Shades of Certainty – Working with Swedish Medical Records and the Stockholm EPR Corpus

Sumithra VELUPILLAI, Ph.D.

Oslo, May 30th 2012

Health Care Analytics and Modeling, Dept. of Computer and Systems Sciences (DSV)

Health Care Analytics and Modeling

- Four professors, two physicians, nine PhDs and three PhD students
- NLP and Text Mining on (Swedish) clinical data -->

The Stockholm EPR Corpus

- Stockholm City Council
- ~ 1 million patients
- ~ 900 clinical units
- ~ 23 000 users
- 2006 – 2008 (plus newer now)

Ethics

- Approval from Regional Vetting board (Etikprövningsnämnden)
- De-identified with respect to names and social security number
  - Personal information still in free-text
  - Secure storage

The medical records

- Structured information
  - Gender, age, admission and discharge date, ICD-10 code, categories (specific to departments)
- Unstructured information (free-text)
  - Documentation – spelling errors, jargon, domain-specific abbreviations, etc.
  - Still a lot of personal information

Stockholm University

Internal Projects

- Automatic de-identification
- Automatic identification of symptoms, diseases, diagnoses
- Automatic assignment of diagnosis codes (ICD-10)
- Co-morbidity networks
- Linguistic characterization

External Projects

- Interlock – Inter-language collaboration in clinical NLP
  - DSV (me) and UCSD School of Medicine's Division of Biomedical Informatics (Dr. Wendy Chapman), Supported by the Stockholm University Academic Initiative
- HEXAnord - HEalth TeXt Analysis in the Nordic and Baltic Countries
  - Sweden, Finland, Norway, Denmark, Estonia, Lithuania

  Supported by Nordforsk – The Nordic Council of Ministers

Stockholm University

External Projects

- NICTA-Australia collaboration
  - Dr. Hanna Suominen and Dr. David Martinez
  - Text mining of invasive fungal infections
    - Decision support for clinicians
- High-Performance Data Mining for Drug Effect Detection (DADEL)
  - Dept. Computer and Systems Sciences, 5 years, 1/5 NLP
  - Funded by the Swedish Foundation for Strategic Research

External Projects

- Detect-HAI
  - Detection of Hospital Acquired Infections through language technology
  - Collaboration with Karolinska University Hospital
- Automatiserad översättning av röntgensvar till allmänsvenska - ett led i demokratiseringen av sjukvården
  - Making medical records understandable for patients

Stockholm University

Shades of Certainty

- My PhD topic..!

Motivation

- Improve information access
  - Reasoning documented in (EPR) free-text
    - speculations, negations, affirmations
    - important to distinguish
  - Accurate and situation-specific information
  - Overviews/summaries: these diagnoses have been affirmed, negated, …
  - Capture reasoning → deepened knowledge

Aim and Objectives

- Build automated information access systems
  - create annotation schema for modeling certainty levels
  - apply on Swedish clinical documentation
  - gain empirical understanding for qualitative analysis and use for automatic classifiers

  - → adverse event surveillance, drug side-effects, decision support, summaries, …

Approaches

- Two annotation initiatives
  - Sentence level, “naive” annotators
    - Different clinical departments
  - Diagnostic statement level, domain-expert annotators
    - One clinical department
- Automatic classification
  - E-health scenarios

