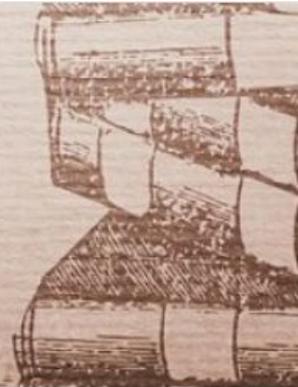




# OCR-BW

*Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen*



## **Stangenware oder Maßanzug? Automatische Texterkennung von Handschriften mit spezialisierten und generischen Modellen in Transkribus**

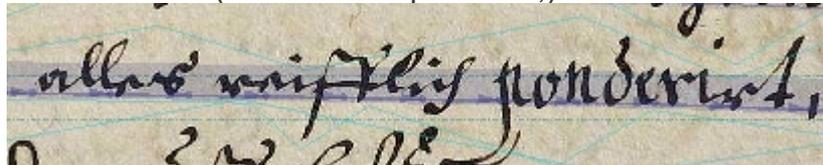
09.06.2021, Dorothee Huff



## Juristische Konsilien

- 269 Bände (1602-1883)
- bearbeitet wurden Bd. 84/13 (1659-1661) und 84/14 (1659-1661)
- Sprache: Deutsch, Latein
- Schrift: deutsche Kurrentschrift, lateinische Schreibschrift

(alles reifflich ponderirt,)





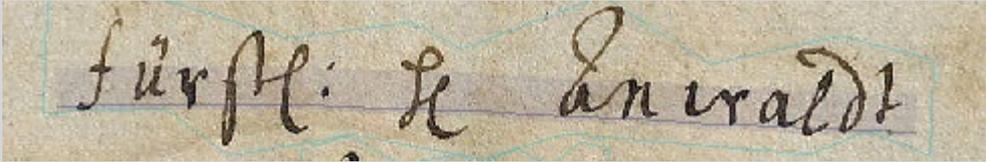
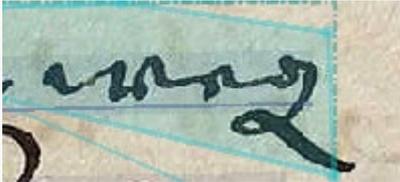
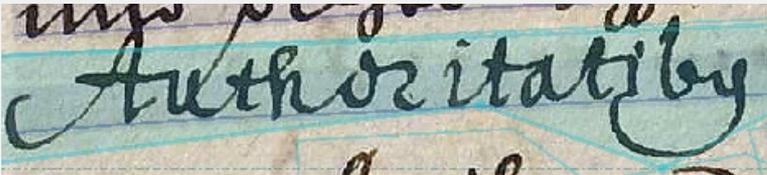
---

## Grundlegendes: Transkriptionsrichtlinien

- Stand: bisher keine Normierung bei der Transkription von Handschriften
- Maxime: möglichst dokumentnah transkribieren
- Orientierung an OCR-D Guidelines → Level 2
- Praktikabilität (nur Unicode kein PUA/MUFI)
- Beratung durch Fachvertreter



## Beispiele

	Fürstl: Hl Anwaldt
	Befond's
	weg_
	Autoritatib <sup>9</sup>



## Herangehensweise: Erstellung der GT

- Transkription auf Grundlage eines Public Models für Kurrentschrift

Documents **HTR Model Data**

All ▾ All engines ▾

1-9 / 9 1 1

Name	Size	Curator	HTR...	nrOfWor...	Date
>  Transkribus German Kurrent M2		guenter.hackl@...	32524	3209689	27.04.21
>  StAZH_RRB_German_Kurrent_XIX		tobias.hodel@...	30919	26026908	14.03.21
>  Transkribus German Kurrent M2		guenter	29820	3209690	28.01.21
>  Land registers (Verfachbücher) Tyrol, 1750-1800		p.andorfer@g...	28866	129152	21.12.20
>  German_Kurrent_17th-18th		alverman@uni-...	27457	1839841	10.11.20
>  VB_1750_1800_M1Base		p.andorfer@g...	26269	106794	24.09.20
>  German_Kurrent_XIX_pylaia		tobias.hodel@...	26068	5100439	11.09.20
>  German_Kurrent_XIX_M2		tobias.hodel@...	23847	5100439	15.05.20
>  German_Kurrent_XVI-XVIII_M1		tobias.hodel@...	19584	1579208	16.12.19
>  Acta 17 (extended)		alverman@uni-...	32677	1632534	04.05.21
>  Acta_17 PyLaia		alverman@uni-...	27337	594628	04.11.20
>  Acta_17 HTR+		alverman@uni-...	26284	594628	25.09.20



