

CDA-Compliant Section Annotation of German-Language Discharge Summaries:

Guideline Development, Annotation Campaign, Section Classification

Christina Lohr,¹ Stephanie Luther,¹ Franz Matthies,¹ Luise Modersohn,¹
Danny Ammon,² Kutaiba Saleh,² Andreas G. Henkel,² Michael Kiehntopf,³ Udo Hahn ¹



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

¹ Jena University Language & Information Engineering (JULIE) Lab,
Friedrich-Schiller-Universität Jena



² Data Integration Center, IT Business Division, Jena University Hospital

³ Department of Clinical Chemistry and Laboratory Diagnostics and Integrated Biobank Jena (IBBJ),
Jena University Hospital

Discharge Summary – Example

PHYSICIAN HOSPITAL DISCHARGE SUMMARY

Provider: Ken Cure, MD

Patient: Patient H Sample **Provider's Pt ID:** 6910828 **Sex:** Female

Attachment Control Number: XA728302

HOSPITAL DISCHARGE DX

- ☐ 174.8 Malignant neoplasm of female breast: Other specified sites of female breast
- ☐ 163.8 Other specified sites of pleura.

HOSPITAL DISCHARGE PROCEDURES

1. 32650 Thoracoscopy with chest tube placement and pleurodesis.

HISTORY OF PRESENT ILLNESS

The patient is a very pleasant, 70-year-old female with a history of breast cancer that was originally diagnosed in the early 70's. At that time she had a radical mastectomy with postoperative radiotherapy. In the mid 70's she developed a chest wall recurrence and was treated with further radiation therapy. She then went without evidence of disease for many years until the late 80's when she developed bone metastases with involvement of her sacroiliac joint, right trochanter, and left sacral area. She was started on Tamoxifen at that point in time and has done well until recently when she developed shortness of breath and was found to have a larger pleural effusion. This has been tapped on two occasions and has rapidly reaccumulated so she was admitted at this time for thoracoscopy with pleurodesis. Of note, her CA15-3 was 44 in the mid 90's and recently was found to be 600.

HOSPITAL DISCHARGE PHYSICAL FINDINGS

Physical examination at the time of admission revealed a thin, pleasant female in mild respiratory distress. She had

Preamble

Diagnoses

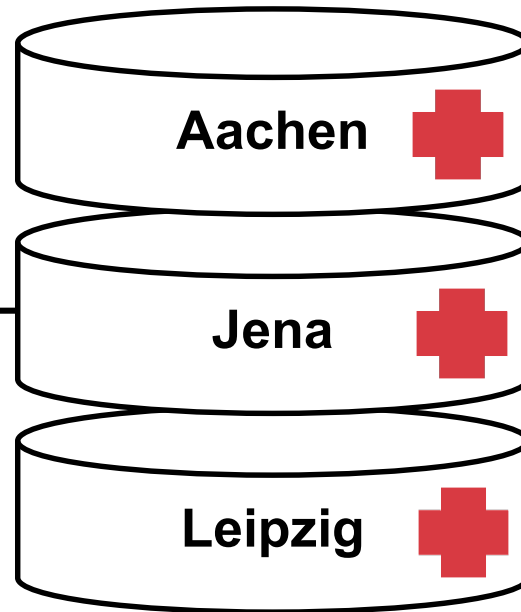
Procedures

Anamnesis

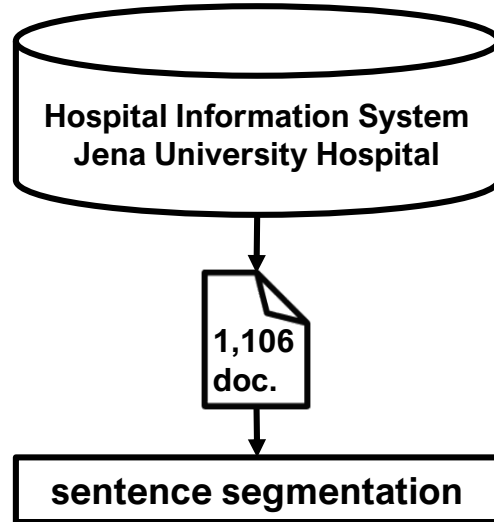
Diagnostics

- **German Clinical Reference Text Corpus**

- 1106 documents
 - discharge summaries
 - transfer letters
- 1.2M tokens
- 170K sentences

