

# Transcription Guidelines of Handwritten Content

Version 1.1 July 2023

## Manual for transcribing learner manuscripts for the use of handwriting recognition and text analysis

### Data Source:

For the transcription and examples within these guidelines, the content of “Textkorpus Scriptoria” of the research database “FD-LEX” by Michael Becker-Mrotzek and Joachim Grabowski of the University of Cologne is used<sup>1</sup>.

### **Note: Protection of Anonymity!**

To ensure the confidentiality of the writers, anonymization is required. Names are already blanked out in the images by FD-LEX. However, if names appear in the handwritten texts, they should be replaced in the transcript with # to ensure anonymity!

### Acknowledgements:

This work was partly conducted at CATALPA - Center of Advanced Technology for Assisted Learning and Predictive Analytics' of the FernUniversität in Hagen, Germany.

We would like to express our sincere gratitude to the annotators who dedicated their time and expertise to transcribing the challenging handwritten documents used in this study. Their contributions were essential to the success of this project. We would also like to thank them for their valuable feedback, suggestions, and work to continually improve the transcription guidelines throughout the course of this study.

We would like to extend our heartfelt appreciation to Sandra Tietjens and the FD-LEX team for generously providing us with the dataset and for their invaluable support and communication throughout our research process.

---

<sup>1</sup> Becker-Mrotzek, Michael/Grabowski, Joachim (2018). Textkorpus Scriptoria. In: Michael Becker-Mrotzek und Joachim Grabowski (Hrsg.). FD-LEX (Forschungsdatenbank Lernertexte). Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Verfügbar unter: <https://fd-lex.uni-koeln.de>. DOI: 10.18716/FD-LEX/861.

Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



### References:

These Guidelines were introduced at:

**Preserving the Authenticity of Handwritten Learner Language: Annotation Guidelines for Creating Transcripts Retaining Orthographic Features**

Gold, Christian and Laarmann-Quante, Ronja and Zesch, Torsten in *Proceedings of the 1st Computation and Written Language (CAWL) Workshop at ACL, 2023*

Applied Use-Case:

**Recognizing Learner Handwriting Retaining Orthographic Errors for Enabling Fine-Grained Error Feedback**

Gold, Christian and Laarmann-Quante, Ronja and Zesch, Torsten in *Proceedings of Innovative Use of NLP for Building Educational Applications (BEA) Workshop at ACL, 2023*

Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 1 General Aspects

The transcribed text must correspond to what the learner wrote. The spelling mistakes may not be corrected. In case of doubt, the following must be brought together: What did the learner write or want to write and what will a machine transcription read?

### 1.1 Format of the transcript

The transcript can be written in MS Excel for simplicity. However, the autocorrection must be switched off, as unintended corrections might be made! The structure of the Excel file should look like this:

# image name	line number	status	transcript	comment
Originaltext-IGS-9_2-027.png	1	ok	am 19.11.15 (15:46 Uhr)	wurde über die eigentliche erste Zeile g
Originaltext-IGS-9_2-027.png	2	dis	Der Fahrradfahrer hat die Straße von Rechts nach	grenzwertige Abstände zw. Fahrrad und
Originaltext-IGS-9_2-027.png	3	ok	Linke überqueren wollen.	
Originaltext-IGS-9_2-027.png	4	dis	aus der Sicht des Fahrradfahres (näher sich ein	Abstand (s. Kommentar oben)
Originaltext-IGS-9_2-027.png	5	dis	Aut von (links auf der rechten fahrbahn).	I oder L?
Originaltext-IGS-9_2-027.png	6	err	###	
Originaltext-IGS-9_2-027.png	7	dis	dio(ch) der Fahrer(fahrrad)fahrer ist trötzdem los	Abstand (s. Kommentar oben)
Originaltext-IGS-9_2-027.png	8	ok	geg(a)ngen, doch der fahrer im (R)oten a(u)t(?) ## (w)ar erschr(o)ken weil	
Originaltext-IGS-9_2-027.png	9	dis	(der fahrer(a)d fahrer(i)n(m) den (W)ege a(b)geschnitten h(a)t	im/in/ih?
Originaltext-IGS-9_2-027.png	10	ok	b(e)im versuch dem Fahrrad(fahrer (aus zu w)e)lichen	
Originaltext-IGS-9_2-027.png	11	ok	h(a)t der fahrrer(r) im Rot(e)n aut(o) ein entgegen	
Originaltext-IGS-9_2-027.png	12	ok	k(j)ommendes blaues a(u)t(F)röntal geramt.	Buchstabe o sieht jeweils aus wie a
Originaltext-IGS-9_2-028.png	1	ok	Ihr (flünde der Räid fahrfier ist (schuld	