# Modellauswahl mit einem Sample-Set

Compare Samples

Documents Samples
— □ ×

Sample Collection

- > 485695 - Sample\_UAT\_084\_14 (14)
- > 485650 - Sample\_UAT\_084\_14\_Sa

Reference : GT

Select hypothesis by toolname :

PyLaia decoding 0.3.5 - Model: 27337

→ Compute

Upper bound : 9.405%

Lower bound : 4.992%

Mean : 7.198%

With the probability of 95% the CER for the entire document will be in the interval [4.992% 9.405%] with the mean : 7.198%

By taking 4 times the number of lines

Results

Created	Status
27.05.21 06:47:18	Completed
27.05.21 06:44:37	Completed
25.05.21 09:56:20	Completed
17.02.21 09:17:10	Completed
25.01.21 13:59:39	Completed
19.11.20 15:15:23	Completed
19.11.20 15:13:49	Completed
19.11.20 15:10:46	Completed
19.11.20 15:08:56	Completed
19.10.20 09:01:47	Completed
19.10.20 09:00:40	Completed



# Modellentwicklung: 1. Modell (26 S. GT)

HTR 'Juristische\_Konsilien\_Tuebingen\_M1a'

Name: Juristische\_Konsilien\_Tuebingen\_M1a Language: German, Latin

Description: First model based on UAT 84/13, one scribe, p. 25-39, 102-112 (=26 p.), automatic selection of validation set 10% (p. 34-35, 107)

Parameters: Nr. of Epochs 200

Document Type: Handwritten Show advanced parameters...

Nr. of Words: 4029 Nr. of Lines: 610

Save Show Train Set Show Validation Set Show Characters

**Learning Curve**

CER on Train Set: 0.12% CER on Validation Set: 6.21%

HTR 'Juristische\_Konsilien\_Tuebingen\_M1a\_basemodel'

Name: Juristische\_Konsilien\_Tuebingen\_M1a\_basem Language: German, Latin

Description: First model based on UAT 84/13, one scribe, p. 25-39, 102-112 (=26 p.), automatic selection of validation set 10% (p. 34-35, 107),

Parameters: Nr. of Epochs 200  
HTR Base Model ID 26284  
HTR Base Model Name Acta\_17\_HTR+

Document Type: Handwritten Show advanced parameters...

