Healthcare IT and NLG – Great synergies ahead?

Dirk Hüske-Kraus
Philips Medical Systems
Böblingen

LT-ACCELERATE
Brussels, 22.11.2016

A personal note

Dirk Hüske-Kraus

- Background in medical informatics
- PhD thesis on multilingual text generation for clinical documents

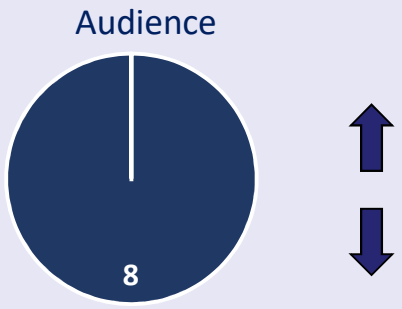
ToC

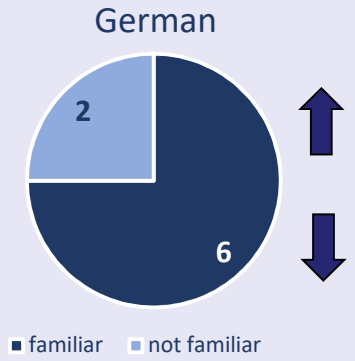
- Problem space
  - The „vicious mandala“ of clinical document generation
  - Presentation of complex data
  - Patient interaction (multilingual)
- Specific requirements for clinical documents
- The desideratum and limits of conventional approaches
- Practical example
- Solution components
- Where‘s the money? (And where are the stumbling blocks?)

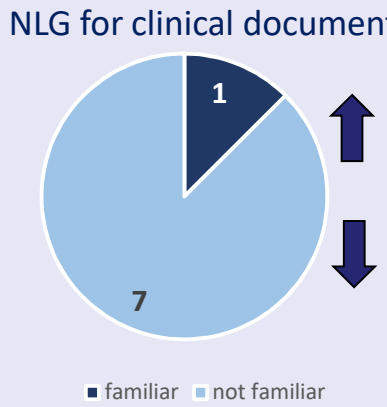
TOC

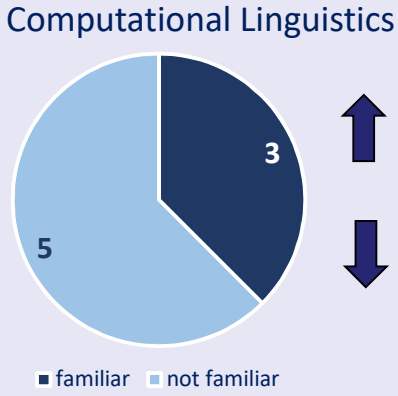
- Problem space
  - The „vicious mandala“ of clinical document generation
  - Presentation of complex data
  - Patient interaction (multilingual)
- Specific requirements for clinical documents
  - Correctness and completeness– conciseness, cohesion and coherence – stylistic adequacy – naturalness and configurability – phenotypology of clinical text types
- The desideratum and limits of conventional approaches
  - Canned text – templates – procedural approaches
- Practical example
- Solution components
- Where‘s the money?

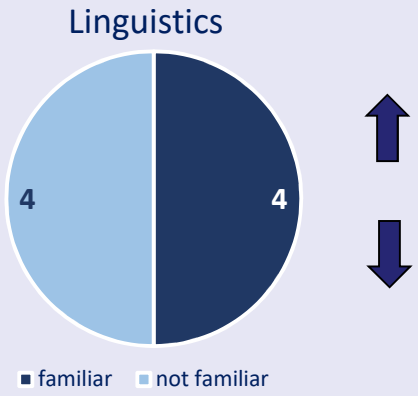
Presenter, know thy audience!





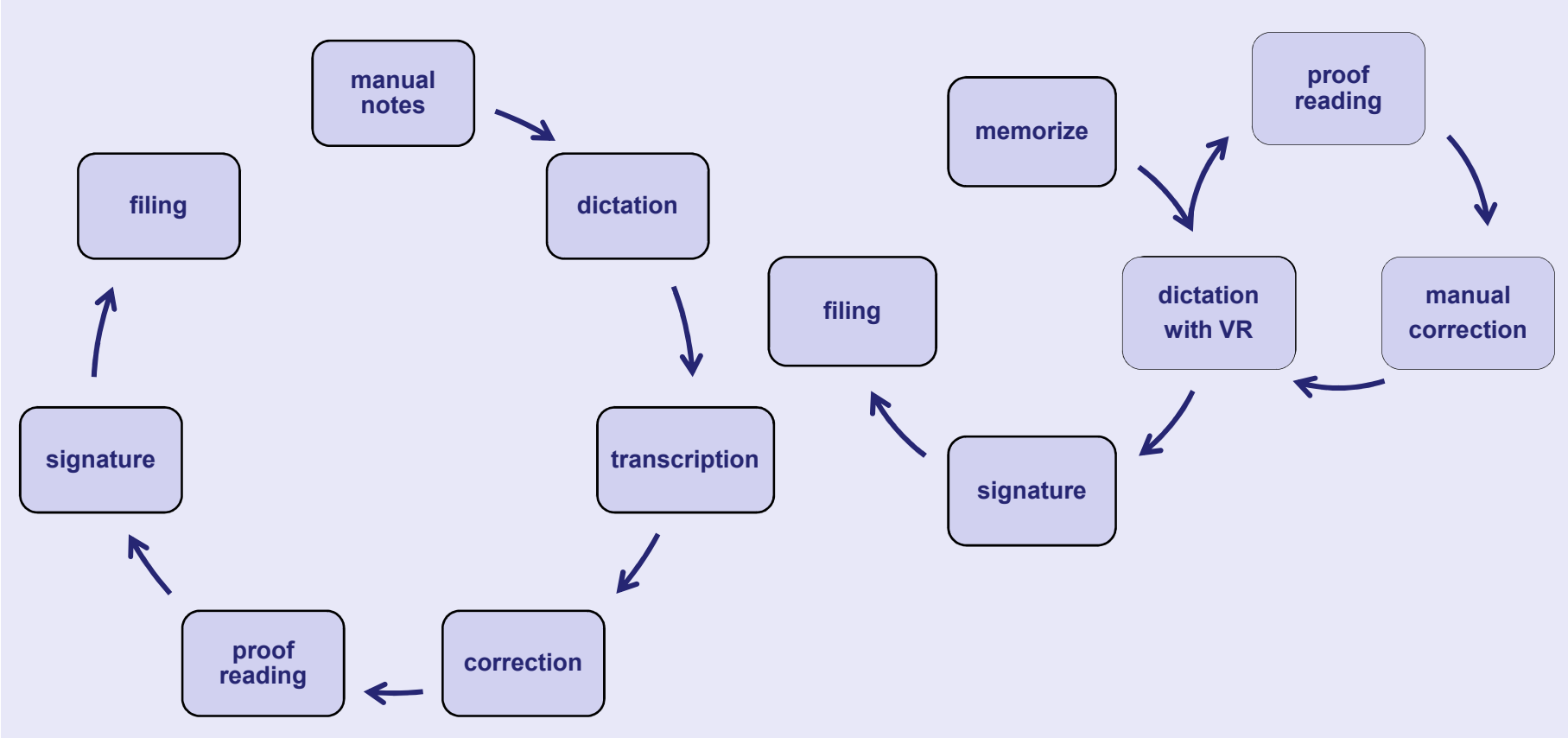






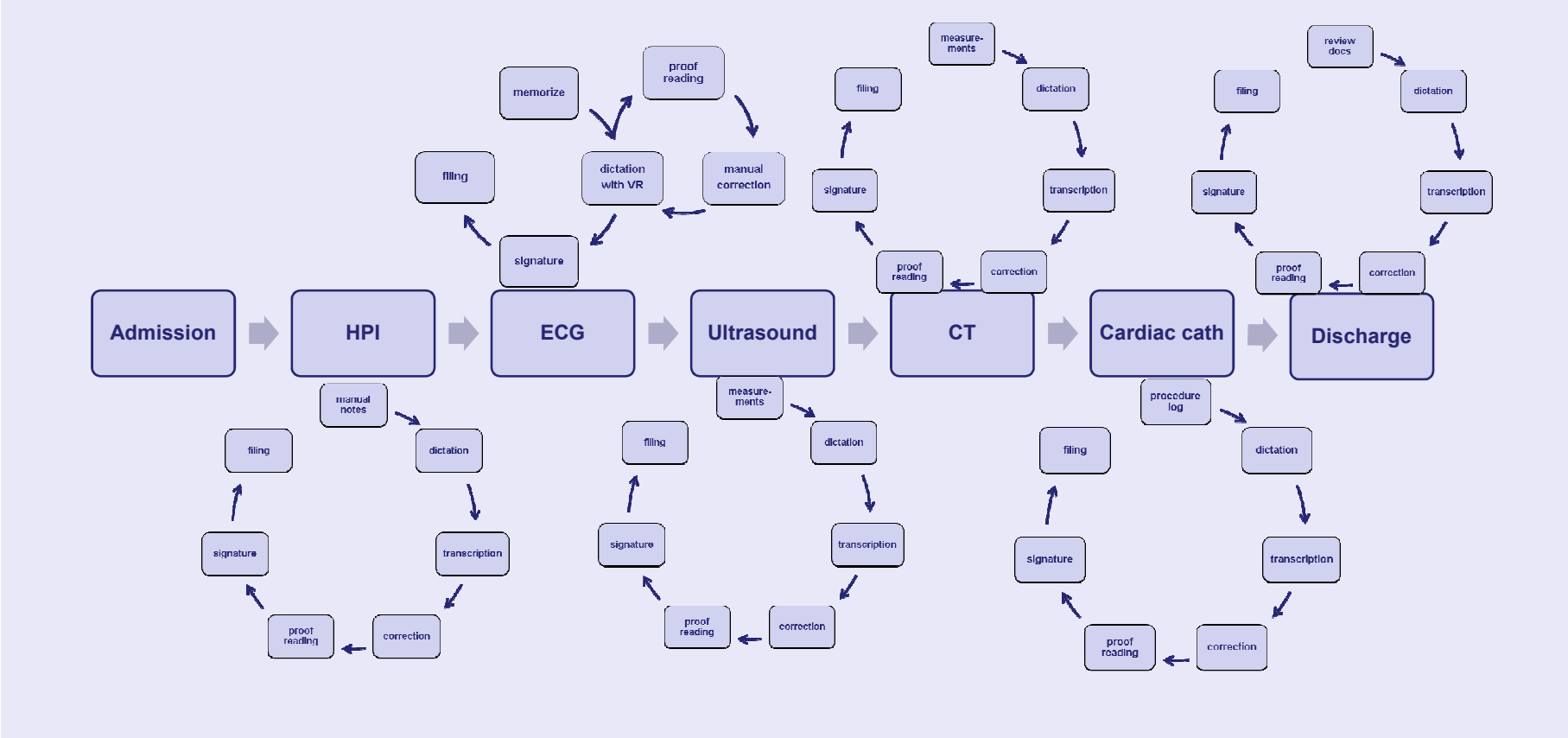
The problem

The vicious “circle“ of clinical document generation



The problem

The „vicious mandala“ of clinical document generation



The problem

The hospital as editorial office

- Approx. 50 documents per inpatient spell[1]
- 60-80 % textual and „human generated“, most of the content available as structured data
- A 350 bed institution with an LOS of 7 d creates ~1600 of these per day

➔ A hospital‘s textual output is comparable to a medium-sized daily newspaper

An example

The problem

Walter O. Whitley:
Interprofessional Communication Pearls for Writing Referral Letters

The "8 Cs" of business communication, which aid in improving both the efficiency and efficacy of interactions with other healthcare providers.

- **Conciseness** – Referral letters should be short and to the point.
- **Correctness** – Accurately state the facts; check your grammar.
- **Clarity** – Make it easy to understand.
- **Completeness** – Cover the essentials to get your intended response.
- **Consideration** – Put yourself in the place of the receiver.
- **Concreteness** – Be specific to reinforce confidence in your skills.
- **Courtesy** – Strengthen your relation
- **Consistency** – Make it a habit in your daily practice.

Example of a Referral Letter

January 20, 2011

John Sheppard, M.D., M.Msc.
241 Corporate Blvd.
Norfolk, VA 23502

Patient: Jane Doe / D.O.B. 1-2-64

Dear Dr. Sheppard,

Thank you for agreeing to evaluate Mrs. Doe for her uveitic condition. Mrs. Doe is a 47-year-old African American who was seen in our clinic for decreased vision and eye pain O.S. > O.D. According to Mrs. Doe, she denies any previous episode of this condition. Her medical history includes hypertension and high cholesterol, which are medically controlled. Upon examination, she presented with a bilateral, granulomatous uveitis, which was treated with prednisolone acetate q.2.h. O.U. and homatropine 5% t.i.d. O.U. Although there has been mild improvement in her condition for the last two weeks, I am referring her to your care for further evaluation and a possible sub-tenon corticosteroid injection and/or systemic steroid therapy. Lab testing revealed elevated ACE levels, which may be indicative of sarcoidosis. We are still waiting for the results of her chest X-ray. Enclosed are the results of recent lab testing for your review.

Once again, thank you for seeing Mrs. Doe. Please keep me updated on her progress and kindly refer her back to my care once her condition resolves.

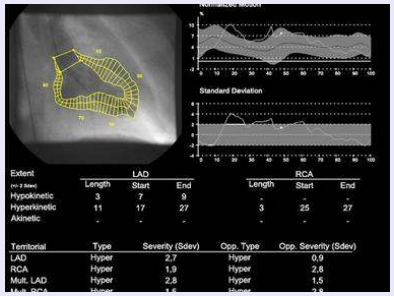
Best Regards,

Walt Whitley, O.D., M.B.A., F.A.A.O.

