



Qingyun Wang*¹, Zhihao Zhou*¹, Lifu Huang¹, Spencer Whitehead¹, Boliang Zhang¹, Heng Ji¹, Kevin Knight²

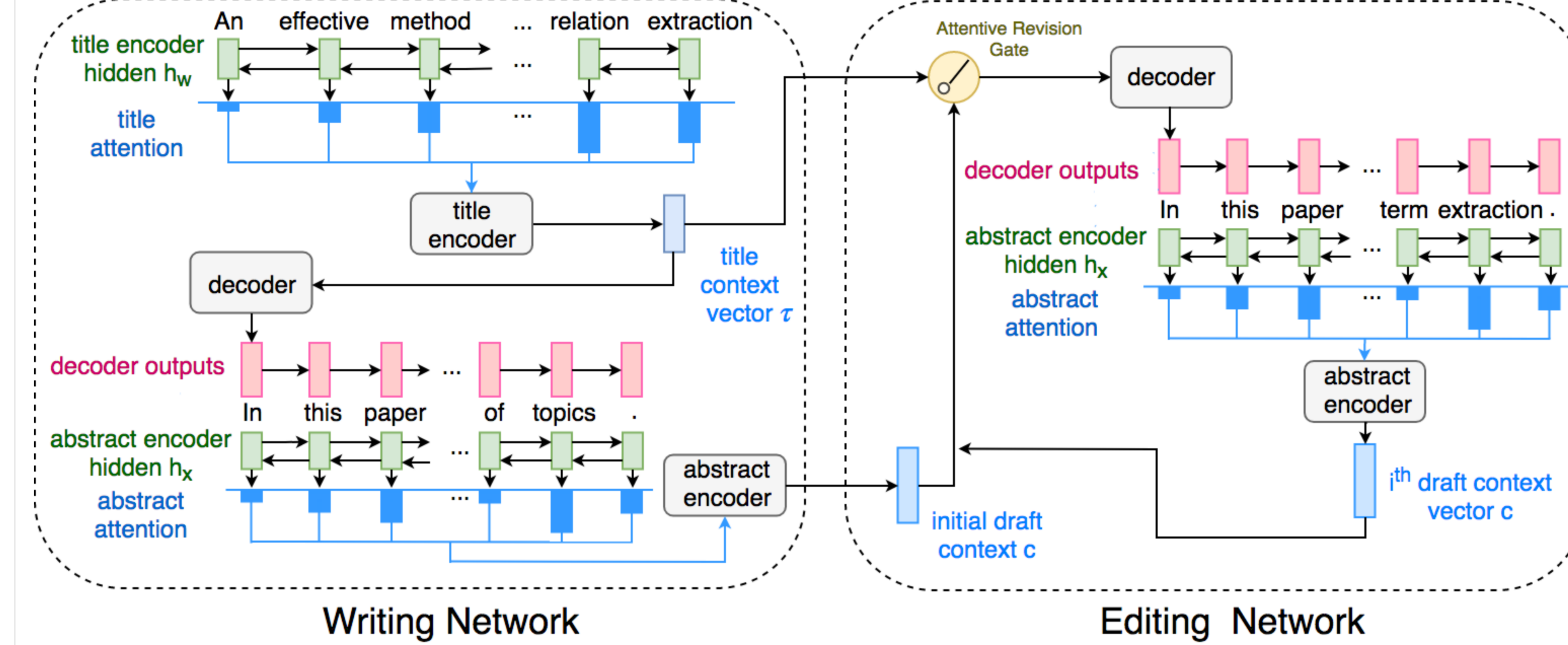
¹ Computer Science Department, Rensselaer Polytechnic Institute