Four of the five Columns must be filled out in each line, containing the *image name* of the transcription document, the current *line number*, the *status* and the *transcript* of the line. An additional *comment* column follows last. The transcription follows a line-wise alignment, meaning each text line in the handwritten image must be transcribed into a transcript text line.

The *status* column can be filled out with *ok*, *err* or *dis*.

- The status 'dis' should be set if the decision remains in doubt and the text should be discussed with a second annotator.
- The *err* (error) status must be set when the selected line may be ignored for further text processing. This can be the case when only crossed-out words are present in the line or an illustration was drawn.

A whole page can be excluded from the transcription in special cases, but this should be decided based on the goals of the project.

### Note: Final Sanity Check

A sanity check is mandatory. When the page has been fully transcribed, the transcription should be read once again completely, so that one's own errors can be eliminated. Furthermore, the number of text lines in the image must match the highest *line number* in the transcript.

### Transcription Guidelines of Handwritten Content

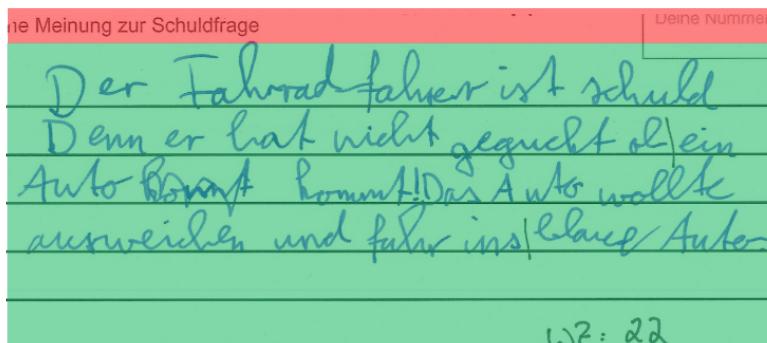
<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

## 1.2 General Rules

The goal is to convert what the learner wrote into a digital transcript. Therefore, anything that is not part of the handwritten text must be ignored. As example, the image below shows in green the handwritten text which must be transcribed, while the red part is part of the overall page layout.



### Transcription Rules:

To emphasize non-normative elements (e.g. unclear letters, special features, hyphens, inserts) they must be transcribed in curly brackets {}. In the case of special features, an additional character is added into the brackets. An overview list of characteristics of the transcriptions follows:

1. Permitted Quotation Marks are: , ‘ ’ “ ”
2. Lines, whose content should not be transcribed (e.g. crossed out lines), are given the status “error”.
3. Tally Marks are transcribed based on their angle followed by a '&': {/&}, {||&}, {\&}
4. Spacing {m }

The spaces between two letters is too big and it appears as a split of a word

5. Uncertainties {m}einung  
It's hard to tell if a letter is spelled uncommonly, incompletely, or with a spelling error.
6. Overlay {M+m}  
Letters were written on top of each other. The first letter marks the correct letter and the second letter marks the letter that is also readable. It can also be more than just one readable letter.
7. Insert {insert1} {insert1 Meinung}  
If the content is at a different position, {insert1} must be written at the place, where the insert should be placed and {insert1 Meinung} at the position, where the insert is located. The inserts need to be counted. If a character like a plus + or asterisk \* was used, it must be added to the transcription.
8. Separator {-}  
If a separation of one word into two was indicated by a vertical line.
9. Emojis {😊}

### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



Regardless of the large variety of emojis, only this one is used in the transcripts to represent an emoji.

