Building A User-Centric and Content-Driven Socialbot

## Hao Fang

Committee: Mari Ostendorf (Chair)   Hannaneh Hajishirzi
Leah M. Ceccarelli (GSR)   Eve Riskin
Yejin Choi   Geoffrey Zweig

Agenda

o Background

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations
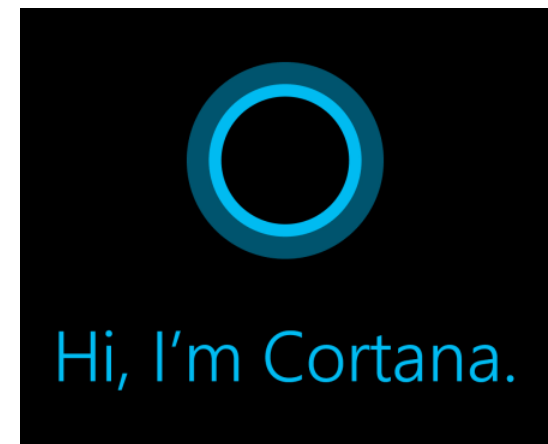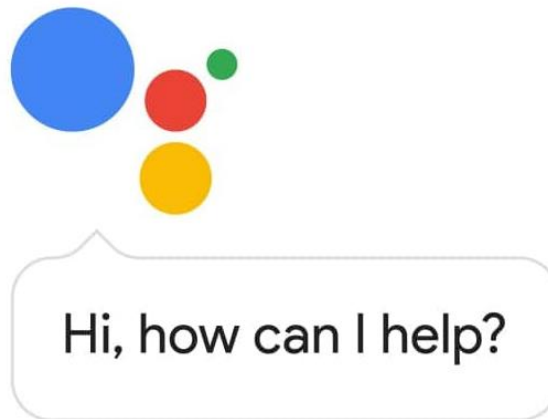
o Summary and Future Directions

Agenda

o **Background**

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations

o Summary and Future Directions
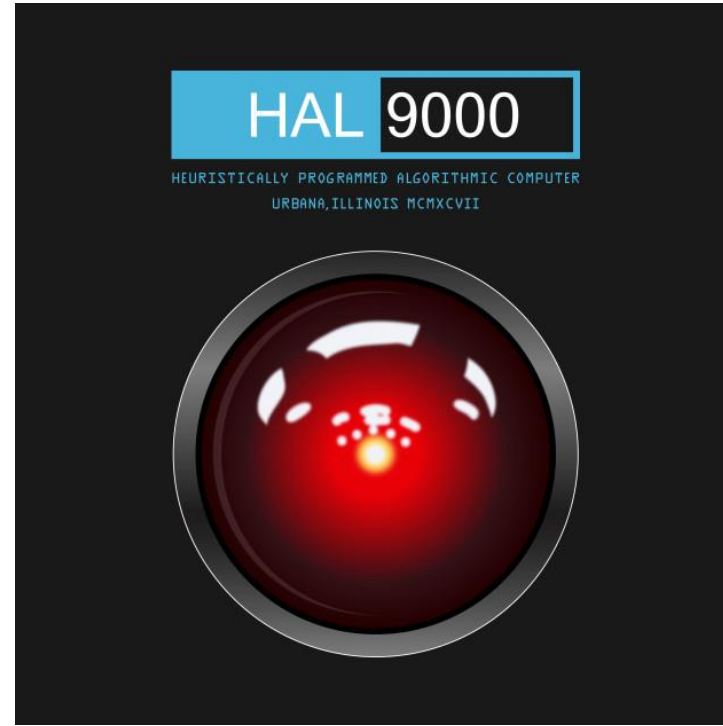
Sci-Fi Movies

Daily Life

Types of Conversational AI

Socialbots *"converse coherently and engagingly with humans on popular topics and current events"*

Task Definition
task-oriented

non-task-oriented

Domain Coverage
single-domain

multi-domain
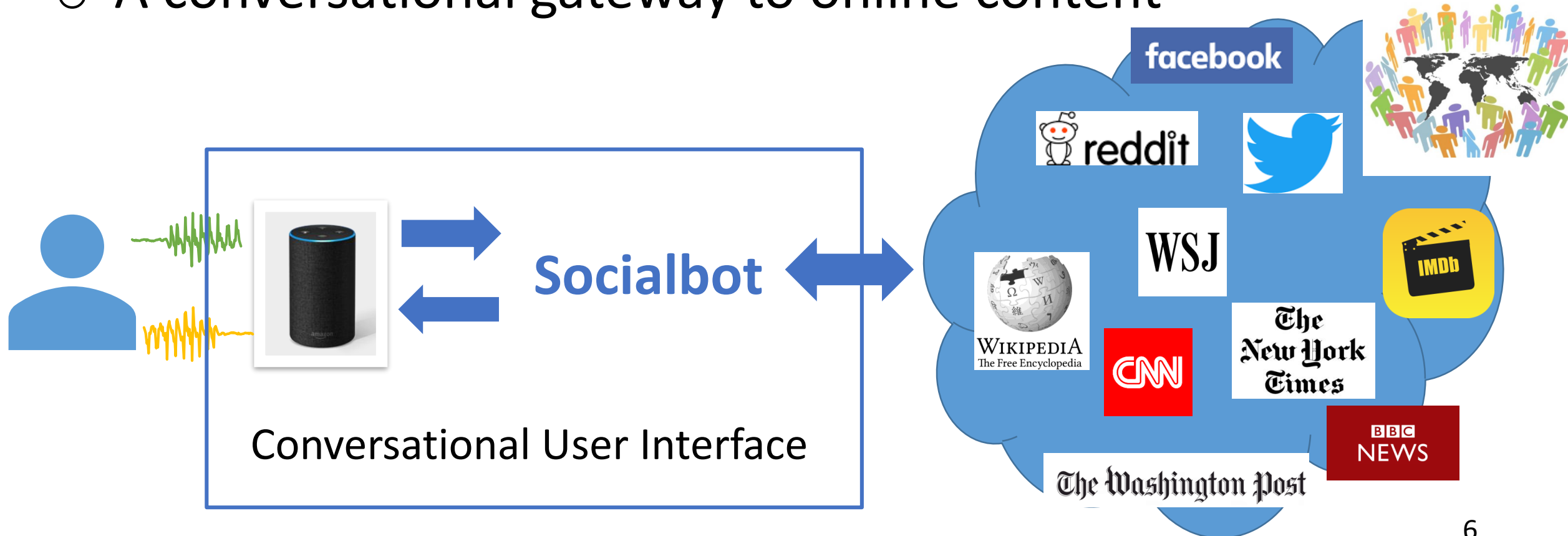
open-domain

Dialog Initiative
system-initiative

user-initiative

mixed-initiative

Socialbot Applications

o Entertainment, education, healthcare, companionship, …

o A conversational gateway to online content

Agenda

o Background

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations

o Summary and Future Directions
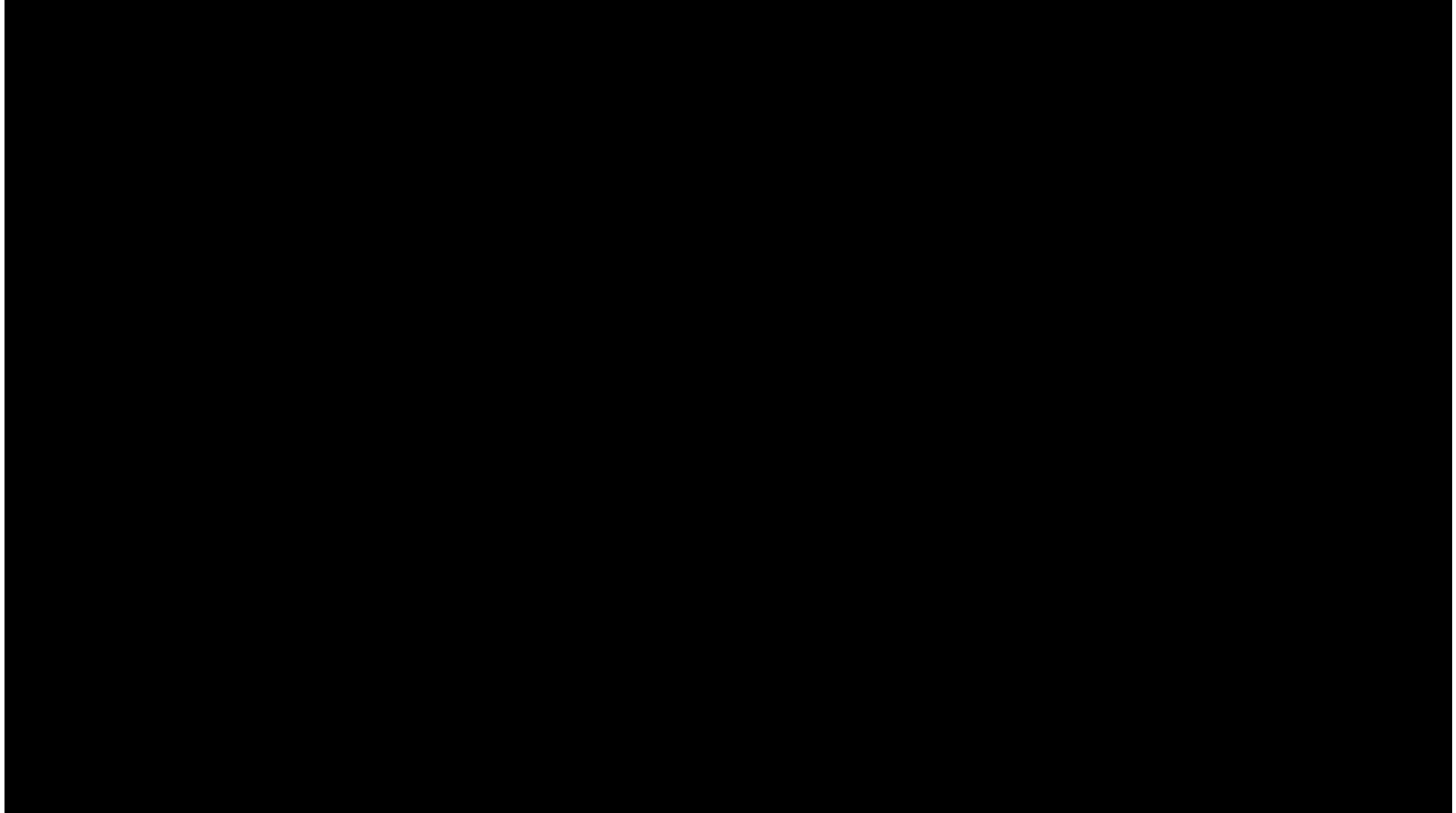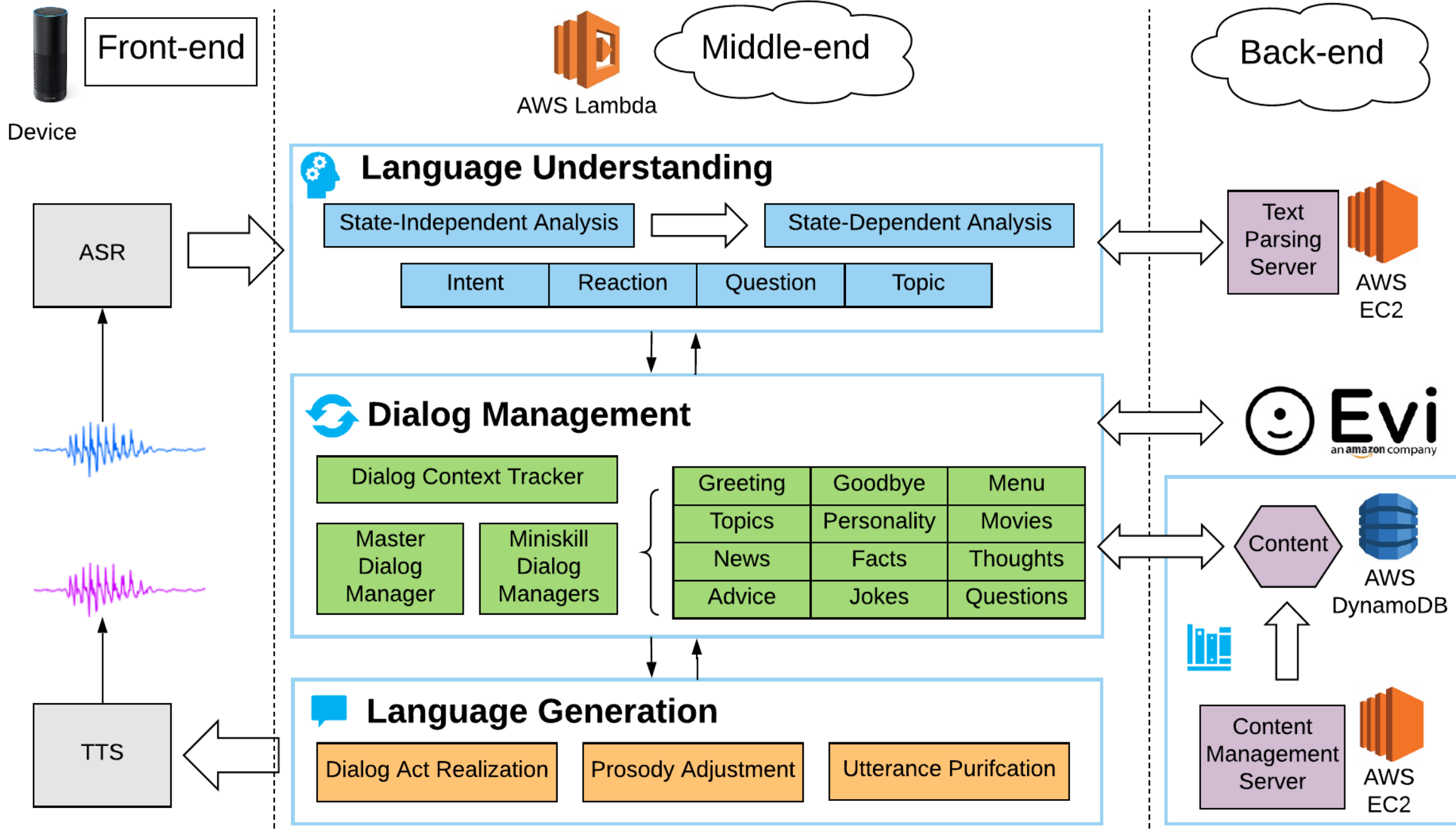
Design Objectives