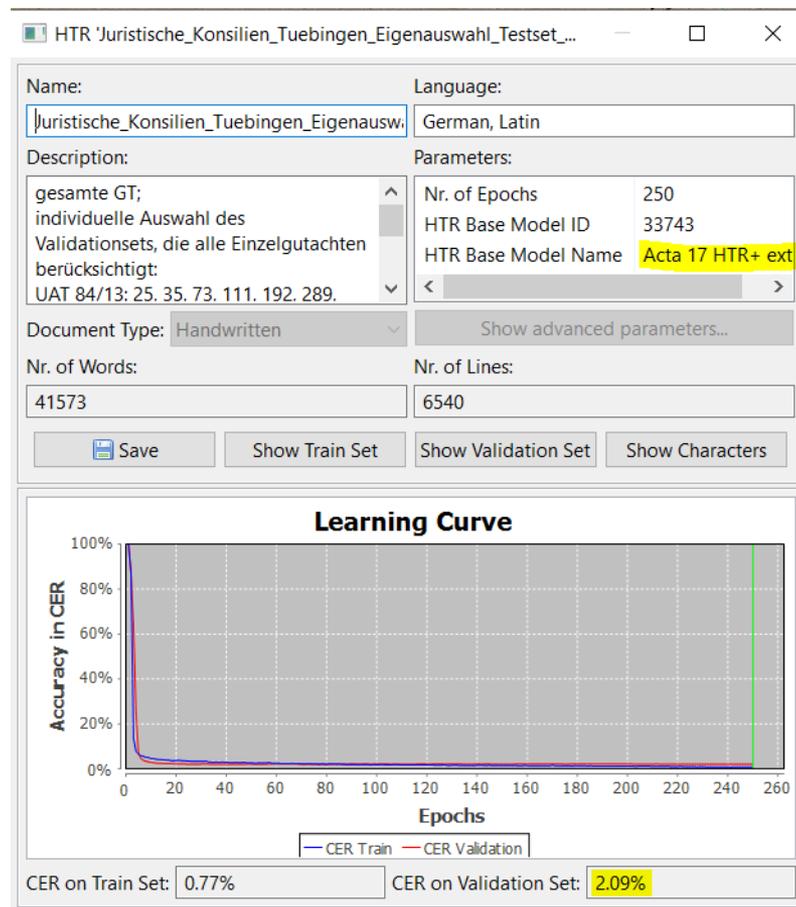
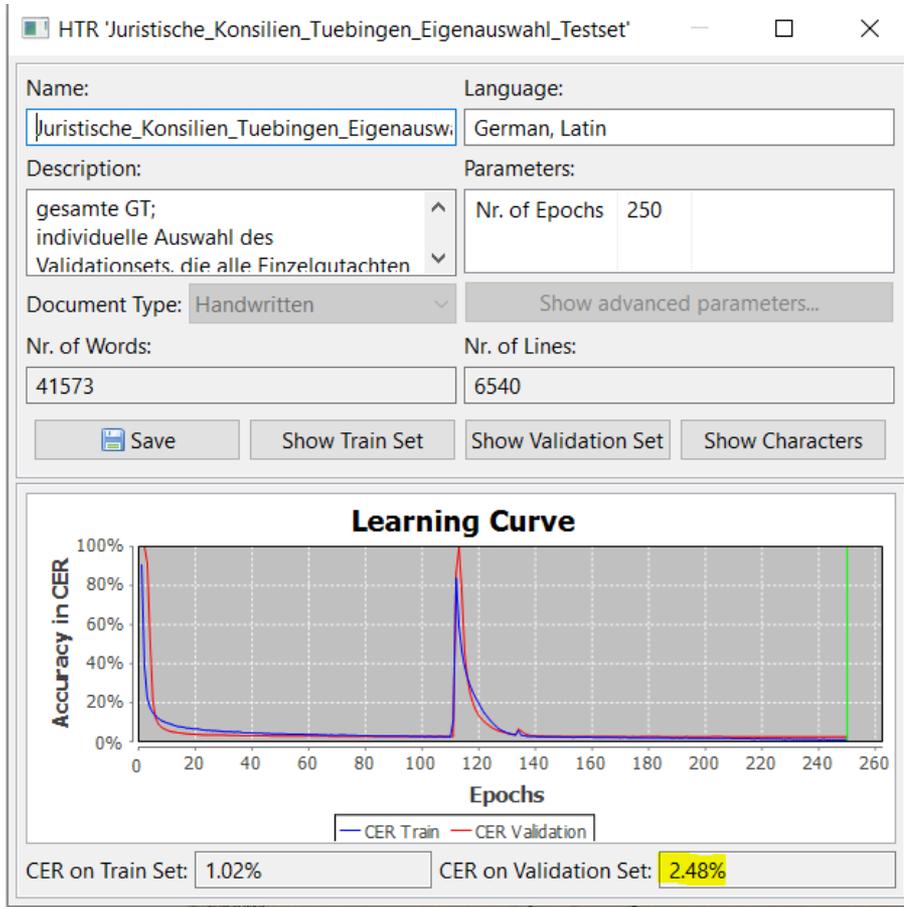
Nr. of Words: 4029 Nr. of Lines: 610

Save Show Train Set Show Validation Set Show Characters

**Learning Curve**

CER on Train Set: 0.07% CER on Validation Set: 4.72%

# Modellentwicklung: Endergebnis (223 S. GT)





# Modell „Juristische\_Konsilien\_Tuebingen\_Eigenauswahl\_Testset\_Basemodel2“

- durchschnittliche CER mit LM: 1,95%

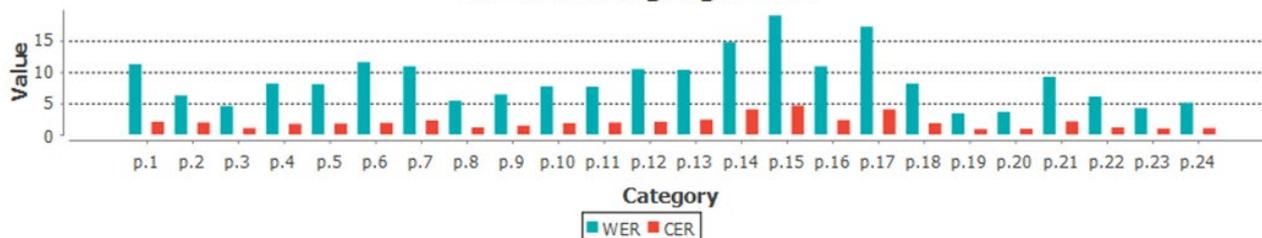
Advanced Statistics

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Overall	8.59 %	1.95 %	86.08 %	96.88 %	0.931	0.934	0.931