10. Mirrored letter {M}einung and Character that was "invented" {M+}einung
11. Swap {swap1.2 Meinungen} {swap1.1 viele}  
"viele Meinungen" instead of "Meinungen viele"  
If indicated, two words shall be swapped to restore the correct order of words.

Usages of vertical bars:

1. Mirrored {M|}
2. Separator {:-}

#### Note: Further Examples

The following examples explain the previously mentioned aspects in more detail. Each example only focuses on one particular aspect. If there are further notable aspects within an example for which there are rules in these guidelines, these are ignored for the moment in order to focus on the one aspect that is currently being explained. In a real transcription, all transcription rules must be obeyed.

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch  
<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,  
<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## Table of Content

1	Target Hypotheses .....	3
1.1	Format of the transcript.....	3
1.2	General Rules .....	4
2	Detailed Rules .....	7
2.1	Layout .....	7
2.1.1	Overlay with layout of sheet.....	7
2.1.2	Empty line.....	7
2.1.3	Change of pen-color.....	8
2.1.4	Text spreading over several lines.....	8
2.1.5	Invalidation of lines.....	8
2.1.6	Underlines .....	Error! Bookmark not defined.
2.2	Punctuation Marks .....	10
2.2.1	Quotation Marks .....	10
2.2.2	Regular and special characters.....	10
2.2.3	Tally Marks.....	10
2.3	Spacing .....	12
2.3.1	Spacing between letters.....	12
2.3.2	Spacing between words.....	12
2.4	Spelling Errors and unclear situations.....	14
2.4.1	Spelling Errors.....	14
2.4.2	Unclear situations.....	14
2.4.3	Intentional overlays .....	16
2.5	Inserts .....	17
2.5.1	Direct inserts .....	17
2.5.2	Indirect inserts .....	18
2.6	Struckthrough.....	Error! Bookmark not defined.
2.7	Word Separation .....	22
2.8	Date and Time.....	22
2.9	Artifacts .....	23
2.10	Emoticons and drawings .....	23
2.10.1	Smileys .....	23
2.10.2	Emoticons .....	23
2.10.3	Drawings .....	24
2.11	Mirrors and Inventions of Letters.....	25
2.11.1	Mirrors.....	25
2.11.2	Additions and Inventions.....	25
2.12	Swap of words.....	27

## Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

## 2 Detailed Rules

### 2.1 Layout

#### 2.1.1 Overlay with layout of sheet

Problem:

Handwritten text was placed over some text that is printed on the sheet.

Handling:

If the handwritten text is overlayed with text of the sheet, it should be transcribed but flagged with an "err" status.

Example:



status:	content:
correct:	err      {*) hatte einen Pferdeschwanz

#### 2.1.2 Empty line

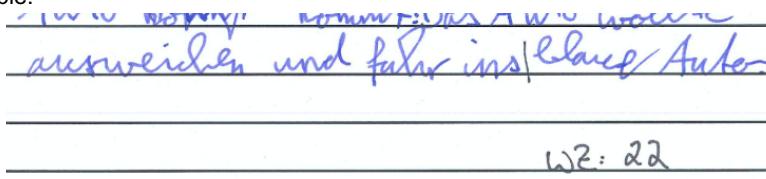
Problem:

An empty line is present and something was written by another author (e.g. teacher).

Handling:

The empty line is ignored, and the content is transcribed but the line status is set to "err". Although we want to focus on the learner content only, those elements are considered for a complete transcript as well.

Example:



status:	content:
correct:	ok      [...] blaue Auto.
	err      WZ: 22

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



### 2.1.3 Change of pen-color

Commented [CG1]: TODO

Problem:

The color of the pen changes within a paragraph.