User-Centric
- Users can control the dialog flow and switch topics at any time
- Bot responses are adapted to acknowledge user reactions

Content-Driven
- Content cover the wide range of user interests
- Dialog strategies to lead or contribute to the dialog flow

2017 Alexa Prize Finals

Dialog Control for Many Miniskills?

Conversation Activities (Miniskills)

- Greet
- List Topics
- Tell Fun Facts
- Tell Jokes
- Tell Headlines
- Discuss Movies
- Personality Test
- …

Hierarchical Dialog Management

o Dialog Context Tracker

    o dialog state, topic/content/miniskill history, user personality

o Master Dialog Manager

    o miniskill polling

    o topic and miniskill backoff

o Miniskill Dialog Managers

    o miniskill dialog control as a finite-state machine

    o retrieve content & build response plan

Social Chat Knowledge

An important type of social chat knowledge is online content.

How to organize content to facilitate the dialog control?

A framework that allows dialog control to be defined in a consistent way.

Knowledge Graph

o Nodes
  o content post (fact, movie, news article, …)
  o topic (entity or generic topic)

o Relational edges between content post and topic
  o topic mention (NER, noun phrase extraction)
  o category tag (Reddit meta-information)
  o movie name, genre, director, actor (IMDB)

o Dialog Control: move along edges

UT Austin and Google AI use machine learning on data from NASA's Kepler Space Telescope to discover an eighth planet circling a distant star.
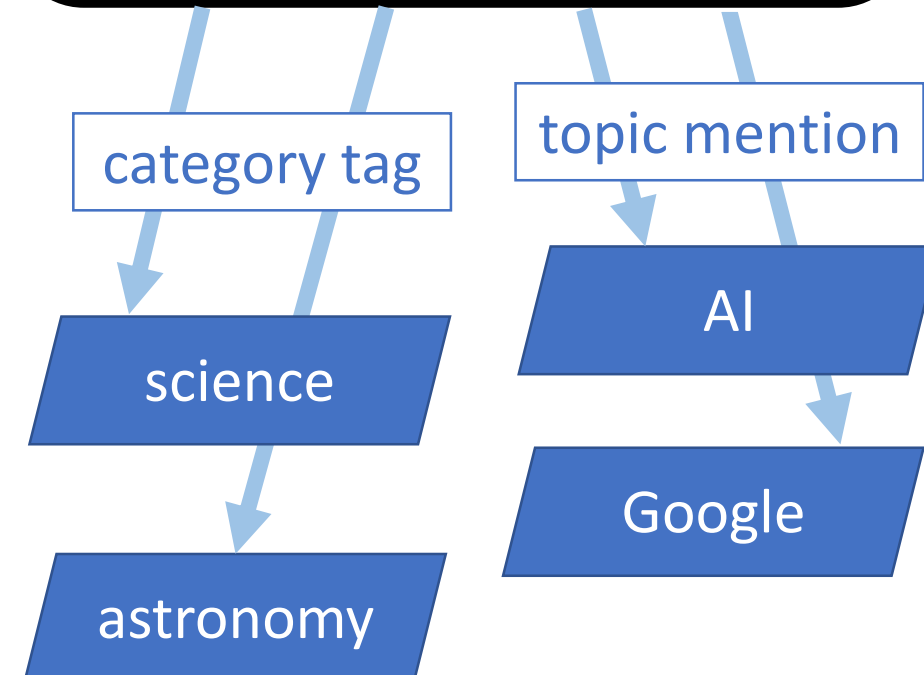
category tag

topic mention

science

astronomy

AI

Google

14

Agenda

o Background

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations

o Summary and Future Directions

Motivation

o Dialog control defined based on moves on the graph
  o lead the conversation
  o handle user initiatives

o Challenges for unstructured document (e.g., news articles)
  o not all sentences are equally interesting to a listener ← Storytelling
  o need to figure out a coherent presenting order
  o answer questions about the document ← Question Answering & Asking
  o need a smooth transition between sentences
  o handle entity-based information seeking requests ← Subject Entity
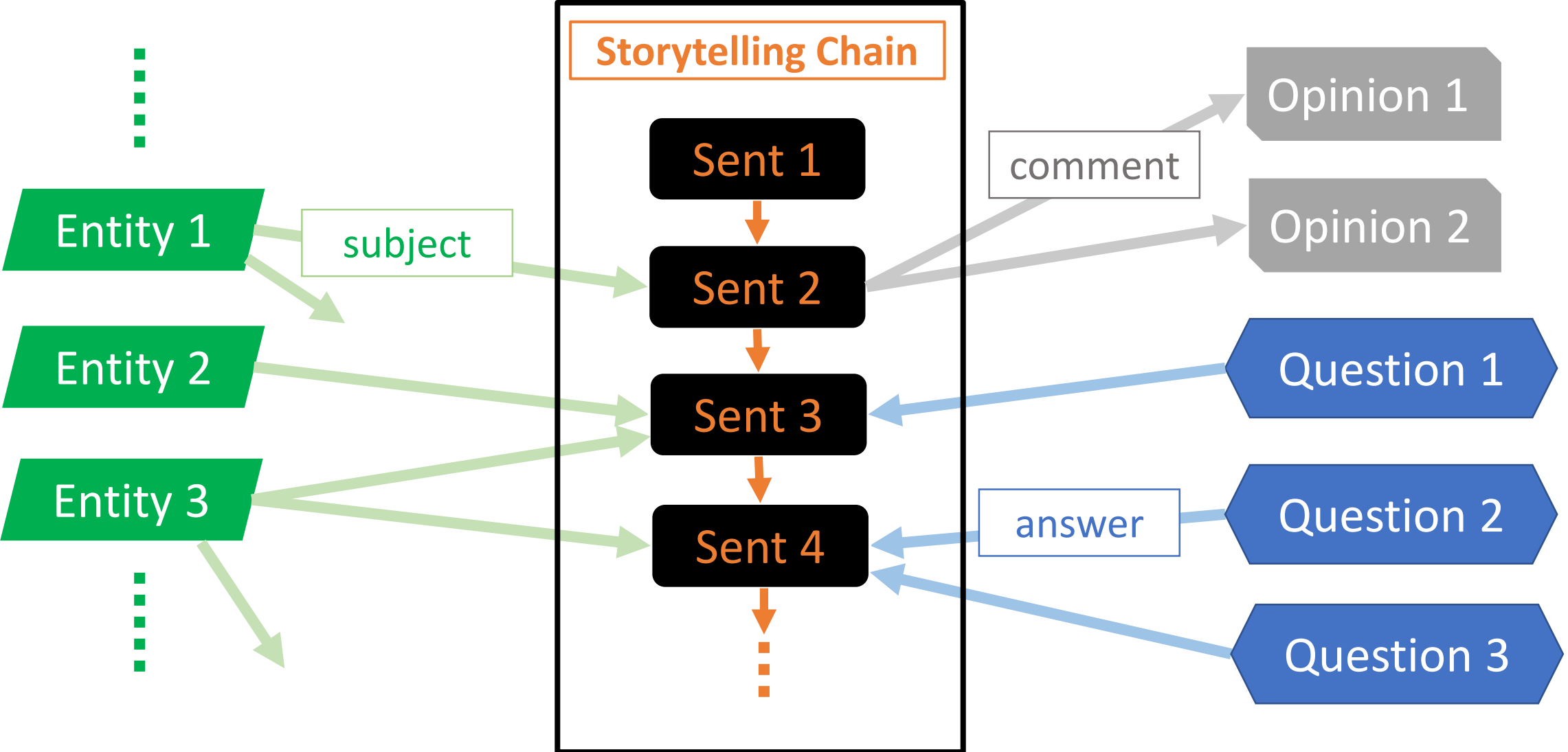 o handle opinion-seeking requests ← Opinion Comment
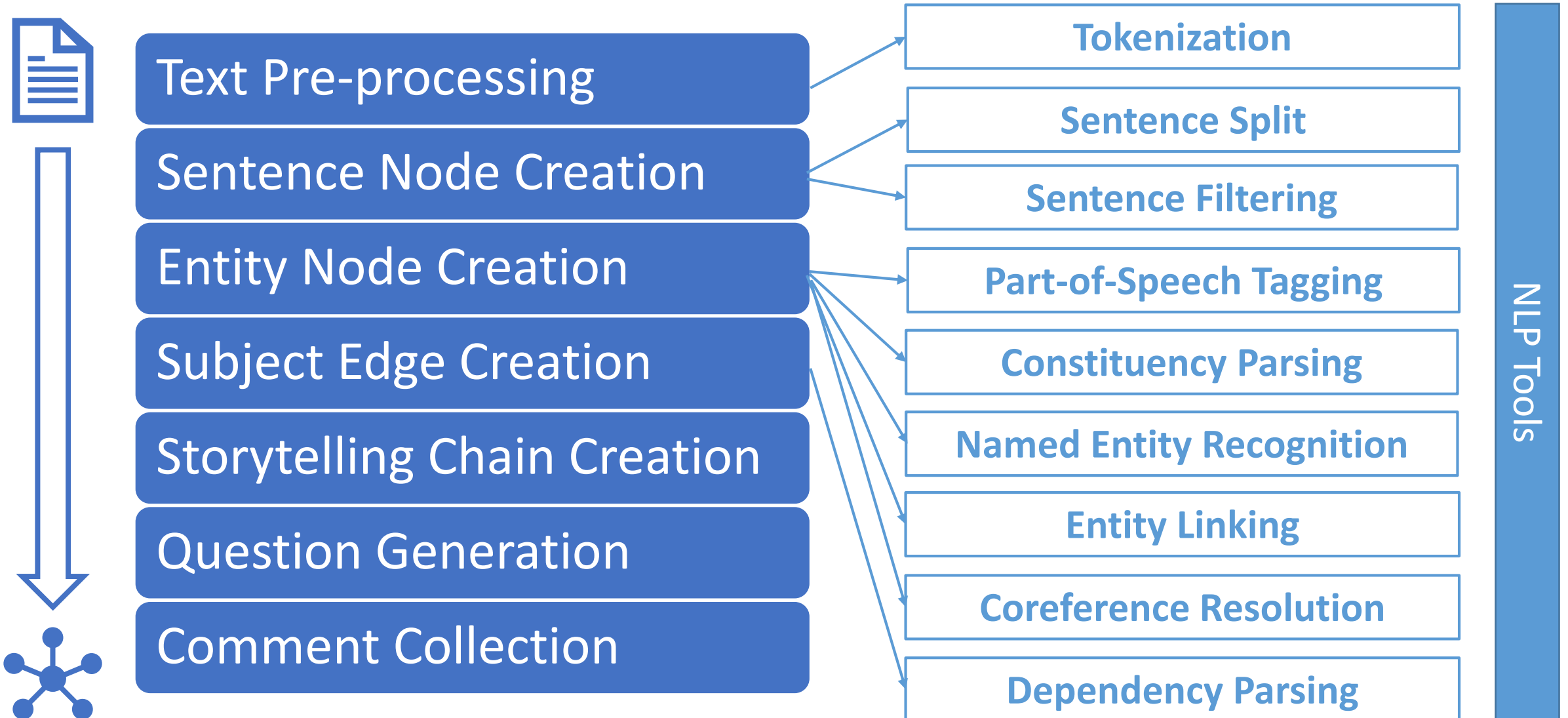
Graph-Based Document Representation
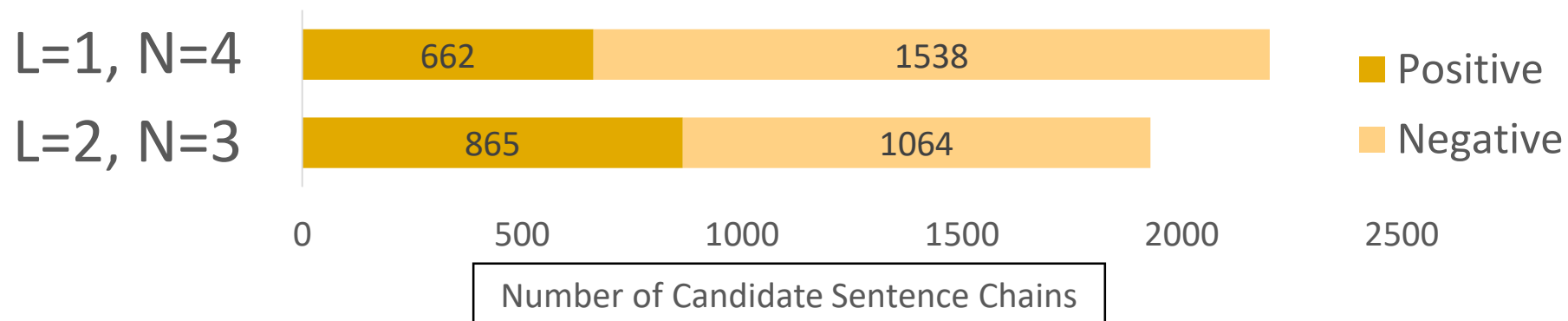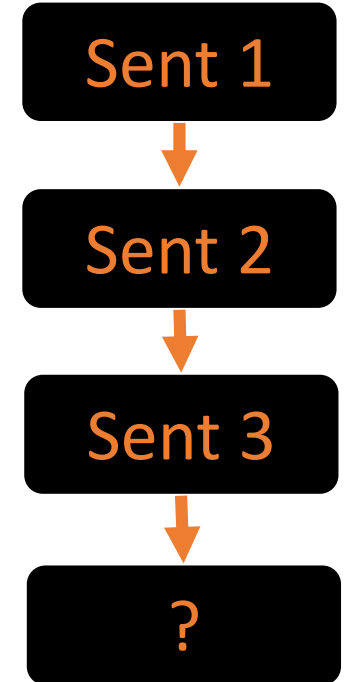
Document Representation Construction

