R·I·T

Annotation Schemes to Encode Domain Knowledge in Medical Narratives

Wilson McCoy	wgm4143@rit.edu
Cecilia Ovesdotter Alm	coagla@rit.edu
Cara Calvelli	cfcscl@rit.edu
Rui Li	rxl5604@rit.edu
Jeff B. Pelz	pelz@cis.rit.edu
Pengcheng Shi	spcast@rit.edu
Anne Haake	arhics@rit.edu

6/12/2012
The 6th Linguistic Annotation Workshop
ACL 2012, Jeju, South Korea

Outline

- Introduction
- Data Collection and Data Set
- Annotation of Thought Units
  - Distributions of Thought Units
  - Agreement Metrics
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Introduction and Motivation

It is important to understand the cognitive processes of physicians.

Annotation schemes can be used to encode such information.

This research analyzes two annotation schemes in the context of dermatology for transcribed verbal medical narratives.

Outline

- ~~Introduction~~
- Data Collection and Data Set
- Annotation of Thought Units
  - Distributions of Thought Units
  - Agreement Metrics
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Data Collection

16 physicians with dermatological experience

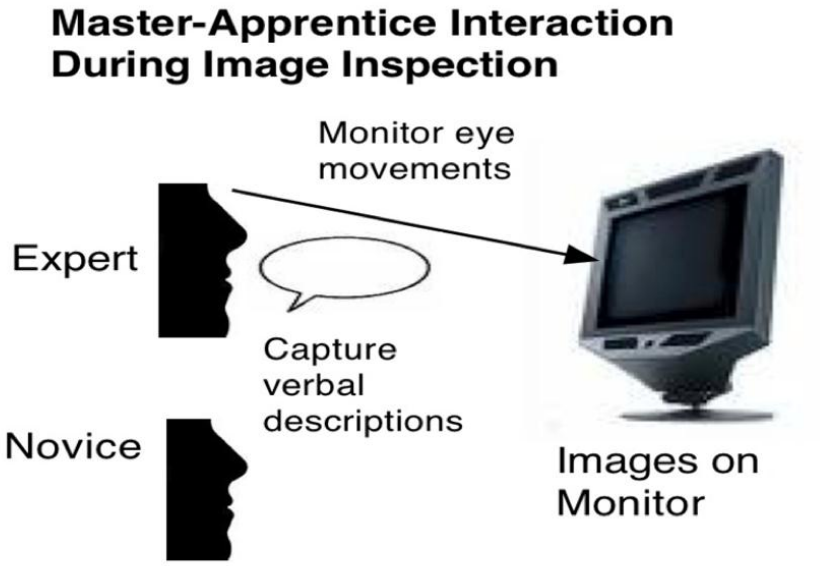
- 12 attending physicians + 4 residents

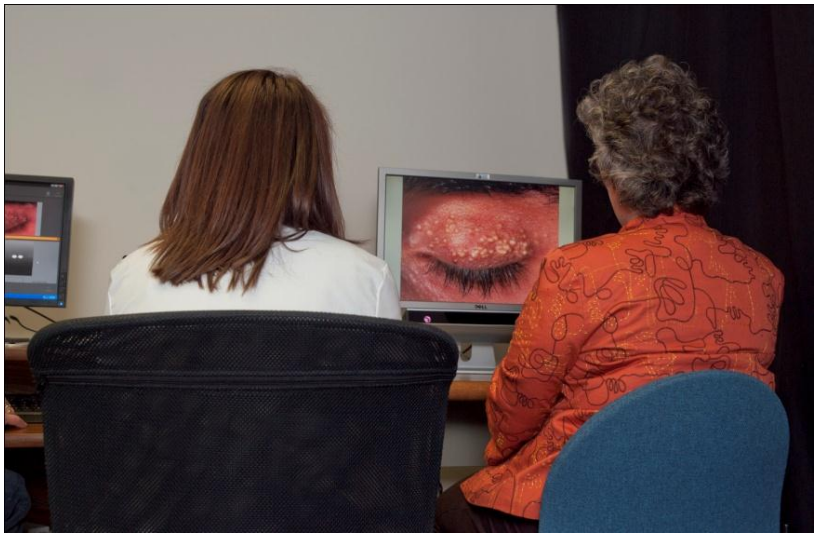
50 images of dermatological conditions

Physicians were instructed to narrate their diagnostic process aloud to a student as each image was presented.

This allowed us to create a “Master-Apprentice” interaction scenario.