Handling:

In each text-line it must be decided which pen was used for most words. The words written with the other pen must be placed within curly brackets ({}).

Example:

Am 21.11.2014, um 13:45 Uhr, ereignete

correct: Am 21.11.2014, um 13:45 Uhr, {ereignete}

### 2.1.4 Text spreading over several lines

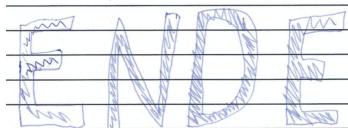
Problem:

A short text that would have fit in a single line was written over several lines.

Handling:

The content will still be written in one line.

Example:



	status:	content:
correct:	err	Ende
incorrect:	err	Ende
	err	Ende

Commented [CG2]: 'err' status

### 2.1.5 Invalidation of lines

Problem:

A whole line of the text is crossed out.

Handling:

The line needs to be tagged with an 'err' in the status and the transcription contains three hashes (###).

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



Example:

~~ICHTYON berne sind schuld.~~  
~~weil der~~ Der Fahrer des Roten

	status:	content:
correct:	err	###
	ok	##### #### ## Der Fahrer des Roten

Commented [CG3]: Linestatus: err

## 2.1.6 Underlining

Problem:

Words were underlined to add emphasis to them.

Handling:

Generally, the underlining are ignored.

Example:

Einbruch genau gesehen.

correct:	Einbruch genau gesehen.
----------	-------------------------

### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

## 2.2 Punctuation Marks

### 2.2.1 Quotation Marks

To reduce the amount of different characters, only these quotation marks are permitted in the transcript: ‘ ’ “ ”. These must be written into the transcript in the same way as they were by the writer.

### 2.2.2 Regular and special characters

Regular punctuation marks and special characters must be transcribed the way they were written by the learner.

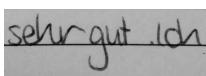
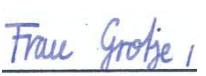
Problem:

Punctuation marks and spacing are not placed according to the norm.

Handling:

Punctuation marks such as the full stop (.), comma (,) and exclamation mark (!) are positioned according to the standard rules.

Examples:

		
correct:	sehr gut. Ich	Frau Grotje,
incorrect:	sehr gut .Ich	Frau Grotje ,

		
correct:	laut § 31!	um den ... {→}

### 2.2.3 Tally Marks

Problem:

To count the number of words in the text, tally marks were placed in the text.

Handling:

The character should be taken over and placed in curly brackets. The direction of the stroke is important, thus allowed characters are: /, |, \. The tally mark should be followed by an ampersand: {&}.

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



Example:

ganz | normal.

correct:

ganz {&} normal.

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 2.3 Spacing

### 2.3.1 Spacing between letters

Problem:

The spaces between two letters is too big and it appears as a split of a word.

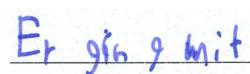
Handling:

Irregular spacing must be marked separately, if it creates a string of letters but that does not create a word, especially for cases where a pattern of the writer becomes evident. That might be the case when the writer repeatedly does not connect capital letters to the rest of the word. In the transcript, this should be represented by using curly braces, the disconnected letter, and a space, e.g. {A }. But if the distance between two letters divides one word into two independent words, the distance must be entered with a space.

Examples:



correct: ! Das {A }uto  
incorrect: !Das A uto; ! Das A uto; ! Das Auto



Er gin{ g} mit  
Er gin g mit



correct: eine {S }traße  
incorrect: eine S traße

### 2.3.2 Spacing between words

Problem:

The spacing between words or punctuation marks to words is too small or too big. It seems there is no spacing, or a word usually written together appears as two separate words. It may occur that a separation was indicated with the help of a separator.

Handling:

Spaces are added if it is recognizable that the writer wrote the words separately or intended to write them separately. Words that are spelled as one word but which are actually separate words should be written the way the writer wrote them. Words that should be written together but were not must be separated by a space. This is especially the case when the second word begins with a capital letter. If a separator was used it must be marked within curly brackets.