Storytelling Chain Creation

o **Problem formulation**

  o context sentence sequence $(s_1, s_2, …, s_L)$

  o candidate sentence set $\{y_1, y_2, …, y_N\}$

  o candidate sentence chain $(y_i \mid s_1, s_2, …, s_L)$

o **Data collection: 550 news articles**

  o Train/Validation/Test: 3/1/1 based on article ID

the next N sentences following s_L in the article

Binary Label

Sent 1

Sent 2

Sent 3

?

L=1, N=4 — 662 — 1538

L=2, N=3 — 865 — 1064

Positive

Negative

0   500   1000   1500   2000   2500

Number of Candidate Sentence Chains

Model and Features

o Model: binary logistic regression

   o input: candidate sentence chain $(y_i \mid s_1, s_2, \ldots, s_L)$

   o output: probability score $s(y_i \mid s_1, s_2, \ldots, s_L) \in \mathbb{R}^{[0,1]}$
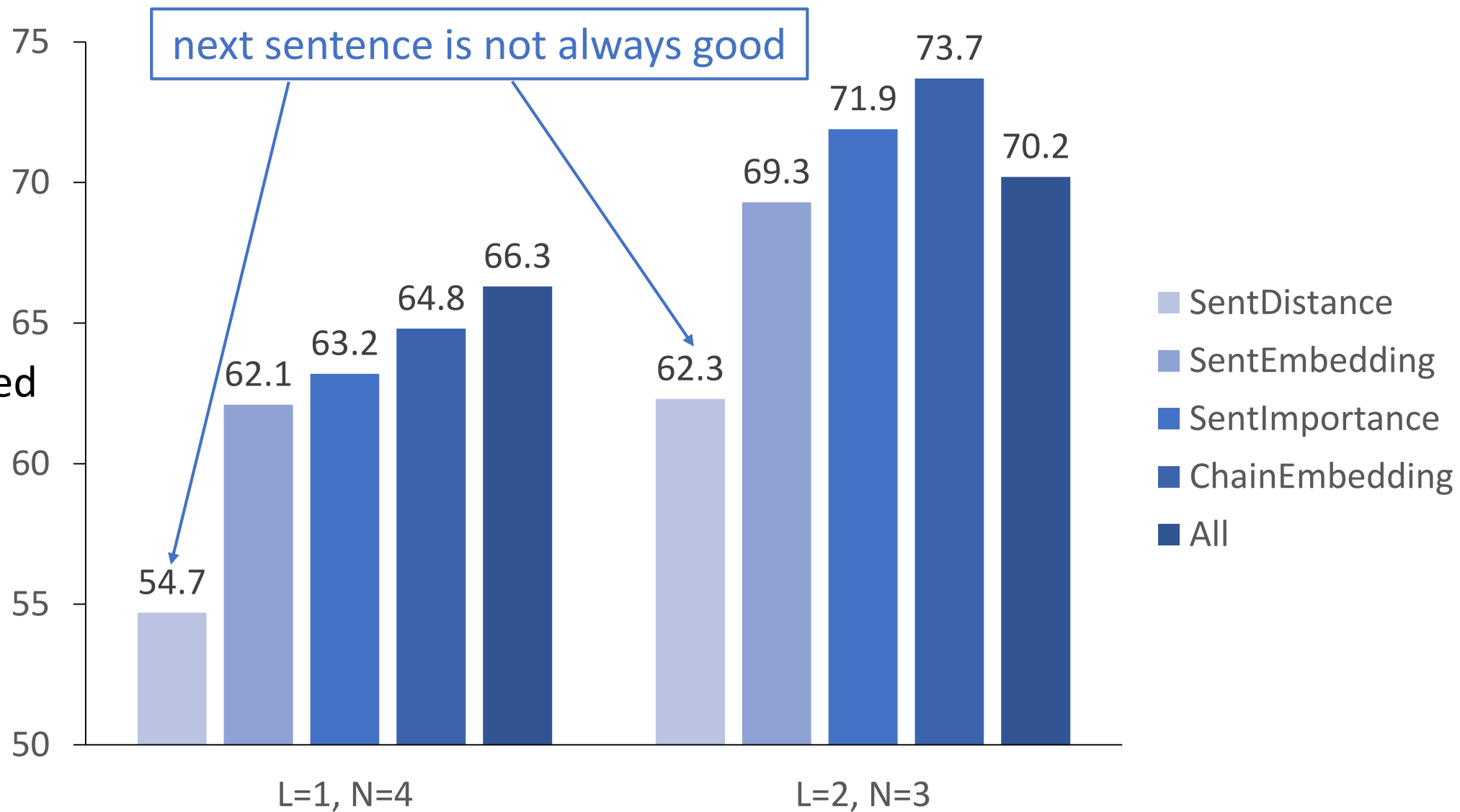
o Features

   o SentImportance: $r(y_i \mid D)$

   o SentDistance: $d(y_i \mid s_1, s_2, \ldots, s_L) = SentIdx(y_i) - SentIdx(s_L)$