The desideratum…

Methods to

- transform existing structured data into



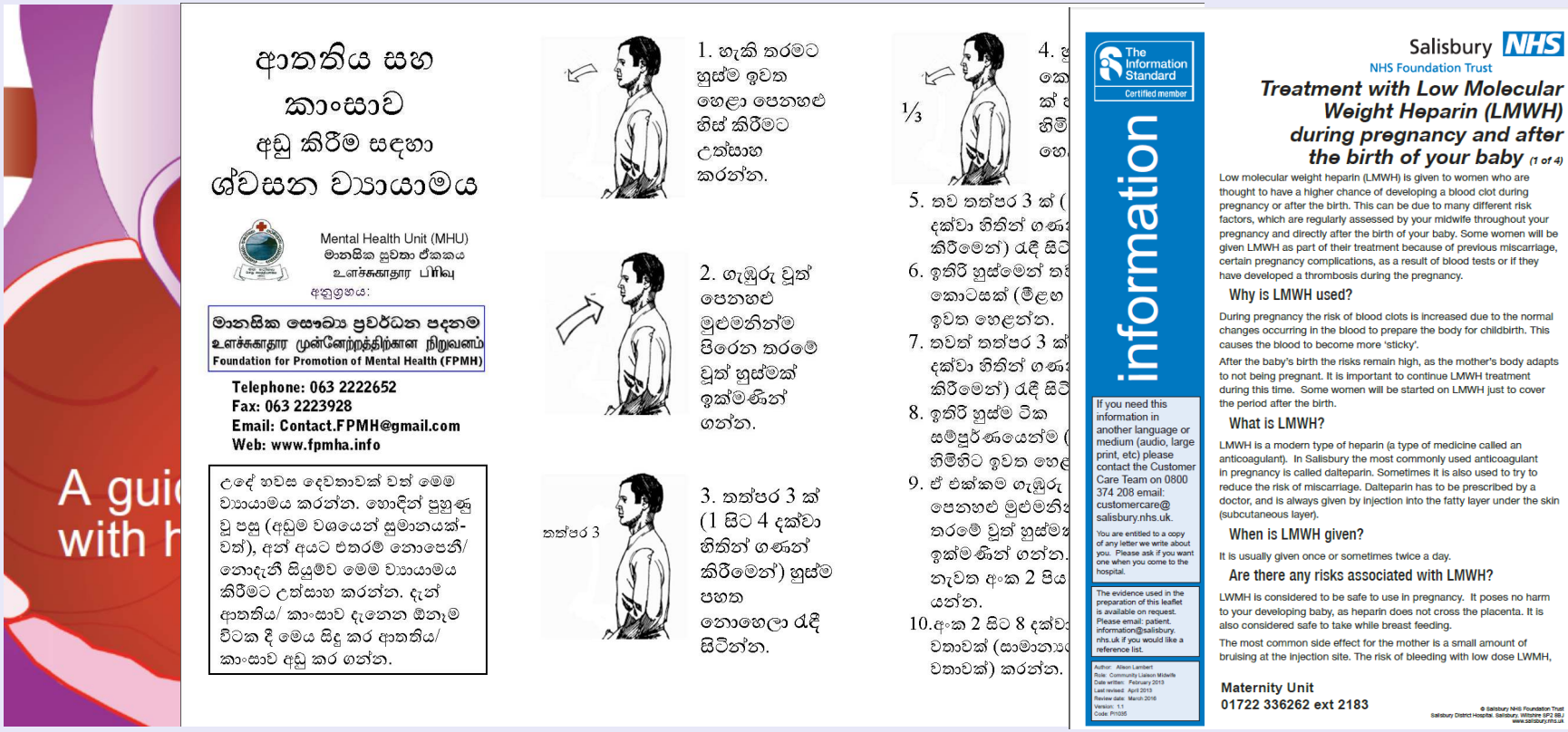
- <u>adequate</u> human-readable text

would reduce efforts significantly (and reduce errors, enforce canonical language…)

Another problem

Patient information/patient education tailored to…

- Current condition and comorbidity
- Language and cultural background
- Gender, age, treatment …

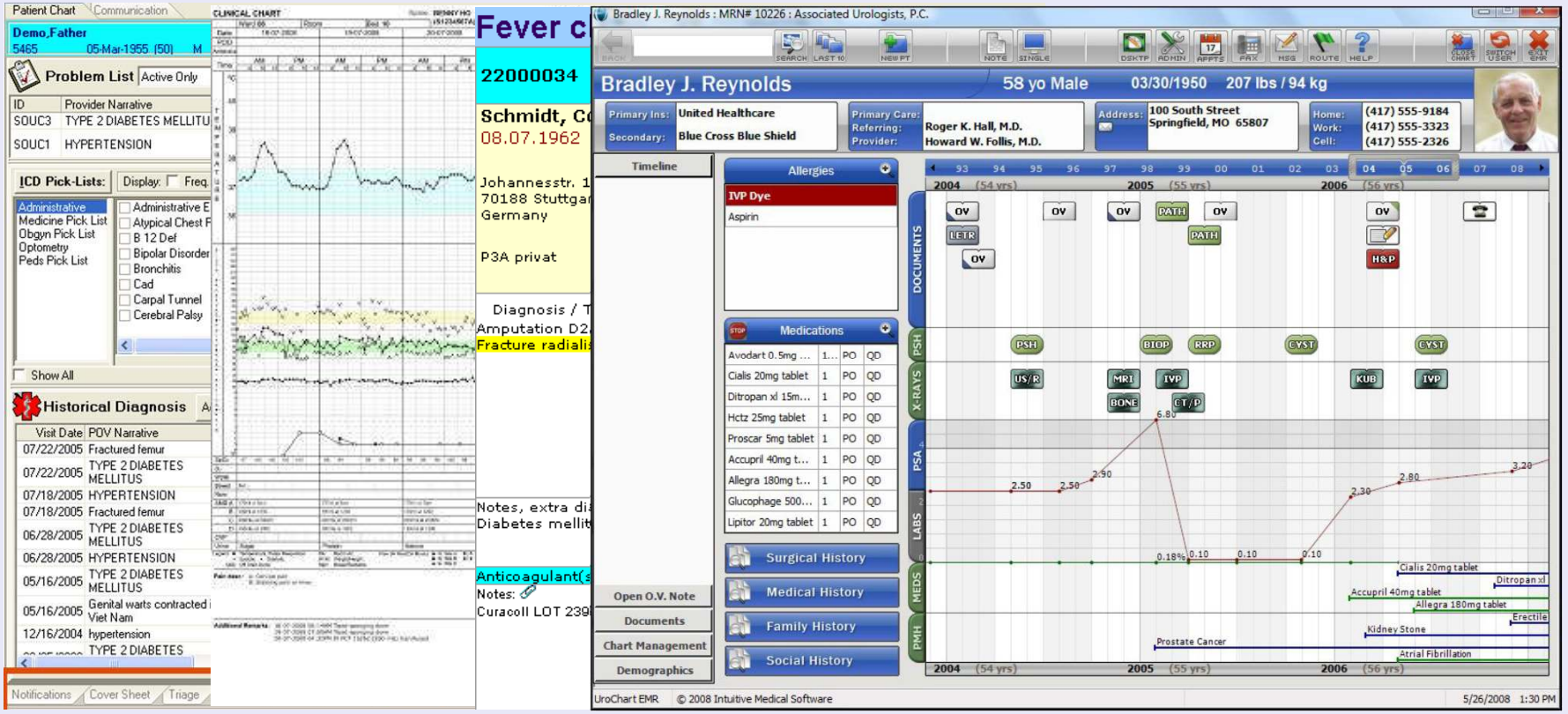


http://fms-itskills.ncl.ac.uk/howto/infoleaflet.html https://fpmh.wordpress.com/downloads/ http://www.vte.salisbury.nhs.uk/Education/Documents/LMWH%20during%20pregnancy%20PI1035.pdf

One more problem

Presentation of complex information

- Structured data
- Vital signs
- Vital signs & therapy
- Lifetime data



https://uxpa.org/jus/article/usability-electronic-medical-records http://www.efoza.com/post_vitalsigns-chart-daily_128512/ http://www.wikiwand.com/en/Health_informatics http://www.cs.umd.edu/hcil/lifelines/healthtronics-lifelineexample.pdf

One more problem

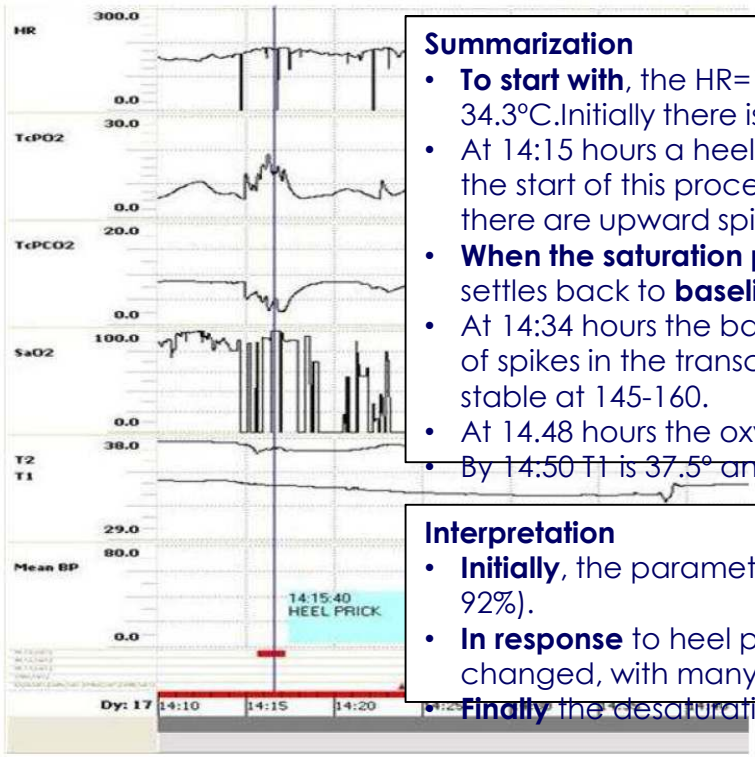
"A picture is worth a thousand words."

http://www.worldpressphoto.org/collection/photo/2016

One more problem

Presentation of complex information

- Sometimes, a hundred words say more than a picture



Summarization

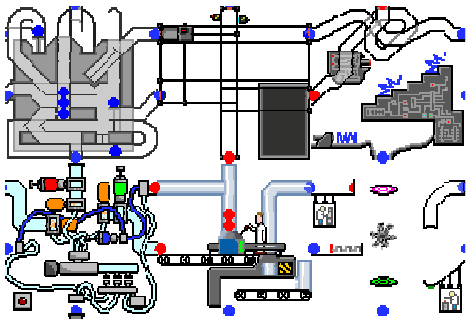
- **To start with**, the HR=152; TcPO2= 6; TcPCO2= 8.6 and SaO2= 92%. T1= 37.7° and T2= 34.3°C.Initially there is an increase in oxygenation, HR and TcPCO2 are static.
- At 14:15 hours a heel prick is done. The HR increases **at this point and for 7 minutes** from the start of this procedure there is a lot of artifact in the oxygen saturation trace. Also there are upward spikes in the TcPO2, each **corresponding** with downward spikes.
- **When the saturation probe is reapplied**, the oxygen saturation is 100%, and the HR settles back to **baseline**.
- At 14:34 hours the baby is examined. For the **following 5 minutes** there is **another series** of spikes in the transcutaneous parameters **while** the saturations fall to 74%. HR remains stable at 145-160.
- At 14.48 hours the oxygen saturation falls to 83%, HR remains steady around 160.
- By 14.50 T1 is 37.5° and T2 is 34.2°C.

Interpretation

- **Initially**, the parameters are normal and stable except for a slight desaturation (SaO2 92%).
- **In response** to heel pricks and examination, respiratory and circulatory parameters changed, with many desaturations down to 70%, but returned to baseline quickly
- **Finally** the desaturation has worsened to 83%, other parameters normal

Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., & Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, *173*(7), 789-816.

To summarize…

data  text

Methods to…

…take in structured data… s,

….and create adequate…

…textual output… r

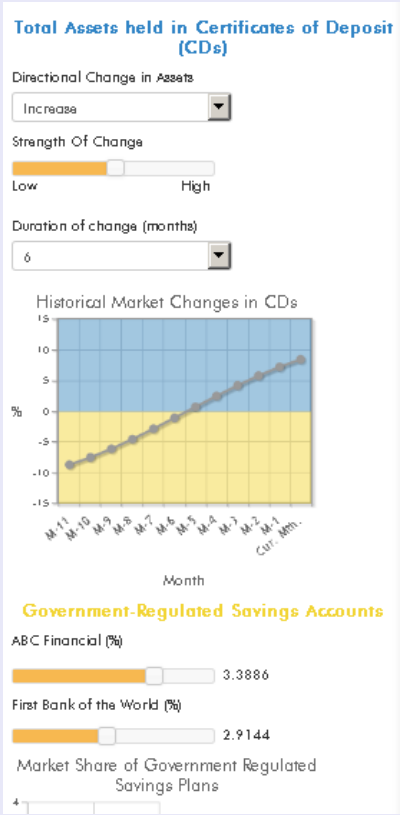
are highly desirable

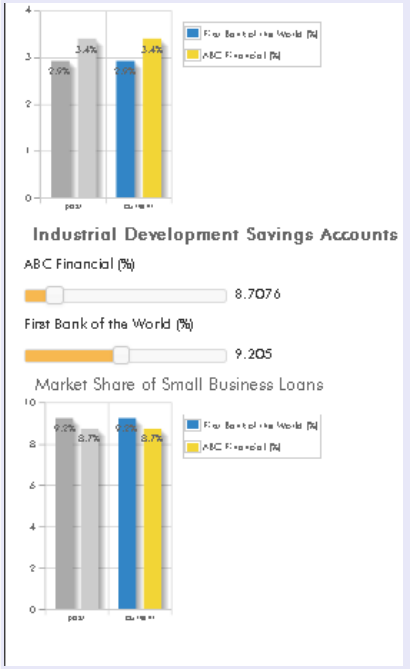
The problem

…but here‘s solutions out there that seem to do just this!