#### Transcription Guidelines of Handwritten Content

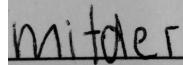
<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

Examples:

2.3.2a) Small distance between words



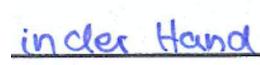
mitder

correct:

mitder

incorrect:

mit der

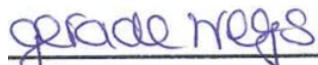


in der Hand

in der Hand

inder Hand

2.3.2b) Big distance within a word



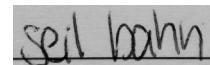
gerade wegs

correct:

gerade wegs

incorrect:

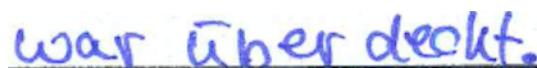
geradewegs



seil bahn

seil bahn

seilbahn



war über deckt.

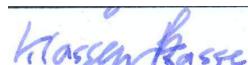
correct:

war über deckt.

incorrect:

war überdeckt.

2.3.2c) Connection



Klassenkasse

correct:

Klassenkasse

incorrect:

Klassen kasse

2.3.2d) Separation



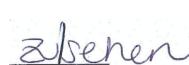
auf die

correct:

auf {-} die

incorrect:

aufdie



zu {-} sehen

zu {-} sehen

zu|sehen



## 2.4 Spelling Errors and Uncertainties

### 2.4.1 Spelling Errors

Problem:

A word contains a spelling error"

Handling:

Spelling errors must be transferred to the transcript without correcting them.

Examples:

<u>Das Rote Auto</u>	<u>kammen</u>
----------------------	---------------

correct: Das Rote Auto  
incorrect: Das rote Auto

kammen  
ka{m}en

### 2.4.2 Uncertainties

Problem:

It's hard to tell if a letter is spelled uncommonly, incompletely, or with a spelling error.

Handling:

Spelling errors must be transferred to the transcript without correcting them. Always assume the correct letter unless there is clear evidence for a wrong letter.

If a letter is unclear or not readable, it must be decided within the context: which word the learner wanted to write and which letter it should be in terms of spelling. An unclear letter must be emphasized by adding curly brackets.

Examples:

#### 2.4.2a) Unclear

<u>ausweichen</u>	<u>Blane</u>	<u>Achtfete</u>
-------------------	--------------	-----------------

correct: ausw{e}ichen      Bl{a}ue      A{c}htete  
incorrect: auswachen; ausweichen      Bloue      Aehtete

#### 2.4.2b) Missing or additional element of a Character

<u>fuh</u>	<u>de</u>
------------	-----------

correct: fuh{r}      {d}e  
incorrect: fulr; fulv      {d}ie; d{i}e; de

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



wolle

correct:  
incorrect:

woll{t}e  
wollle; wollte

Faschenlampe

{T}aschenlampe  
Faschenlampe, Taschenlampe

#### 2.4.2c) Possibility of various interpretations

hinein

correct:  
incorrect:

{h}inein  
lunein

schnell

schnell{e}  
schnell; schnelle

#### 2.4.2d) Unusual scaling

denke

correct:  
incorrect:

den{k}e  
den{K}e; denke

überqueren

über{q}ueren  
überQueren; überqueren

nach

correct:  
incorrect:

meiner {m}einung  
meiner meinung; meiner Meinung

nac{h}  
nacn

Moment

correct:  
incorrect:

{M}oment  
Moment; {m}oment

Kann

{k}ann  
Kann; kann

#### 2.4.2e) Overlay of characters

wenn

correct:  
incorrect:

{w}{e}nn  
wenn

#### 2.4.2f) Squeezed Letter

meineR

correct:  
incorrect:

me{h}{r}ere; me{h}{r}{e}re  
mehere; me{h+r+e}

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

### 2.4.3 Intentional overlays

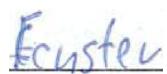
Problem:

Instead of striking out a letter, it was overwritten and allows several transcriptions to be correct.