   o SentEmbedding: $e(y_i)$

   o ChainEmbedding: $c(y_i \mid s_1, s_2, \ldots, s_L)$

used for ranking sentences given $s_1, s_2, \ldots, s_L$

TextRank unsupervised summarization on the document D

Pre-trained BERT

Test Set Results

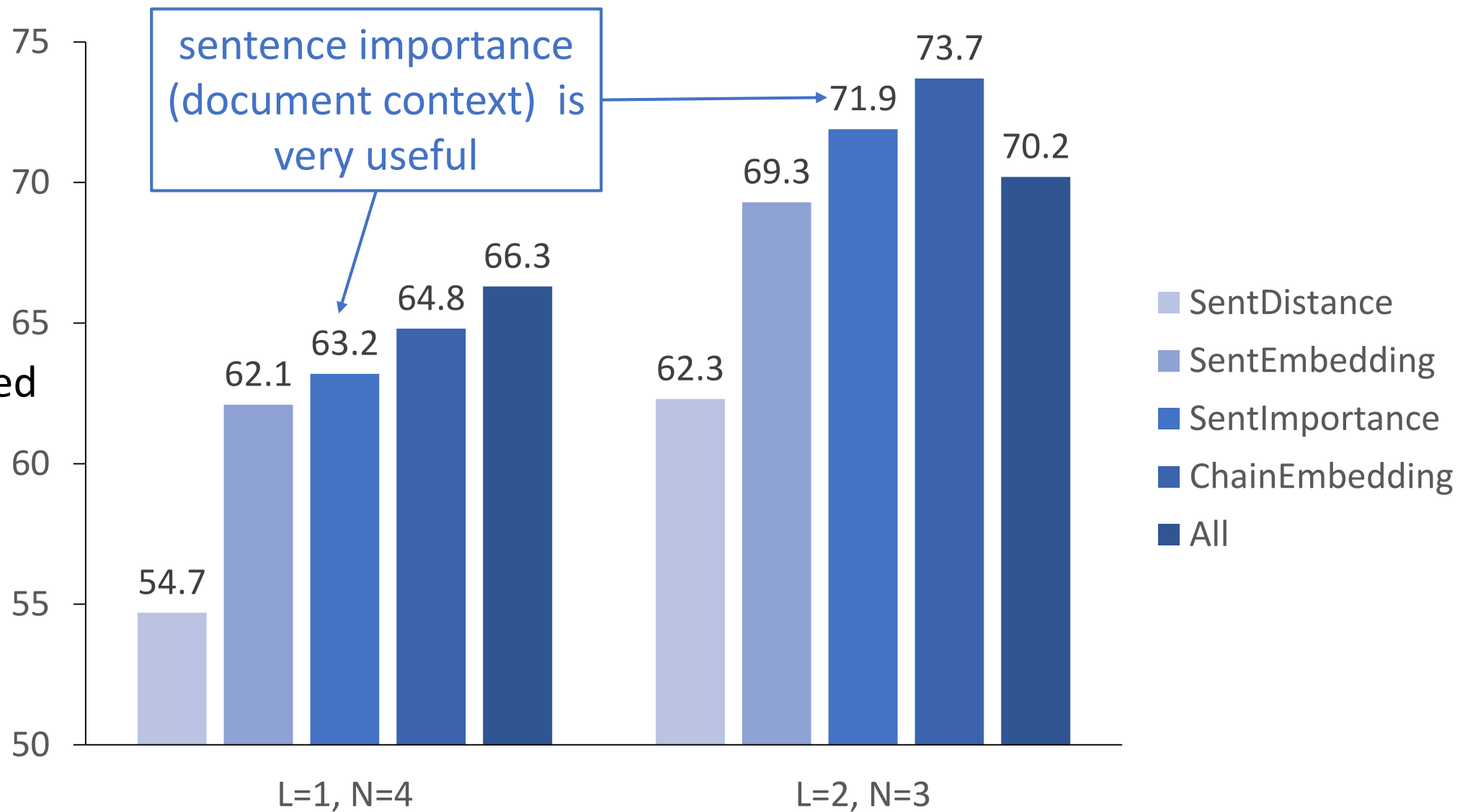% the highest-ranked sentence has a positive label

next sentence is not always good

Legend:
- SentDistance
- SentEmbedding
- SentImportance
- ChainEmbedding
- All

L=1, N=4: 54.7, 62.1, 63.2, 64.8, 66.3

L=2, N=3: 62.3, 69.3, 71.9, 73.7, 70.2

21

Test Set Results

% the highest-ranked sentence has a positive label

Chart legend:
- SentDistance
- SentEmbedding
- SentImportance
- ChainEmbedding
- All

Callout box: sentence embedding alone may capture some features about importance / style (e.g., length, informativeness)

L=1, N=4:
- SentDistance: 54.7
- SentEmbedding: 62.1
- SentImportance: 63.2
- ChainEmbedding: 64.8
- All: 66.3

L=2, N=3:
- SentDistance: 62.3
- SentEmbedding: 69.3
- SentImportance: 71.9
- ChainEmbedding: 73.7
- All: 70.2

22

Test Set Results

% the highest-ranked sentence has a positive label

sentence importance (document context) is very useful

Legend: SentDistance, SentEmbedding, SentImportance, ChainEmbedding, All

L=1, N=4: 54.7, 62.1, 63.2, 64.8, 66.3
L=2, N=3: 62.3, 69.3, 71.9, 73.7, 70.2

Test Set Results

% the highest-ranked sentence has a positive label

dialog context is important as the chain gets longer

+4.4

+2.7

L=1, N=4: 54.7, 62.1, 63.2, 64.8, 66.3

L=2, N=3: 62.3, 69.3, 71.9, 73.7, 70.2

Legend:
- SentDistance
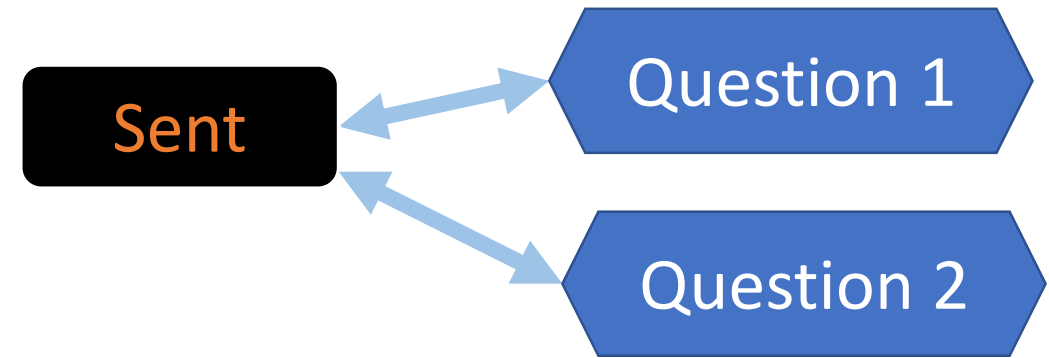- SentEmbedding
- SentImportance
- ChainEmbedding
- All

Test Set Results

using all features (2050-dimensional) overfits for L=2 (1239 training samples)

% the highest-ranked sentence has a positive label

Legend:
- SentDistance
- SentEmbedding
- SentImportance
- ChainEmbedding
- All

L=1, N=4: 54.7, 62.1, 63.2, 64.8, 66.3

L=2, N=3: 62.3, 69.3, 71.9, 73.7, 70.2

Question Generation

Sent → Question 1

Sent → Question 2
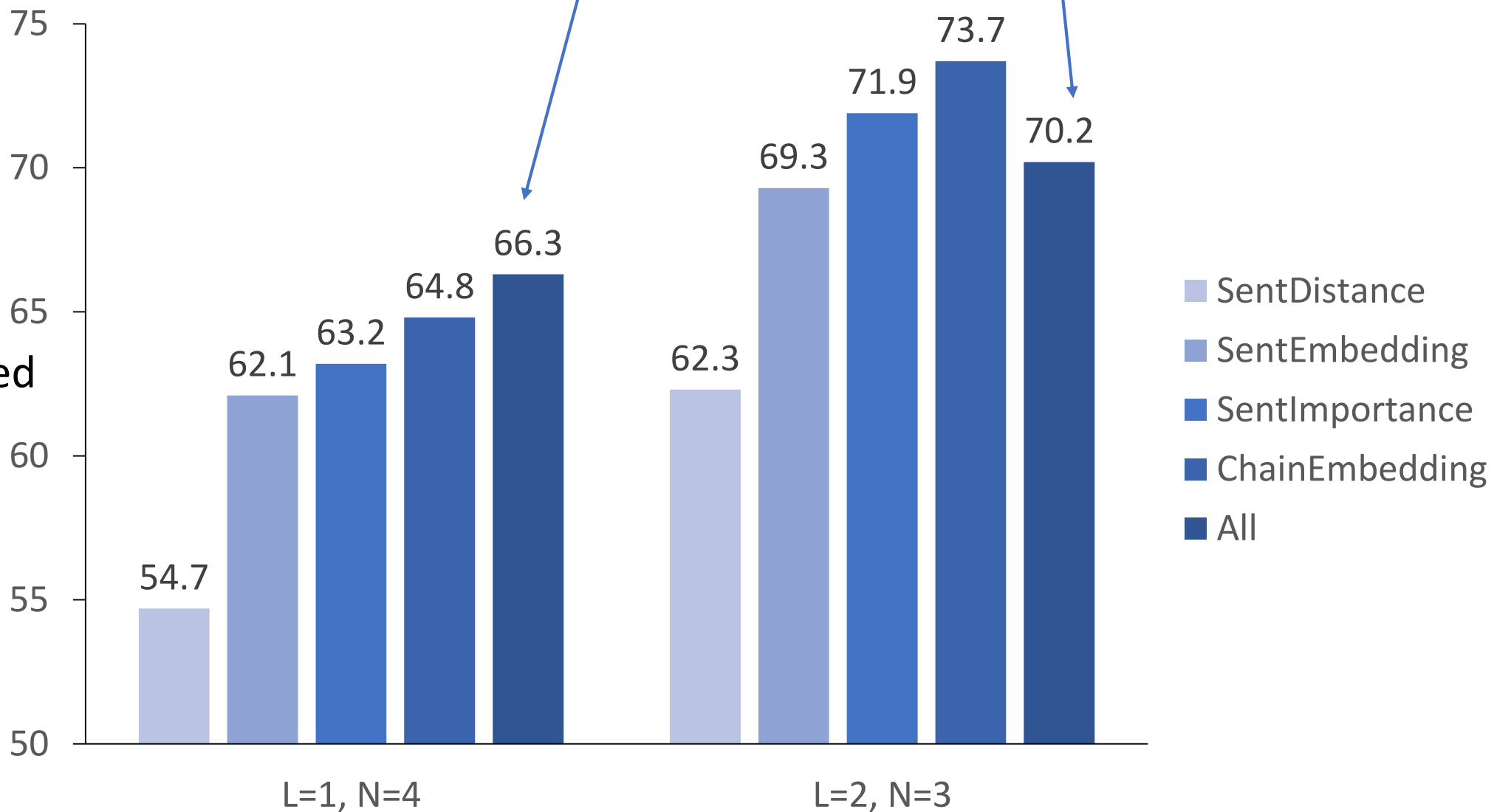
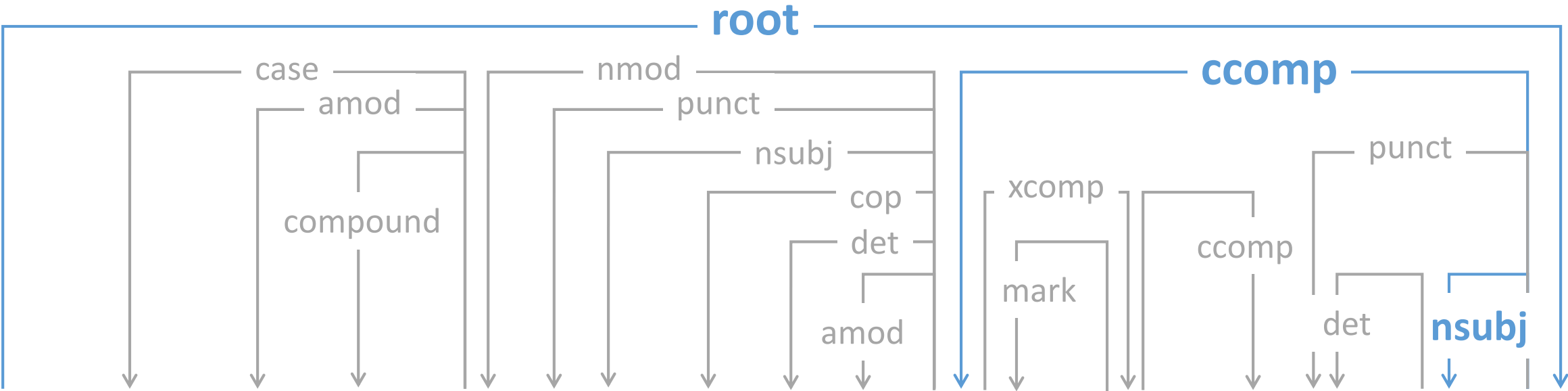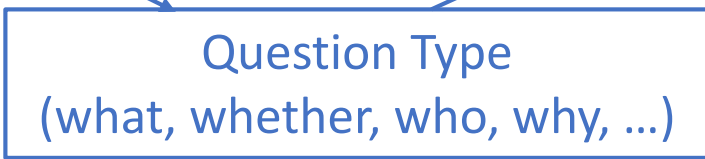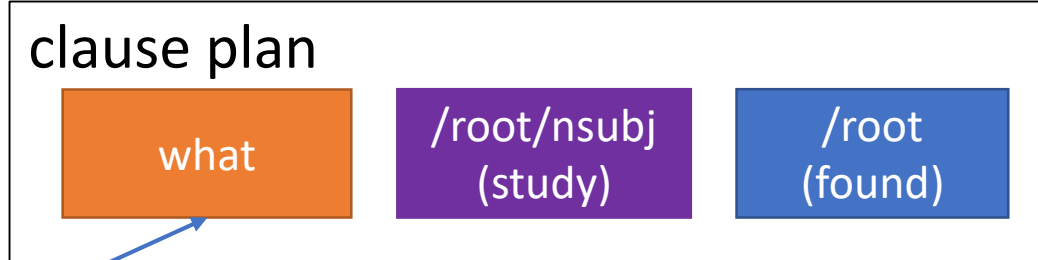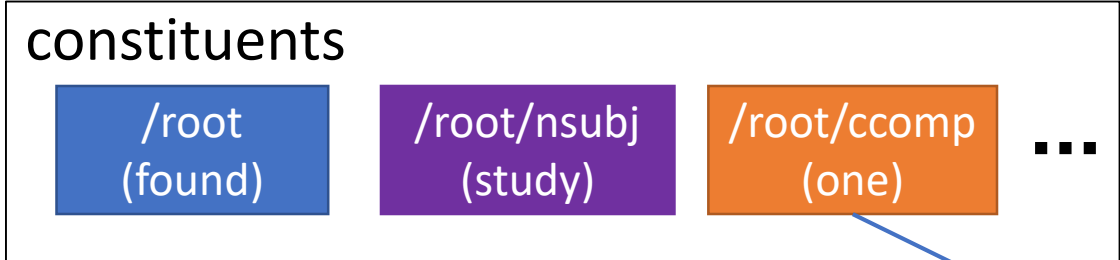| Dependency Parsing	Universal Dependencies
Dependent Selection for Answer	Question Interestingness/Importance
Question Type Classification	Hand-Crafted Decision Tree
Clause/Question Planning	Template-Based Planning
Clause/Question Realization	Dependency-Based Realization

