly classify errors along graphematic dimensions. Our error categorization is rather inspired by the approaches of Thelen (2010) and Fay (2010).

- 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**, 8 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 this error (*true, false, colloquial*)
morph_const: role of morpheme constancy for the correct spelling (*necessary, hypercorrection, redundant, n.a.*)

Examples

- (1) *schpielen → spielen '(to) play' **PG:literal**
phon_orig_ok = true, morph_const = na
learner used GPC-compliant spelling, ignoring the exceptional spelling of a particular phoneme combination
- (2) *kommt → kommt '(he) comes' **SL:Cdouble_beforeC**
phon_orig_ok = true, morph_const = neces
learner ignored consonant doubling before other consonants
- (3) *Sahnd → Sand 'sand' **SL:rem_Vlong_short**
phon_orig_ok = false, morph_const = na
learner marked a long vowel for a phonetically short vowel
- (4) *Hunt → Hund 'dog' **MO:final_devoice**
phon_orig_ok = true, morph_const = neces
learner ignored that final devoicing is not reflected in the spelling
- (5) *räd → rät 'he guesses' **MO:hyp_final_devoice**
phon_orig_ok = true, morph_const = neces
learner incorrectly assumed final devoicing
- (6) *hund → Hund 'dog' **SN:low_up**
learner ignored capitalization

Data Representation: LearnerXML

```
<?xml version="1.0" ?>
<tokens id="test">
  <token id="tok1" orig="fäld" target="fällt"
    foreign_target="false" exist_orig="false">
    <characters_orig>
      <char_o id="o1">f</char_o>
      <char_o id="o2">ä</char_o>
      <char_o id="o3">l</char_o>
      <char_o id="o4">d</char_o>
    </characters_orig>
    <characters_target>
      <char_t id="t1">f</char_t>
      <char_t id="t2">ä</char_t>
      <char_t id="t3">l</char_t>
      <char_t id="t4">t</char_t>
      <char_t id="t5"> </char_t>
    </characters_target>
    <characters_aligned>
      <char_a id="a1" o_range="o1" t_range="t1"/>
      <char_a id="a2" o_range="o2" t_range="t2"/>
      <char_a id="a3" o_range="o3" t_range="t3..t4"/>
      <char_a id="a4" o_range="o4" t_range="t5"/>
    </characters_aligned>
    <phonemes_target>
      <phon_t id="p1" t_range="t1">f</phon_t>
      <phon_t id="p2" t_range="t2">ä</phon_t>
      <phon_t id="p3" t_range="t3..t4">l</phon_t>
      <phon_t id="p4" t_range="t5">t</phon_t>
    </phonemes_target>
    <graphemes_target>
      <gra id="g1" range="t1"/>
      <gra id="g2" range="t2"/>
      <gra id="g3" range="t3"/>
      <gra id="g4" range="t4"/>
      <gra id="g5" range="t5"/>
    </graphemes_target>
    <syllables_target>
      <syll id="s1" range="t1..t5" type="stress" plausible_orig="true"/>
    </syllables_target>
    <morphemes_target>
      <mor id="m1" range="t1..t4" type="NN"/>
      <mor id="m2" range="t5..t5" type="INFL"/>
    </morphemes_target>
    <errors>
      <err range="a3" cat="SL:Cdouble_beforeC" phon_orig_ok="true"
        morph_const="neces"/>
      <err range="a4" cat="MO:hyp_final_devoice" phon_orig_ok="true"
        morph_const="neces"/>
    </errors>
  </token>
</tokens>
```

Outlook

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

¹www.exmaralda.org

Eisenberg, P. (2006). *Grundriss der deutschen Grammatik Band 1: Das Wort* (3rd ed.). Stuttgart: J.B. Metzler.

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

Frieg, H. (2014). *Sprachförderung im Regelunterricht der Grundschule: Eine Evaluation der Generativen Textproduktion* (Doctoral dissertation). Retrieved from <http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/FriegHendrike/diss.pdf>

Herné, K.-L., & Naumann, C. L. (2002). *Aachener Förderdiagnostische Rechtschreibfehler-Analyse* (4th ed.). Aachen: Alfa Zentaurus.

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

Thelen, T. (2010). *Automatische Analyse orthographischer Leistungen von Schreibanfängern* (Doctoral dissertation). Retrieved from https://repositorium.uos.de/bitstream/urn:nbn:de:gbv:700-201006096307/1/thesis_thelen.pdf

Thomé, G., & Thomé, D. (2004). *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.