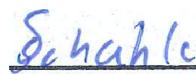
Handling:

All recognizable letters must be transcribed, starting with the correct letter in terms of spelling inside of curly brackets ({} ) followed by a plus sign (+) with the incorrect letter.

Examples:



correct: {F+f}enster  
incorrect: fenster



{S+s}chuhle  
Schuhle



correct: Au{ß+ss}erdem  
incorrect: Außerdem



Spa{ß+ss}  
Spa{ss+ß}



## 2.5 | Inserts

Commented [CG4]: Ab hier weiter!

### 2.5.1 Direct inserts

Problem:

The writer indicated that something should be inserted between two words or above a crossed out word. The word or letter that should be inserted was placed directly at the insert position.

Handling:

If the insert is positioned where these letters or few words should be inserted, they must be written within curly brackets added by an less than (e.g. {<and}). If a symbol, such as an asterisk, has been placed, it must be inserted too. Furthermore, an less than symbol (<) must be added to indicate a direct insertion. If many words should be inserted, {insert1} must be placed at the insert location and the words to be inserted must be transcribed at the end of the text line alike: {insert1 text}.

Examples:

#### 2.5.1a) Insertion of a letter

correct:	üb{<e}r	St{<r>aße	###{<her-}aus
incorrect:	übre	Straße	###her-aus

correct:	scho#{<n} selber	Spielgeräte geka#{<uft}
incorrect:	scho#n selber	Spielgeräte geka#uft

#### 2.5.1b) Insert of a word

correct:	des {<blauen} Autos	Der Fahrad {<zusammen} – fahrer
incorrect:	des blauen Autos	Der Fahrad-Fahrer;
incorrect:		Der Fahrradfahrer

correct:	Verkehrsunfall ### {<in der Nähe}
incorrect:	Verkehrsunfall ### in der Nähe

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



### 2.5.1c) Insert of a many words

Verletzten <sup>nun auch bekannt als</sup> Michael U. zu helfen.

correct: Verletzten {insert1} zu helfen. {insert1 nun auch bekannt als Michael U.}

incorrect: Verletzten nun auch bekannt als Michael U. zu helfen

### 2.5.2 Indirect inserts

#### Problem:

A word, several words, or even sentences should be inserted at a certain location but were written at the end of the page. This is typically indicated by a character like an asterisk (\*) or plus (+). When several indirect inserts were made the sign is usually followed by a number or the signs differ.

#### Handling:

At the location where the text should be inserted a {insert} must be set followed by numbering starting with 1 and the sign writer used: e.g. {insert1 \*}. Additionally, if the writer wrote numbered signs, the corresponding number must be added at the sign: {insert1 \*1}. Likewise, at the place where the content of the insert is written, curly bracket must be used: {insert1 \*1 that's what I wanted to add}.

#### Examples:

### 2.5.2a) Inserts

Commented [CG5]: Overkill - denke das muss man nicht unterteilen. Einfach 2 Beispiele

Gestern <sup>\*</sup> stand ich

correct: Gestern, {insert1 \*} stand ich  
incorrect: Gestern, stand ich

and

\* am 11.03.2015 um 12<sup>30</sup> Uhr,

correct: {insert1 \* am 11.03.2015 um 1230 Uhr,}  
incorrect: # am 11.03.2015 um 1230 Uhr,

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



Hand, und eine Taschenlampe. Erging zu einem  
an. Die Frau ~~ha~~ trug ein hellblaues Pullover, und

correct: Hand, {insert1 \*} und [...]  
[...]  
an. [...] Pullover, {insert2 \*} und  
and

\* hatte ein dunkel grüne ~~inf~~ ~~shirt~~ Pullover an und

~~(P)~~ hatte einen Pferdeschwanz

correct: {insert1 \* hatte ein du{<c>kel [...] und}  
{insert2 \* hatte einen Pferdeschwanz}