Question Generation

Evaluation of Generated Questions

- As a transition clause for introducing Sent2 given Sent1
  - *do you want to know _____?*

- 4 question generation methods
  - generic: *more about this article*
  - constituency-based (Heilman, 2011)
  - dependency-based
  - human-written

- Human judgments on question pairs (A, B, cannot tell)
  - 134 sentences, 5 judgments per pair

Sent 1

Sent 2

Do you want to know _____?
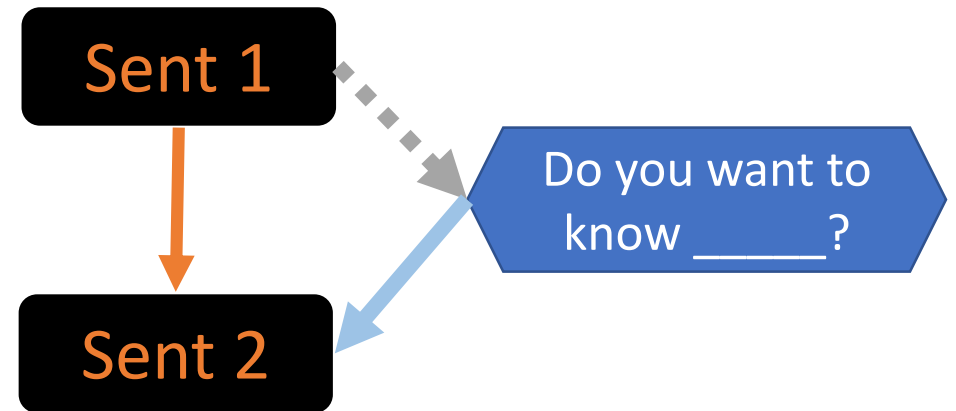
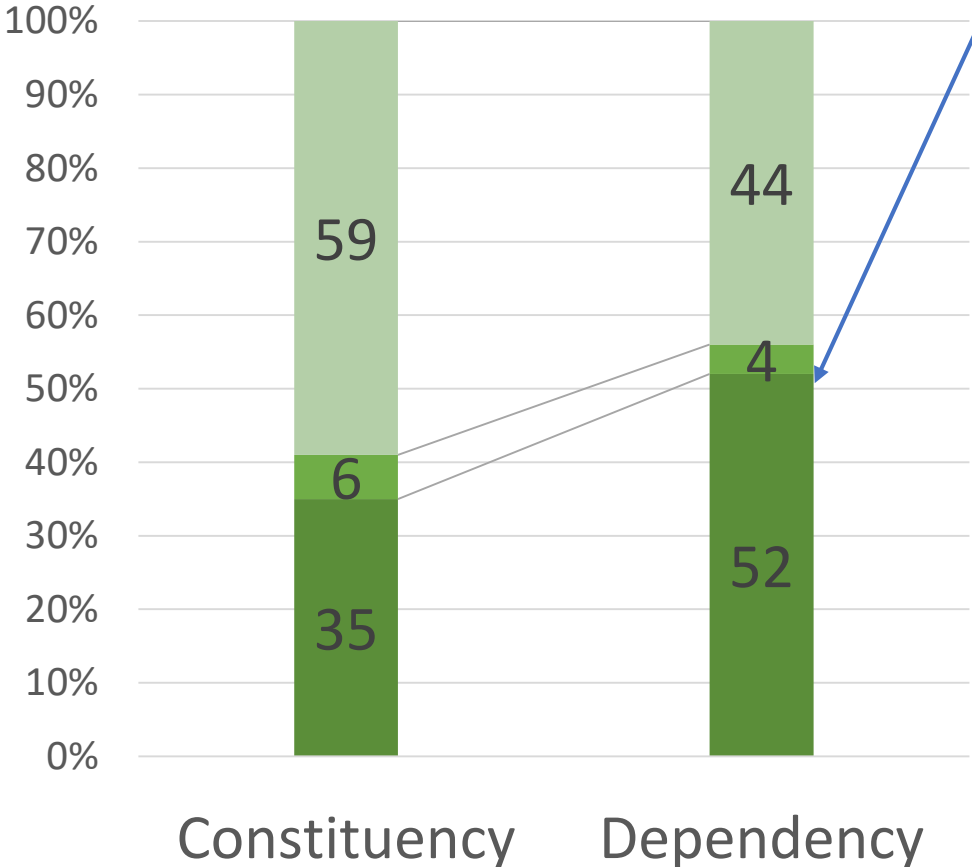# Overall Quality

dependency-based outperforms constituency-based, but does not achieve "human performance"
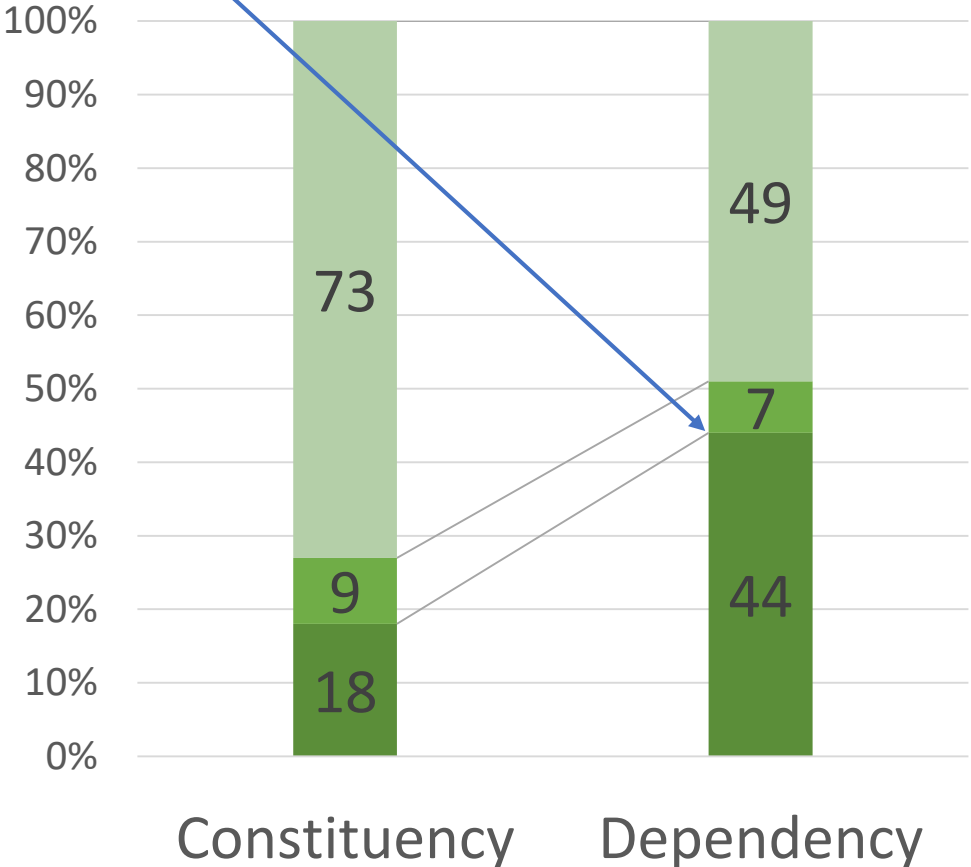
Informativeness

vs. Generic

dependency-based method generates much more informative questions (better than human)

