The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing

Tutorial Abstracts

November 20 - 21, 2022
Message from the Tutorial Chairs

Welcome to the Tutorials Session of AACL-IJCNLP 2022.

The AACL-IJCNLP tutorials session is organized to give conference attendees a comprehensive introduction by expert researchers to some topics of importance drawn from our rapidly growing and changing research field.

The call, submission, reviewing and selection of tutorials were carried out by the tutorial chairs. The selection criteria were based on clarity and preparedness; novelty or timely character of the topic, interest for Asia-Pacific NLP community; instructors’ experience; likely audience interest; open access of the tutorial instructional material; and diversity and inclusion.

A total of 14 tutorial submissions were received, of which 6 were selected for presentation at AACL-IJCNLP. We solicited two types of tutorials, namely cutting-edge themes and introductory themes. The 6 tutorials for ACL include an introductory tutorial, a tutorial combining introductory and cutting edge contents, and six cutting-edge tutorials. The introductory tutorial is dedicated to pre-trained models for Cantonese language. The tutorial on knowledge graph construction deals with both kinds of theme. The cutting-edge tutorials address situated reasoning, the effectiveness of pre-trained language models, multilingual semantics, and misinformation and media bias.

We would like to thank the authors of the tutorials for their contribution and flexibility in organizing the conference, which went from initially face-to-face to hybrid mode and finally to fully online mode. Our thanks go to the conference organizers for effective collaboration, in particular to the general chair Yulan He.

We hope you enjoy the tutorials.

AACL-IJCNLP 2022 Tutorial Co-chairs
Miguel A. Alonso
Zhongyu Wei
Organizing Committee

Tutorial Co-Chairs

Miguel A. Alonso, Universidade da Coruña, Spain
Zhongyu Wei, Fudan University, China
Table of Contents

Efficient and Robust Knowledge Graph Construction  
Ningyu Zhang, Tao Gui and Guoshun Nan ................................................................. 1

Recent Advances in Pre-trained Language Models: Why Do They Work and How Do They Work  
Cheng-Han Chiang, Yung-Sung Chuang and Hung-yi Lee ........................................... 8

When Cantonese NLP Meets Pre-training: Progress and Challenges  
Rong Xiang, Hanzhuo Tan, Jing Li, Mingyu Wan and Kam-Fai Wong .............................. 16

Grounding Meaning Representation for Situated Reasoning  
Nikhil Krishnaswamy and James Pustejovsky .............................................................. 22

The Battlefront of Combating Misinformation and Coping with Media Bias  
Yi Fung, Kung-Hsiang Huang, Preslav Nakov and Heng Ji ........................................... 28

A Tour of Explicit Multilingual Semantics: Word Sense Disambiguation, Semantic Role Labeling and Semantic Parsing  
Roberto Navigli, Edoardo Barba, Simone Conia and Rexhina Blloshmi .......................... 35
Efficient and Robust Knowledge Graph Construction

Ningyu Zhang¹, Tao Gui², Guoshun Nan³,
¹Zhejiang University & AZFT Joint Lab for Knowledge Engine, China
²Institute of Modern Languages and Linguistics, Fudan University, China
³Beijing University of Posts and Telecommunications, China
zhangningyu@zju.edu.cn, tgui@fudan.edu.cn, nanguo2021@bupt.edu.cn

Abstract

Knowledge graph construction which aims to extract knowledge from the text corpus, has appealed to the NLP community researchers. Previous decades have witnessed the remarkable progress of knowledge graph construction on the basis of neural models; however, those models often cost massive computation or labeled data resources and suffer from unstable inference accounting for biased or adversarial samples. Recently, numerous approaches have been explored to mitigate the efficiency and robustness issues for knowledge graph construction, such as prompt learning and adversarial training. In this tutorial, we aim to bring interested NLP researchers up to speed on the recent and ongoing techniques for efficient and robust knowledge graph construction. Additionally, our goal is to provide a systematic and up-to-date overview of these methods and reveal new research opportunities to the audience.

1 Introduction

Motivation: Knowledge Graphs (KGs) regard the knowledge as fact triples in the form of \(<\text{subject}, \text{predicate}, \text{object}>\), which can benefit a wide range of natural language processing tasks including question answering (Jia et al., 2021; Fei et al., 2022; Zhang et al., 2021a), fact verification (Zhou et al., 2019), data-to-text generation (Li et al., 2021), commonsense reasoning (Bosselut et al., 2019) and so on. Knowledge graph construction tasks including Named Entity Recognition (NER) (Gui et al., 2019), Relation Extraction (RE) (Zeng et al., 2015) target to extract structural information from unstructured texts, have appealed to researchers in NLP community. While those researchers have largely separated approaches from tasks, they have encountered similar issues such as efficiency and robustness.