Master-Apprentice Interaction Scenario (Beyer and Holtzblatt, 1997)





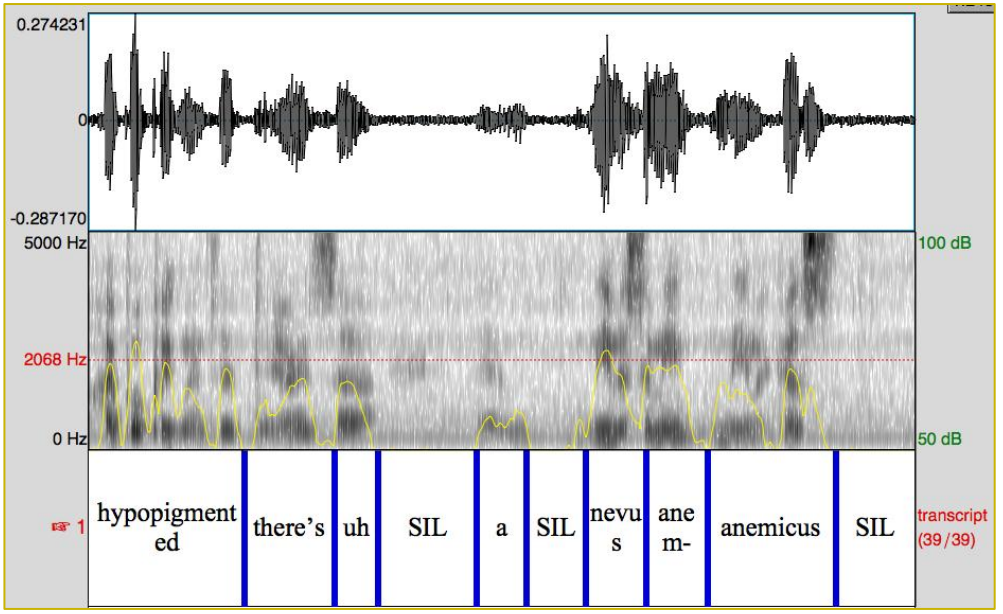
This scenario allows us to extract information about the Master's (i.e. physician's) cognitive process by coaxing them to vocalize their thoughts in rich detail.

This scenario is a monologue, however, it induces a feeling of dialogic interaction in the Master.

Data Set

Audio of the physician's speech was recorded as well as a scan path of their eye movements.*

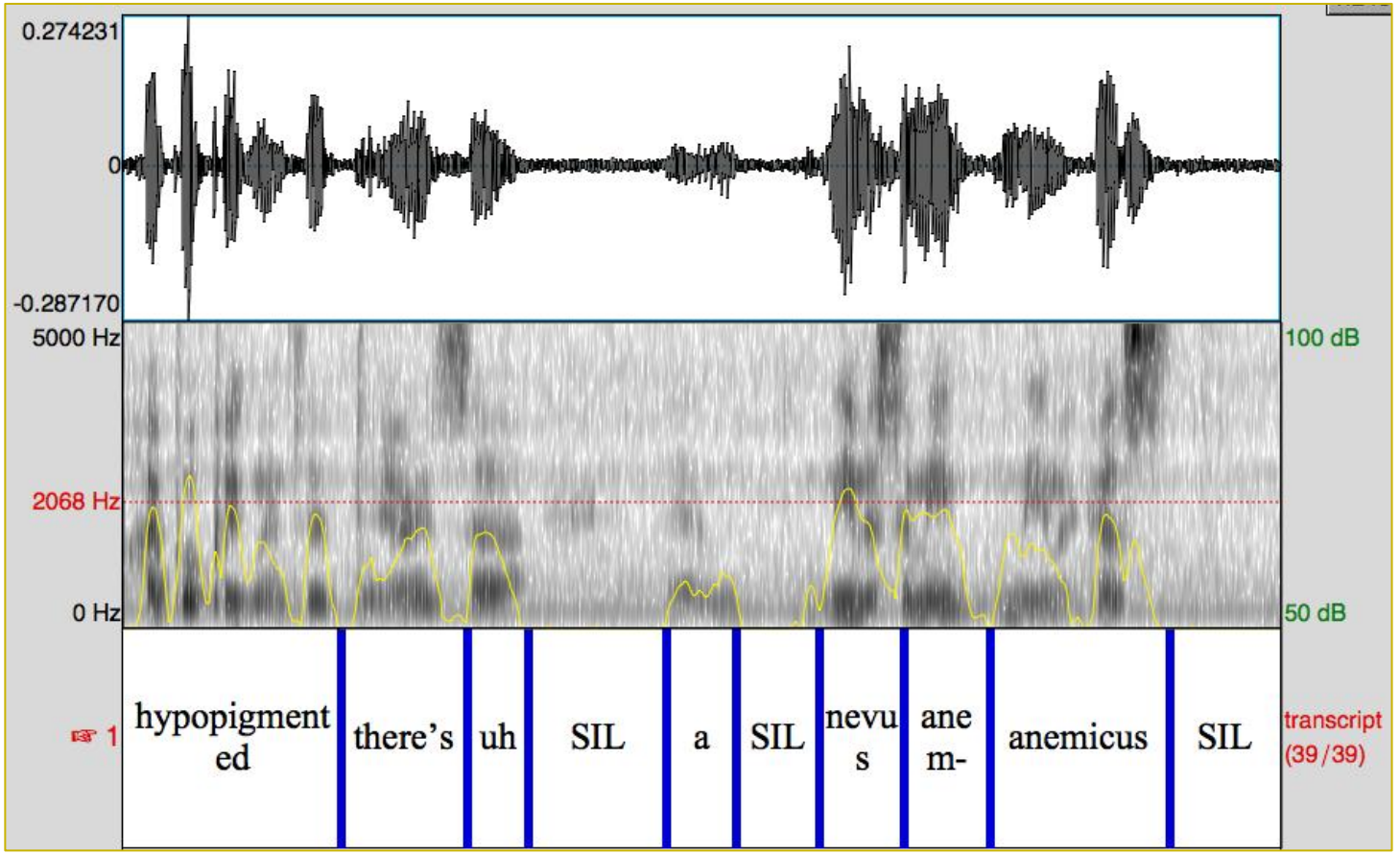
Praat (Boersma, 2001) was used to time-align each *narrative* (one physician inspecting one image).