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Page 1	11.22 %	2.06 %	82.81 %	97.0 %	0.9347	0.9073	0.92079207...
Page 2	6.23 %	1.92 %	90.75 %	96.97 %	0.96829999...	0.9506	0.95937090...
Page 3	4.59 %	1.01 %	92.89 %	98.39 %	0.9769	0.96790000...	0.97235023...
Page 4	8.17 %	1.68 %	86.81 %	97.57 %	0.9439	0.9346	0.93924466...
Page 5	8.03 %	1.77 %	87.59 %	97.3 %	0.93599999...	0.93980000...	0.93787575...

Compare Text Versions for Page ..

**Error Rate Chart | Ref: GT | Hyp: CITlab HTR:  
Juristische\_Konsilien\_Tuebingen\_Eigenauswahl\_Testset\_Basemodel2 | Dictionary:  
trainDataLanguageModel**





## 0,88% CER

Cum tamen Reus, maxime in  
 gorie, perspicue & rotunde  
 nis membra respondere debeat  
 fieri possit, quidnam ab actore  
 tenus atq; in quibus membris on  
 gere debeat; adeo, ut, si reus  
 spondere nolit, & in hac contu  
 ad veritatem eruendam tortu  
 Carpz. prax. crim. part. 3. q. 1  
 Ubicumq; autem clara & rotunda  
 ibi non sufficiunt talia verba d  
 in specie de litis contestatione, q

459.

Cum tamen Reus, **maximè maxime** in criminalibus, cate,  
 gorie, **perspicie perspicue** & rotunde ad singula accusatio,  
 nis membra respondere debeat, ita ut Judex certus  
 fieri possit, quidnam ab actore probandum sit, & qua,  
 tenus atq; in quibus membris onus probandi ei injun,  
 gere debeat; adeo, ut, si reus categorice & **clare clarè** re,,  
 spondere nolit, & in hac contumacia persistat, ipse  
 ad veritatem eruendam torturæ subjici **posfit posfit.**

Carpz. prax. crim. part. 3. q. 106. n. 72. & seqq.

Ubicunq; autem clara & rotunda **requiritur, responsio,, requiritur responsio,**

ibi non sufficiunt talia verba dubia & ambigua; quod  
 in specie de litis contestatione, quod sc: illa per talia  
 verba ambigua fieri nequeat, tradit

Wesemb. in Cod. tit. de litis contestat. n. 6. in f. per  
 l. si **debitor. debitor,** ff. de heredit. petit.

Unde lis non censetur contestata, si reus respon,,  
 derit: credo **ine-me** non teneri; vel credo insti,,  
 tutam actionem non esse veram.

Accurf. in l. un. C. de lit. contestat.

**Quod maximè Quod maxime** in criminalibus procedit, et in  
 illis talem responsionem litis contestationem non  
 inducere, docet

Obrecht. tr. de litis contest. cap. 7. n. 6. & 7. ubi  
 subjicit in criminalibus non sufficere credulita,,  
 tis responsionem, sed veritatis exigi. Carpzov.

in Jurisp. forens. part. 1. const. 16. def. 6. Umm. **ad-ad**  
 procesf. disp. 12. th. 4. n. 14. Brunnemann. de procesf.  
 cap. 14. n. 15.

Jdem tradunt Dd. si reus dicat: neq; nego, neq;