„Real life“ NLG:

- Yseop: At the end of July 2016, the sustained expansion of **total deposits** (+5.5%) resulted from the positive change of deposits made towards **certificates of deposit**, **checking accounts** and **savings accounts** over the last 12 months (+8.5%, +7.4% and +7.3%, respectively). This growth occurred in spite of the decrease of deposits made towards **money market accounts**, in comparison to July 2015 (a major decline of -22.7%).
  In this instance, the deposits made into **checking accounts**, **savings accounts** and **certificates of deposit** recorded a uniform increase over the previous 10 months, the past 7 months and the past 6 months, respectively. On the contrary, the deposits made into **money market accounts** displayed a steady decrease over the previous 11 months, and have clearly





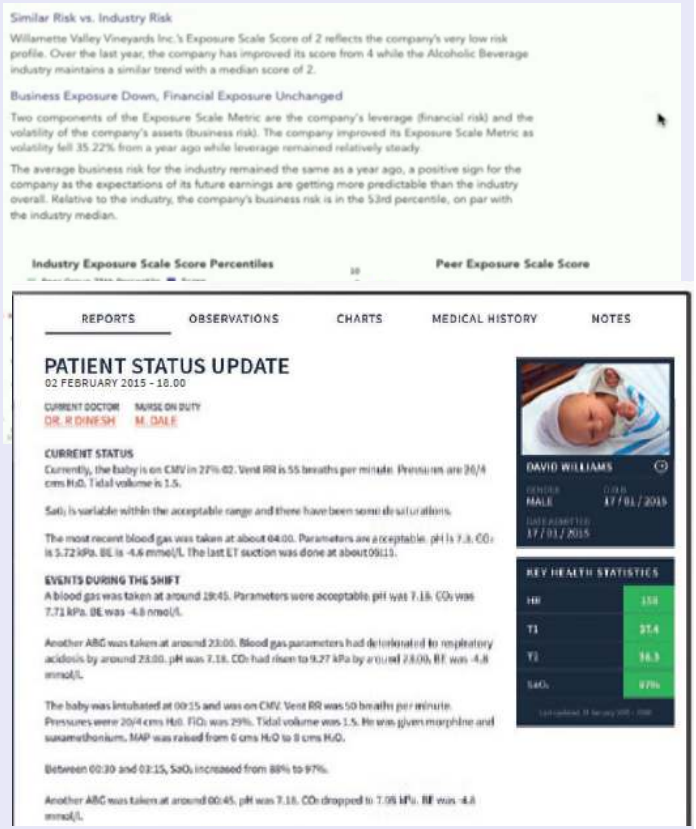
The problem

…but here‘s solutions out there that seem to do just this!

„Real life“ NLG:

- NarrativeScience („Quill“): Combination of data analytics and NLG output, English only, technology? Parameters and report content/structure predetermined

- ARRIA NLG: Spin-off by Reiter & Dale (“Data2Text”), experiences in Healthcare (“BabyTalk”, scientifically validated, not commercialized), English only, rule-based, no end-user configurability



Requirements for clinical documents

Example: ECG

“Telegraphic” style: asyndetic chain of nominal phrases

- Linkstyp; Sinusrhythmus mit einer Frequenz von 90 /min; regelrechte Erregungsüberleitung; kompletter Linksschenkelblock mit einer QRS-Dauer von 156 ms; vereinzelt monotope ventrikuläre Extrasystolen; Repolarisationsstörungen anterolateral
- LAD, NSR 90, complete LBBB, ST elevations I, aVL, V1-V6; rare unifocal VPC’s

Requirements for clinical documents

Example: Discharge letter – HPI & summary

"Narrative" style: Maximally concise full sentences

- Bei Herrn Wicht ist eine koronare Herzerkrankung bekannt. Vor knapp 3 Jahren erfolgte eine operative Myokardrevaskularisation, nachdem der Patient einen Monat zuvor einen Hinterwandinfarkt erlitten hatte. Herr Wicht stellt sich zu einer Routinekontrolle beschwerdefrei vor. Er verneint insbesondere pectanginöse Beschwerden und Dyspnoe. Der Patient kann 3-4 Stockwerke ohne Pause Treppensteigen. Seit der letzten ambulanten Vorstellung hier am 1.1.2009 keine Synkope, keine Operation, keine Krankenhausaufenthalte.
  Herr Wicht ist gelernter Bergmann, er arbeitete dann im Strafvollzug und ist berentet.
  Kein Nikotinabusus, ein Diabetes mellitus ist seit 1970 bekannt und wird seit 1995 mit Insulin behandelt. Herr Wicht ist verheiratet, hat 2 Kinder. In seiner Freizeit fährt er vor allen Dingen Fahrrad.[…]
- Bei dem 81jährigen Herrn Wicht finden sich knapp 3 Jahre nach operativer Myokardrevaskularisation keine Hinweise auf eine Progression der koronaren Herzerkrankung oder eine Bypassdysfunktion. Der Patient ist beschwerdefrei und belastet sich normal. Echokardiographisch zeigte sich wie vor 2 Jahren eine leichte Dilatation des linken Vorhofs ohne Nachweis eines hämodynamisch bedeutsamen Vitiums. Die globale LV-Funktion ist gut.
  Herr Wicht ist übergewichtig, anzustreben ist ein Gewicht von 70 kg. Das LDL-Cholesterin sollte unter 100 mg/dl liegen. Die Blutdruckwerte sind gut eingestellt. Grundsätzlich empfehlen wir die bisherige Medikation beizubehalten, möglicherweise kann unter dem Aspekt der Nierenprotektion bei bekanntem Diabetes mellitus die ACE-Hemmer-Dosis noch gesteigert werden. Beschwerdefreiheit vorausgesetzt empfehlen wir eine fachkardiologische Kontrolle im Abstand von 2 Jahren.

Requirements for clinical documents

Example: Discharge letter – Hospital Course

"Narrative" style: Maximally concise full sentences

- 1. Fall: The patient was admitted and ruled out for syncopal episode. Echocardiogram was normal, and when the patient was able, her orthostatic blood pressures were within normal limits. Any serious conditions were quickly ruled out.
  2. Status post fall with trauma: The patient was unable to walk normally secondary to traumatic injury of her knee, causing significant pain and swelling. Although a scan showed no acute fractures, the patient's frail status and previous use of cane prevented her regular abilities. She was set up with a skilled nursing facility, which took several days to arrange, where she was to be given daily physical therapy and rehabilitation until appropriate for her previous residence.
- The patient is a 50-year-old right-handed Caucasian female, who works as an independent contractor and as a human resources consultant.
  Her neurological history first begins in December of 1987, when she had a rather sudden onset of slurred speech and the hesitancy when she started to walk.
  The slurred speech resolved after a few weeks, but her gait hesitancy persisted for a number of years and then finally partially improved. She also began to note that she

Requirements for clinical documents

How is the clinical sociolect different?

Everyday language

- Few foreign words
- Inaccuracy admissible, qualitative expressions preferred
- Figurative language, pseudodialogue, freedom of illustration
- Balance between conciseness and liveliness
- …

Language of clinical documents

- Foreign words galore
- Accuracy preferred, quantitative expressions whenever possible
- Monologic language, descriptive with restricted metaphors and comparisons
- Maximum conciseness
- …

Plus…

Requirements for clinical documents

How is the clinical sociolect different (and language/culture specific)?

- ad-hoc neologisms↗
  - *„Präkollaps“*
  - *C0-syndrome*
- (nominal-) compounds↗
  - *Pseudopseudoparahyperthyreoidism*
  - *non-Hodgkin lymphoma*
- New lexemes, proper names
  - *SARS*
  - *Viagra*
  - *„selektive Septalastokklusion“*
- idiomatic expressions/ collocations/phrasal verbs
  - *„echokardiographisch zeigt sich…“*

- Pluralis hippocraticus
  - „Ihr freundliches Einverständnis vorausgesetzt vereinbarten wir mit Herrn Hüske eine erneute ambulante Wiedervorstellung in ca. 1 ½ Jahren.“
- Illocutionary acts always indirect
  - „Anticoagulation treatment should be started with an INR of…“
  - „Bei Durchsicht der Vorbefunde, insbesondere bei Durchsicht unseres Arztbriefes vom xx.xx.xx gewinnt man den Eindruck, dass unsere therapeutischen Empfehlungen hinsichtlich der Antikoagulation nicht berücksichtigt werden.“
- Passive voice dominant
  - Two Hemovac drains **were** then **placed** inferiorly at the deltoid. The deltopectoral interval **was** then **closed** with 0 Vicryl sutures. A third drain **was placed** in the subcutaneous tissues to prevent any infections or any fluid collections. This **was sewn** into place with the drain **pulled out** superiorly. Once all the sutures **have been secured** and the drain **visualized** throughout this part of the closure, the drain **was pulled** distally until it **was** completely **covered**.

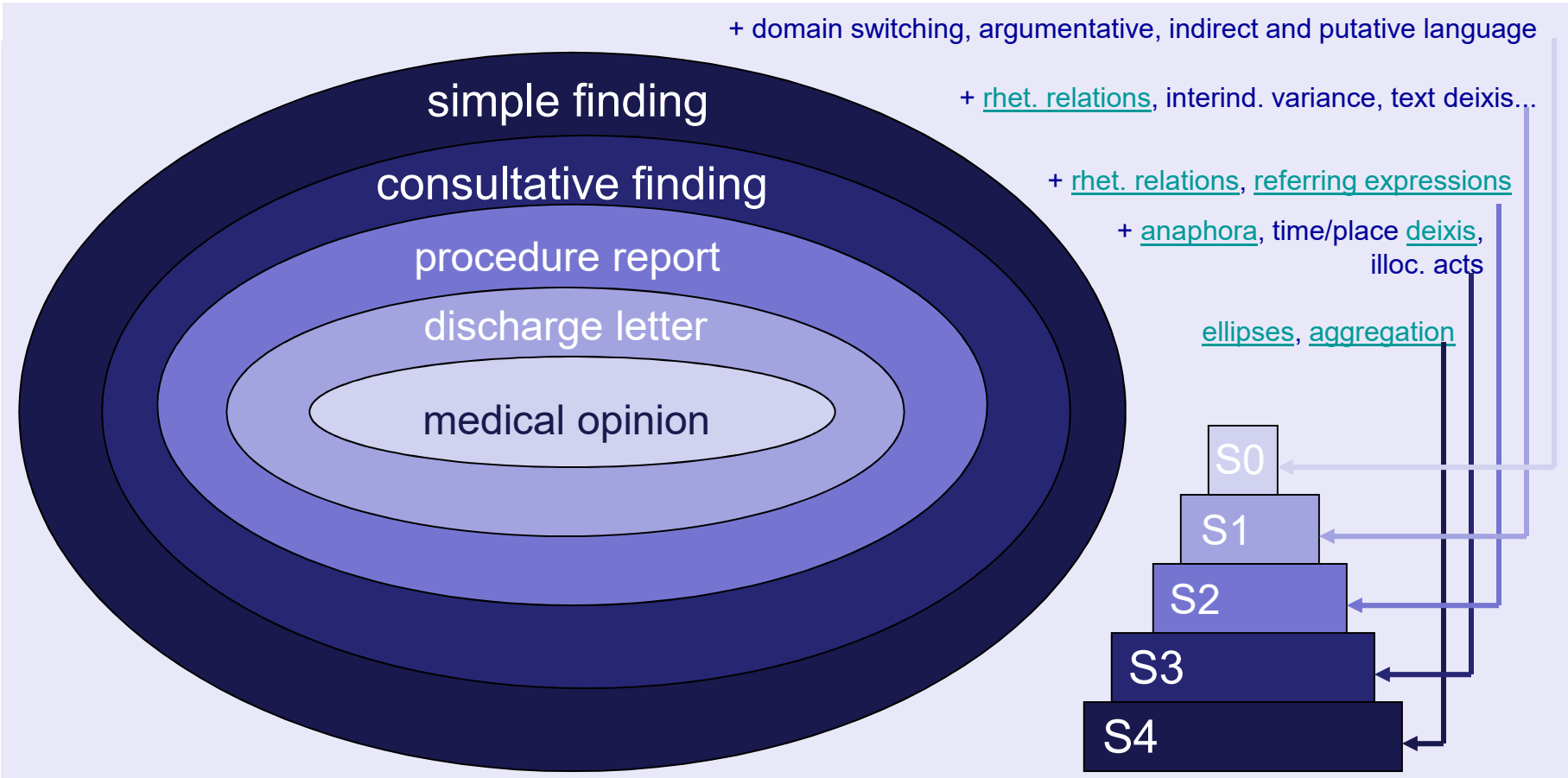
Requirements for clinical documents

Summary

- Completeness (Caveat: may need to specify normal and missing information!)
- Correctness
- Adequacy: Relevance, usage of appropriate register and collocations/phraseology, lexemes, units of measurement, and familiar structure/reading cues; coherence…, practical: Inter- & intra-document variance
- Conciseness: Maximum information in minimal amount of text
  - aggregation
  - ellipses
- Cohesion
  - anaphora, referring expressions
  - text deixis
  - rhetorical relations

Requirements for clinical documents

A linguistic phenotypology of clinical documents



The desideratum…

A not entirely unrealistic example…

Cardiac cath report

- Pre exam: Document known conditions (diseases, risk factors, allergies…)
- During exam: Document vascular access, stenoses, maneuvers, implants…
- Post exam: Coding for Accounting (stenoses, activities…)
- Post exam: Coding for QA (stenoses, activities, complications…)
- Post exam: Dictation of report
- Transcription, proof reading, correction…
- At discharge: (partial) dictation of cath report, transcription, proof reading…

…and limits of conventional approaches

Canned text (aka „Autotext“)

PROs	CONs
Easy and cheap to create	Order dependent
	Error prone
	No aggregation, deixis, rhetorical relations, cohesion…
	No inflection
	Content determination?
➔ Usable only for almost trivial text types (telegraphic, enumerative)	

…and limits of conventional approaches

Canned text (aka „Autotext“)

When it works:

- Uniformly structured, telegraphic texts, e.g. routine normal physical exam:

GENERAL: The patient was lying in bed in no acute distress.
HEENT: Atraumatic and normocephalic. Pupils are equal. No conjunctival hemorrhage. No sinus tenderness. No oral lesion.
NECK: Supple. No cervical lymphadenopathy.
LUNGS: Clear to auscultation.
HEART: S1, S2 audible. No murmurs are heard.
ABDOMEN: Soft. Bowel sounds audible. No organomegaly appreciated. No tenderness noted.
EXTREMITIES: No edema, clubbing or cyanosis.

- Straightforward, simple procedure reports

The patient was prepped and draped. We identified the old incision; this was elongated. Loculations of fibrous tissue were broken up. An inflammatory rind was identified and this was sent for culture. Seropurulent, somewhat bloody fluid was noted. The infection appeared contained to a golf ball-sized area in the subcutaneous tissues above the fascia. There was no evidence of myonecrosis, penetration of the fascia or significant extent along the fascia of the infection. We cleaned the area with Betadine and then packed the wound with Betadine-soaked Kling. Dry dressings were applied. The patient appeared to tolerate the procedure well.