vs. Human

30
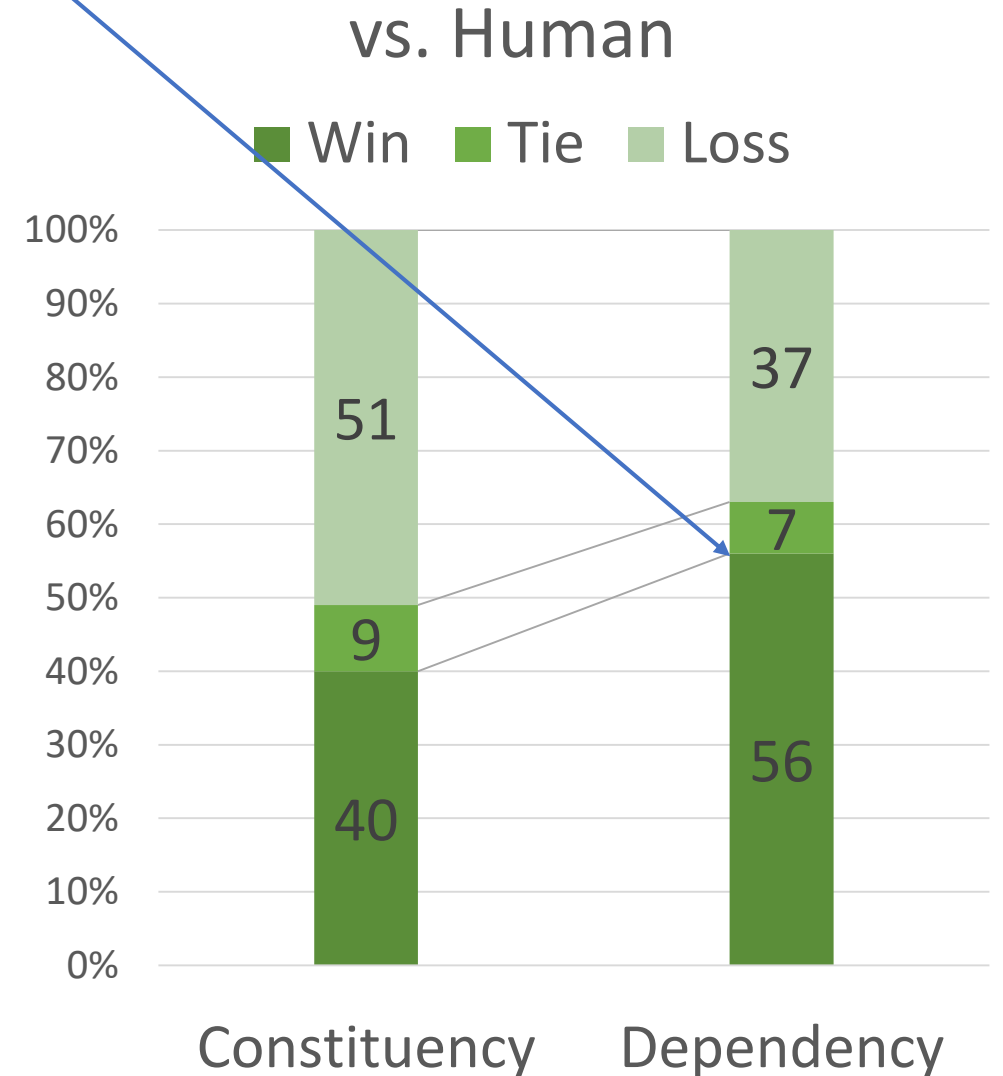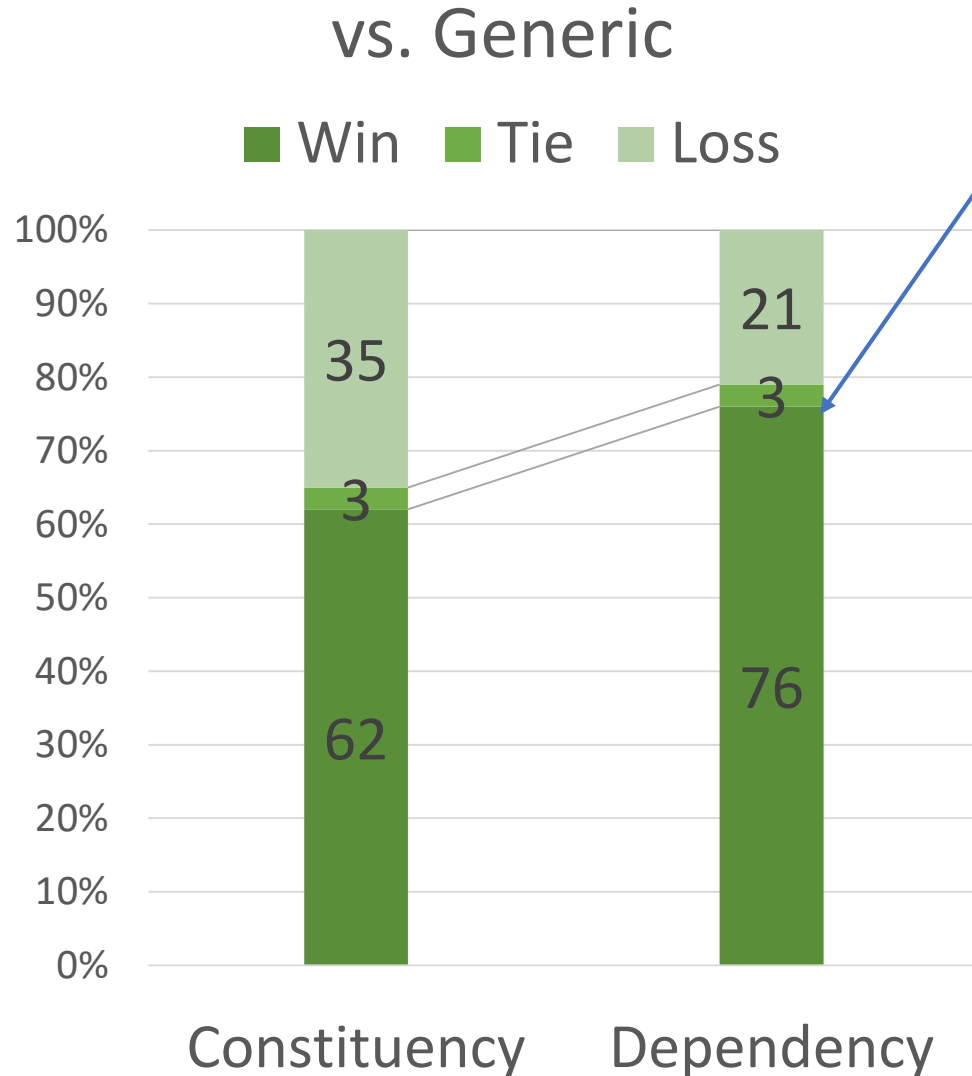
Transition Smoothness
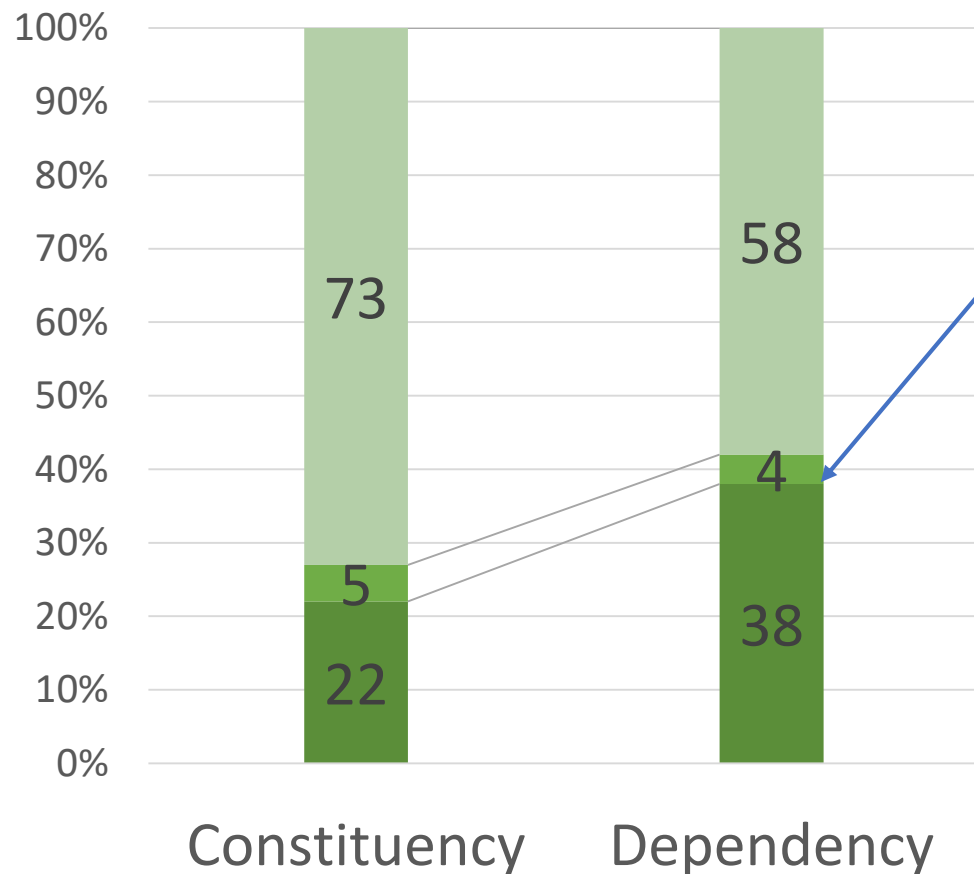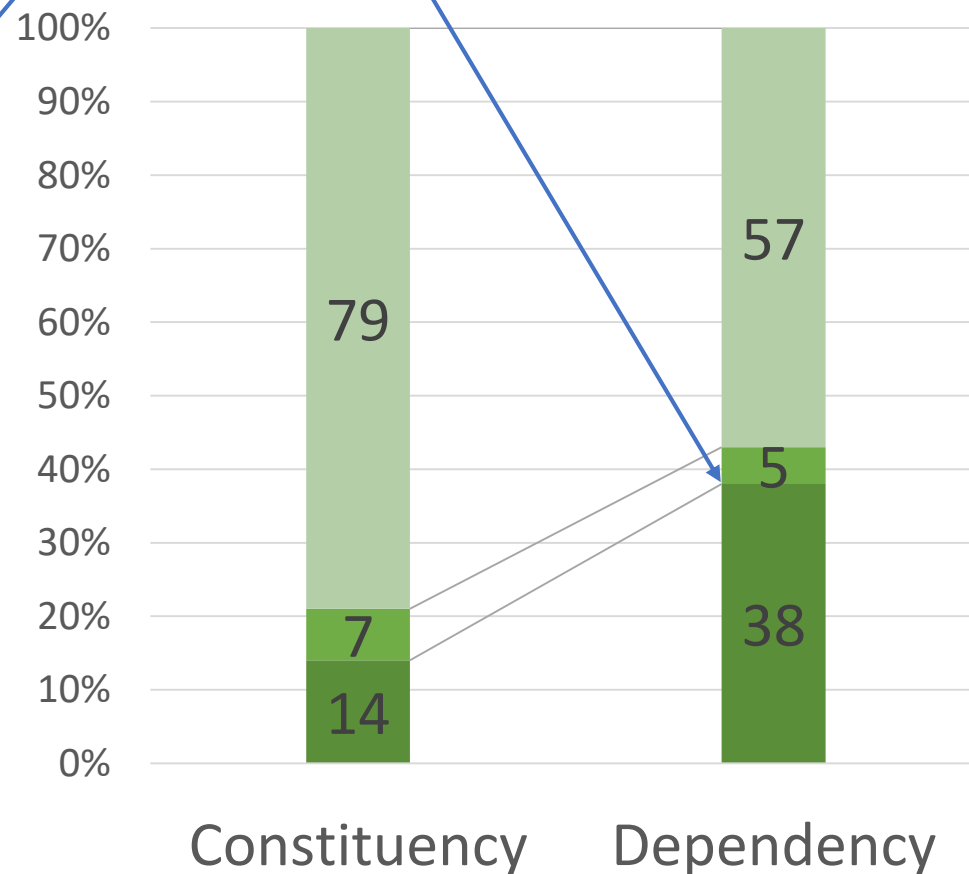
vs. Generic

vs. Human

dialog context is important!

Agenda

o Background

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations

o Summary and Future Directions

Motivation: Evaluation & Diagnosis

o Users only give an optional conversation rating

o Aspects that influence user ratings?

    o prior model-free metrics do not outperforms conversation length

o Structure of socialbot conversations?

    o prior models of dialog structure are not suitable

o Diagnosis calls for more than conversation scores

    o a conversation can involve good and bad segments/topics/policies/...