4,64%

112.  
übel tractirt und gehalten, d3 nicht  
ein geringer argwohn sich ereigen  
wollen, als ob folches vier Jähriges  
Knäblein, Knäblein, durch übele fractation-tractation  
und haltung, sein leben-Leben eingebueft  
hette, und verstorben were, were.  
Wir Wie dann hiervon gleichsam fama publica publica,  
und ein allgemein geschrey entständen-entstanden  
ist, auch die verhörte zeug-zeug uno ore  
bekundtschafften thuen. Gestaldt-dae Gestaldtame, in  
specie specie, Michel Klenckh sagt: d3 Kind  
puta (puta d3 verstorbene Knäblein) habe nit  
adlein-allein  
nicht  
igung „gnug zu bfen-Eben bekoimen mögen,  
fond'n feye auch über die maßen hart  
Ghalten, Gehalten, und immer-immer zu von des Jacob  
Brunen weib gehaft, geschlagen, geschleiffet,  
und im-un Christlich tractirt worden. Mit  
welchem zeug auf-auch  
Durchhaß-durchhaß  
übereinstimmen, Michel  
dung-Jung der 13. wie auch Melchior Jung  
deri4-der 14. Item Hanß Georg Schon-Schön der 18  
Zeug, welche alle, nebenland'n mehr,, neben and'n mehr,  
einhelliglich, von folcher des Kinds  
ußgestandtener böfen tractation  
attestiren thuen. Margaretha  
hauß wollen Schneids-Tochter Hanß Wolfen Schneid's tochter die  
25 gengin-zeugin zeigt an, Mann habe d3 Kind  
aufgehungert, aufgehüngert, biß weihn ihme Brot-brot  
eingebrocket, und darüber nur  
bloß waßer geschittet, welches d3 Kind  
nicht genüßen Können, fondern wid'  
von sich geben müßen. dahero vonhard-müßen. Dahero Leonhard  
se vor-Ruopp vermeinen undt mittmaßen-muttmaßen  
thutt, d3 Kindt feye im Leiblein-leiblein undt  
mäglin, daureh-durch gegebene harte Steiß-Speiß  
undt tränkch, tranckh, ganzt verderbt worden,

112.  
übel tractirt und gehalten, d3 nicht  
ein geringer argwohn sich ereigen  
wollen, als ob folches vier Jähriges  
Knäblein, Knäblein, durch übele fractation-tractation  
und haltung, sein leben-Leben eingebueft  
hette, und verstorben were, were.  
Wir Wie dann hiervon gleichsam fama publica publica,  
und ein allgemein geschrey entständen-entstanden  
ist, auch die verhörte zeug-zeug uno ore  
bekundtschafften thuen. Gestaldt-dae Gestaldtame, in  
specie specie, Michel Klenckh sagt: d3 Kind  
puta (puta d3 verstorbene Knäblein) habe nit  
adlein-allein  
nicht  
igung „gnug zu bfen-Eben bekoimen mögen,  
fond'n feye auch über die maßen hart  
Ghalten, Gehalten, und immer-immer zu von des Jacob  
Brunen weib gehaft, geschlagen, geschleiffet,  
und im-un Christlich tractirt worden. Mit  
welchem zeug auf-auch  
Durchhaß-durchhaß  
übereinstimmen, Michel  
dung-Jung der 13. wie auch Melchior Jung  
deri4-der 14. Item Hanß Georg Schon-Schön der 18  
Zeug, welche alle, nebenland'n mehr,, neben and'n mehr,  
einhelliglich, von folcher des Kinds  
ußgestandtener böfen tractation  
attestiren thuen. Margaretha  
hauß wollen Schneids-Tochter Hanß Wolfen Schneid's tochter die  
25 gengin-zeugin zeigt an, Mann habe d3 Kind  
aufgehungert, aufgehüngert, biß weihn ihme Brot-brot  
eingebrocket, und darüber nur  
bloß waßer geschittet, welches d3 Kind  
nicht genüßen Können, fondern wid'  
von sich geben müßen. dahero vonhard-müßen. Dahero Leonhard  
se vor-Ruopp vermeinen undt mittmaßen-muttmaßen  
thutt, d3 Kindt feye im Leiblein-leiblein undt  
mäglin, daureh-durch gegebene harte Steiß-Speiß  
undt tränkch, tranckh, ganzt verderbt worden,

## Modell „Acta\_17 HTR+“

- durchschnittliche CER mit LM: 4,61%

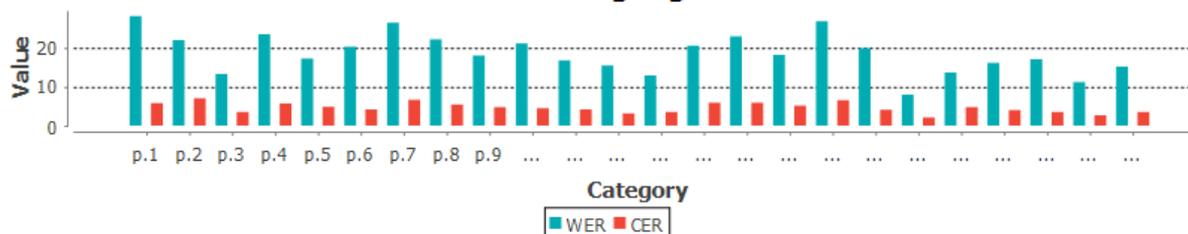
Advanced Statistics

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Overall	18.35 %	4.61 %	71.62 %	92.62 %	0.846	0.849	0.846