* This eye tracking data will not be considered here.

Data Set

Disfluencies and pauses were also annotated.



Example Narrative

"SIL hm SIL uh involving a SIL an older patient's infraorbital area is a a pearly SIL papule with overlying telangiectasia SIL uh suggestive of a basal cell carcinoma."

Narrative Statistics

Average Narrative Length	55.9 seconds
Average Words per Narrative	105 words
Average Pauses per Narrative	15.4 pauses
Average Pause Length	1.28 seconds
Percent Pause in Narrative	35.4% silent

Outline

- ~~Introduction~~
- ~~Data Collection and Data Set~~
- Annotation of Thought Units
  - Distributions of Thought Units
  - Agreement Metrics
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Annotation of Thought Units

An annotation scheme was created to reveal the cognitive decision making processes used by physicians.

Narratives were annotated for *thought units*: words or sequences of words that received a descriptive label based on its role in the diagnostic process.

A set of nine basic thought units were created for annotation.

Provided Thought Units

Thought Unit Label	Abbreviation	Example
Patient Demographic	DEM	young male
Body Location	LOC	arm
Configuration	CON	linear
Distribution	DIS	acral
Primary Morphology	PRI	papule
Secondary Morphology	SEC	scaly
Differential Diagnosis	DIF	X, Y, or Z
Final Diagnosis	DX	this is X
Recommendations	REC	P should Q

Thought Unit Annotations

Of the narratives collected, 60 were chosen to be annotated for thought units.*

- 6 narratives from each of 10 images

10 images were chosen for their diverse representation of the image set.

6 narratives from each of these images were chosen based on length.

- 3 shortest and 3 longest

*Only 59 narratives annotated for thought units were used in the study, however.

Thought Unit Annotations

Transcripts were then manually cleaned to remove disfluencies and ungrammatical structures that could confuse the annotators.

"SIL hm SIL uh involving a SIL an older patient's infraorbital area is a a pearly SIL papule with overlying telangiectasia SIL uh suggestive of a basal cell carcinoma."	➔	Involving an older patient's infraorbital area is a pearly papule with overlying telangiectasia suggestive of a basal cell carcinoma.

Thought Unit Annotations

Cleaned transcripts were then shuffled and given to two dermatologists (MD 1 and MD 2)

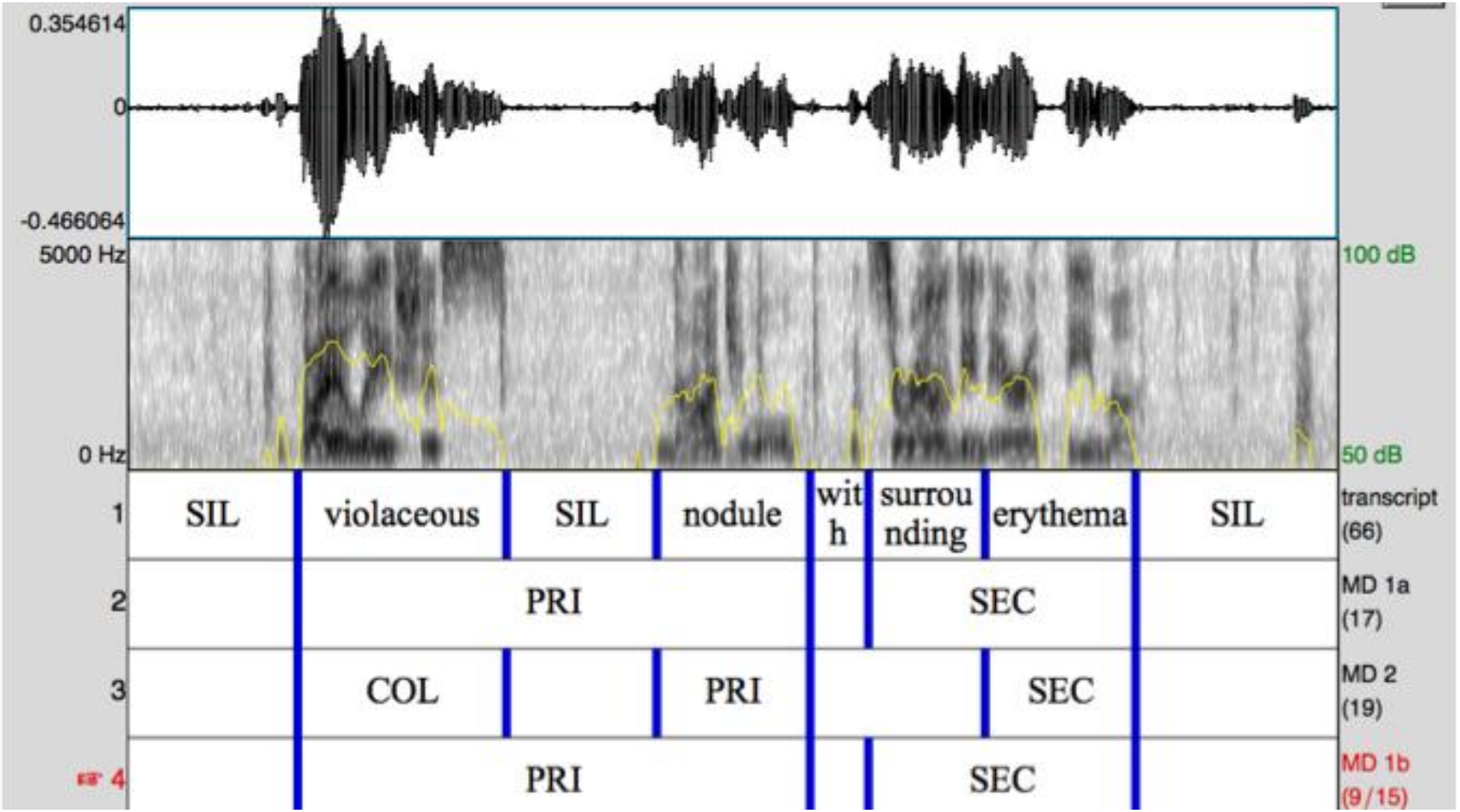
- Annotate using the nine provided thought units
- Add other thought unit labels if necessary