Correlation Analysis

Multi-Level Scoring

Conversation Acts for User Turns

- AskQuestion
- RequestHelpOrRepeat
- ProposeTopic
- AcceptTopic
- RejectTopic
- FollowAndNonNegative

Rule-Base Tagging

- InterestedInContent
- NotInterestedInContent
- PositiveToContent
- NegativeToContent
- PositiveToBot
- NegativeToBot

Model-Base Tagging

Correlation Analysis

For each act A

- number of turns N_A
- percentage of turns P_A

N_A cannot tell any negative correlation

Conversation Length $r = 0.15$

Pearson r with conversation user ratings

r_{num} r_{pct}

-0.2   -0.1   0   0.1   0.2

AskQuestion
RequestHelpOrRepeat
ProposeTopic
AcceptTopic
RejectTopic
FollowAndNonNegative
InterestedInContent
NotInterestedInContent
PositiveToContent
NegativeToContent
PositiveToBot
NegativeToBot

35
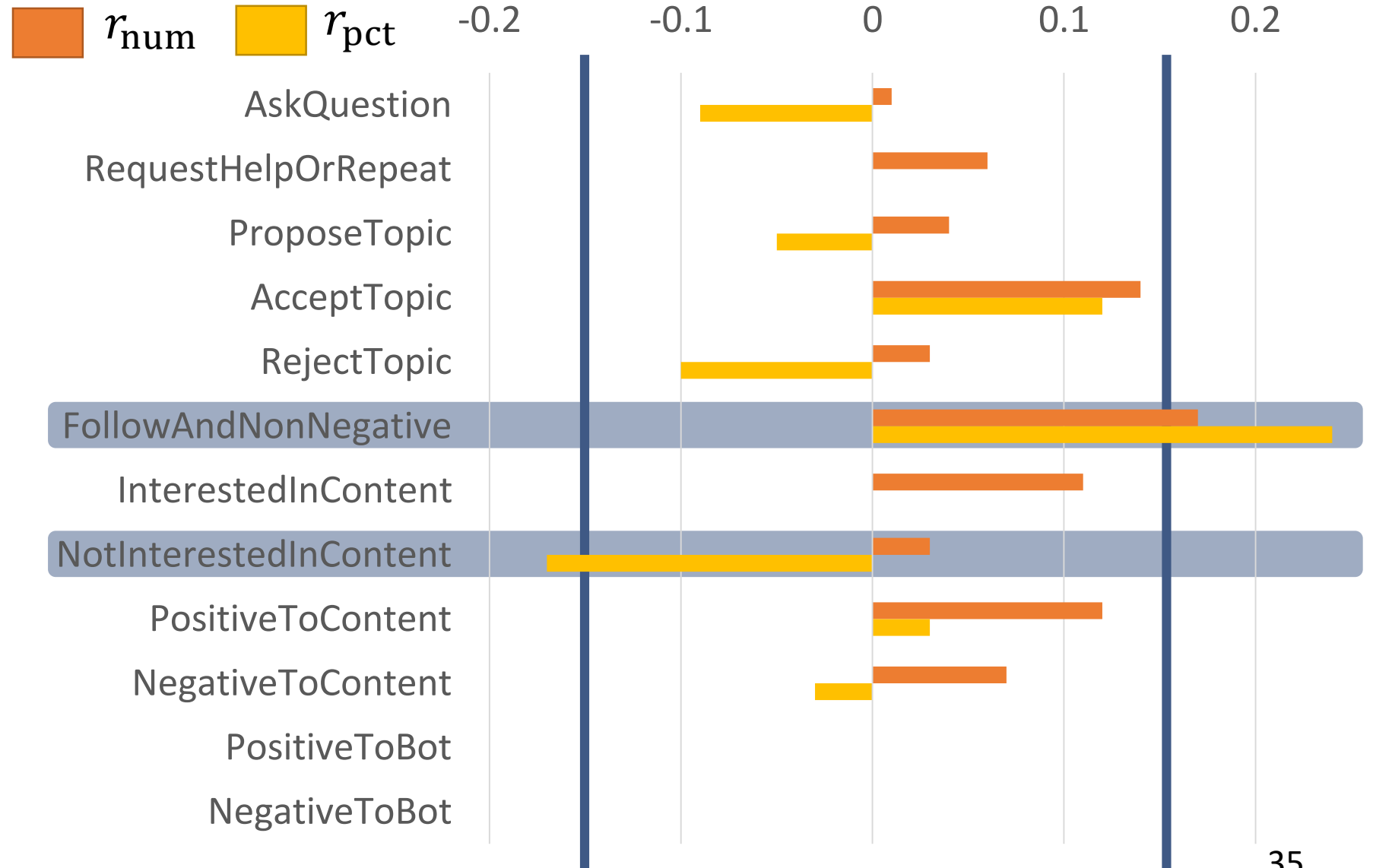
Correlation Analysis

It is a good sign that user follows the conversation flow when the bot is the primary speaker

Design, learn, & maintain engaging conversation flows (≠ system-initiative)

r_{num} r_{pct}

-0.2   -0.1   0   0.1   0.2

AskQuestion
RequestHelpOrRepeat
ProposeTopic
AcceptTopic
RejectTopic
FollowAndNonNegative
InterestedInContent
NotInterestedInContent
PositiveToContent
NegativeToContent
PositiveToBot
NegativeToBot

36
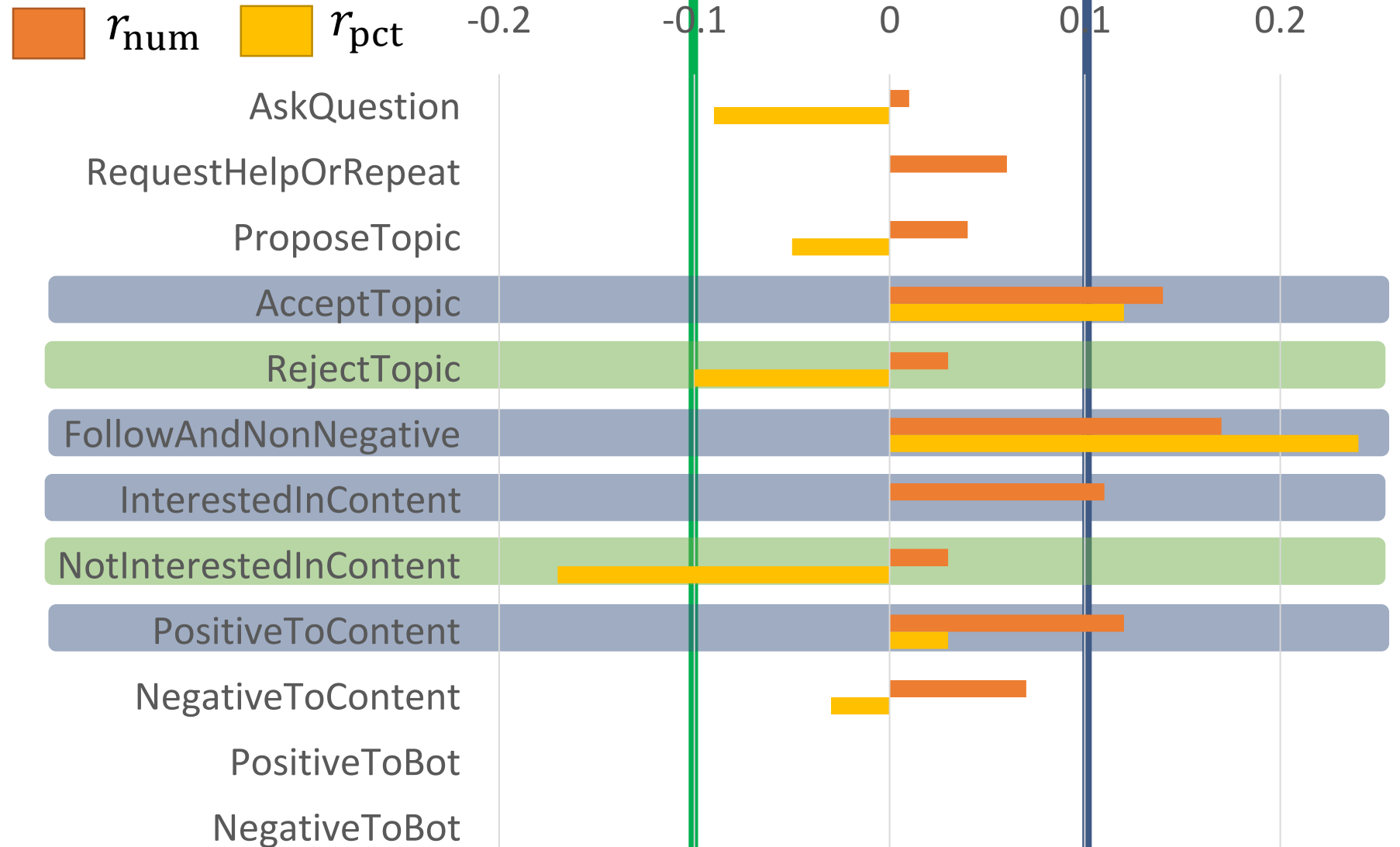
Correlation Analysis

AskQuestion and ProposeTopic slightly impact user ratings in the negative direction

Improve the bot's capability of handling user questions and topic requests

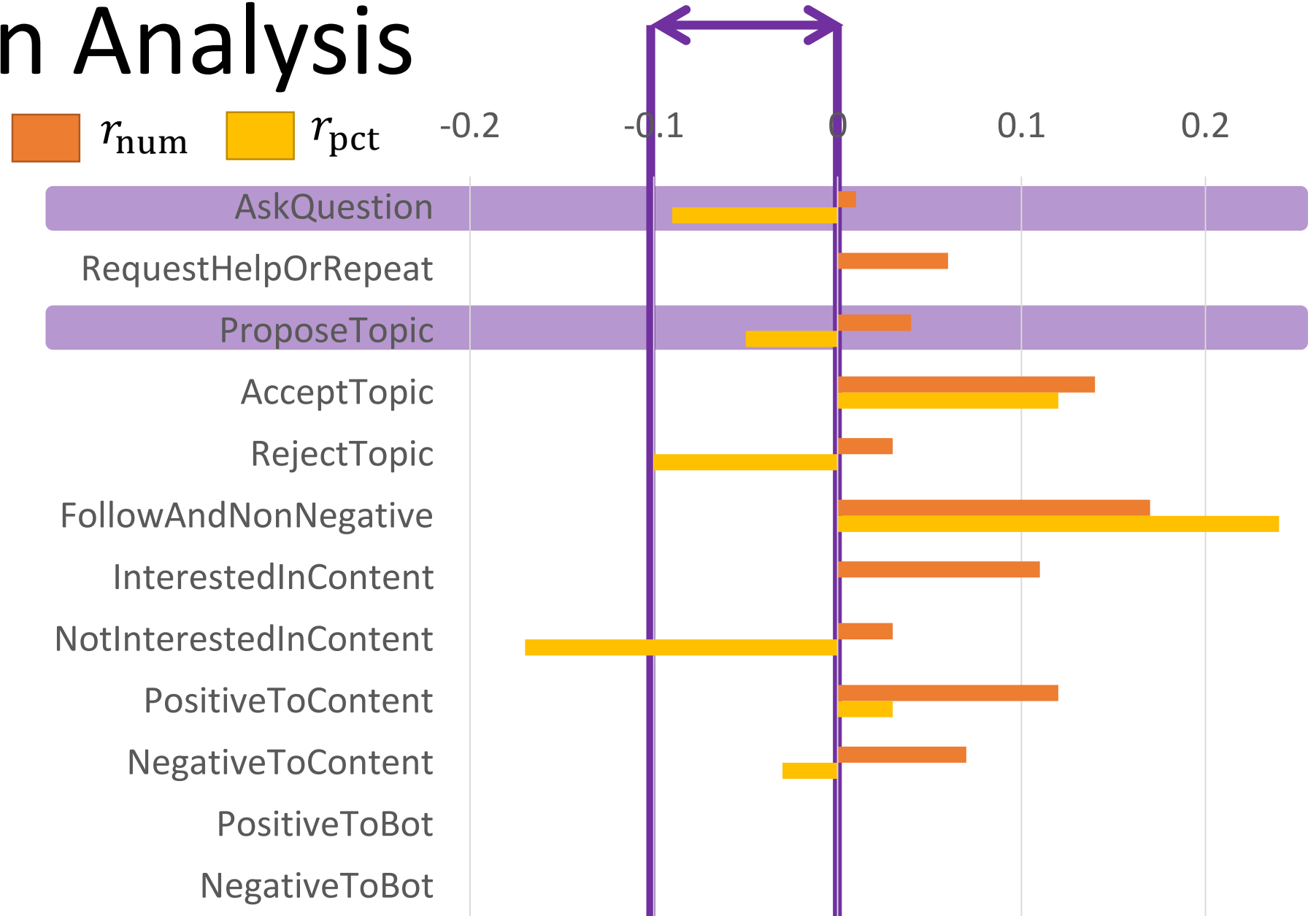# Limitations

o Conversation ratings and conversation-act-based metrics do not tell

- o which topics are handled badly by the bot
- o which dialog policies need improvement
- o which content sources have less suitable quality

o Segment-level scores can tell us more, but

- o how to segment a socialbot conversation?
- o how to compute a segment-level score?