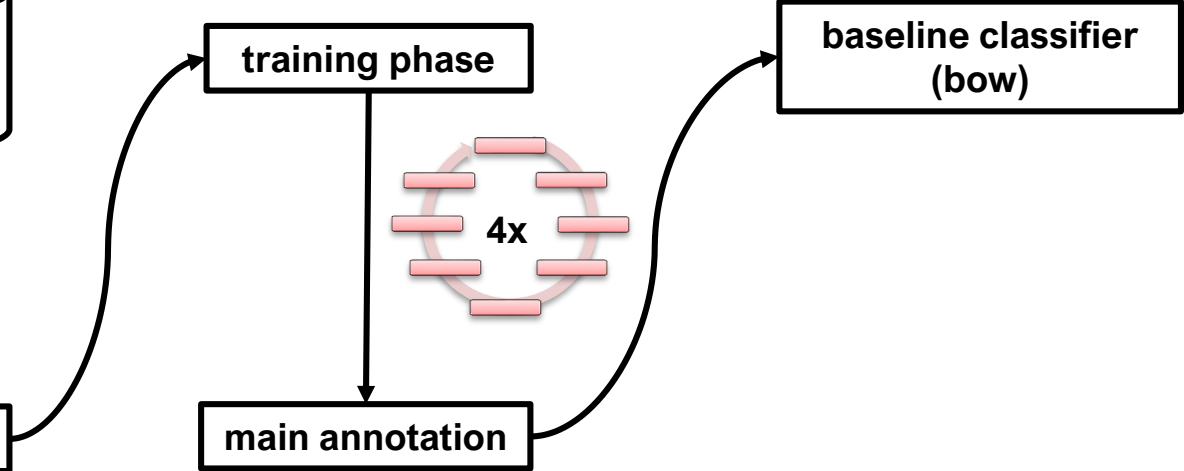


extraction



annotation

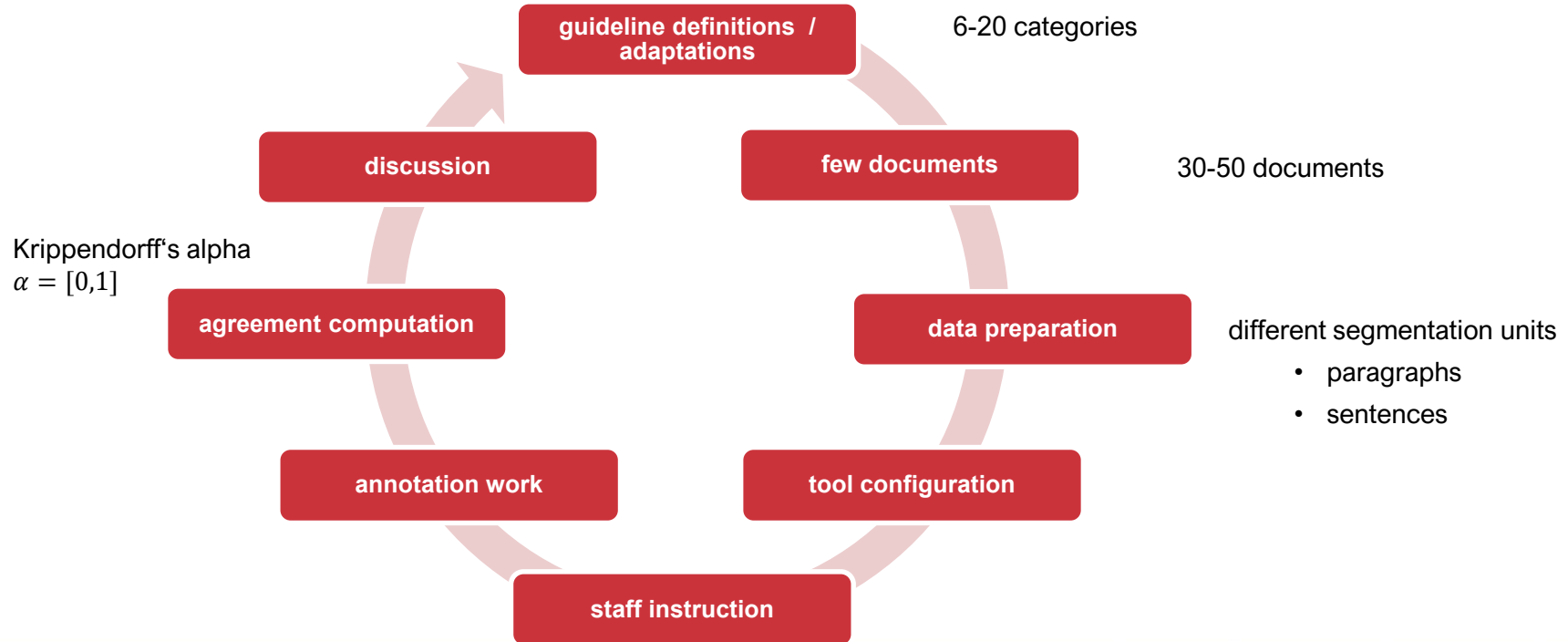
- 8 annotators (medical students)
- tool: WAT-SL (Kiesel et al., EACL 2017)