Reshuffled transcripts were given to MD 1 to re-annotate. The initial annotation became MD 1a and the re-annotation became MD 1b.

Example Thought Unit Annotated Narrative

Involving an [older patient's]$_{DEM}$
[infraorbital area]$_{LOC}$ is a
[pearly papule]$_{PRI}$ with
[overlying telangiectasia]$_{SEC}$
suggestive of a
[basal cell carcinoma]$_{DX}$.

Thought Unit Annotations



Outline

- ~~Introduction~~
- ~~Data Collection and Data Set~~
- ~~Annotation of Thought Units~~
  - Distributions of Thought Units
  - Agreement Metrics
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Distributions of Thought Units

- There are a total of 1608 thought unit tokens in our data set.
- Only MD 2 created tags beyond the 9 provided.
  - Examples are COL (Color), UDX (Underlying Diagnosis), and SEV (Severity)
- The PRI (Primary Morphology) thought unit was found in all narratives by at least one annotator.

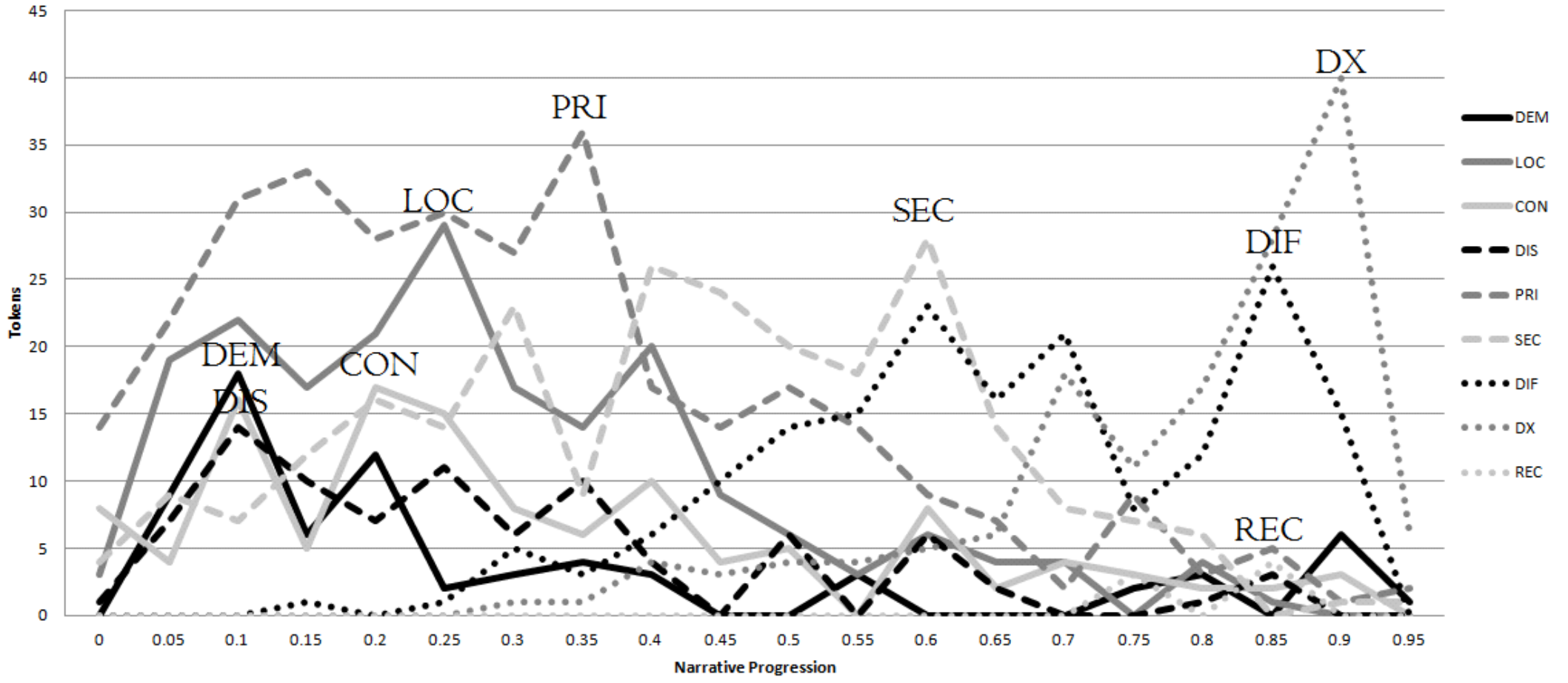
Thought Unit Word Clouds

Created with Wordle
www.wordle.net

Thought Unit Word Clouds

Created with Wordle
www.wordle.net