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Page 1	28.16 %	5.85 %	60.08 %	90.57 %	0.76379999...	0.73790000...	0.75061728...
Page 2	22.08 %	7.06 %	67.11 %	88.43 %	0.8249	0.8078	0.81627296...
Page 3	13.3 %	3.54 %	78.1 %	93.91 %	0.892	0.87159999...	0.88167053...
Page 4	23.53 %	5.75 %	63.98 %	90.62 %	0.8092	0.80390000...	0.80655737...
Page 5	17.27 %	4.84 %	74.48 %	92.21 %	0.856	0.85939999...	0.85771543...

Compare Text Versions for Page ..

**Error Rate Chart | Ref: GT | Hyp: CITlab HTR: Acta\_17 HTR+ | Dictionary: trainDataLanguageModel**





## 2,13%

Cum tamen Reus, maxime in  
gorice, perspicue & rotunde  
nis membra respondere debeat  
fieri possit, quidnam ab actore  
tenus atq; in quibus membris on  
gere debeat; adeo, ut, si reus  
spondere nolit, & in hac contu  
ad veritatem eruendam tortu  
Carpz. prax. crim. part. 3. q.  
Ubicunq; autem clara & rotunda  
ibi non sufficiunt talia verba  
in specie de litis contestatione,

45 D. 459.

Cum tamen Reus Reus, maxime in criminalibus, cate  
gorice, perspicue perspicue & rotunde ad singula accusatio  
nis membra respondere debeat, ita ut Judex certus  
fieri possit, quidnam quidnam ab actore probandum sit, & qua  
tenus atque in quibus membris onus probandi et ei injun  
gere debeat, debeat; adeo, ut, si reus categorice & clare re  
spondere nolit, & in hac contumacia persistat, ipse  
ad veritatem eruendam torturae subijci possit possit.

Carpz. prax. crim. part. 3. q. 106. n. 22. et 72. & seqq.

Ubicunq; autem clara & rotunda requiritur responsio,  
ibi non sufficiunt talia verba dubia & ambigua, ambigua; quod  
in specie de litis contestatione, quod se. sc. illa per talia  
verba ambigua fieri nequeat, tradit

Wesemb. in Cod. tit. de litis contestat. n. 6. in f. per  
I. l. si debitor. debitor, ff. de heredit. petit.

Unde lis non censetur contestata, si reus respon  
derit. Credo derit: credo me non teneri teneri; vel credo insti  
tutam actionem non esse veram veram.

Accur. Accurs. in I. l. un. C. de lit. contestat.

Quod maxime in criminalibus procedit, et in  
illis talem responsionem litis contestationem non  
inducere, docet

Obrecht. tr. de litis contest. cap. 7. n. 6. et 7. ubi

subijcit in criminalibus non sufficere credulita  
tis responsionem, sed veritatis exigi exigi. Carpzov.

in Jurisp. forens. part. 1. const. 6. 16. def. 6. Umm. ad  
process. disp. 12. th. 4. n. 14. Brunemann. de process.  
cap. 14. n. 15.