classification

Annotation – Iterative Training Process

- 4 iterations prior to annotation of whole text corpus

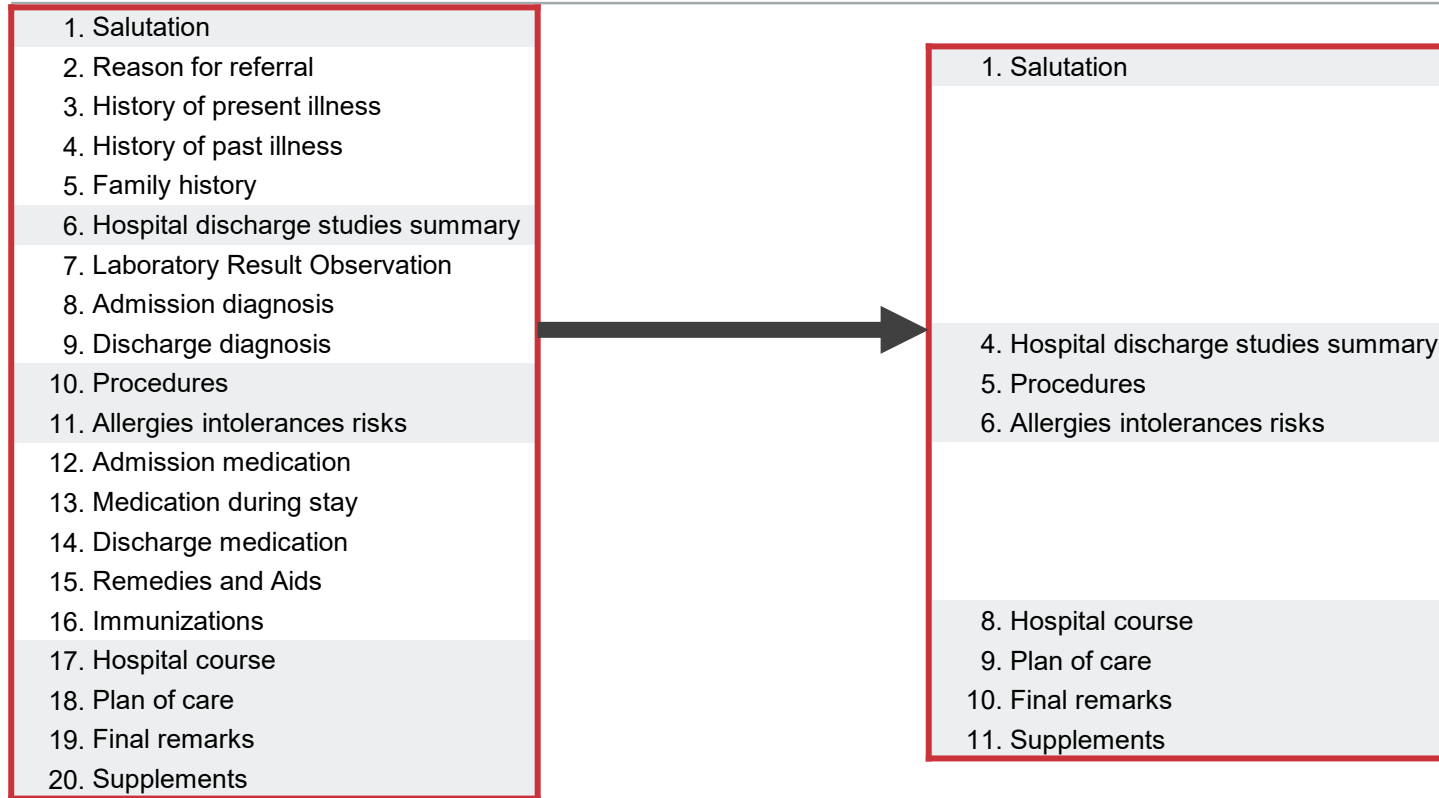


- **Self-defined categories**
 1. Preamble
 2. Anamnesis
 3. Diagnostics
 4. Therapy
 5. Future
 6. Appendix
 7. Mix (only in 2nd iteration)
- **Agreement (α)**
 - 1st iteration: .89
 - 2nd iteration: .80