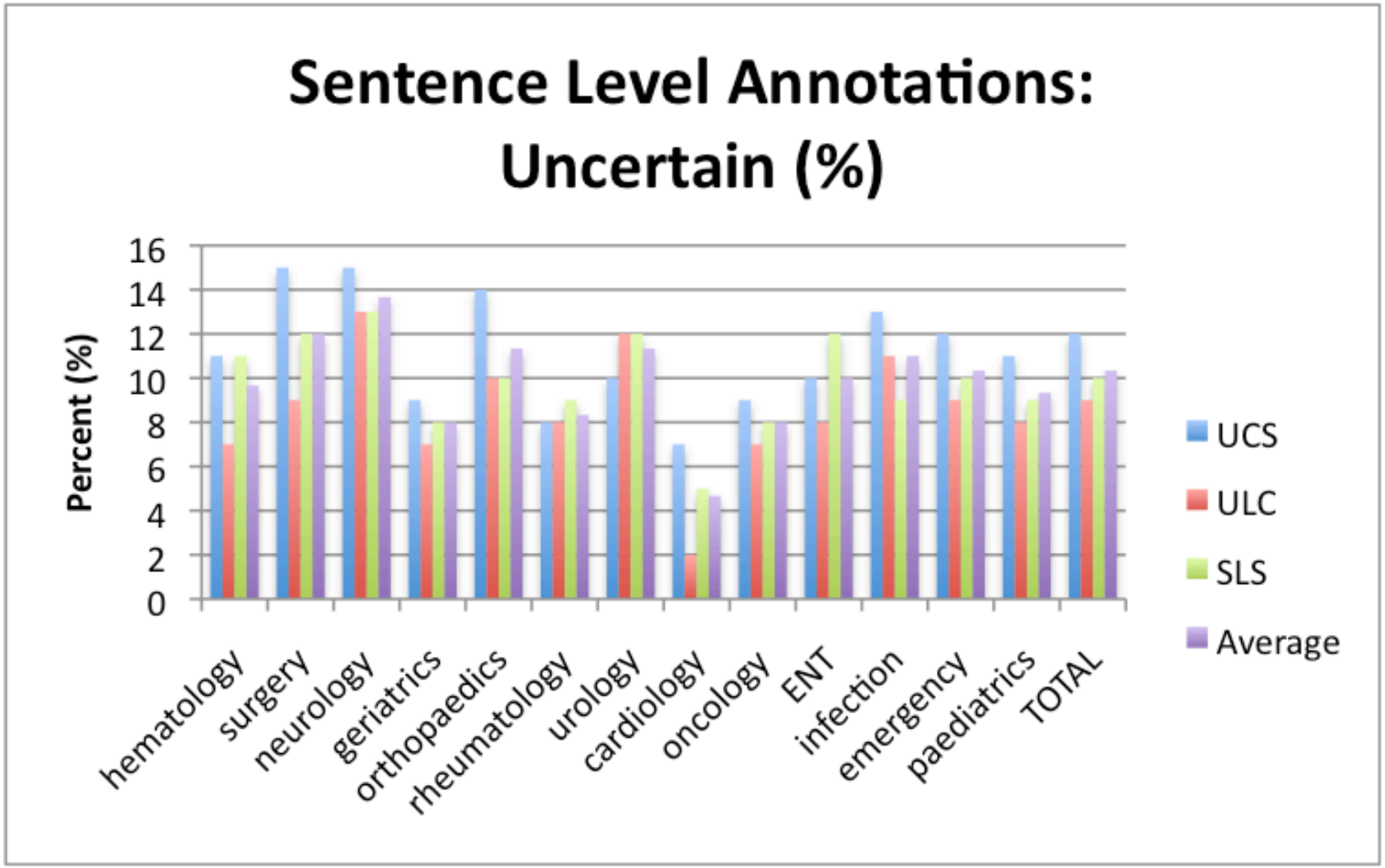
Stockholm University

Sentence level: Comparison over clinics

Geriatric clinics contain less uncertain expressions

Neurology clinics contain most amount of uncertain expressions

Uncertain expressions are often longer


Sentence Level Annotations:
Uncertain (%)
Percent (%)
16
14
12
10
8
6
4
2
0
hematology
surgery
neurology
geriatrics
orthopaedics
rheumatology
urology
cardiology
oncology
ENT
infection
emergency
paediatrics
TOTAL
UCS
ULC
SLS
Average

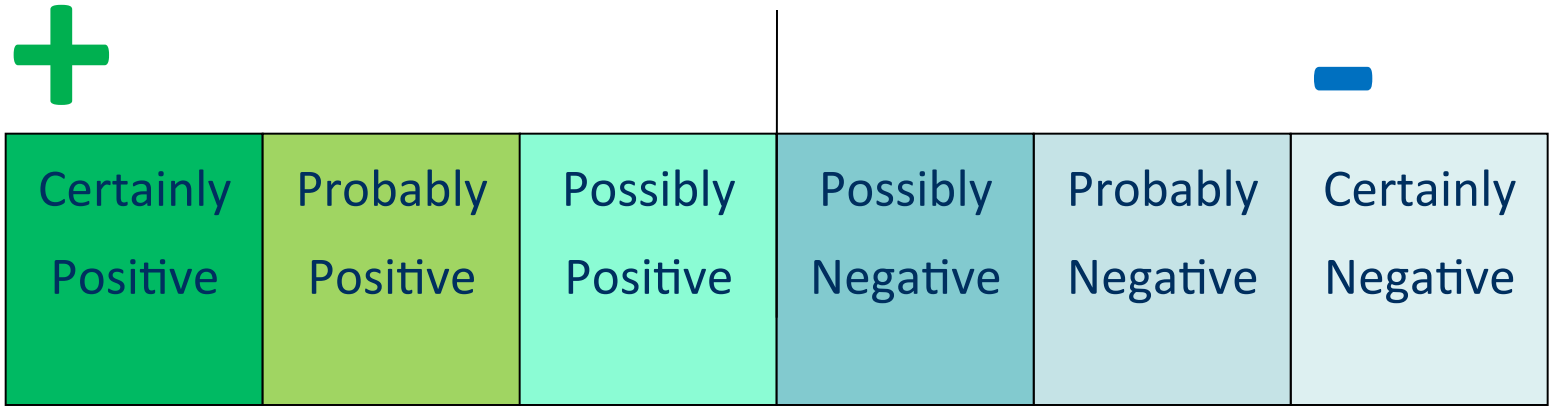
Diagnostic statement level annotation

- Creation of diagnosis list (~300 diagnostic statements)
- Annotation guidelines and annotations
  - Two senior physicians (“domain experts”)
- Emergency ward, assessment entries
  - Stockholm EPR Corpus

Diagnostic statement level: example

Oklart vad pats symtom kan komma av. Ingen säker <D>infektion</D>. Inga tecken till inflammatorisk sjukdom eller <D>allergi</D>. Reflux med irritation av luftrör och sledes hosta? Dock har pat ej haft några symtom på <D>refluxesofagit</D>. Ingen ytterligare akut utredning är befogad. Hänvisar till pats husläkare för fortsatt utredning.