Thought Unit Temporal Distribution

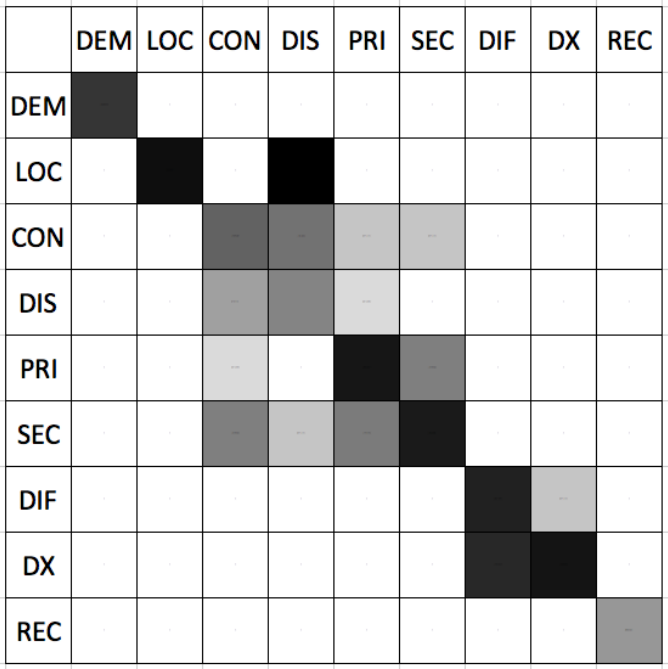


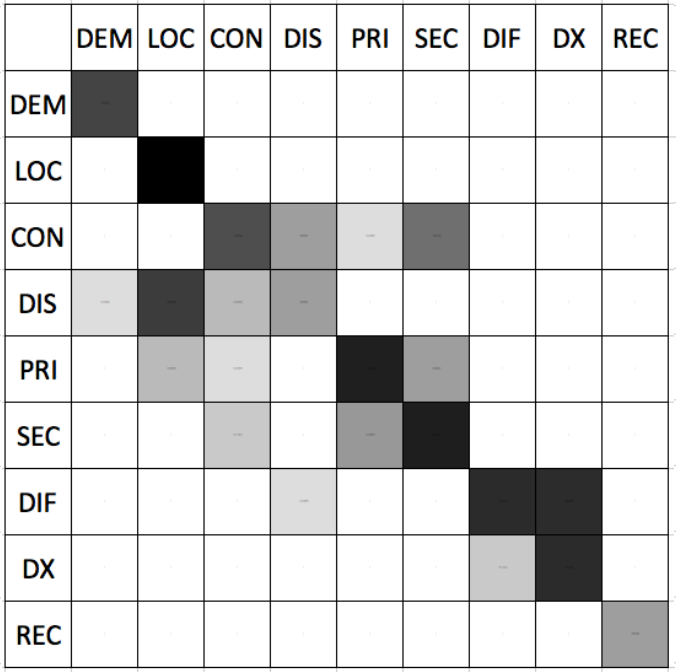
Outline

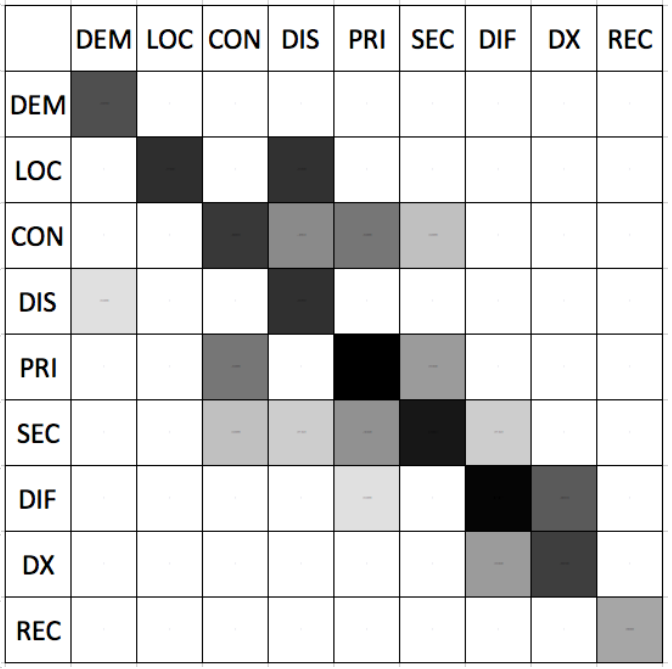
- ~~Introduction~~
- ~~Data Collection and Data Set~~
- ~~Annotation of Thought Units~~
  - ~~Distributions of Thought Units~~
  - Agreement Metrics
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Agreement metrics

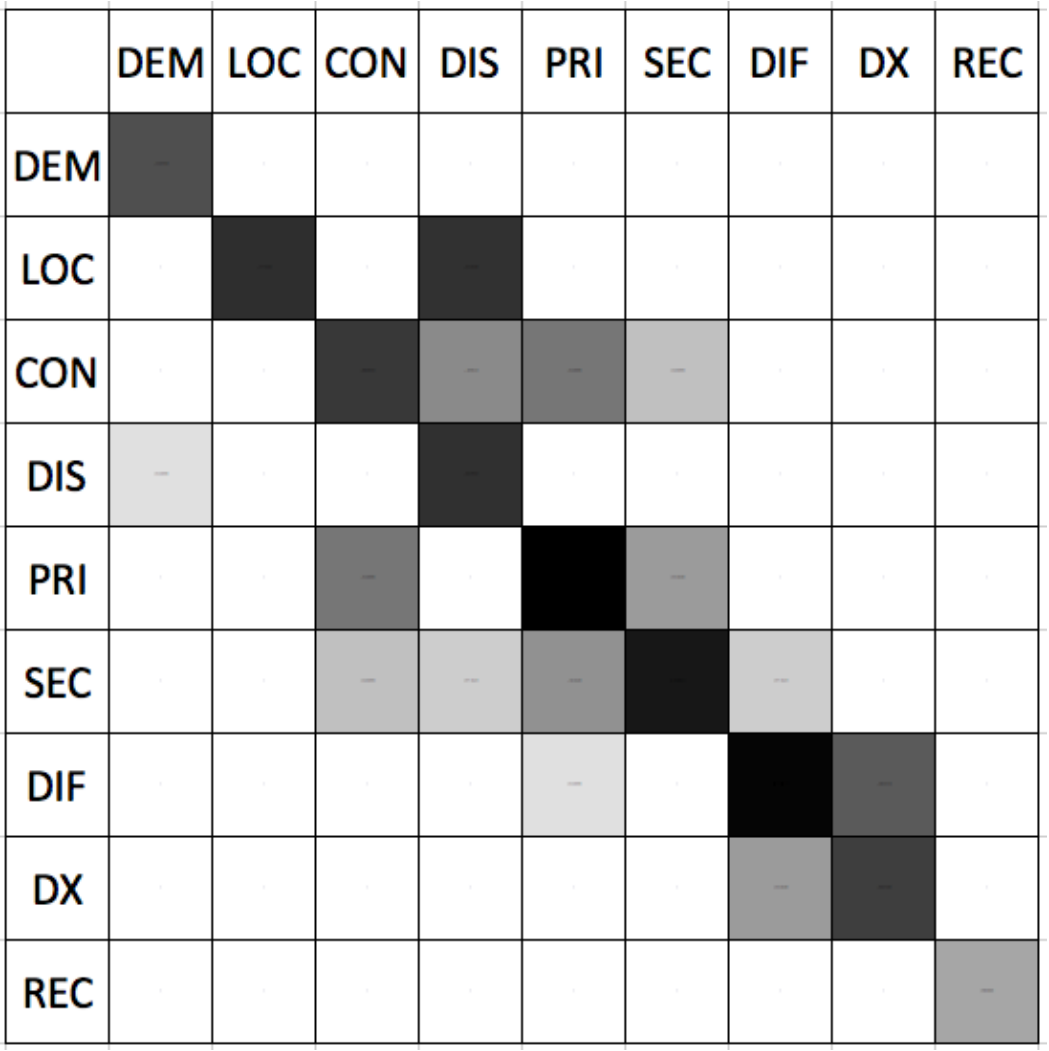
Measure	MD 1a – MD 2	MD 2 – MD 1b	MD 1a – MD 1b
% Agreement	80.69	77.72	80.98
Cohen (1960) Kappa	.56	.54	.62