- **Different levels of segmentation granularity**
 - Paragraphs
 - ***Sentences (used in all subsequent annotations)***

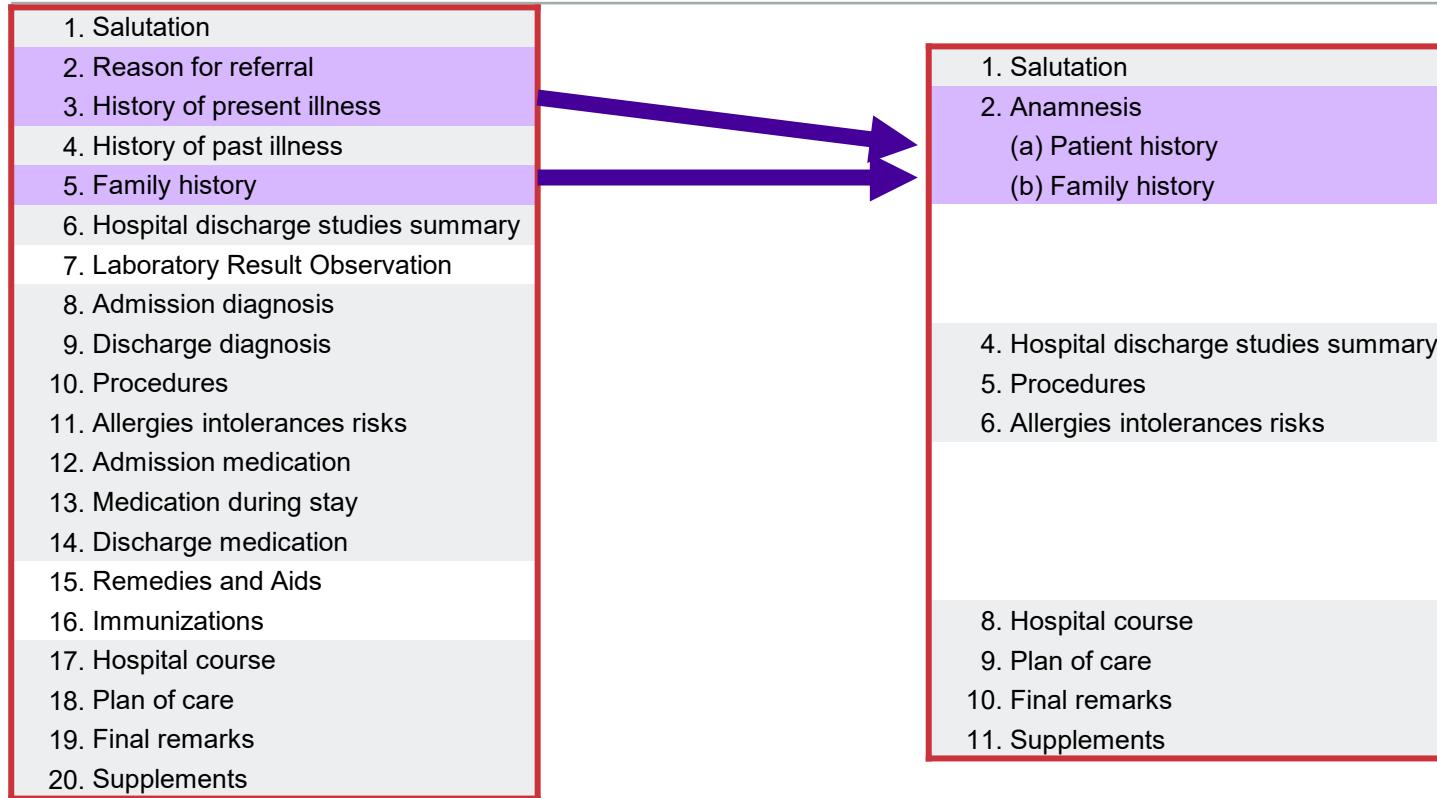
- **Clinical Document Architecture (CDA)**
 - XML-based Health Level 7 (HL7) interoperability standard
 - In Germany: not fully operational, soon mandatory
 - CDA suited for clinical documentation, not for annotation
- **Agreement (α): .70**