#### 2.5.2b) Numbered insert

*Auto +1 wollte*

correct: Auto {insert1 +1} wollte  
incorrect: Auto +1 wollte  
and

*x1 sah ihn und*

correct: {insert1 x1 sah ihn und}  
incorrect: x1 sah ihn und

#### 2.5.2c) Word in other line

*" mit + freundlichen Grüßen*

correct: {insert1 Mit}  
{insert1 freundlichen Grüßen}

*am, oberarm verletzt.  
linken*

correct: ### am {insert1} oberarm verletzt.  
{insert1 linken}

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



überqueren wollte. Der Fahrer im roten Auto, links  
ab um den Fahrradfahrer nicht zu treffen. **bog**

correct: überqueren [...] Auto {insert1} links  
ab [...] zu treffen. {insert1 bog}

dass ich ihr immer bestehe und sie immer zu mir  
kommen  
Wenn sie wirklich meine beste Freundin wäre, dann  
würde sie mit mir zusammen zum Lehrer gehen und

correct:  
dass [...] mir {insert1} {insert2}  
{insert1 kommen}  
Wenn [...] dann {insert2 kann.}  
würde [...] und

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 2.6 Struck-through Characters

Problem:

A letter or word was struckthrough.

Handling:

For each struck-through character, a hash (#) must be placed in the transcript. If a correction was placed, it must be inserted either to the left or right of the hash.

Examples:

correct:		
incorrect:	Blut	v erpetzen

special case/ exception:

correct:			
incorrect:	{<jedoch} ###	austo#ben	j{u}nge

correct:		
incorrect:	###	der ##

correct:		' von der Zeugin'
		rachförderung und Deutsch als Zweitsprache

correct: #### von der Zeugin

correct:	
	also {<##### kein} Veräter

correct:	
incorrect:	in ### {<den} gemütlichen Sofas

### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 2.7 Word Separation

Problem:

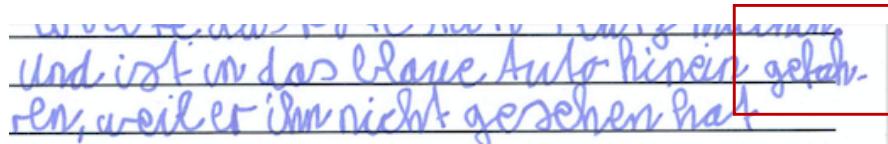
The word separation does not appear correctly (syllabification).

Handling:

The transcript must represent what is written. Therefore, the used separation of words must be transcribed.

Example:

It should be noted that no suitable example with incorrect syllable separation was found. However, this example shall indicate how to proceed.



und ist in das blaue Auto hinein gefahren, weil er ihn nicht gesehen hat

correct:      und ist in das blaue Auto hinein gefahren, weil er ihn nicht gesehen hat

## 2.8 Date and Time

Problem:

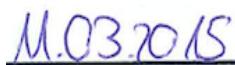
Date and time can be given in various forms.

Handling:

The date and time should be transcribed as the learner wrote them. Underlinings are ignored.

Examples:

### 2.8a) Date



11.01.2015

correct:

11.01.2015



13. November. 2014

13. November. 2014

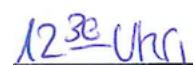
### 2.8b) Time



14:30h

correct:

14:30h



12:30 Uhr

1230 Uhr,

### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

## 2.9 Artifacts

Problem:

Letter artifacts may be present that do not form a valid letter.

Handling:

The artifact should not be interpreted as a letter. Instead, a hash (#) will be placed.

Examples:



correct:	frei # ist	Fundbüro # und
incorrect:	frei c ist	Fundbüro cund

## 2.10 Emoticons and Drawings

### 2.10.1 Smileys

Problem:

Smileys were written in form of horizontal compositions of characters.

Handling:

The used characters must be transcribed as they appear.

Examples:



correct:	:D	(-:-)
----------	----	-------

### 2.10.2 Emoticons

Problem:

Emoticons are included in the text in form of vertical small drawings.