Hierarchical Dialog Model

Subdialog

Microsegment
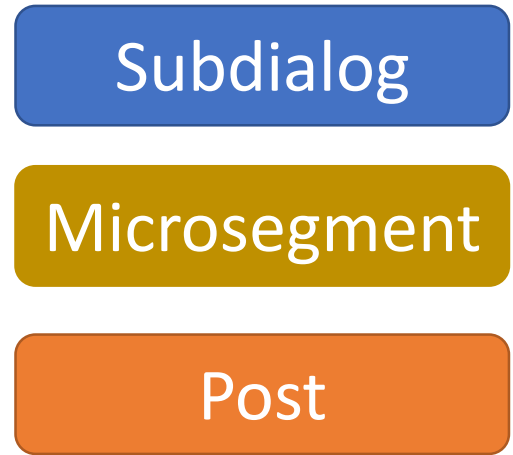
Post

o A conversation is a sequence of topical subdialogs, each of which is a sequence of microsegments, each of which contains posts

| SmallTalk | Cats | Batman | Robots |

| Batman vs. Superman | Henry Cavill | Ben Affleck |

| fun fact | amusing thought | news headline |
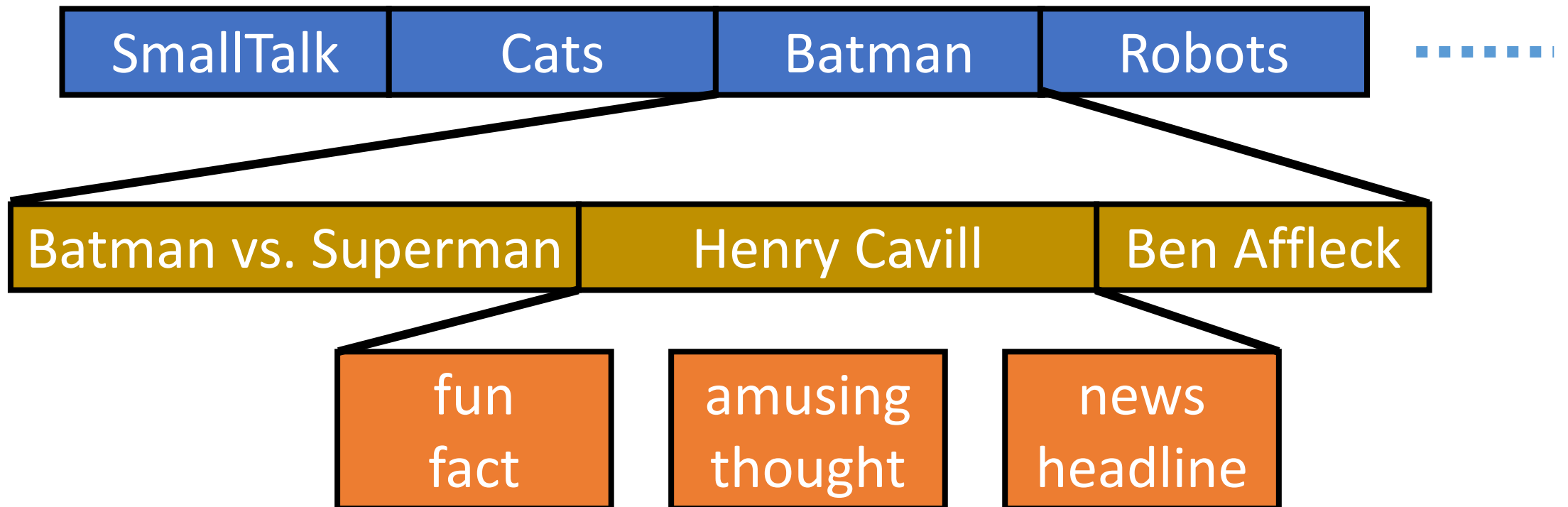
Automatic Segment Scoring

o Labels: conversation-level user ratings

o Features

  o conversation-act-based metrics

  o other features such as bag-of-words, verbosity, …

o Two different model hypotheses

  o H1: segment scores are predicted just like conversation scores

  o H2: a conversation score is some aggregation of segment scores

Automatic Segment Scoring

o H1: Linear Scoring Model

  o segment score = f(segment features)

  o conversation score = f(conversation features)

  o $f(x_1, \dots, x_d) = \sum_{i=1}^{d} u_i x_i + u_0$

o H2: BiLSTM Scoring Model

  o segment score $s_t = h_t$(segment features)

    o $h_1, h_2, \dots, h_T$: BiLSTM over individual segments

    o $s_{mean} = \text{mean}(s_1, s_2, \dots, s_T)$, ...

  o conversation score = $g(s_{mean}, s_{max}, s_{min})$

  o $g(s_{mean}, s_{max}, s_{min}) = \sum v_i s_m + v_0$

Both learned from conversation-level rating regression

■ NumTurns
■ Linear
■ Subdialog BiLSTM

Pearson r

0.4

0.2

0

41
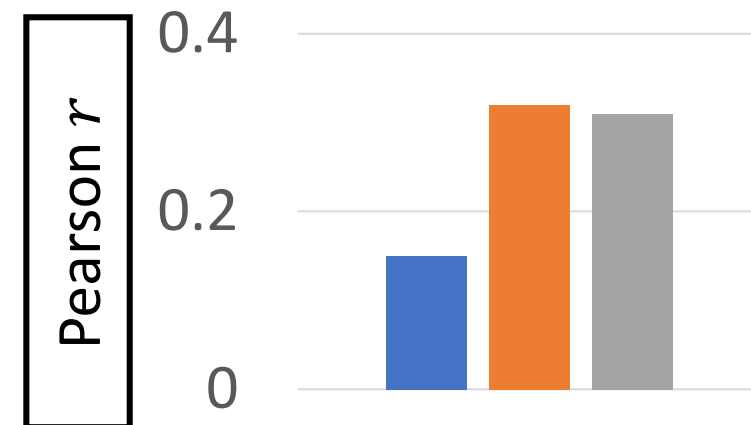
Evaluation of Subdialog Scores

o Human judgments on subdialog pairs (A, B)

- o 250 within-conversation pairs (same user)
- o 250 cross-conversation pairs (same topic)
- o 5 judgments per pair

o Spearman rank correlation ρ between x and y

- o x = votes on A – votes on B
- o y = score of A – score of B
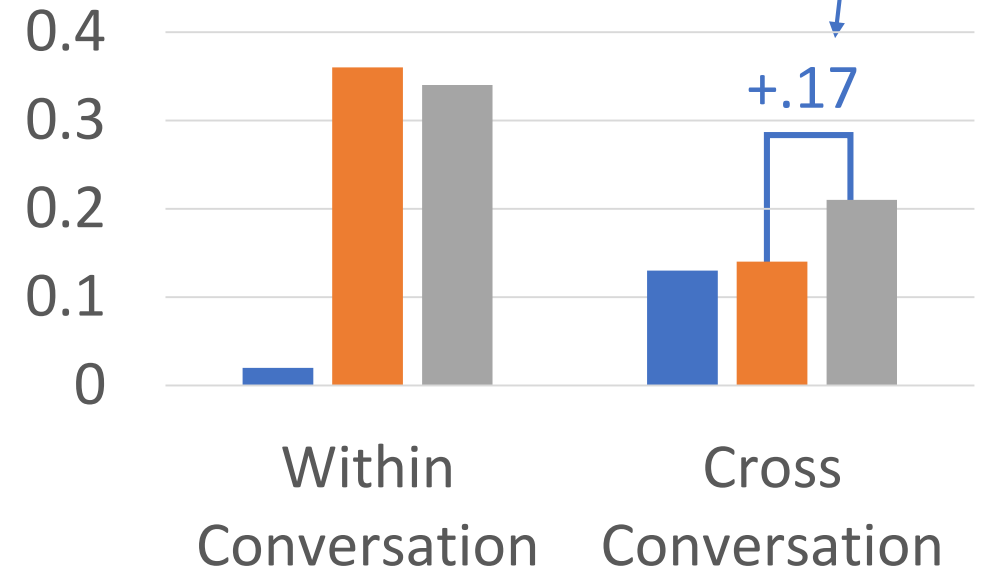
42

Agenda

o Background

o Sounding Board System – 2017 Alexa Prize Winner

o A Graph-Based Document Representation for Dialog Control

o Multi-Level Evaluation for Socialbot Conversations

o Summary and Future Directions

Summary: Sounding Board System

o A mixed-initiative and open-domain socialbot

    o user-centric and content-driven dialog strategies

    o it is a new and fast-growing area and we are one of the pioneers

    o several strategies have influenced 2018 socialbots

o System architecture

    o a hierarchical DM framework for efficient dialog control

    o social chat knowledge graph

    o several 2018 socialbots follow a similar DM architecture and acknowledge the importance of content

Summary: Graph-Based Representation

o Extended conversations grounded on a document

    o a graph-based document representation

    o bridge machine reading and dialog control

o Automatic document representation construction

    o a model for storytelling chain creation

    o an unsupervised dependency-based question generation

    o new NLP tasks that emphasize both dialog context and sentence/question interestingness

Summary: Multi-Level Evaluation

o In-depth analysis on aspects that influence user ratings

    o conversation acts for socialbot conversations

    o valuable insights for socialbot evaluation

    o better metrics than the conversation length baseline

o Automatic segment scoring for system diagnosis

    o a new hierarchical dialog model for socialbot conversations

    o two scoring models with different hypotheses for segments scores

Future Directions

o Open-domain and mixed-initiative conversational AI
  o large-scale knowledge base & computational dialog control
  o switch between two roles (primary speaker & active listener)

o Document/content analysis for conversational AI
  o unstructured text to structured representation
  o understand interestingness and socially appropriateness

o Human-in-the-loop for conversational AI
  o data collection & evaluation
  o crowd-powered system

Acknowledgements

o PhD Advisor: Mari Ostendorf

o Committee Members

   o Leah M. Ceccarelli, Yejin Choi, Hannaneh Hajishirzi, Eve Riskin, Geoffrey Zweig

o Sounding Board Team & TIAL Lab Members & Alumni

   o Hao Cheng, Elizabeth Clark, Ari Holtzman, Maarten Sap, Noah Smith

   o Amittai Axelrod, Sangyun Hahn, Ji He, Jingyong Hou, Brian Hutchinson, Aaron Jaech, Yuzong Liu, Roy Lu, Yi Luan, Kevin Lybarger, Alex Marin, Julie Medero, Farah Nadeem, Nicole Nichols, Sining Sun, Trang Tran, Ellen Wu, Victoria Zayats

o Mentors and collaborators during Internships

o Amazon Alexa Prize organizers

Thank You