Unclear what patient's (abbr.) symptoms arise from. No certain <D>infection</D>. No signs of inflammatory disease or <D>allergy</D>. Reflux with irritation of airways and therefore cough? But pat has not had any symptoms of <D>refluxoesophagitis</D>.No further urgent investigation required. Refer to pats GP for continued investigation..



Patient has Parkinsons disease.

Physical examination strongly suggests Parkinson.

Patient possibly has Parkinson.

Parkinson cannot yet be outruled.

No support for Parkinson.

Parkinsson can be excluded.

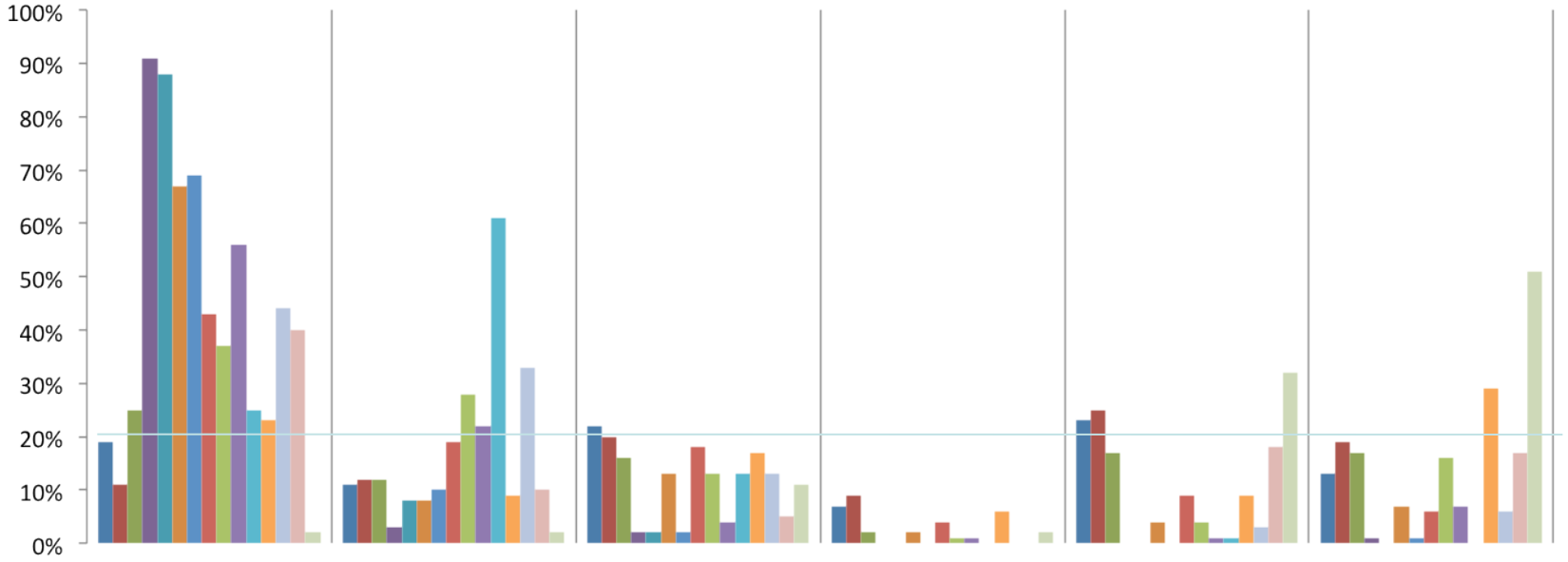
Stockholm University

Results

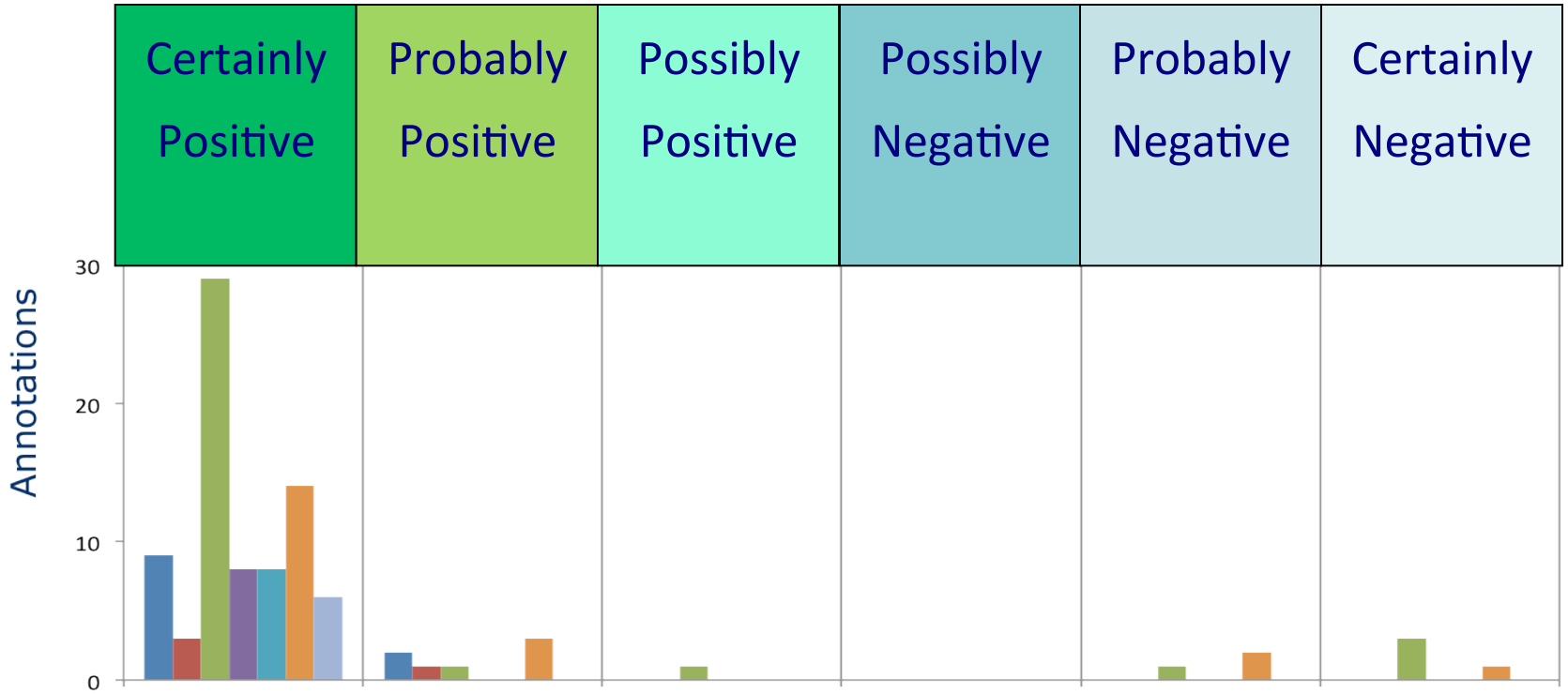
- Intra- and Inter-Annotator Agreement
  - 0.7/0.58 F-measure, 0.73/0.6 Cohen's κ, 0.88/0.82 κ_w
- *Certainly Positive* clear majority (approx. 50%)
  - High IAA: 0.9 F-measure
- *Possibly Negative* very rare
- Discrepancies in intermediate classes
  - 1-step most common

Stockholm University

Certainly Positive	Probably Positive	Possibly Positive	Possibly Negative	Probably Negative	Certainly Negative