² Information Sciences Institute, University of Southern California

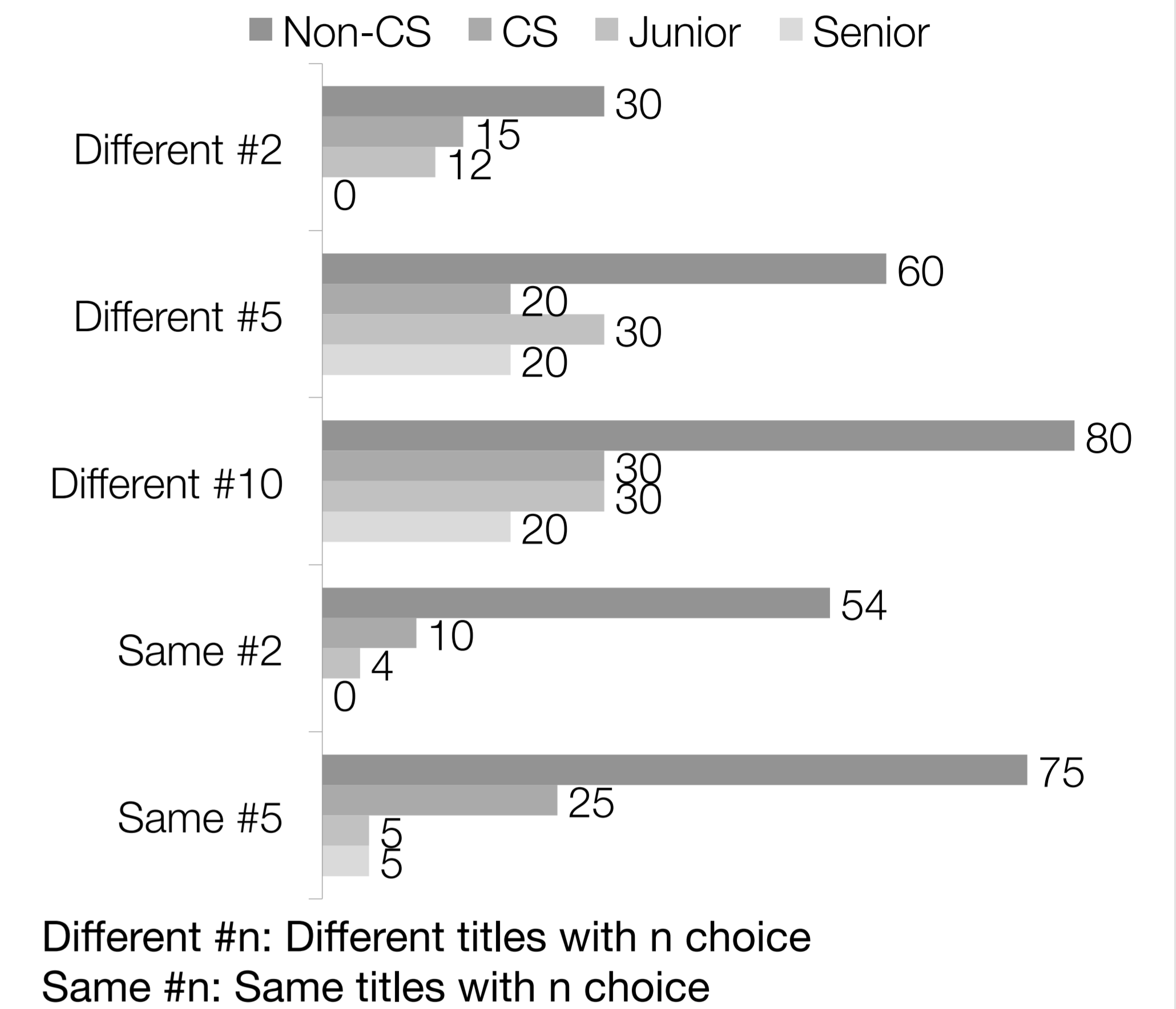
➤ Overview

- **Task:** given a title, we aim to automatically generate a paper abstract.
 - **Motivation 1:** a scientific paper abstract should **focused on the topics** specified in the title. – **Seq2seq with Attention**
 - **Motivation 2:** human writers usually start with a draft and **keep polishing and revising** it. – **Writing-Editing with Attentive Revision Gate**
- “It is perfectly okay to write garbage — as long as you **edit** brilliantly.”
— by C. J. Cherryh