…and limits of conventional approaches

Canned text (aka „Autotext“)

When it does NOT work:

- many quantitative variables, conceptual aggregation for conciseness:

  LAD, NSR 90, complete LBBB, ST elevations I, aVL, V1-V6; rare unifocal VPC’s

- Entire sentences (with hypotaxes, aggregation, deixis, enumeration, flexion…):

  The patient was put on IV heparin because of ultrasound of the upper extremity showing deep venous thrombosis in the cephalic vein on the right.

  Frau Hüske arbeitet seit 2001 als Landwirtin. Sie klagt über punktförmige und ziehende, präkordiale Schmerzen, die in linken Arm und Schulter ausstrahlen. Diese Beschwerden treten bei körperlichen und psychischen Belastungen auf, darüber hinaus bei Kälte.

…and limits of conventional approaches

Templates

PROs	CONs
More „natural“ phrases and sentences	Harder to design and maintain
	No recursive nesting?[2]
	No aggregation, deixis, cohesion[2]…
	No inflection
	Rhetorical relations[2]?
	Little generalization and variation
	Content determination?
➔ Usable only for limited domains where limited expressivity (and high efforts for design/maintenance) are acceptable	

…and limits of conventional approaches

Templates

When it works:

- Uniformly structured texts, high proportion of fixed parts, e.g. normal findings,

INDICATIONS FOR EXAMINATION: The patient is a (XX)-year-old female with family history of colon cancer.
PROCEDURE IN DETAIL: A physical examination was performed. The major risks and benefits associated with the procedure were explained to the patient in detail. The patient verbalized understanding and agreement with the same. The patient was connected to the appropriate monitoring devices and an IV was started. Continuous oxygen was provided via nasal cannula and intravenous sedation was administered in divided doses throughout the procedure.
After adequate sedation was achieved, the patient was placed in the left lateral decubitus position and a digital rectal exam was performed. This examination was within normal limits. A well-lubricated colonoscope was then inserted into the rectum and advanced under direct visualization to the level of the cecum. […]
ENDOSCOPIC DIAGNOSIS: Normal colonoscopy.
RECOMMENDATIONS: Follow up in the clinic in 5 years..

…and limits of conventional approaches

Templates

When it does NOT work:

- Texts which require flexion, deixis, aggregation etc. to be concise and

There was a torturous (sic!) sigmoid colon, but with slow insertion, we were able to get through that area, and with moderate pressure, we got to the right hemicolectomy anastomotic site and we identified small bowel. Right next to the anastomosis was a diverticuli (sic!). As we withdrew the scope in the distal transverse, there were two 6 mm sessile polyps, removed by cold biopsy forceps. No other lesions were seen aside from sigmoid diverticuli, which were moderate in amount, and at 20 cm was an 8 mm sessile polyp, approached by a small snare and snared and removed with 20 watt coagulation current and recovered.

- LAD, NSR 90, complete LBBB, ST elevations I, aVL, V1-V6; rare unifocal VPC's

…and limits of conventional approaches

Procedural logic

PROs	CONs
Theoretically unlimited expressiveness	• Next to impossible to design and maintain
Lower requirements for developers	• Significant marginal effort for inflection, aggregation, deixis, cohesion…
	• Content determination?
➔ Usable only for specialized applications where high maintenance cost or low configurability is acceptable	

The story so far…

…in a nutshell

Basic tenets

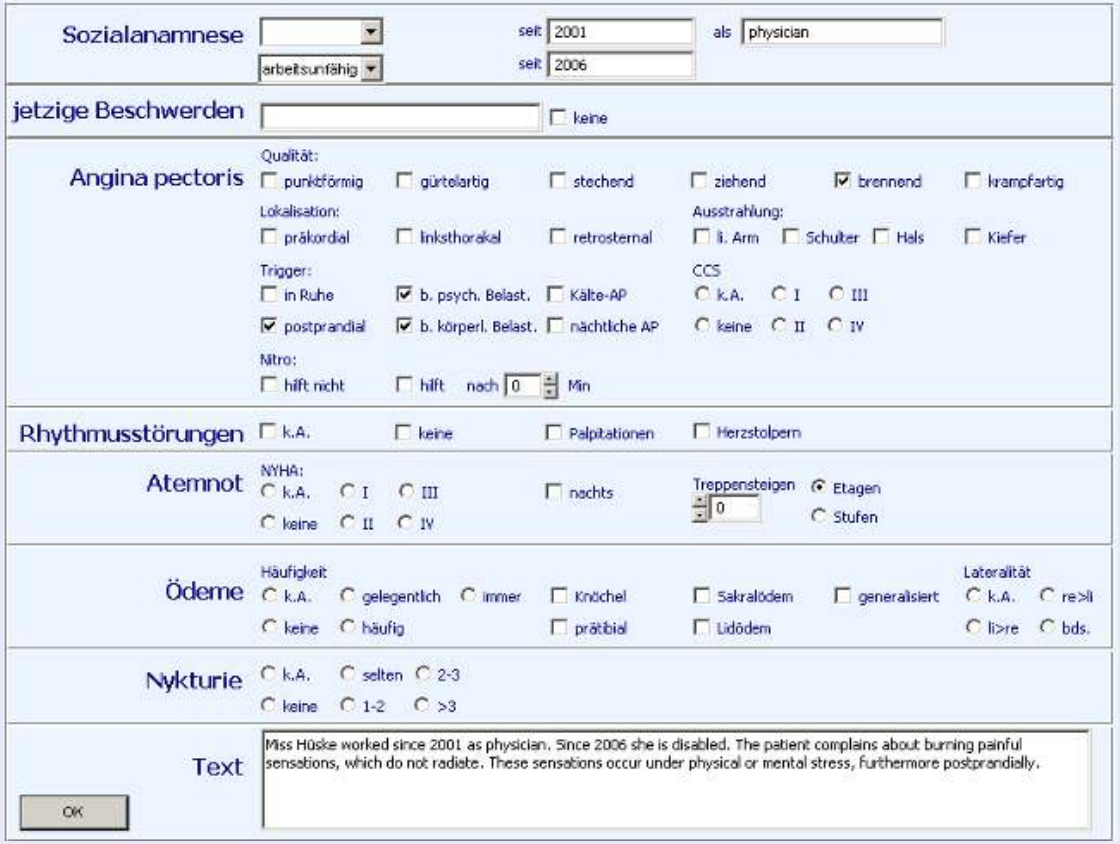
- It is highly desirable to utilize structured clinical data to create appropriate corresponding text for clinical documents
- Approaches based on canned text with variables generally cannot fulfill the „appropriateness“ criteria
- Commercial solutions – e.g. for weather reports, stock market analyses or even “quasi-journalistic” accounts of sports events – do not seem to be up to the task either, both functionally and non-functionally

Practical example

A bilingual NLG-application

Demo

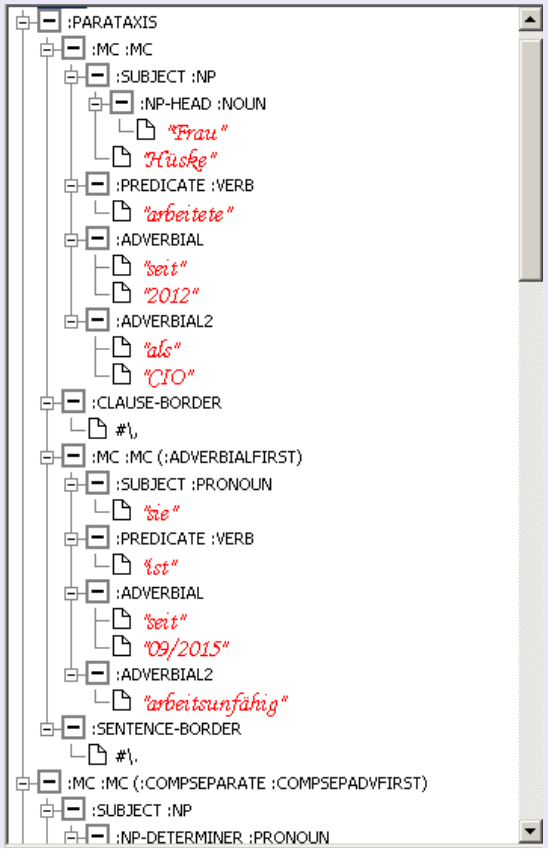
Download:
www.suregen.de



What did we see?

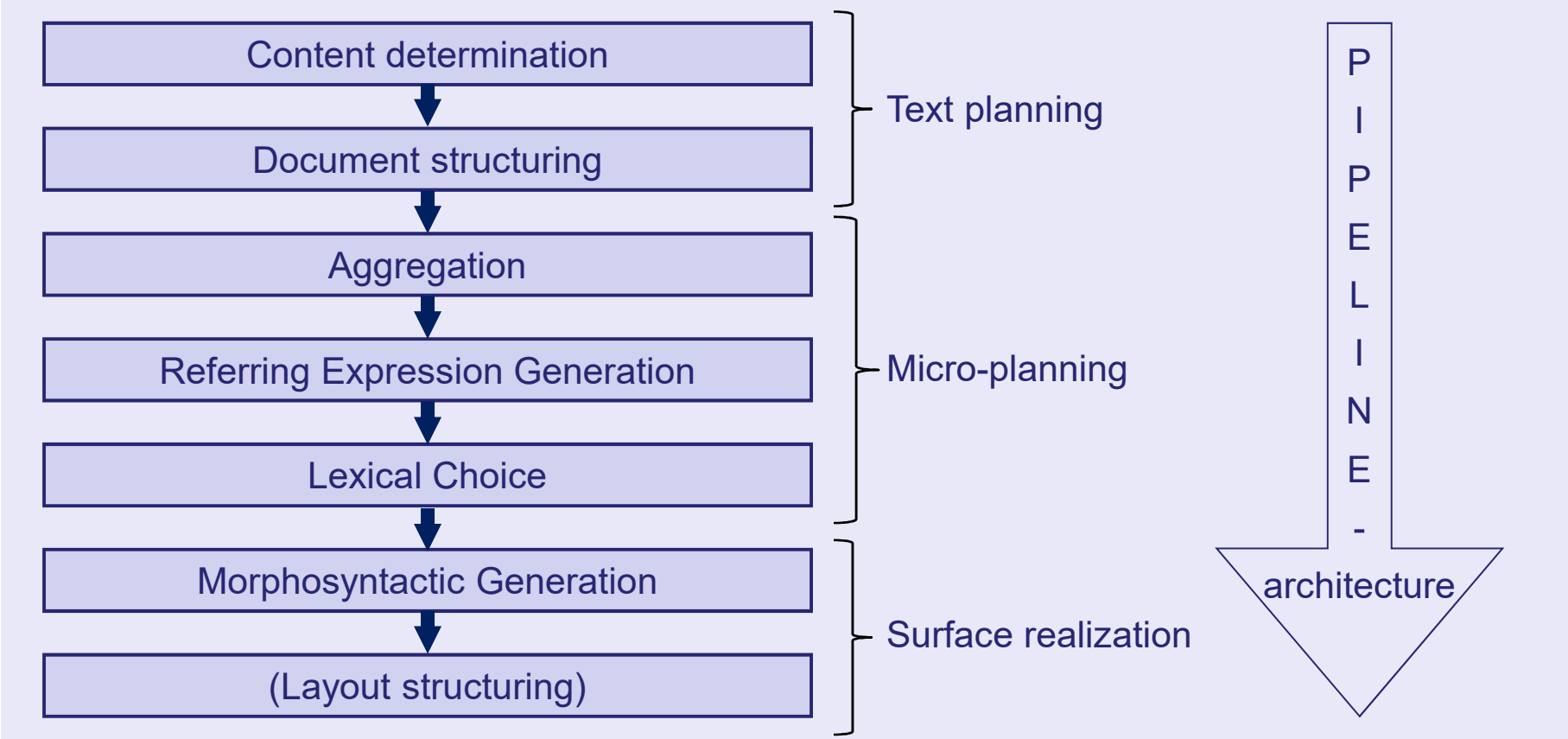
Key takeaways from the demo

- Synchronous,
- „from scratch“-generation with
  - Anaphora
  - Aggregation
  - Enumerations
  - Whole sentences with para- and hypotaxis
- Structure:
  - HIP
  - Current complaints
    - Precordial pain
    - Rhythm
    - Dyspnea
    - Edemata
    - …



Functional structure of „linguistic“ generation

aka „component tasks“



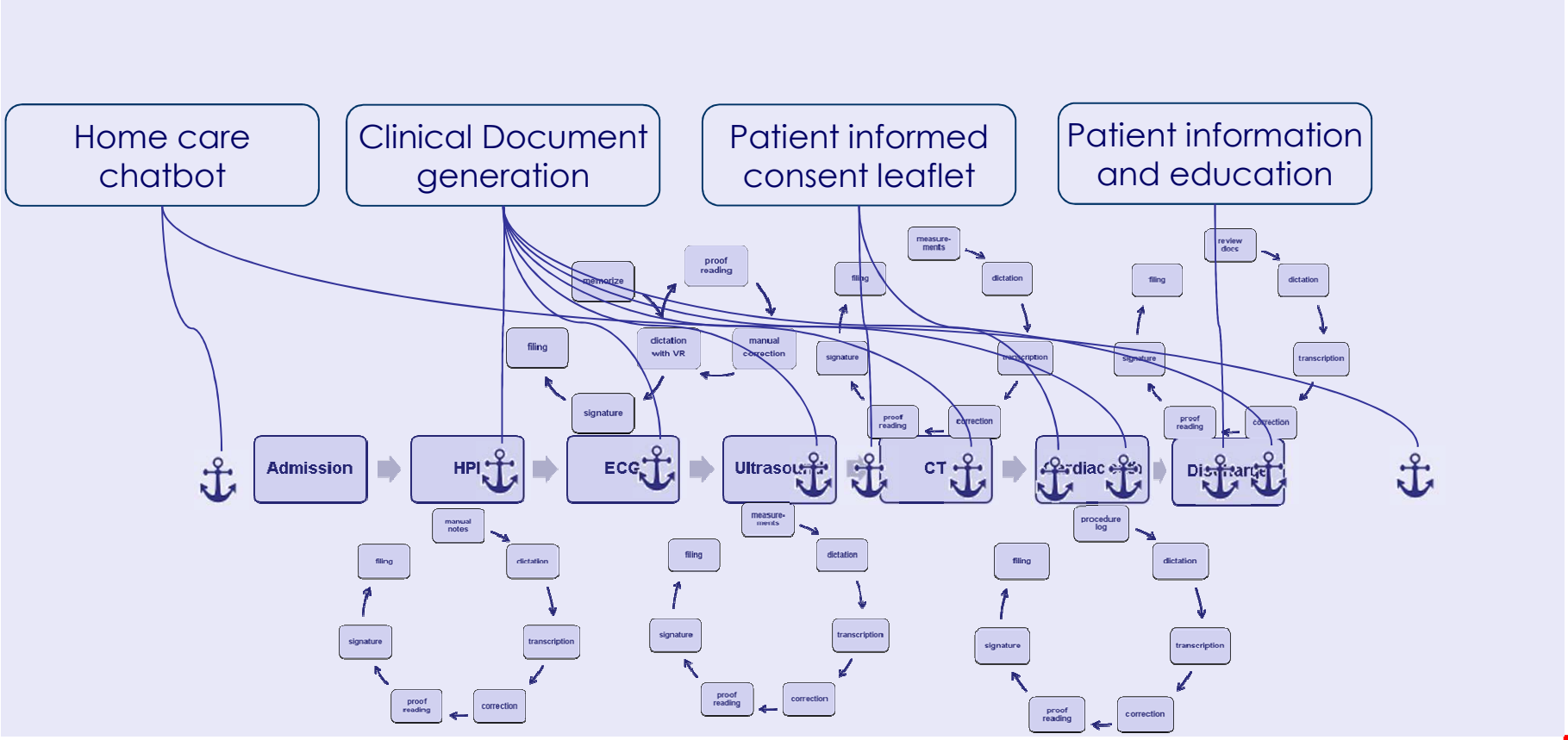
Further needed

Need to have and nice to have (?!)