Handling:

Any type of emoticon should be expressed by the Unicode "U+1F642" within curly brackets and will appear in the transcript as "{{:)}".

#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany

Examples:

correct:      {😊}      {😊}  
 incorrect:     😊      <3

but

*Freundin*  
 Freund{ɪ}n  
 Freund{😊}n

### 2.10.3 Drawings

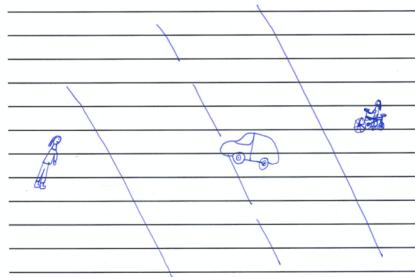
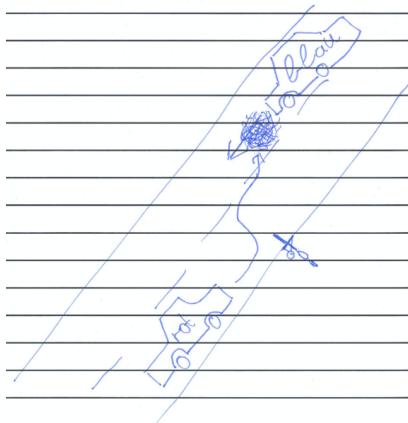
Problem:

Instead of a text, a drawing is found.

Handling:

For each line that includes parts of the drawing, put an empty line in the transcript and set the line status to "err". In addition, a comment with a hint must be written in the line where the image appears for the first time. If the picture finishes the written text, only the first line of the drawing must appear in the transcript (including error status and comment).

Examples:



#### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 2.11 Mirror-inverted Letters and Inventions of Letters

Problem:

A letter was written mirrored, is "invented" or added by other elements.

Handling:

The mirror-inverted letter must be followed by a vertical bar (|) and placed within curly brackets. If a new letter was "invented" or if additional artifacts were added to a letter, choose the letter that the writer wanted to write and add a plus sign (+). In both cases, curly brackets ({} ) must be placed around the transcript.

### 2.11.1 Mirrors

Examples:

correct:      fahrer {a|}us  
incorrect:    fahrer ous

but  
abge{b}ogen

a{b}ge{b}ogen  
a{a}ge{a}ogen

## 2.11.2 Additions and Inventions

Examples:

### 2.12.2a) Addition of Elements on Letters

correct:      {H+}err  
incorrect:    {H}err

be{w+}iesen  
be{w+}iesen  
beuwiesen

### 2.12.2c) Connection of Letters

correct:      gege{n+}spur  
incorrect:    geg{u+}spur

### Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany



## 2.12.2d) No Setting Down of the Pen

correct:  
incorrect:{F+}ahrer  
Fahrer{F+}ahrseite  
Fahrseite

## 2.12.2e) Uncommon Version of a Letter

correct:  
incorrect:{L+}eute  
{L}eute{D+}adurch  
Dadurch{k+}önnen  
können{f+}est  
fest



## 2.12 Swap of Words

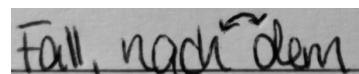
Problem:

Two words should be swapped. This was marked with arrows or a left right arrow ( $\leftrightarrow$ ).

Handling:

The previous word will be marked as the second word and the following word will be marked as the first word. This will be done with two separate curly brackets starting with the keyword: swap followed by an index number and the intended position of the word. For example: {swap1.2 text} and {swap1.1 text}

Example:



correct: Fall, {swap1.2 nach} {swap1.1 dem}  
incorrect: Fall, nach dem

Transcription Guidelines of Handwritten Content

<sup>1</sup>Christian Gold, <sup>2</sup>Ronja Laarmann-Quante, and <sup>1</sup>Torsten Zesch

<sup>1</sup>CATALPA, FernUniversität in Hagen, Germany,

<sup>2</sup>Ruhr University Bochum, Faculty of Philology, Department of Linguistics, Germany