1. Salutation
2. Reason for referral
3. History of present illness
4. History of past illness
5. Family history
6. Hospital discharge studies summary
7. Laboratory Result Observation
8. Admission diagnosis
9. Discharge diagnosis
10. Procedures
11. Allergies intolerances risks
12. Admission medication
13. Medication during stay
14. Discharge medication
15. Remedies and Aids
16. Immunizations
17. Hospital course
18. Plan of care
19. Final remarks
20. Supplements

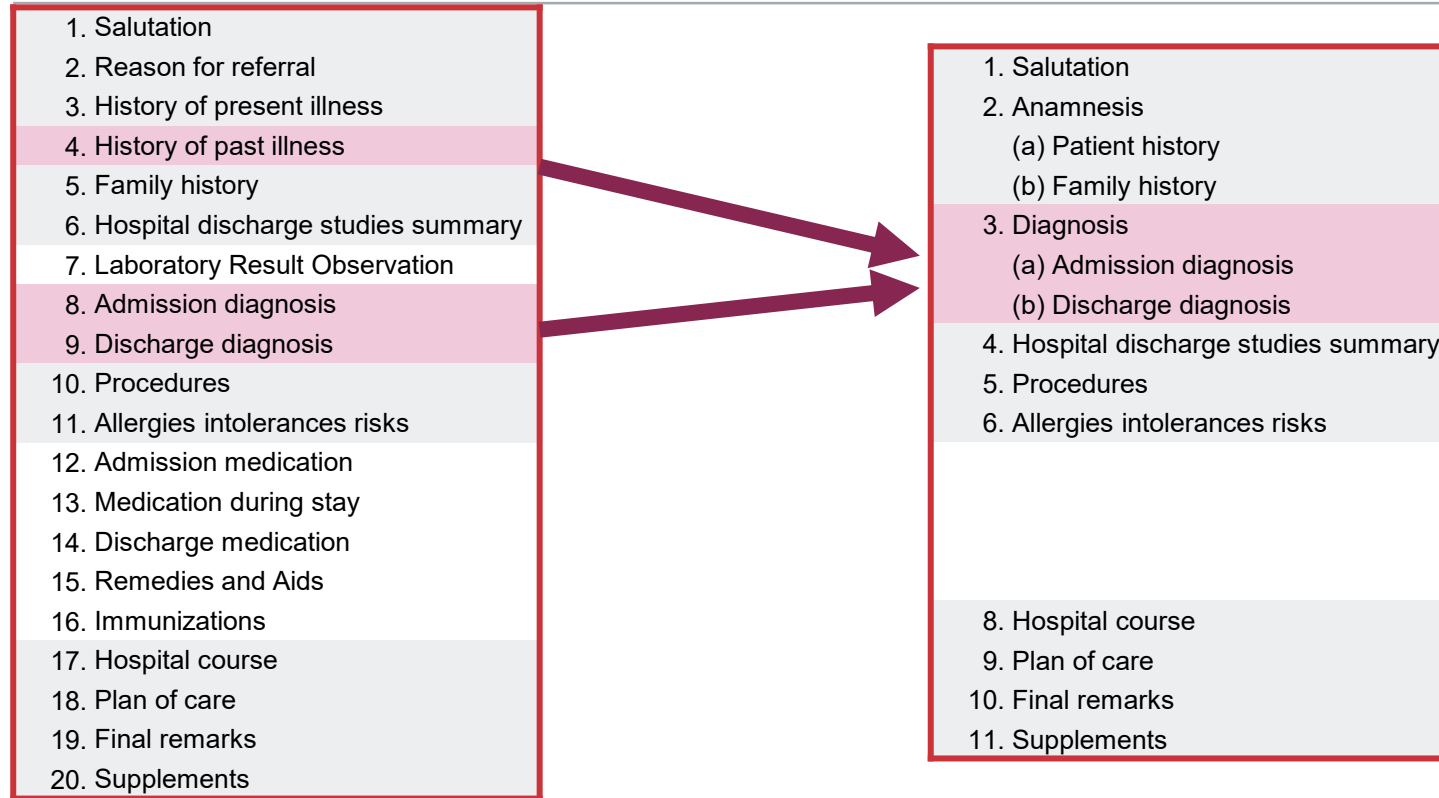
Categories – Redefinition of CDA



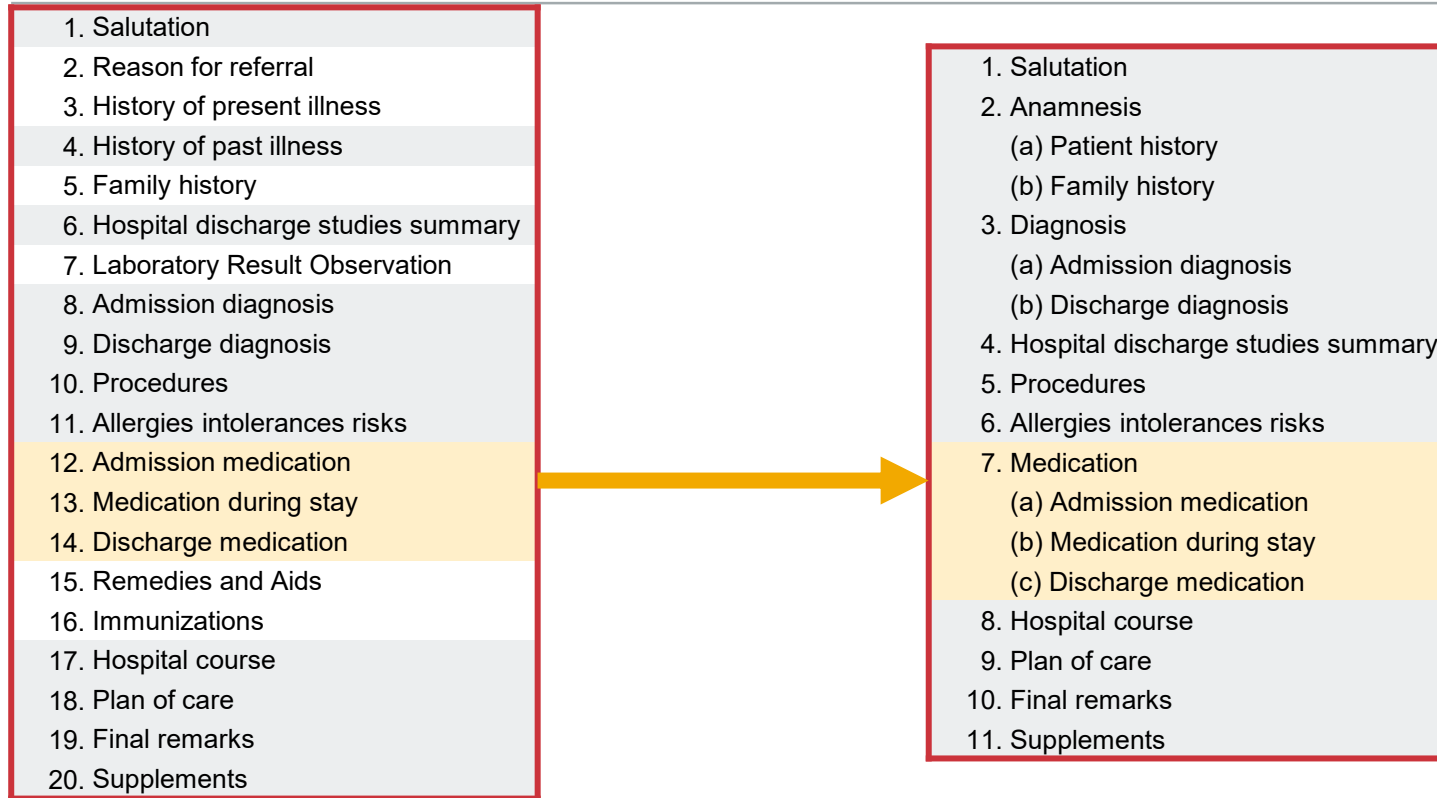
Categories – Redefinition of CDA



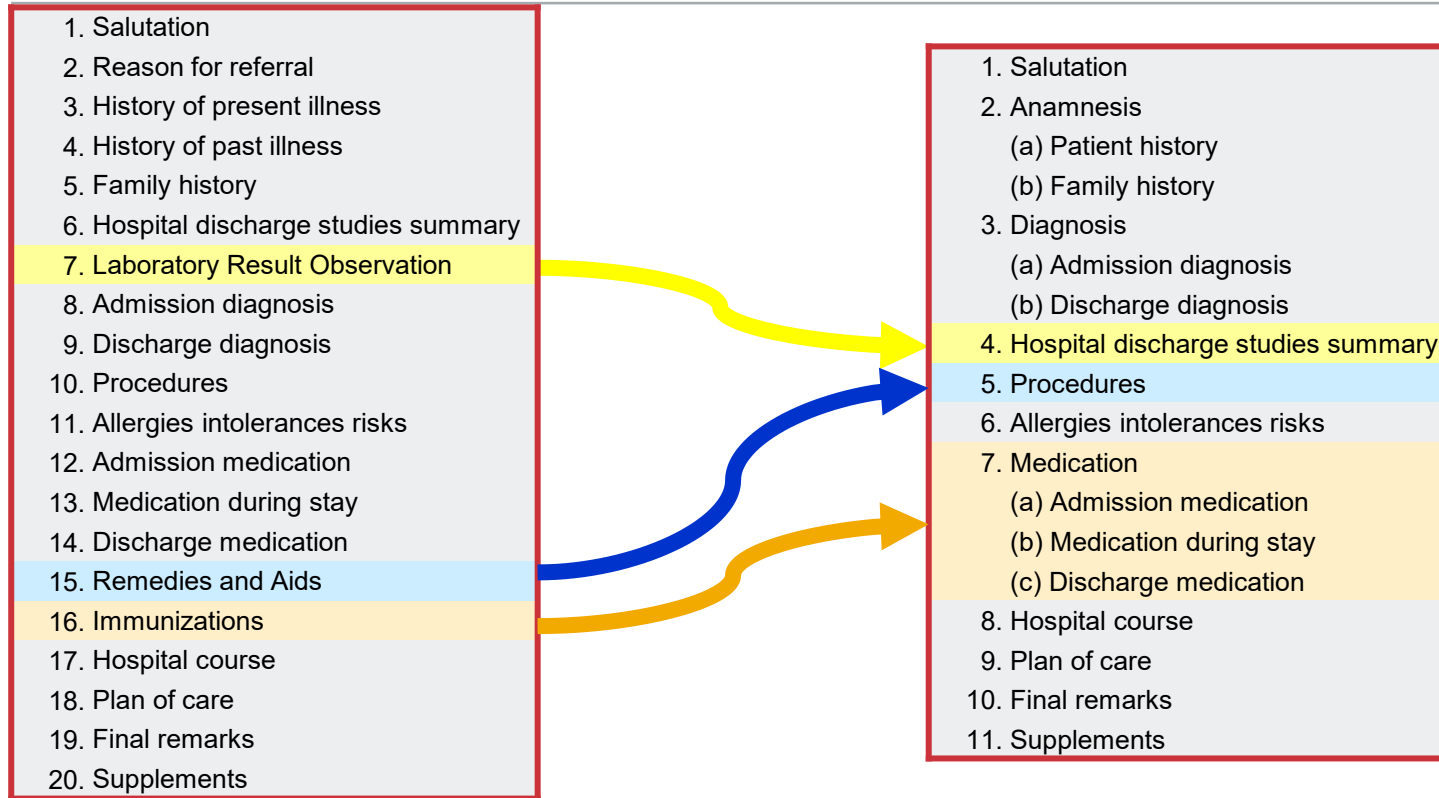