- A very smart „lexicon“
  - Words („lexemes“, including word class, syntactical properties, flexion information…)
  - Simple semantic relations: Synonyms (antonyms, hyponyms, hyperonyms & meronyms…)
  - Pragmatics: Register, collocations, phraseologisms
- Linkage from „meaning“ to „lexeme“
- A model of the domain (not necessarily a full-fledged ontology!)
  - as anchor point for meaning-lexeme mapping, in particular
    - a better model for Hyponym-hyperonym relationship
    - to support lexical choice („A pathology with typeOf(stenosis) an locationOf(aorticValve) isCalled „aortic stenosis““)
  - to support (limited) inference

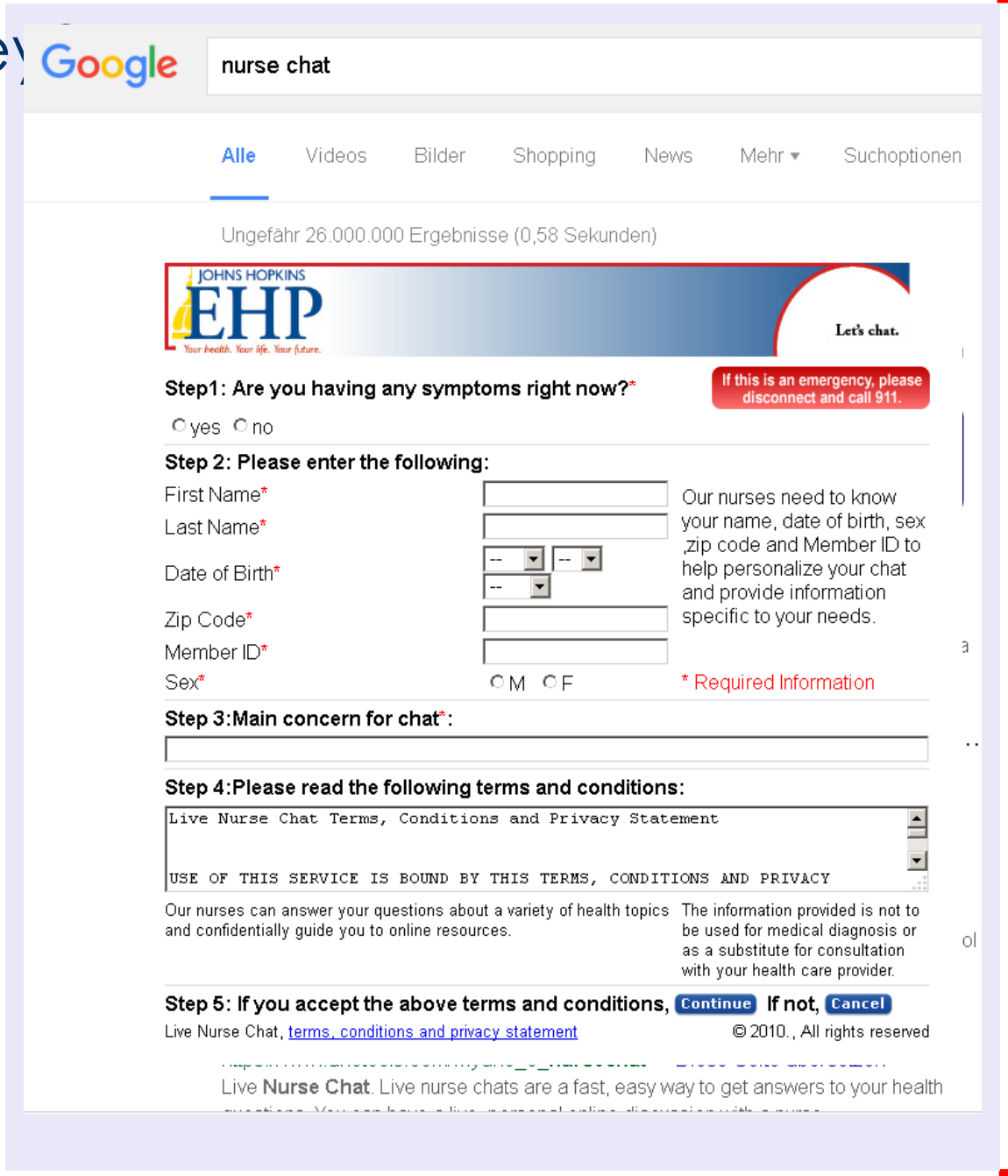
Where‘s the money?

Anchor points for NLG functionalities in Healthcare IT systems



Where‘s the money

Home care chatbot?



Where‘s the money?

Some stumbling blocks

- End-user configurability?
- Medical device?
- Evaluation/Validation?

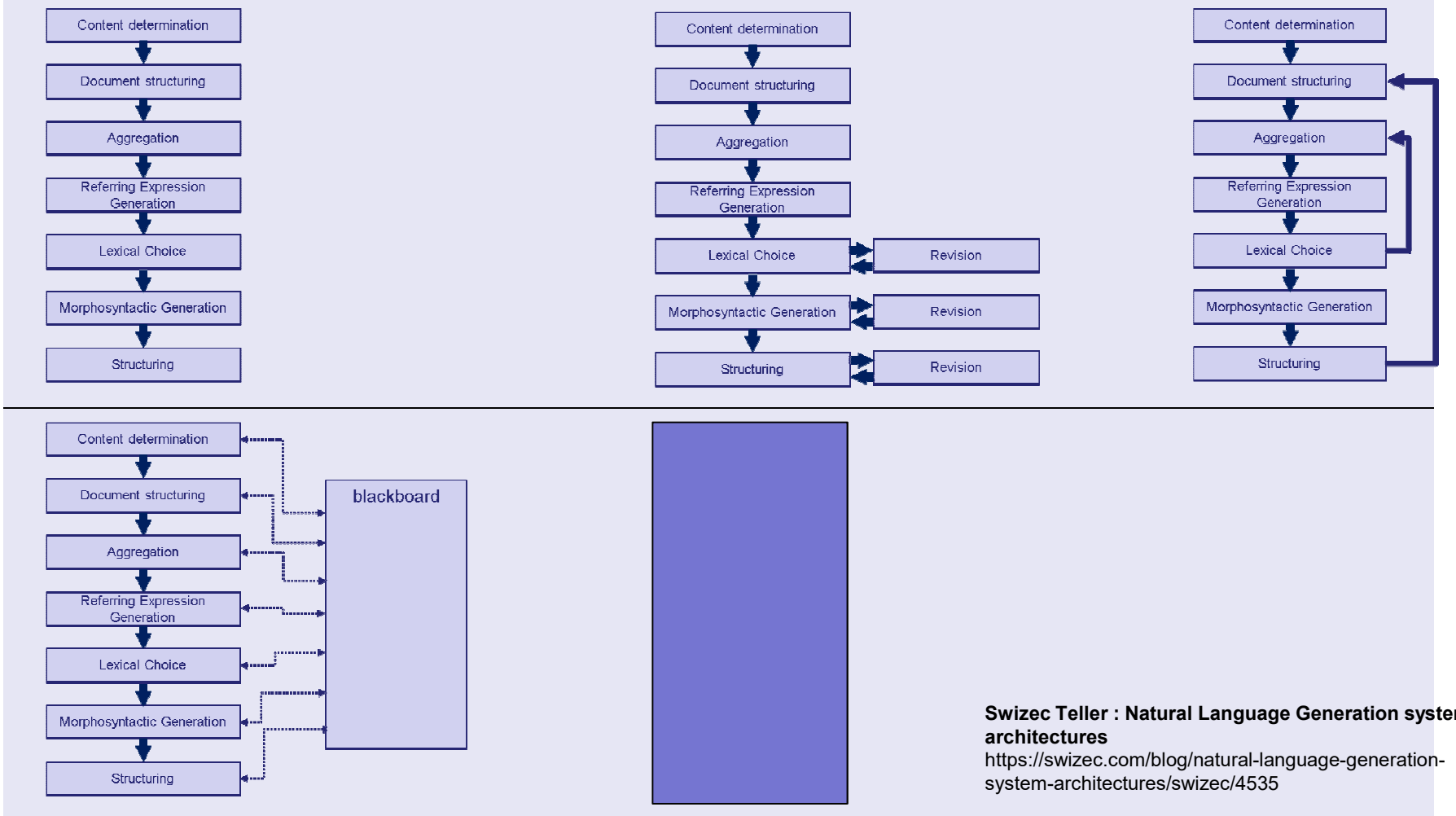
The End.

Thank you for your attention!

What are your questions?

Alternative „architectures“

(selected)



Swizec Teller : Natural Language Generation system architectures
https://swizec.com/blog/natural-language-generation-system-architectures/swizec/4535

Alternative „architectures“

Author-Reviewer and feedback

Author-Reviewer

Content determination

Document structuring

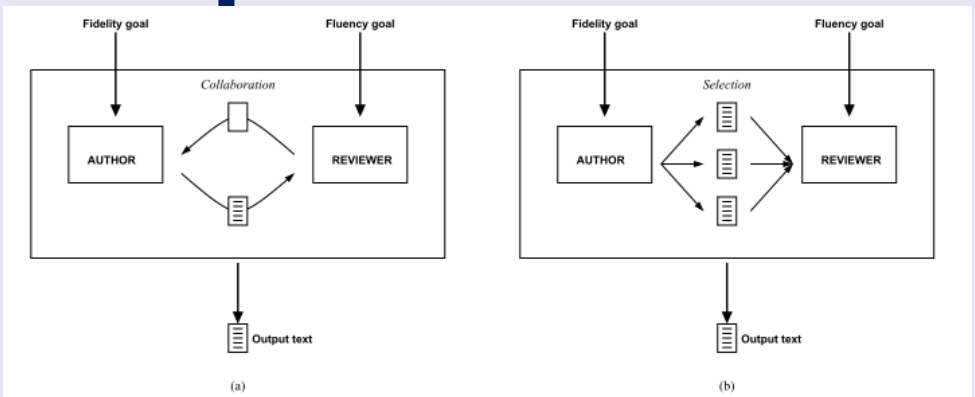


Figure 3.3: (a) The author-reviewer architecture of Oberlander and Brew (2000), and (b) as it is implemented in a NITROGEN-like system

(Layout structuring) ⇄ Revision

Feedback

Content determination

Document structuring

Aggregation

Referring Expression Generation

Lexical Choice

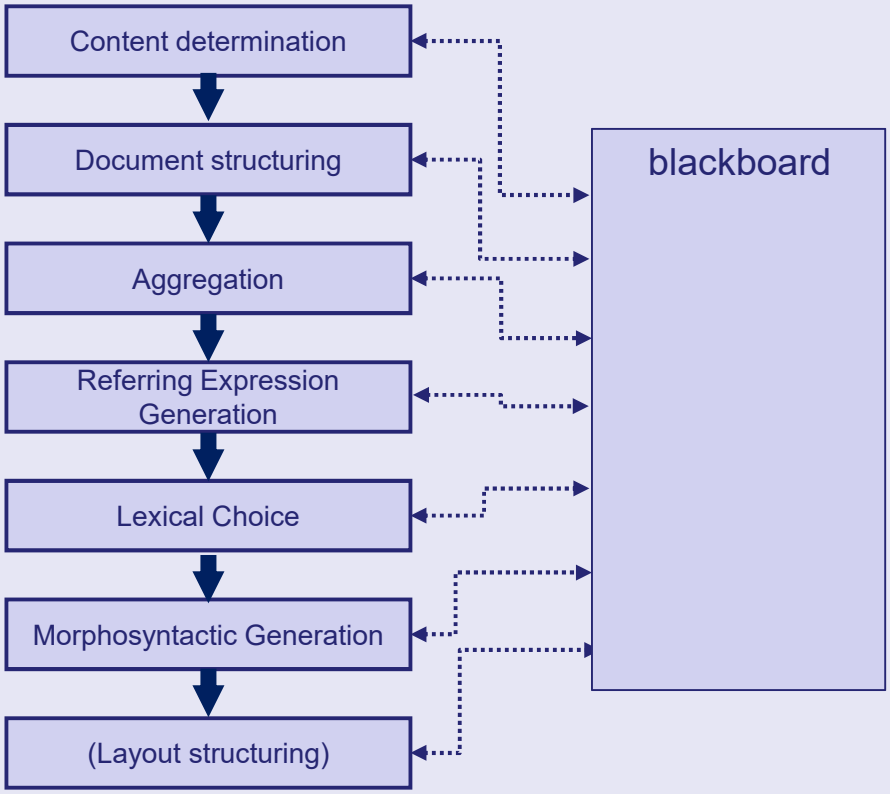
Morphosyntactic Generation

(Layout structuring)