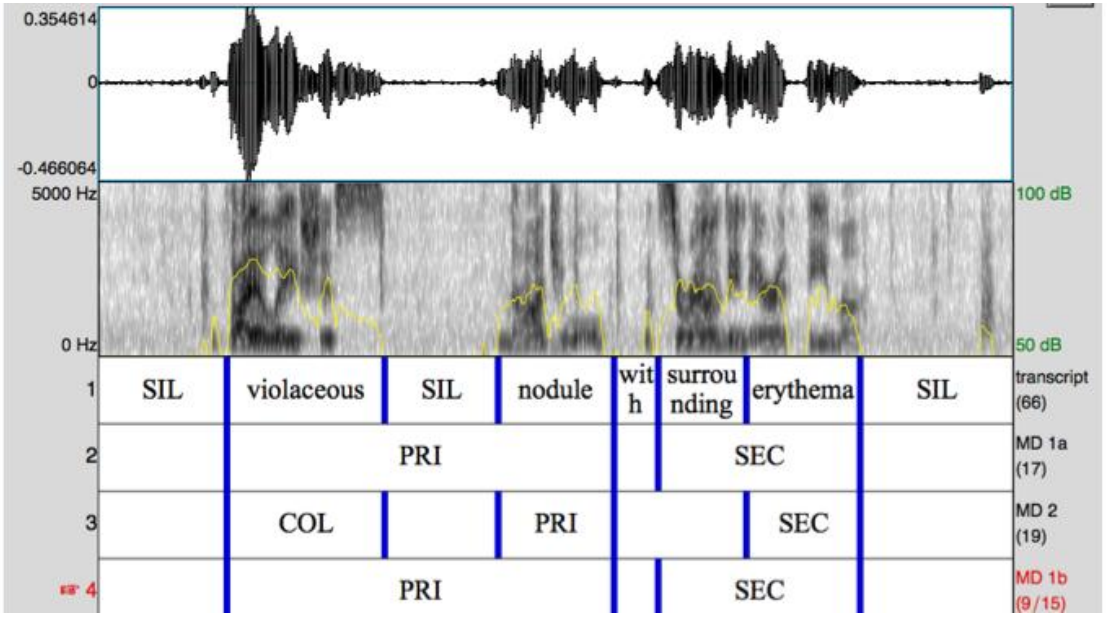
Closer Examination of Confusion Matrices



Outline

- ~~Introduction~~
- ~~Data Collection and Data Set~~
- ~~Annotation of Thought Units~~
  - ~~Distributions of Thought Units~~
  - ~~Agreement Metrics~~
  - External Validation with the UMLS
- Annotation of Correctness
- Conclusion