Categories – Redefinition of CDA



Categories – Redefinition of CDA



Categories – Redefinition of CDA



- **Annotation instruction**

- **Redefined CDA**
- **Hierarchy**

- **Agreement (α): .75**

1. Salutation
2. Anamnesis
 - (a) Patient history
 - (b) Family history
3. Diagnosis
 - (a) Admission diagnosis
 - (b) Discharge diagnosis
4. Hospital discharge studies summary
5. Procedures
6. Allergies intolerances risks
7. Medication
 - (a) Admission medication
 - (b) Medication during stay
 - (c) Discharge medication
8. Hospital course
9. Plan of care
10. Final remarks
11. Supplements

- 1106 documents annotated
- Agreement
 - 50 documents
 - (α): .82

Categories	Sentences
4. Hospital discharge studies summary	51%
8. Hospital course	12%
1. Salutation	8%
7. (c) Discharge medication	7%
...	
5. Procedures	2%
11. Supplements	<1%
...	
6. Allergies intolerances risks	<1%
7. (a) Admission medication	<1%
2. (b) Family history	<1%

Baseline Classifier

- **Features: bow statistics**
- **Logistic regression model with a 10-fold cross-validation**
- **Results**
 - Average f-score: **0.82**
 - Precision: 0.82
 - Recall: 0.84
 - Average accuracy 83.7%
- **Ganesan et al. (IEEE Big Data, 2014)**
 - 9 categories
 - Logistic regression
 - Average accuracy: 93%
 - English language

Categories	Sentences	F-Score
4. Hospital discharge st. sum.	51%	0.93
8. Hospital course	12%	0.83
1. Salutation	8%	0.85
7. (c) Discharge medication	7%	0.94
...		
5. Procedures	2%	0.21
11. Supplements	<1%	0.38
...		
6. Allergies intolerances risks	<1%	0.43
7. (a) Admission medication	<1%	0.22
2. (b) Family history	<1%	0.00

Acknowledgements

- Research conducted within in the STAKI²B² project, funded by the German Research Foundation (DFG)
- In cooperation with the SMITH project, funded by the German Federal Ministry of Education and Research (BMBF)



Kindly contact me at:

christina.lohr@uni-jena.de

www.julielab.de