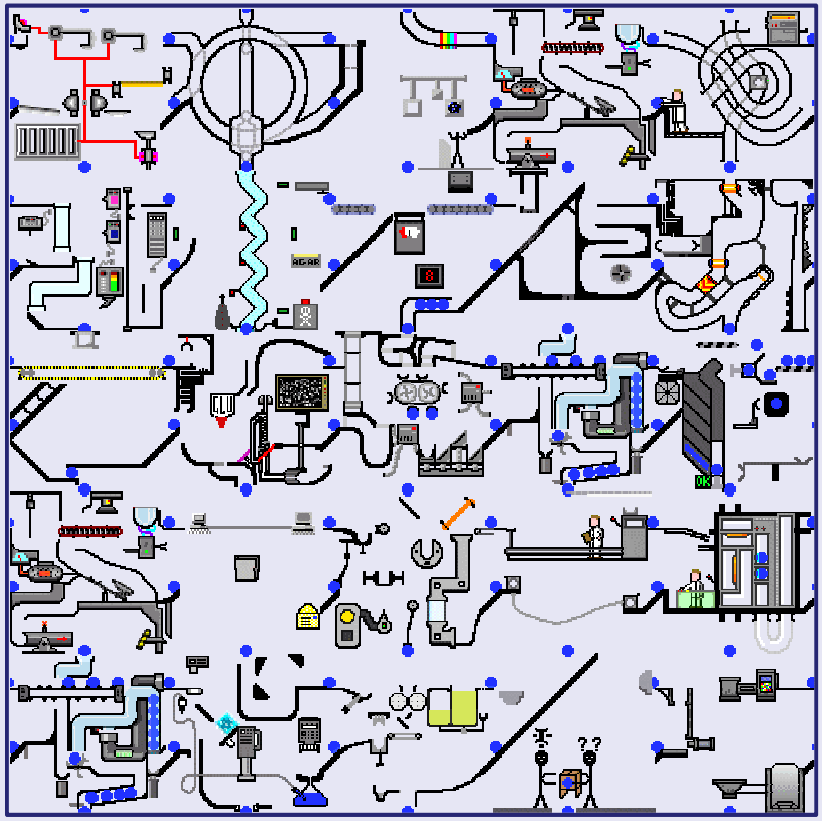
Alternative „architectures“

Blackboard and “Integrated”

Blackboard

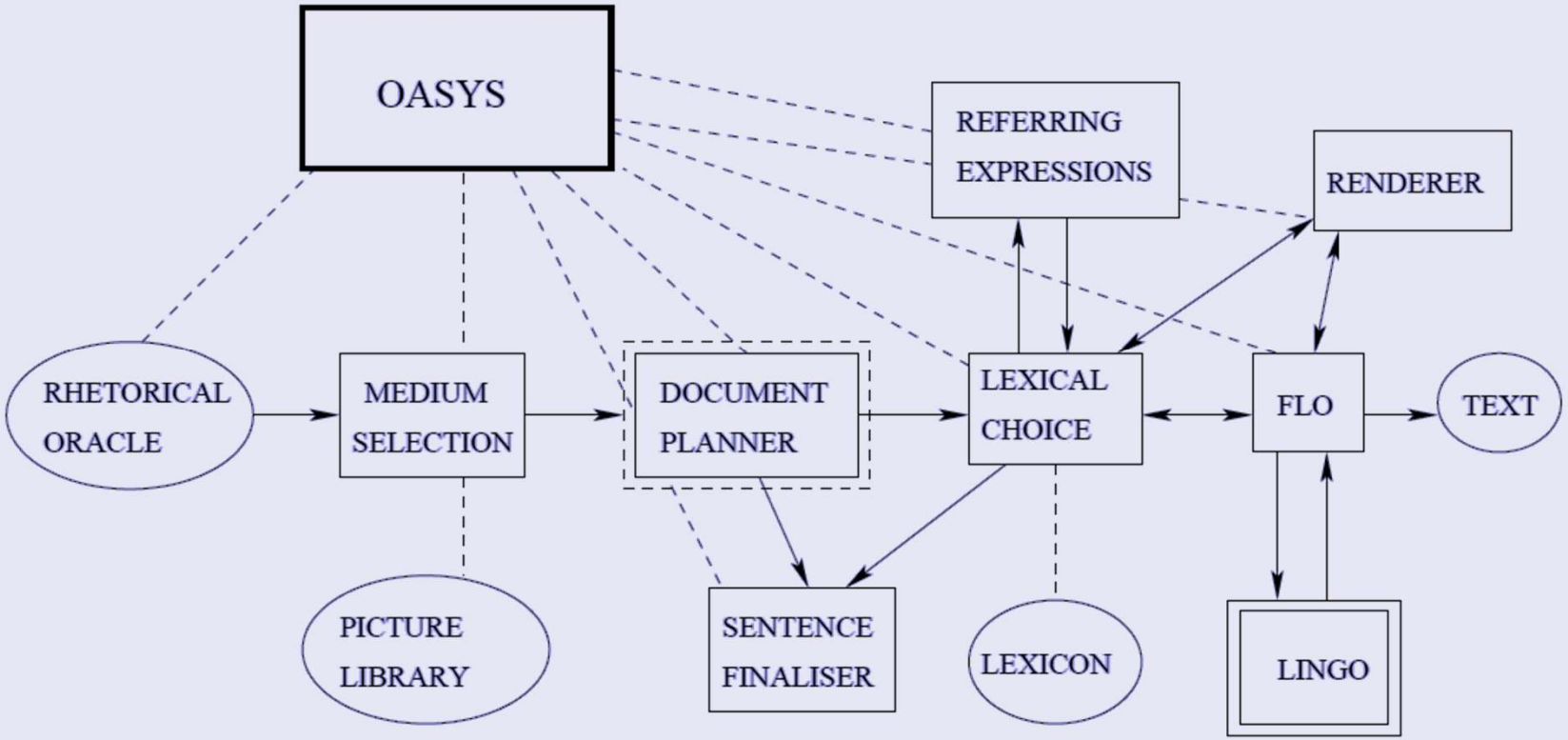


Integrated



Real life architectures

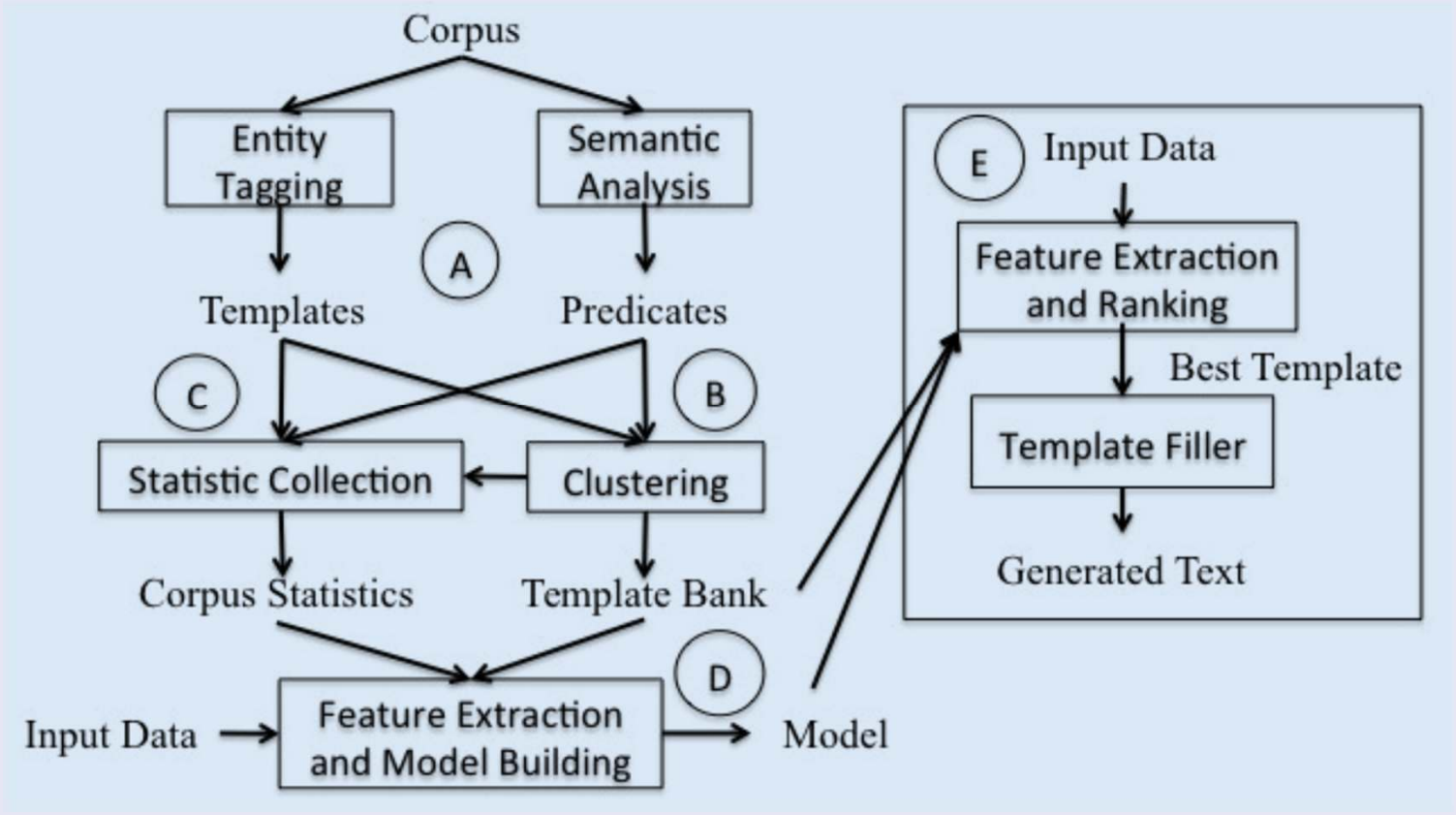
Real life



Cahill, L., Carroll, J., Evans, R., Paiva, D., Power, R., Scott, D., & van Deemter, K. (2001, July). From RAGS to RICHES: exploiting the potential of a flexible generation architecture. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics* (pp. 106-113). Association for Computational Linguistics.

Real life architectures

Real life



Schilder, F., Howald, B., & Kondadadi, R. (2013, August). Gennext: A consolidated domain adaptable nlg system. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 178-182).

Research questions
(selection)

- „Full NLG“ vs. Templates?
- Other use cases?
- Lexicon - Ontology
- Reusability, Extendability, Configurability
- Evaluation

Tell us more about Suregen-II

Literature

(by topic)

Introductory

- Reiter, E. and Dale, R., 2000. Building Natural Language Generation Systems. Studies in Natural Language Processing, 16. Cambridge University Press, Cambridge, 248 pp.
- Natural Language Generation: An Introduction http://www.inf.ed.ac.uk/teaching/courses/nlg/lectures/2012/NLG2012Lect1.pdf

Architecture

- Dealing with Dependencies between Content Planning and Surface Realisation in a Pipeline Generation Architecture http://www.dcs.shef.ac.uk/~kalina/papers/updated-ijcai01.pdf
- Natural Language Generation and Semantic Web Technologies http://www.semantic-web-journal.net/sites/default/files/swj315.pdf

Ontology-Lexicon

- WordNet as an Ontology for Generation https://hal.inria.fr/hal-01195793/document
- Automatic Report Generation from Ontologies: the MIAKT approach http://www.dcs.shef.ac.uk/~kalina/papers/nldb04.pdf

Templates?

- van Deemter, K., Krahmer, E., & Theune, M. Real vs. template-based natural language generation: a false opposition?. *May I Speak Freely*. http://utrecht.cs.utwente.nl/~theune/PUBS/templates-squib.pdf
- Schilder, F., Howald, B., & Kondadadi, R. (2013, August). Gennext: A consolidated domain adaptable nlg system. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 178-182).

Aggregation

- Reape, M., Mellish, C. (1999). *Just what is aggregation anyway?* in: *7th European Workshop on Natural Language Generation (EWNLG-99)*, (Toulouse), S. 20-29.
- Shaw, J. (1995). *Conciseness through Aggregation in Text Generation*. in: *33rd ACL (Student Session)*, S. 329-331.
- Shaw, J., McKeown, K. (1997). *An Architecture for Aggregation in Text Generation*. *IJCAI, Poster Session*.

Literature

(NLG for medical applications)

- Hüske-Kraus, D., 2003. Text Generation in Clinical Medicine - a Review. Methods of Information in Medicine, 1(42): 51-60.
- Kraus, D. (2000). *Suregen2 - a model based generator for surgical reports*. in: *Medical Informatics Europe (MIE-2000)*, (Hannover), IOS Press.
- Reiter, E., Robertson, R. and Osman, L., 2003. Lessons from a failure: Generating tailored smoking cessation letters. (to appear in:) Artificial Intelligence. www.csd.abdn.ac.uk/~ereiter/papers/aij03.pdf
- Bouayad-Agha, N., Power, R., Scott, D. and Belz, A., 2002. PILLS: Multilingual generation of medical information documents with overlapping content. ITRI-02-04, Information Technology Research Institute, University of Brighton, Brighton.
- Abella, A., Kender, J. and Starren, J., 1995. Descriptions generation of abnormal densities found in radiographs, Annual Symposium on Computer Applications in Medical Care. Hanley and Belfus, Philadelphia, pp. 542-546.
- Li, P.-Y., Evens, M. and Hier, D., 1986. Generating Medical Case Reports with the Linguistic String Parser, Fifth National Conference on Artificial Intelligence (AAAI-86). Morgan Kaufmann, Philadelphia, PA, pp. 1069-1073.
- Torgersson, O. and Falkman, G., 2002. Using Text Generation to Access Clinical Data in a variety of Contexts, Medical Informatics Europe (MIE-02). IOS Press.
- Wagner, J.C., Rogers, J.E., Baud, R.H. and Scherrer, J.-R., 1999. Natural language generation of surgical procedures. International Journal of Medical Informatics, 53: 175-192.
- Cawsey, A.J., Webber, B.L. and Jones, R.B., 1997. Natural language generation in health care. Journal of the American Medical Informatics Association, 4(6): 473-482.
- Pianta, E. and Tovena L.M.,1999. Mixing representation levels: The hybrid approach to automatic text generation, COLING-99

Literature

(referenced in this talk)