Mapping of Thought Units to UMLS Semantic Types

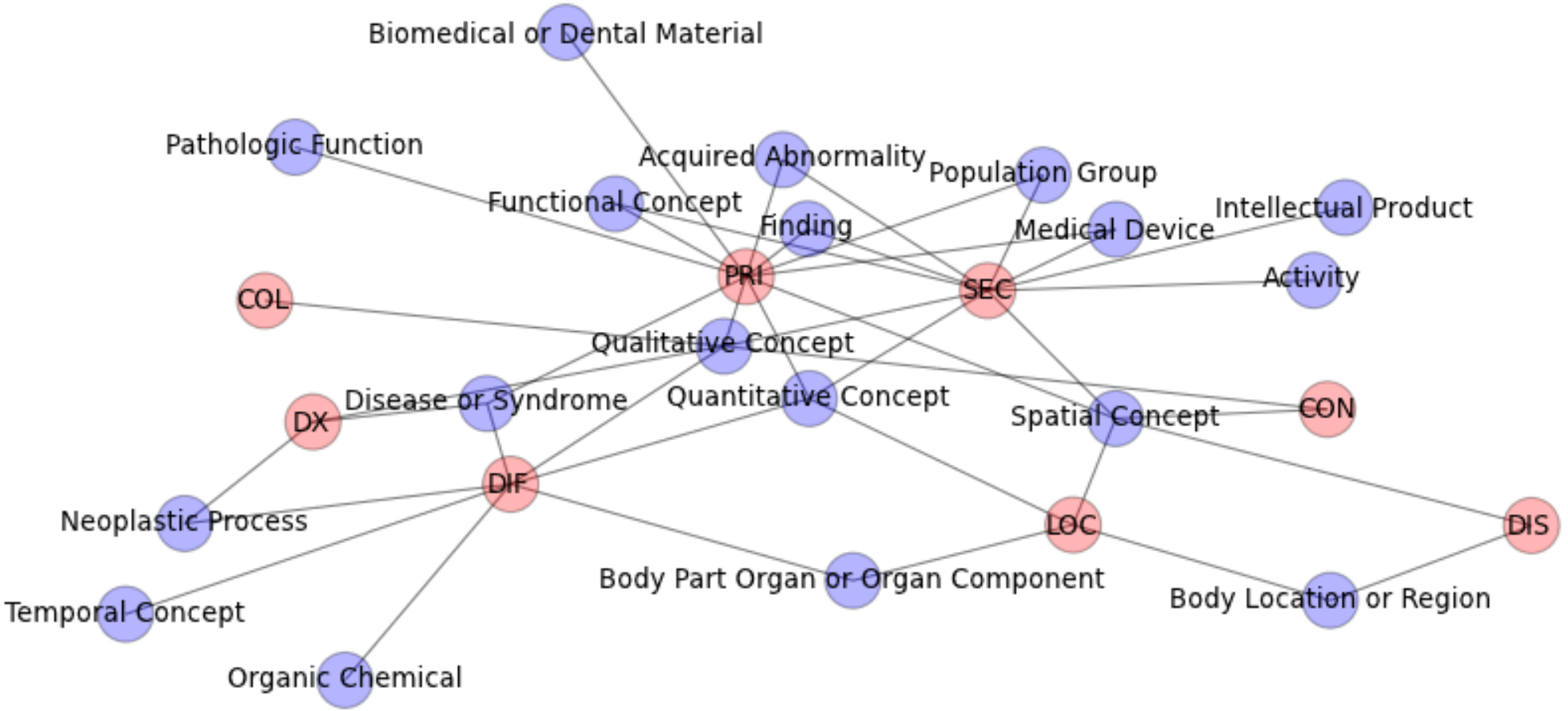


"Finding", "Symptom",
"Disease or Syndrome",
"Qualitative Concept"

"violaceous nodule", PRI

Unified Medical Language System

Mapping of Thought Units to UMLS Semantic Types



Outline

- ~~Introduction~~
- ~~Data Collection and Data Set~~
- ~~Annotation of Thought Units~~
  - ~~Distributions of Thought Units~~
  - ~~Agreement Metrics~~
  - ~~External Validation with the UMLS~~
- Annotation of Correctness
- Conclusion

Annotation of Correctness

A set of 800 cleaned transcripts were sent to three dermatologists to be annotated for correctness

- MD A, MD B, MD C

Narratives were assigned labels of correctness at three steps of the diagnostic process:

- Medical Lesion Morphology
- Differential Diagnosis
- Final Diagnosis

Annotation of Correctness

These stages of correctness give insight to where problems arise in the diagnostic processes.

Medical Lesion Morphology	• Correct • Incorrect • None Given • Incomplete
Differential Diagnosis*	• Yes • No • No Differential
Final Diagnosis	• Correct • Incorrect • None Given

* The dermatologists were asked if the correct diagnosis was present in the differential diagnosis.

Example Narrative Annotated for Correctness

Involving an older patient's infraorbital area is a pearly papule with overlying telangiectasia suggestive of a basal cell carcinoma.

Medical Lesion Morphology:	Correct
Differential Diagnosis:	No Differential
Final Diagnosis:	Correct

Agreement Metrics

% Agreement	MD A – MD B	MD B – MD C	MD A – MD C
Medical Lesion Morphology	67.75	72.40	71.56
Differential Diagnosis	91.81	88.46	88.71
Final Diagnosis	88.21	91.97	83.56

Kappa	MD A – MD B	MD B – MD C	MD A – MD C
Medical Lesion Morphology	.24	.22	.39
Differential Diagnosis	.85	.79	.79
Final Diagnosis	.79	.84	.70

Correctness Scores

t = a thought unit in set T

N_t = all narratives n with t present

$$t_{score} = \frac{\text{“Correct” or “Yes” tags on } N_t}{3 \times |N_t|}$$

Correctness by Thought Unit

Thought Unit	% Present	Medical Lesion Morphology	Differential Diagnosis	Final Diagnosis
PRI	100	.66	.26	.61
LOC	88	.64	.29	.60
- LOC	12	.81	0	.71
DX	86	.67	.24	.66
- DX	14	.58	.42	.29
SEC	85	.70	.27	.67
- SEC	15	.44	.07	.30
DIS	66	.63	.26	.69
- DIS	34	.72	.25	.45
CON	64	.71	.28	.67
- CON	36	.54	.16	.51
DIF	61	.60	.43	.44
- DIF	39	.75	0	.87

* Only thought units present in over 50% of the narratives are shown

Outline

- ~~Introduction~~
- ~~Data Collection and Data Set~~
- ~~Annotation of Thought Units~~
  - ~~Distributions of Thought Units~~
  - ~~Agreement Metrics~~
  - ~~External Validation with the UMLS~~
- ~~Annotation of Correctness~~
- Conclusion

Conclusion

This work contributes to the understanding of linguistic expression of cognitive decision-making in a clinical domain as well as appropriate annotation processes that capture such phenomena.

This study additionally furthers research in linguistically annotated corpora by creating and validating schemes with future potential applications in the medical industry.

Acknowledgements

- NIH grant 1 R21 LM010039-01A1
- NSF grant IIS-0941452
- An RIT Seed Funding grant
- A GCCIS Seed Funding grant
- Lowell A. Goldsmith
- Preethi Vaidyanatha
- Transcribers
- Anonymous reviewers

Questions?