Idem Idem tradunt Dd. si reus dicat. dicat: neque nego, nequi neque



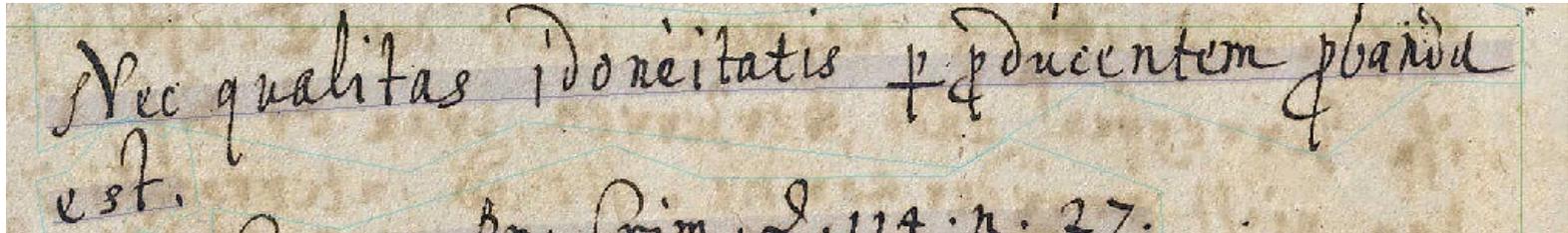
5,90%

112.  
 über tractirt und gefaltet, es nicht  
 ein geringen argwohn sich ereigen  
 wollen, als ob solches vier Jähriges  
 Knäblein, durch übele tractation  
 und faltung, sein Leben eingebueßt  
 hette, und Max Jordan argwoh,  
 wie dem Jacob gleichsam fama publica  
 und ein allgemein geschrey entstanden  
 ist, auch die verhörte Zeugen zeugen uno ore  
 bekundtschafften thuen. Gestalt same, Gestaltsame, in  
 specie specie, Michel Kleuth-Klenckh sagt: das Kindt  
 puta (puta das verstorbene Knäblein) habe nit  
 allein  
 nicht  
 genug zu Eßen bekommen bekommen mögen,  
 sondern wege seye auch über die maßen hart hart  
 gehalten, und immer zu von des Jacob  
 Brunen weib gehaft, gehast, geschlagen, geschleiffet,  
 und im un Christlich tractirt worden, worden. Mit  
 welchen Zeugen welchem zeugen auch  
 Durchaus-durchauß  
 übereinstimmen, übereinstimmen, Michel  
 Jung der 13. Wie wie auch Melchior Jung  
 den der 14. Item Item Hanß Georg Schön-Schön der 18  
 Zeuge, Zeug, welche alle, neben andern mehr= andern mehr,  
 einhelliglich, von solcher des Kinds  
 außgestandener außgestandener bösen tractation  
 attestiren thuen. Margaretha  
 hauß wolten-Schneids-Hanß Wolfen Schneiders tochter die  
 25 Zeugin zeugin zeigt an, Mann habe das Kin-Kind  
 außgehüngert, außgehüngert, biß weiln ihme Grot-brot  
 eingebrochet, eingebrochet, und darüber nur  
 bloß waßer geschittet, welches das Kind  
 nicht genußen können, genußen Können, sondern wider  
 von sich geben mußten. Dahero vorhand-Leonhard  
 Quorg-Ruopp vermeinen undt mittmaßen muttmaßen  
 thut, thut, das Kindt stege-seye im leiblein undt  
 mäglin, durch gegebene harte Speiß  
 undt frauen, gantz Verderbt-tranckh, gantz verderbt worden,

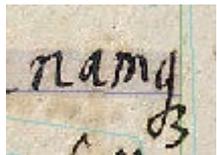
112.  
 übel tractirt und gehalten. Gehalten, das nicht  
 ein geringen-gering argwohn sich ereigen  
 wollen, als ob solches vier Jähriges  
 venablen-Knäblein, durch übele tractation  
 und haltung, kein leben sein Leben eingebueßt  
 hette, und verstorben were etc. were.  
 Wie dann hier von-hiervon gleichsam fama publica-publica,  
 und ein allgemein geschrey entstanden  
 ist, auch die verhörte Zeugen zeugen uno ore  
 bekundtschafften thuen. Gestalt same, Gestaltsame, in  
 specie-specie, Michel Kleuth-Klenckh sagt: das Kindt  
 puta-(puta das verstorbene Knäblein) habe nit  
 allein  
 nicht  
 genug zu Eßen bekommen bekommen mögen,  
 sondern wege-seye auch über die maßen hart hart  
 gehalten, und immer zu von des Jacob  
 Brunen weib gehaft, gehast, geschlagen, geschleiffet,  
 und im un Christlich tractirt worden, worden. Mit  
 welchen Zeugen welchem zeugen auch  
 Durchaus-durchauß  
 übereinstimmen, übereinstimmen, Michel  
 Jung der 13. Wie wie auch Melchior Jung  
 den der 14. Item Item Hanß Georg Schön-Schön der 18  
 Zeuge, Zeug, welche alle, neben andern mehr= andern mehr,  
 einhelliglich, von solcher des Kinds  
 außgestandener außgestandener bösen tractation  
 attestiren thuen. Margaretha  
 hauß wolten-Schneids-Hanß Wolfen Schneiders tochter die  
 25 Zeugin zeugin zeigt an, Mann habe das Kin-Kind  
 außgehüngert, außgehüngert, biß weiln ihme Grot-brot  
 eingebrochet, eingebrochet, und darüber nur  
 bloß waßer geschittet, welches das Kind  
 nicht genußen können, genußen Können, sondern wider  
 von sich geben mußten. Dahero vorhand-Leonhard  
 Quorg-Ruopp vermeinen undt mittmaßen muttmaßen  
 thut, thut, das Kindt stege-seye im leiblein undt  
 mäglin, durch gegebene harte Speiß  
 undt frauen, gantz Verderbt-tranckh, gantz verderbt worden,