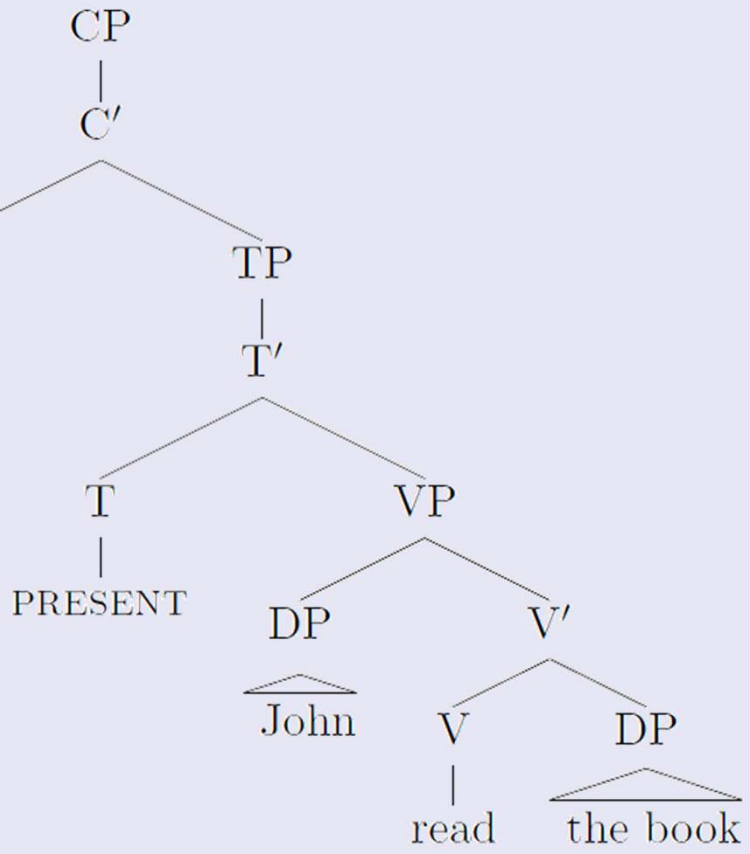
- Meteer, M.W. et al., 1987. MUMBLE-86: Design and Implementation. 87-87, COINS, University of Massachusetts.
- Pianta, E. and Tovena L.M.,1999. Mixing representation levels: The hybrid approach to automatic text generation, COLING-99
- Busemann, S., & Horacek, H. (1997, September). Generating air quality reports from environmental data. In *Proceedings of the DFKI Workshop on Natural Language Generation* (pp. 15-21).
- Pianta, E. and Tovena L.M.,1999. Mixing representation levels: The hybrid approach to automatic text generation, COLING-99

„Linguistic NLG“ vs. Templates

„Linguistic NLG“

John reads the book.

$T_1(X, Y)$ = “X reads the Y”.



„Linguistic NLG“:
The text is generated based (solely) on a syntactical model, taking into account sentence structures (taxis), word order, structure of nominal/verbal/adverbial phrases, governance, and flexion.

„Linguistic NLG“ vs. Templates

Templates

„Pure templates“:
The text is generated based on predefined („canned“) text chunks with slots into which variable text chunks are inserted

$$\begin{bmatrix} \text{LANGUAGE} & \text{german} \\ \text{COOP} & \text{threshold-exceeded} \\ \text{TIME} & \begin{bmatrix} \text{PRED} & \text{season} \\ \text{NAME} & \begin{bmatrix} \text{SEASON} & \text{winter} \\ \text{YEAR} & 1996 \end{bmatrix} \end{bmatrix} \\ \text{THRESHOLD-VALUE} & \begin{bmatrix} \text{AMOUNT} & 180 \\ \text{UNIT} & \text{mkg-m3} \end{bmatrix} \\ \text{EXCEEDS} & \begin{bmatrix} \text{STATUS} & \text{yes} \\ \text{TIMES} & 7 \end{bmatrix} \\ \text{DURATION} & \begin{bmatrix} \text{DAY} & 3 \end{bmatrix} \end{bmatrix}$$

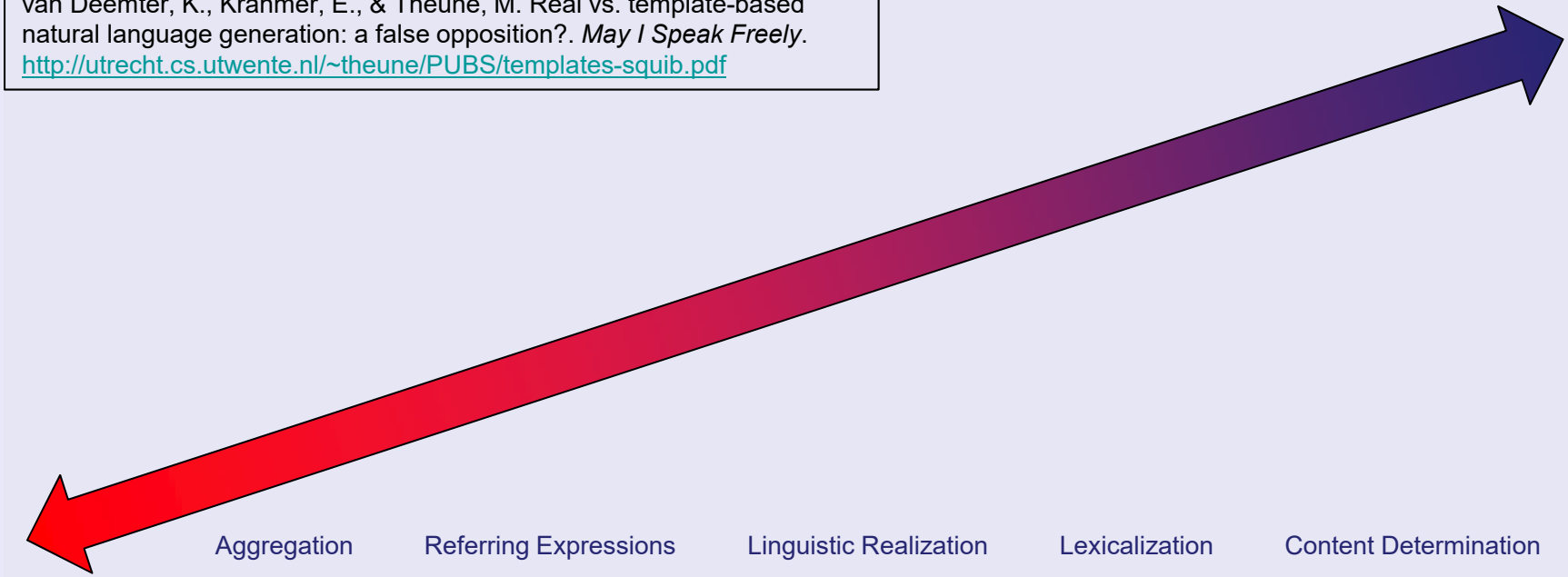
Der gesetzlich zulässige Grenzwert von 0.8 bar für Vortragslangeweile wurde am 7.9.2016 in Potsdam für 60 Minuten deutlich überschritten.

```
(defproduction wertueberschreitung "WU06"
  (:PRECOND (:CAT DECL
             :TEST ((pred-eq 'threshold-exceeded)
                    (not (threshold-type-p))))
   :ACTIONS (:TEMPLATE  "Der gesetzlich zulaessige Grenzwert von "
                        (:RULE VAL (get-param 'threshold-value))
                        (:OPTRULE POLL (get-param 'pollutant))
                        "wurde "
                        (:OPTRULE PPtime (get-param 'time))
                        (:RULE DUR (get-param 'duration))
                        (:OPTRULE SITE (get-param 'site))
                        (:RULE EXCEEDS (get-param 'exceeds))
                        ".")))
```

„Linguistic NLG“ vs. Templates

There‘s a continuum

Recommended reading:
van Deemter, K., Krahmer, E., & Theune, M. Real vs. template-based natural language generation: a false opposition?. *May I Speak Freely*.
http://utrecht.cs.utwente.nl/~theune/PUBS/templates-squib.pdf



Evaluation

Recall the criteria

- Completeness
- Correctness
- Adequacy: Relevance, usage of appropriate register and collocations/phraseology, lexemes, units of measurement, and familiar structure/reading cues; coherence…
- Conciseness: Maximum information in minimal amount of text
  - aggregation
  - ellipses
- Cohesion
  - anaphora, referring expressions
  - text deixis
  - rhetorical relations
- Practical: Inter- & intra-document variance

Evaluation

Types of evaluation

- **Task-based (extrinsic) evaluation:** "Give the generated text to a person, and assess how well it helps him perform a task (or otherwise achieves its communicative goal). For example, a system which generates summaries of medical data can be evaluated by giving these summaries to doctors, and assessing whether the summaries helps doctors make better decisions."
  - DHK: Works for some characteristics (conciseness, coherence), certainly not for other others (completeness, correctness)
- **Human ratings:** "Give the generated text to a person, and ask him or her to rate the quality and usefulness of the text."
  - DHK: Works for some characteristics (conciseness, coherence, adequacy), certainly not for others (completeness, correctness)
- **Metrics:** "Compare generated texts to texts written by people from the same input data, using an automatic metric such as BLEU."
  - DHK: Works for some characteristics (conciseness, adequacy?), certainly not for others (completeness, coherence, correctness) unless a substantial

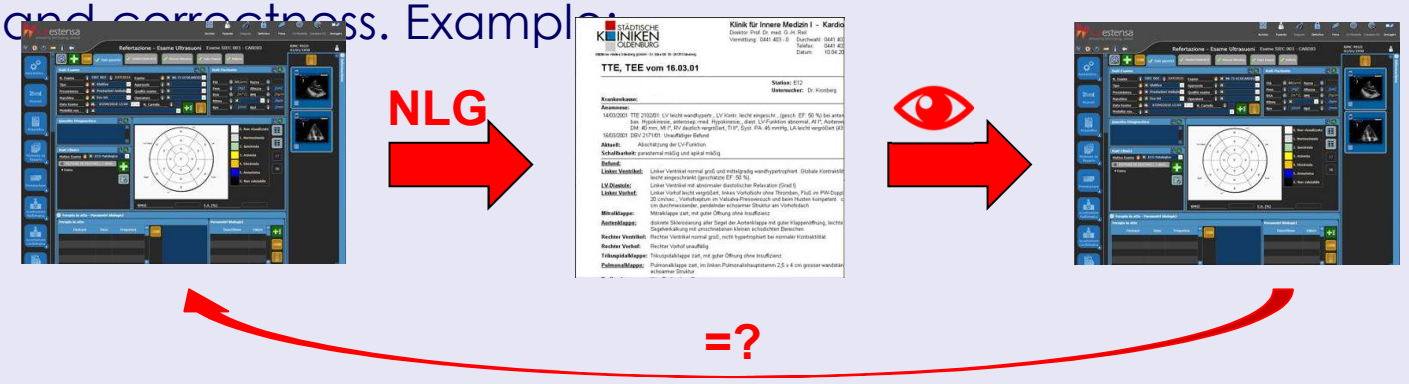
Evaluation

Conclusion & suggestions

- There doesn't seem to be a simple method to cover all relevant aspects. Some methods seem to be of questionable practicality.
- Suggestion 1 DHK: Use CLOZE-method to gauge adequacy, coherence, cohesion

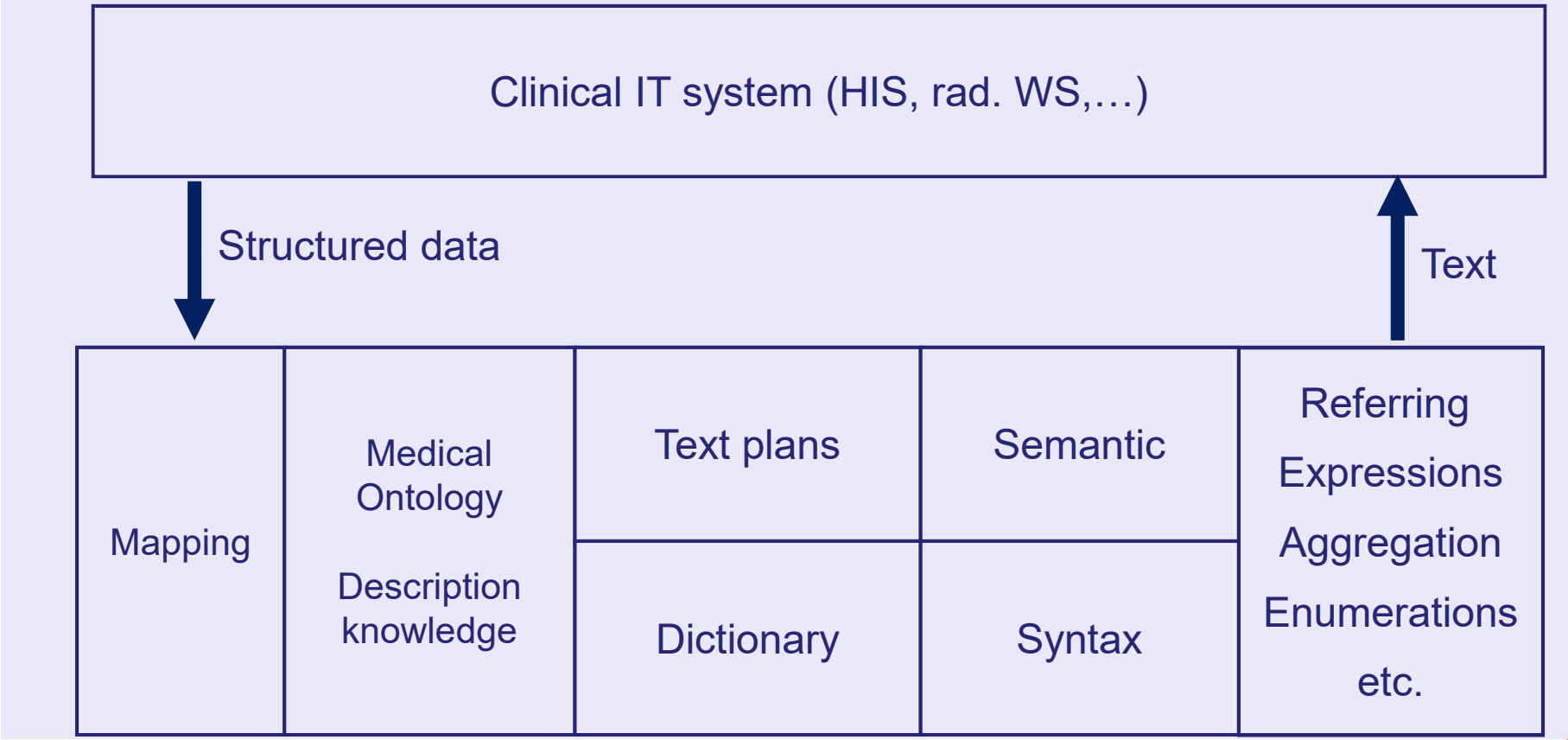
There was a tortuous sigmoid colon, but with slow insertion, we were able to get through that area, and with moderate pressure, we got to the right hemicolectomy anastomotic site and we identified small bowel. Right next to the anastomosis was a diverticulum. As we withdrew the scope in the distal transverse, there were two 6 mm sessile polyps, removed by cold biopsy forceps.