➤ Approach Overview



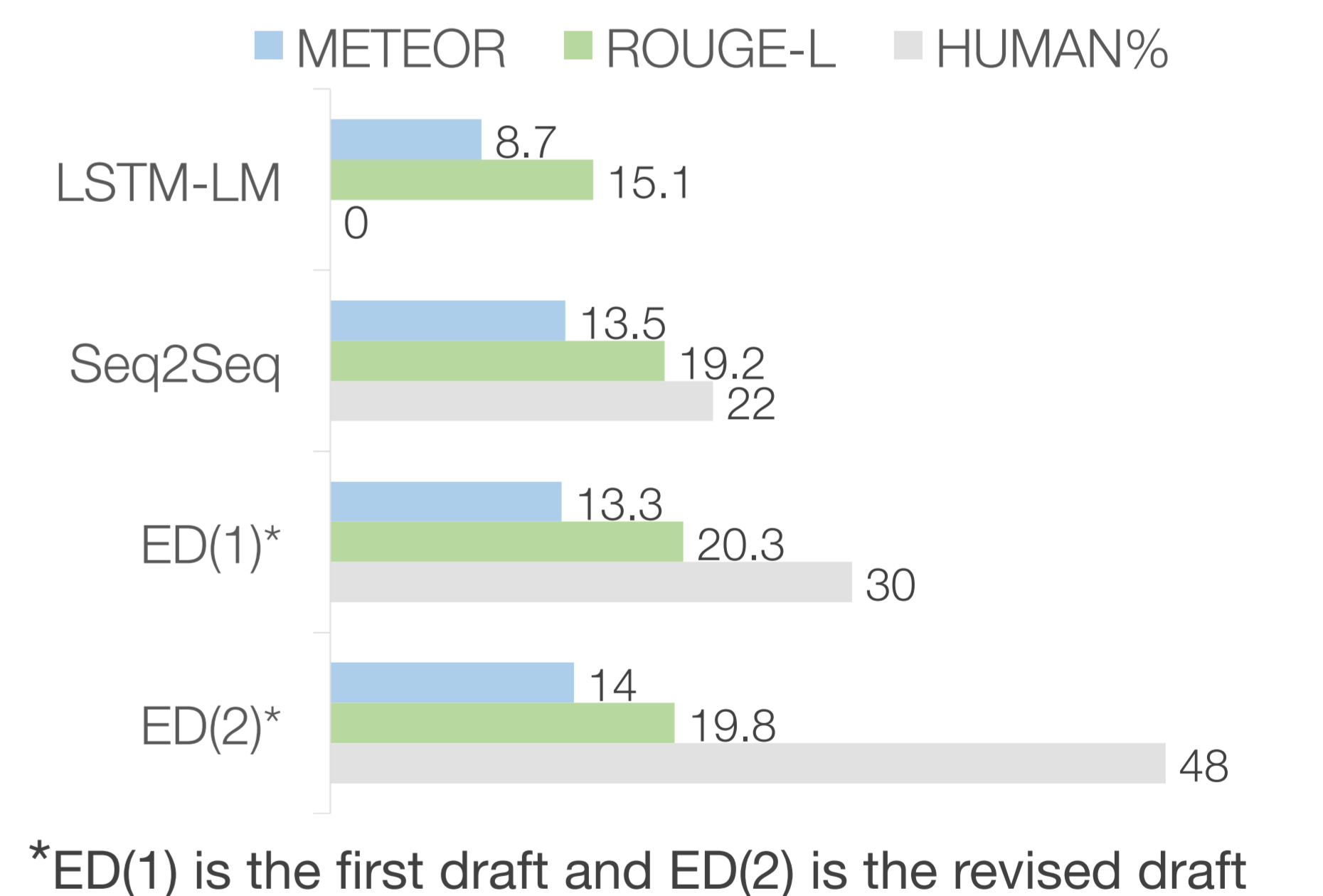
➤ Turing Test



➤ Example

- **Title:** An effective method of using **Web** based **information Relation Extraction**^[1]
- **Human written abstract:** We propose a method that incorporates paraphrase **information** from the **Web** to boost the performance of a supervised **relation extraction** system. Contextual **information** is extracted from the **Web** using a semi-supervised process, and summarized by skip-bigram overlap measures over the entire **extract**. This allows the capture of local contextual **information** as well as more distant associations. We observe a statistically significant boost in **relation extraction** performance.
- **LSTM_LM:** This paper proposes a method for automatic **extraction** of salient **information** from an original text. Our method shows promising results in which a noun clustering is employed. The **relation** is then be applied to the **relation extraction** task.
- **Initial draft:** In this paper we focus on the task of **extracting** the most commonly used **Web** pages in named entity recognition (**IE**) **information** from a large corpus of news articles. In particular, we propose a new kernel method that can be applied to a large collection of **Web** pages by automatic detection of labeled and unlabeled data. Our method is very simple and because the related method of using the data to be used as a guide as a source of **information** for the topic detection and classification can be applied, but often fail at improving the performance by combining **relations** and previous methods. In this method, the co-occurrence counts and the occurrences of each **relation** are not statistically significant, but also that it is possible to contribute to entity **relations** with a term in a variety of topics.
- **Final draft:** In this paper we focus on the task of identifying the most commonly relevant features of **Web** documents. In particular, we propose a generic, automated **IE** algorithm that can be applied to a large collection of **Web** pages containing full large documents. This is a first step in helping a wide range of collaborative works for **relation extraction**. We show that it is possible to eliminate a good number of errors in relation extraction from a variety of documents, but that it is difficult to define a problem of term **extraction**.

➤ Method Comparison



➤ Remaining Challenges

- Lack knowledge of deep connections among scientific knowledge elements
- Lack specification, conciseness
- Lack common sense knowledge
- Lack logical coherence