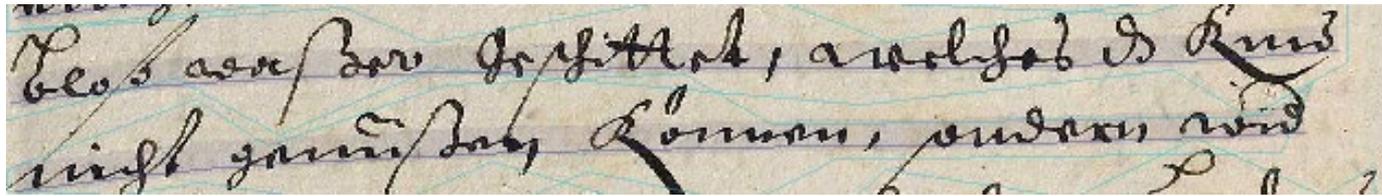
## Auswirkung der Transkriptionsrichtlinien



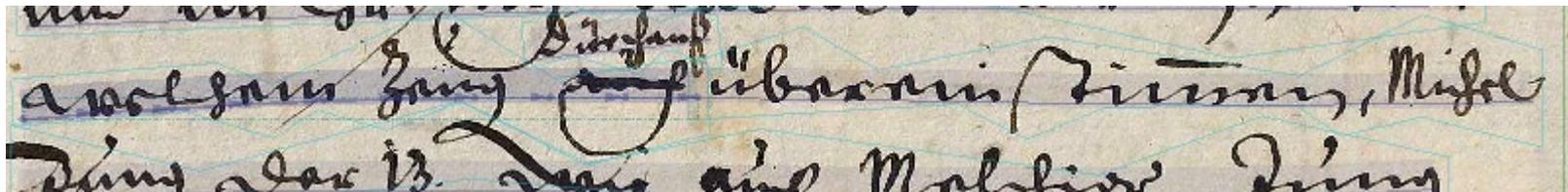
- **Jur\_Kons\_Tue:** Nec qualitas idoneitatis p pducentem pbandu
- **Acta\_17:** Nec qualitas idoneitatis per producentem pbanda



- **Jur\_Kons\_Tue:** namq<sup>3</sup>
- **Acta\_17:** namque



- **Jur\_Kons\_Tue:** bloß waſſer geſchittet, welches **d3** Kind nicht genüßen können, ſondern wid
- **Acta\_17:** bloß waſſer geſchittet, welches **das** Kind nicht genußen können, ſondern wid**er**



- **Jur\_Kons\_Tue:** welchem zeug **u** auf Durchanß **übereinſtimmen**, Michel
- **Acta\_17:** welchen Zeugen **en** auch Durchaus übereinstim**en**, Michel



---

## Fazit

- Zwar führt ein auf ein Textkorpus spezialisiertes Modell aktuell noch zu spürbar besseren Ergebnissen, jedoch können gerade bei Dokumenten, die relativ standardisiert sind auch mit generischen Modellen z.B. für Kurrentschriften bereits sehr gute Ergebnisse erzielt werden.
- Davon abgesehen können generische Modelle die Erstellung von Ground Truth erheblich erleichtern und zudem bei einer Nutzung als Basismodell trotz unterschiedlicher Transkriptionsrichtlinien ein spezialisiertes Modell noch einmal verbessern.



# Danke.

Kontakt: Dorothee Huff

**Universitätsbibliothek Tübingen**  
Wilhelmstraße 32, 72074 Tübingen  
Telefon: +49 7071 29-72852  
[dorothee.huff@uni-tuebingen.de](mailto:dorothee.huff@uni-tuebingen.de)