- Suggestion 2 DHK: Use "feed back"-method to gauge completeness and correctness. Example:



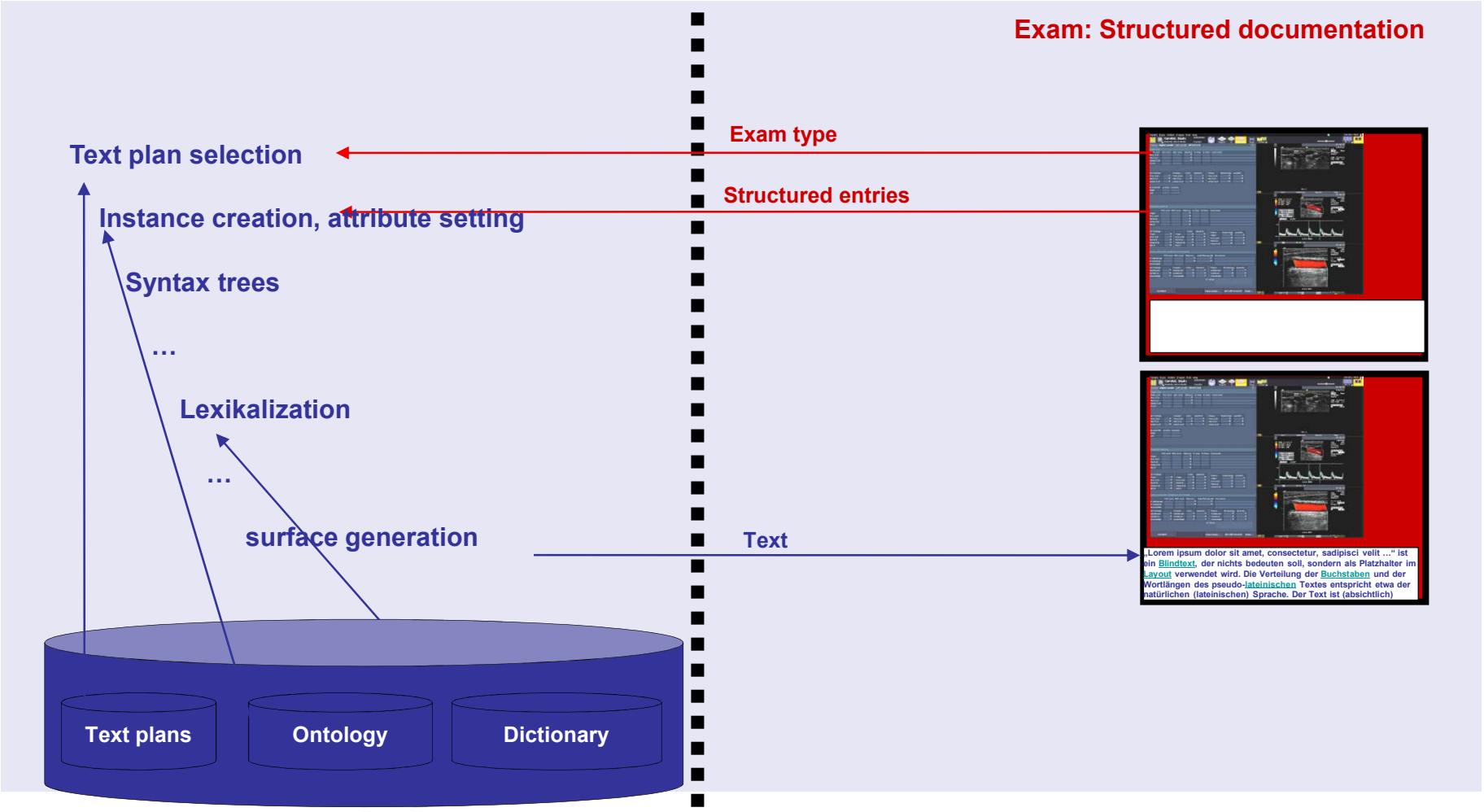
Suregen-II

Building blocks and integration



Suregen-II

Dynamic 50.000 ft. view



Suregen-II

Necessary assets

Steps:

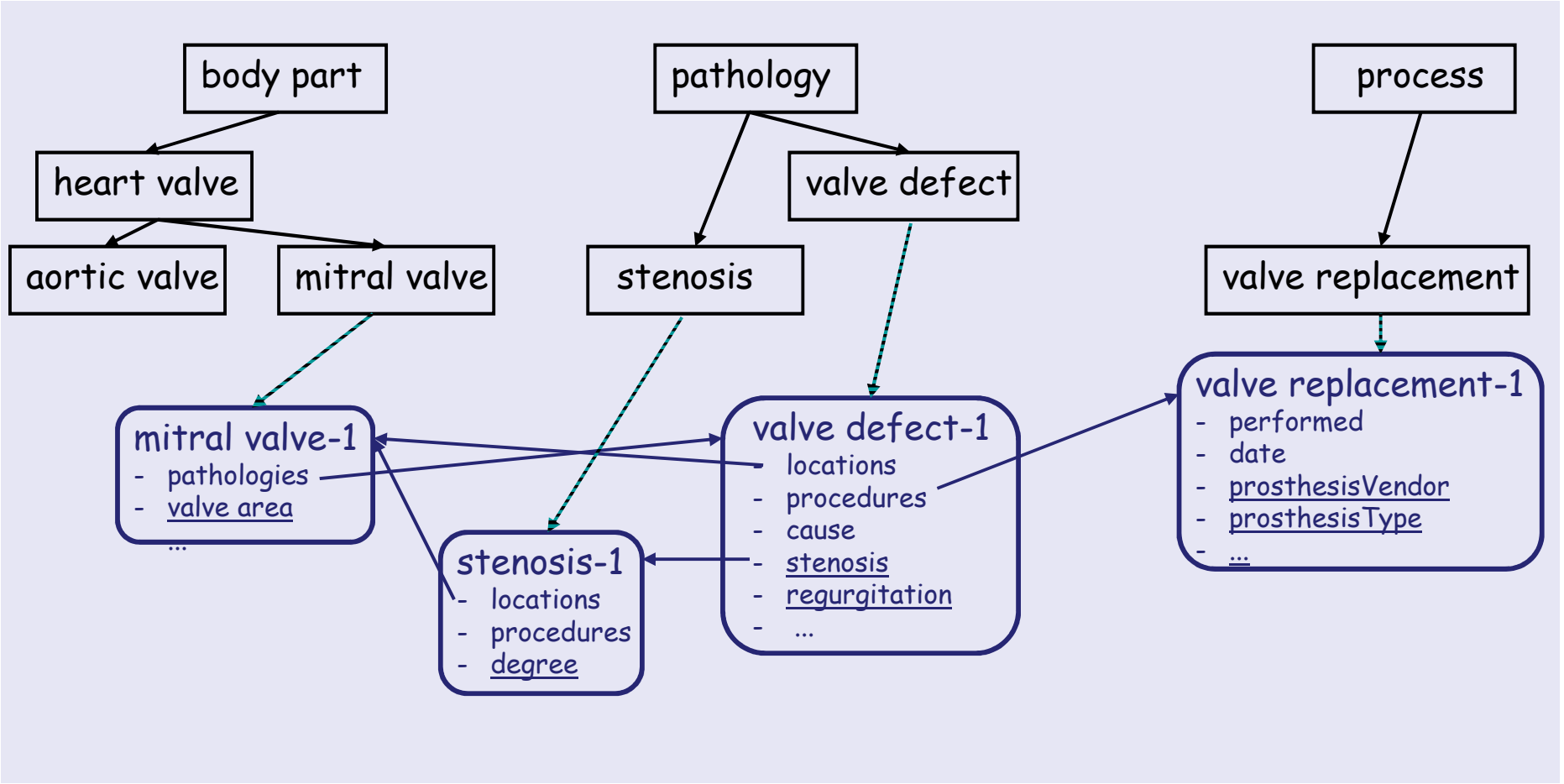
- Exam type drives selection of text plan
- Structured entries create instances of an ontology and set their attributes/relations
- With every update the textplan asks the referenced instances to describe themselves – recursively.
- In the resulting syntax tree referring expression, inflected forms etc. are inserted
- The text is sent back

Preconditions:

- Text plans for all exam types
- Medical ontology for all relevant entities
- Description knowledge, semantic schemes, dictionary…
- Flexion, Enumerations, References…

Suregen-II

Classes, Instances and Attributes



Suregen-II

Classes carry „description knowledge“ expressed in a simple formalism

```
(general-clause
    :head (CHASES/S-V-O_two-explicit-args
          (general-np
            :head (np-proper-name „Fluffy“)
            :accessories (:number singular :gender masculine
                   :person third :determiner-policy no-determiner))
          (general-np
            :head (np-common-noun “mouse”)
            :accessories (:number plural :gender neuter
                  :person third :determiner-policy initially-indefinite)
            :further-specifications
              ((:attachment-function restrictive-modifier
                :specification (predication-to-be *self*
                                (adjective “little”))))))
    :accessories (:tense-modal present :progressive :unmarked))
```

Specification of the sentence “Fluffy is chasing little mice” in MUMBLE, Example from {Meteer, 1987 #375}

```
(main-clause
    :subject „Fluffy“
    :predicate “to chase”
    :object (Noun-Phrase  :noun “mouse”
                :adjective “little” :numerus :plural)
    :tense :presentProgressive)
```

Specification of the sentence “Fluffy is chasing little mice” in SUREGEN-2

Suregen-II

Classes carry „description knowledge“, examples

Class definition

„A heart valve replacement is a iatrogenic process with the attributes prosthesisVendor and prosthesisSize“

```
(SuregenConcept HeartValveReplacement
                :is-a SuregenIatrogenicProcess
                :has ((prosthesisVendor :default NIL)
                      (prosthesisSize :default NIL)))
```

➔ Inheritance of attributes like „date“, „institution“ etc.

Description in natural language

„To describe a heart valve replacement as a noun phrase do:
If it has been performed already then use a noun phrase built with „*Klappenersatz*“ as head, „*Z.n.*“ as preposition, the date of the procedure as an attribute and, if any of the prosthesis vendor or prosthesis size attributes are given, put these as an apposition in parentheses.

Description in Suregen-2

```
(ToDescribe :a HeartValveReplacement :as :NP
   :use (IF (my :performed)
            (Noun-Phrase :noun "Klappenersatz"
                :preposition "Z.n."
                :attribute (DescribeS (my :endDate)
                 :facet :date :as :ISO-date)
                 :apposition (Parenthesized (my :prosthesisVendor)
                                            (my :prosthesisSize)))))
```

```
(ToDescribe :the :valveReplacement
      :of-a ValveDefect :as :NP
      :use (DescribeS
      (GiveMe :the
'HeartValveReplacement
            :ofthe 'procedures :of it)
            :as :NP))
```

Suregen-II

Overview: Generation algorithm

- Text plan selection
- Incremental substitution of Description Templates
- Aggregation, referential expressions, enumerations etc.
- Morphosyntactic Realisation

Suregen-II

Generation algorithm

incremental substitution of description templates in the text plan

- `(DescribeS anamnesis :as :text)`
- ```
  (SyntacticalSequence
    (DescribeS occupationalStatus :as :MC)
    (DescribeS CurrentComplaints :as :MC)
    (DescribeS HistoryOfDiseases :as :MCs))
  ```
- ```
  (SyntacticalSequence
    (Main-Clause :subject (Make-Reference :patient)
       :predicate „arbeiten“
       :adverbial (DescribeS (Profession *patient* :as :NP)))
    (Main-Clause ......))
  ```
- `...`

Suregen-II

Requirements for the ontology (or the lexicon?)

- **Inheritance (also of description knowledge)**
- **Defaults**
  - A human being has two arms and two legs
- **Inference**
  - When there's an instance of type „valveReplacement", there must have been a previous valve defect.
- **Conceptual Aggregation**
  - Any pathology present in the lower arm and the upper arm is present in the entire arm
- **Terminological Aggregation (?)**
  - When both a regurgitation and a stenosis is present, the defect is call „combined defect"
- **„Pragmatic" interpretation of partitive relations**
  - When both left and right atria as well as left and right ventricles are enlarged, the entire heart is said to be enlarged.

Suregen-II

Requirements for the ontology

- **Restricted transitivity of„pathology-at“-relation along the „part-of“-relation**
  - exhaustivePartition(arm, upper_arm, lower_arm) ∧ pathology(phlegmon, upper_arm) ∧ pathology(phlegmon, lower_arm) ⇒ pathology(phlegmon, arm)

  but

  - exhaustivePartition(arm, upper_arm, lower_arm) ∧ pathology(fracture, upper_arm) ∧ pathology(fracture, lower_arm) ⇏ pathology(fracture, arm)

Suregen-II

Requirements for the ontology (or the lexicon?)

- Terminological and conceptual knowledge needs to be represented (language dependent)
  - A Stenosis which is located at the aortic valve is called „aortic stenosis“
    `(AssertThat :a HeartValveStenosis which (Is-Located l-AV) :is-called "Aortenstenose")`
  - An edema which is located at the tibia is called „pretibial edemea“
- Collocations/phraseologisms needs to be represented (language dependent)
  - Führt man eine Prozedur bei der Randbedingungung durch, dass eine andere Prozedur läuft, verwendet man „unter“ mit der Prozedurbeschreibung: „Unter Intubationsnarkose zeigte der Patient zunächst…“, „Unter extrakorporaler Zirkulation…“
  - A complaint mentioned by the patient in the first encounter of a hospital visit is described by „<the patient> presented with <complaint>“
- Many more peculiarities (language dependent)
  - E.g.: A missing pathology can be described in singular form („no organomegaly“) or in plural form („no heart murmurs“). There seems to be a regularity which needs to be modelled.

Fin

```
(main-clause
   :subject (Make-Reference :speaker)
   :predicate “to thank for”
   :object (Make-Reference :audience)
   :attribute (Noun-Phrase :noun „attention“
              :preposition „for“ :article :possessive))

"I thank you for your attention“
```

Backup slides

Backup slides