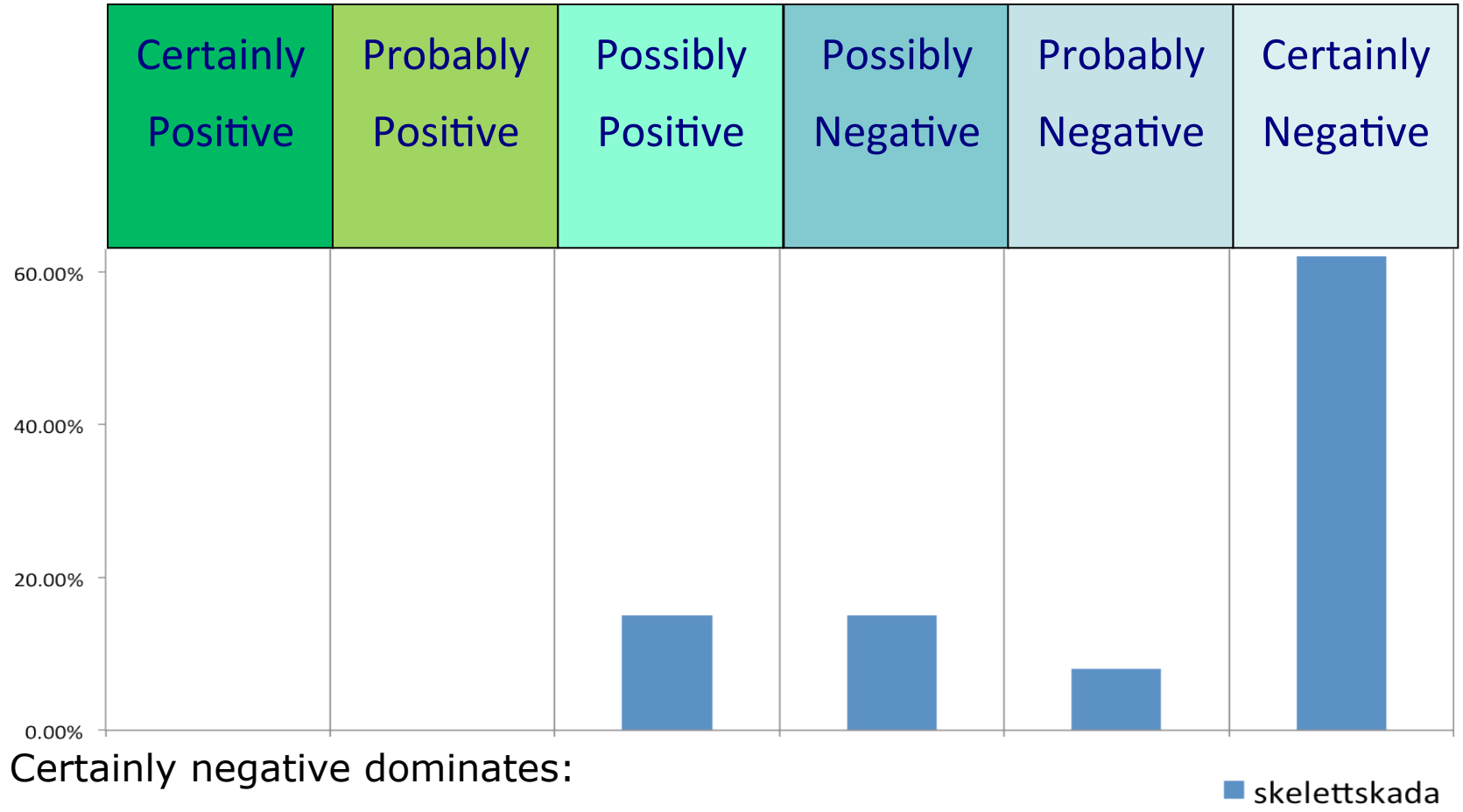


Annotation classes for 15 diagnoses



Certainly positive dominates:

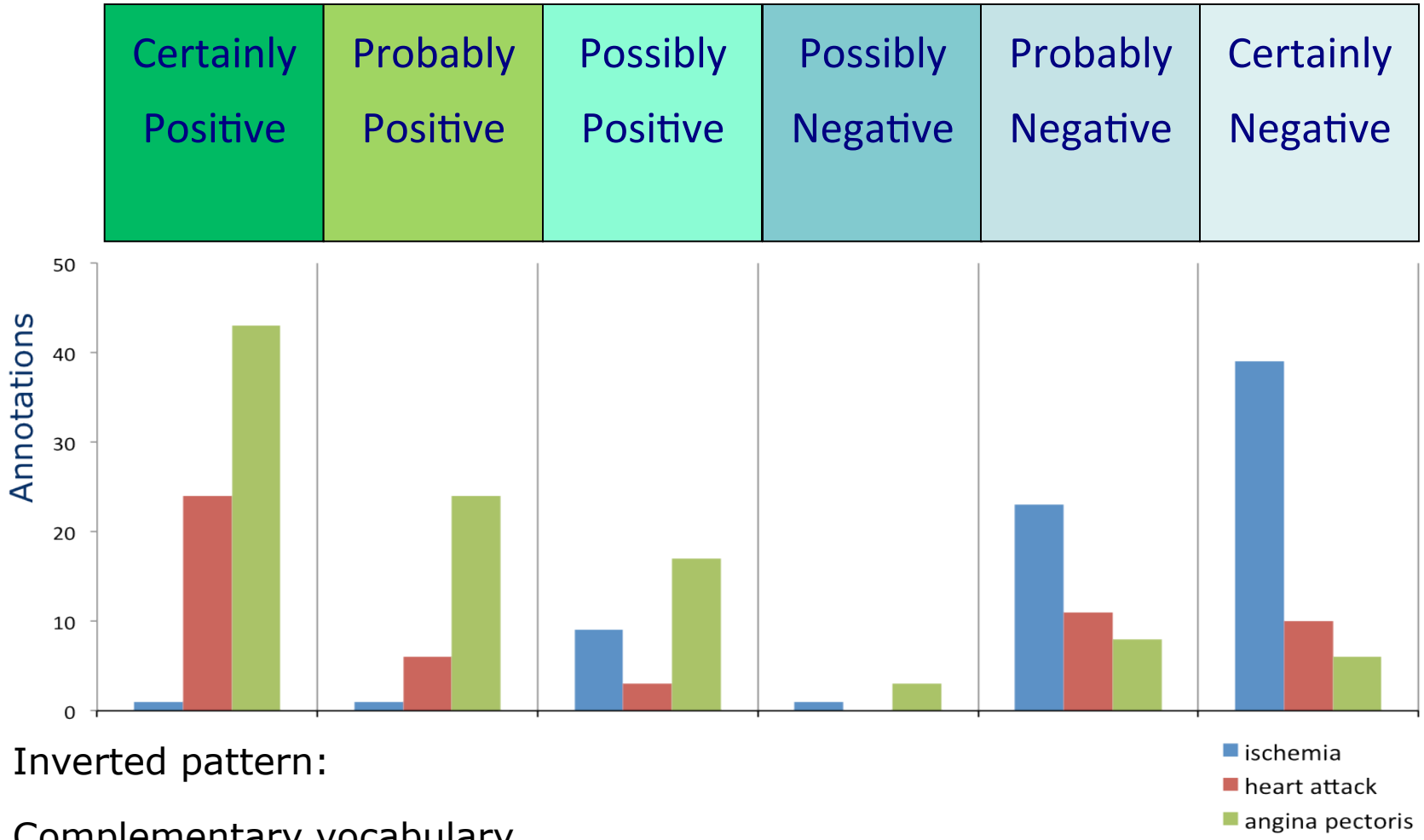
diagnosis shows on the outside



Certainly negative dominates:

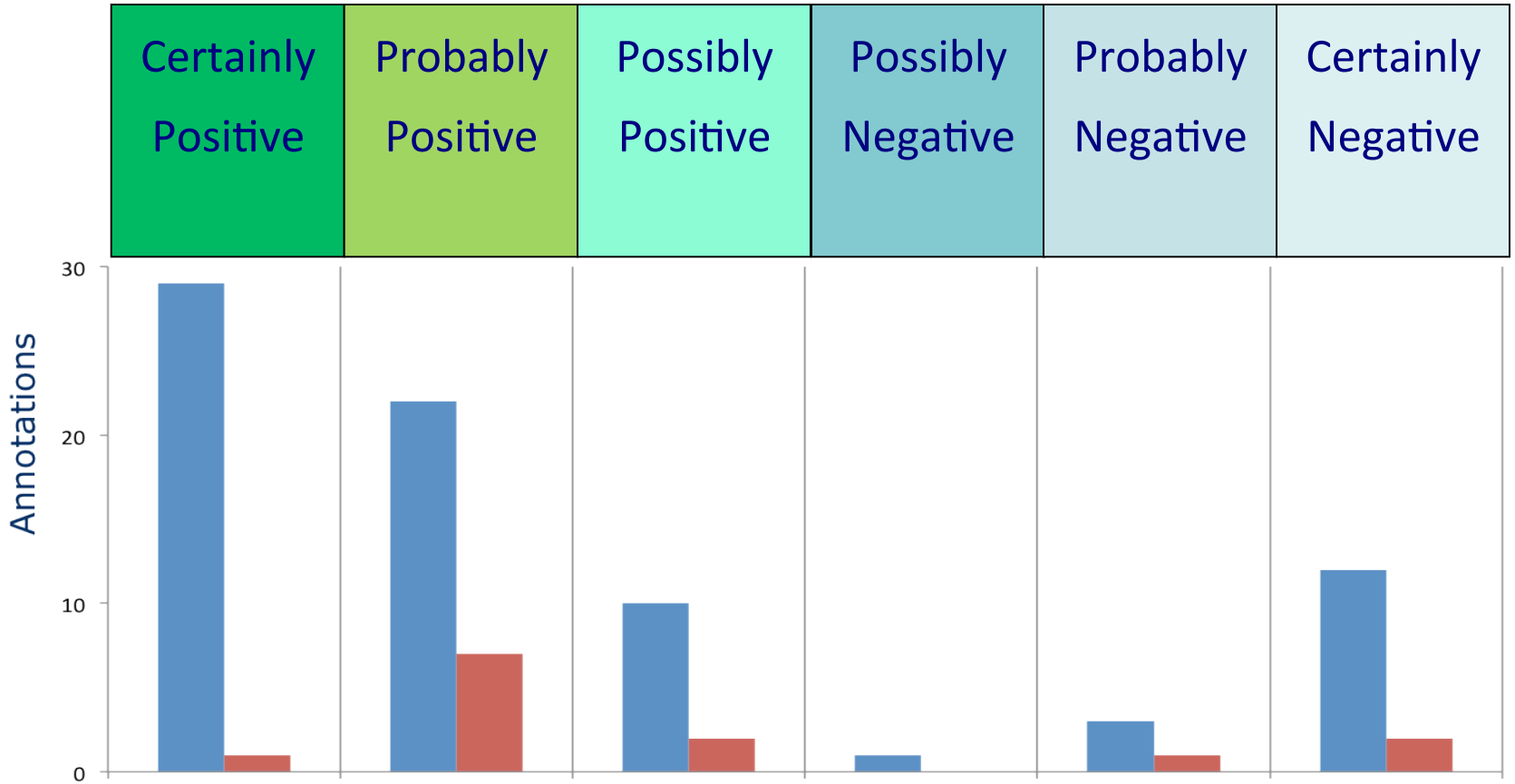
Generic skeleton injuries

Stockholm University



Inverted pattern:

Complementary vocabulary



Shift pattern:

Speculation in Swedish

Stockholm University

Annotation model: Conclusion

- Functional and agreeable model for annotators
- IAA results suggest that this model can be used for developing automated systems
- Different types of “cues” (not only linguistic)

Automatic classification

- Conditional Random Fields
- Local (simple) context features
  - Window +/- 4
  - Words, lemmas, PoS

Stockholm University

Automatic classification

- all classes (8)
- merged classes (5)
  - probably/possibly, nd+o
- Evaluation: 80/20% training/testing split
  - stratified class distribution
  - precision, recall, f-score (micro-average)
    - conlleval

Automatic classification - results

- 0.699 F-measure (all classes)
- 0.762 F-measure (merged classes)

	P_a (95% CI)	R_a (95% CI)	F_a	P_m (95% CI)	R_m (95% CI)	F_m	Merged
CP	0.826 ± 0.03	0.814 ± 0.03	0.82	0.839 ± 0.03	0.818 ± 0.03	0.828	CP
PrP	0.64 ± 0.07	0.576 ± 0.07	0.604	0.825 ± 0.04	0.72 ± 0.05	0.769	PrPoP
PoP	0.643 ± 0.08	0.437 ± 0.08	0.521				
PoN	0.636 ± 0.20	0.304 ± 0.18	0.412	0.58 ± 0.08	0.55 ± 0.08	0.564	PrPoN
PrN	0.504 ± 0.09	0.528 ± 0.09	0.516				
CN	0.789 ± 0.06	0.584 ± 0.08	0.716	0.79 ± 0.06	0.604 ± 0.08	0.686	CN
O	0.444 ± 0.19	0.16 ± 0.14	0.25				
ND	1.0 ± 0.0	0.6 ± 0.18	0.75	0.885 ± 0.08	0.418 ± 0.13	0.568	O-ND
Avg	0.744 ± 0.02	0.66 ± 0.03	0.699	0.805 ± 0.02	0.723 ± 0.02	0.762	Avg_m

Automatic classification - results

- Preceding context important
  - Låg sannolikhet för (*low probability for*)
  - Tolkas som *(interpreted as)*, sannolikt (*likely*)
- Lower levels of certainty more difficult
  - Rare classes, lower IAA
  - Conjunctions and other higher-level features

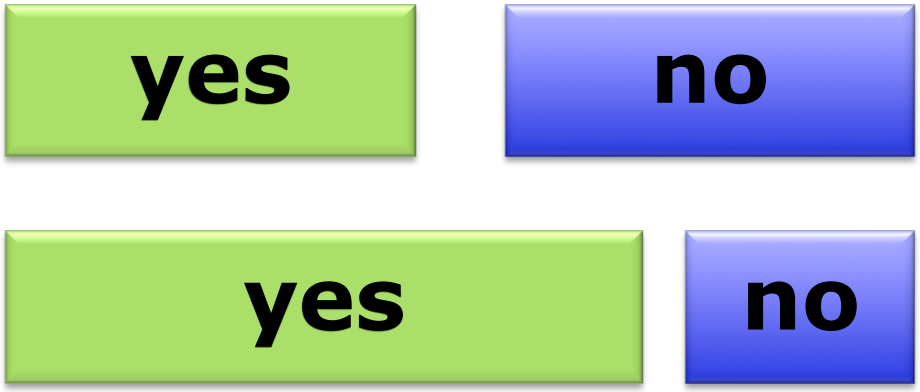
E-Health Scenarios

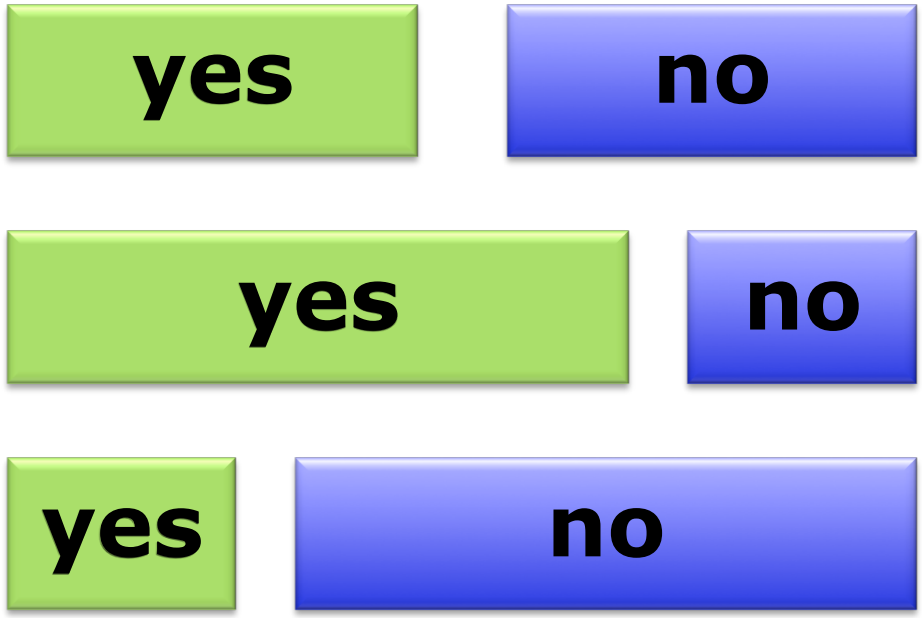
- Are these fine-grained levels needed/practical?

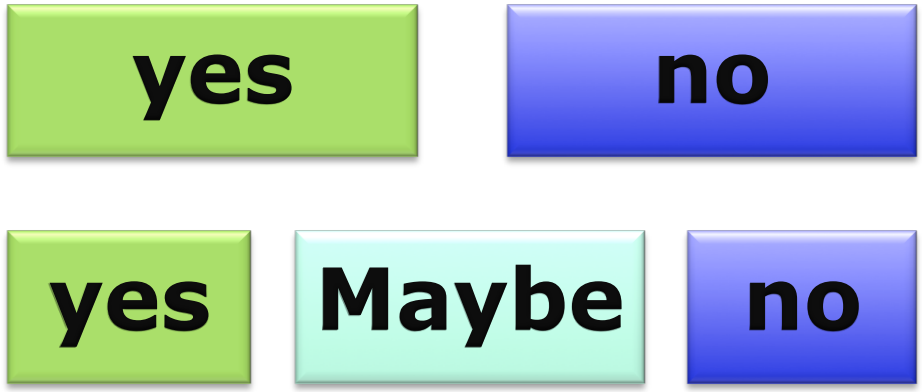
Stockholm University

yes

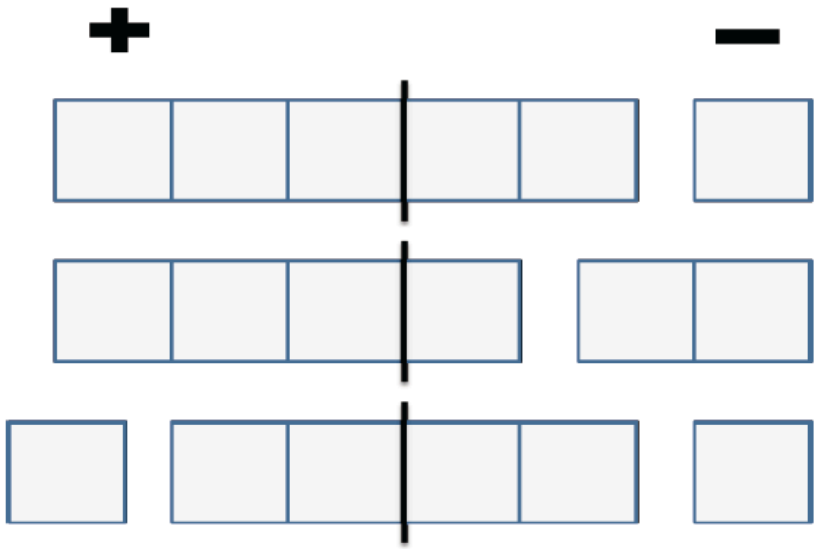
no


yes
no
yes
no


yes
no
yes
no
yes
no


yes
no
yes
Maybe
no

E-Health Scenarios



Adverse event surveillance

Decision support alerts

Summary/overview

E-Health Scenarios: Automatic classification

- Adverse event surveillance
  - *existence, no existence*
- Decision support alerts
  - *plausible existence, no plausible existence*
- Automatic summaries/overview
  - *affirmed, maybe, negated/excluded*

E-Health Scenarios: Automatic classification

- Overall average results
  - Adverse event surveillance: 0.89 F-score
  - Decision support alerts: 0.91 F-score
  - Automatic summaries/overviews: 0.8 F-score

- Improvements over baselines
  - majority class + no context

E-Health Scenarios: Error analysis

- Difficulties in distinguishing *probably negative* and *certainly negative*
  - *Inga hållpunkter för* (*no indicators of*)
- Local or global context
- Modifier emphasis
  - *liten misstanke* (*small suspicion*)

E-Health Scenarios: Error analysis

- Clinical exclusion difficult
  - e.g. DVT, important severe consequences if missed
- Test results
  - performing a test in itself is indication of risk, but surrounding context suggests otherwise
- Chronic diseases
  - e.g. *probably stress triggered asthma*

Stockholm University

Conclusion & Discussion

- One fine-grained annotation model for several purposes/scenarios
- Annotation discrepancies need to be analyzed → refine annotation task
- Further studies on classification algorithms, representation and features needed

Stockholm University

Conclusion & Discussion

- Valuable corpora created for further studies
- Feasibility studies of automatic classification
- Evaluation – involve users

Thank you for your attention

Ideas and comments welcome! sumithra@dsv.su.se