Intuitively, efficient and robust knowledge graph construction has been widely investigated due to its potential value of making models scenario-adaptable, data-efficient, and particularly convenient for real-world applications with cold-start issues. In this tutorial, we take a holistic view of knowledge graph construction, introducing the commonalities in the issues and solutions regarding efficiency and robustness. We will explore the approaches of named entity recognition and relation extraction with few-shot labeled data, limited computation resources and approaches to improve the model robustness.

Note that our tutorial is related to green deep learning (Xu et al., 2021) that appeals to researchers to focus on carbon emission and energy usage during model training and inference, and relevant to robust NLP (Omar et al., 2022) which focuses on addressing issues in current models’ language understanding capabilities with adversarial attacks. Meanwhile, trends within knowledge graph construction have shifted toward low-resource rather than considering massive labeled data and reliable & trustworthy knowledge graph construction. Notably, it is worth considering the knowledge graph construction tasks as a whole to develop methodologies for efficiency and robustness issues. We will discuss these works and suggest avenues in the future.

Tutorial Content: We will start this tutorial by defining tasks of knowledge graph construction, including named entity recognition relation extraction from sentences or documents. Then, we will give introductions to the basic models, open datasets and tools used in knowledge graph construction covering both English and Chinese (Zhang et al., 2022). We plan to focus on methods that enable efficient knowledge graph construction, such as the distant supervision (Wang et al., 2022b) and data augmentation paradigms of creating training data (Liu et al., 2021), model enhancement methods like meta-learning (Yu et al., 2020), transfer learning (Ma et al., 2022a) and prompt learn-
ing (Chen et al., 2022d,c,b), parameter-efficient approaches (Ma et al., 2022b; Chen et al., 2022a), including adaptor-based tuning. We will then explore research focusing on robust knowledge graph construction for stable learning with adversarial attacks and selection or semantic biases.

During the tutorial, we plan to deliver lessons learned from the diverse communities involved in knowledge graph construction research and will introduce insights from the industry when building a business knowledge graph in low-resource settings. Section 3 has an outline of tutorial content.


Relevance to AACL: Knowledge graphs benefit many crucial NLP tasks, and knowledge graph construction tasks such as relation extraction and named entity recognition are core tasks in information extraction. A 2018 NAACL tutorial, “Scalable Construction and Reasoning of Massive Knowledge Bases" (Ren et al., 2018), introduced a summary of recent KB, and IE works. More recently, an ACL tutorial, "Multi-modal Information Extraction from Text, Semi-structured, and Tabular Data on the Web" (Dong et al., 2020), provided an overview of information extraction (IE) from Web data with two vital dimensions: the thrust to develop scalable approaches and the diversity in data modality. However, previous tutorials mainly focus on models with rich resources of labeled data and computation, and recent years have witnessed the fast development of efficient and robust knowledge graph construction. On the other hand, the NLP community has paid much attention to robust NLP, such as an EMNLP 2021 tutorial, "Robustness and Adversarial Examples in Natural Language Processing" (Chang et al., 2021). Different from this tutorial in general NLP, we target a small, focused domain of knowledge graph construction and introduce the detailed latest work in limited 3 hours.

2 Type of this Tutorial

This tutorial contains cutting-edge approaches in general knowledge graph construction approaches regarding efficiency and robustness issues. However, our coverage of this tutorial will contain introductory material of knowledge graph construction for widespread audiences of the NLP community. Besides, we will introduce methods of Chinese knowledge graph construction for Asia audiences.

3 Outline

1. (1 hour) Introduction and Applications
   - Named Entity Recognition (NER)
     - Flat NER (Li et al., 2020)
     - Nested NER (Straková et al., 2019)
     - Joint Flat and Nested NER (Wang and Lu, 2020)
   - Relation Extraction
     - Supervised Relation Extraction (Lin et al., 2016; Song et al., 2018; Nan et al., 2021)
     - Distance-supervised Relation extraction (Zeng et al., 2015)
     - Open Relation Extraction (Wu et al., 2019; Zhao et al., 2021)
   - Knowledge Graph Construction
     - Introduction (Bosselut et al., 2019)
     - Industry Examples
     - Resource Applications and Toolkits
     - Importance of the Efficiency and Robustness

2. (1 hour) Efficient KG Construction
   - Data Efficiency
     - Data Augmentation (Chaudhary et al., 2019)
     - Model Enhancement (Chen et al., 2022d)
     - Hybrid Approaches (Hu et al., 2021)
   - Model Efficiency
     - Parameter-efficient Learning (Zhou et al., 2021)
     - Efficient Architecture (Zhu, 2021)
   - Inference Efficiency
     - Generative Inference (Yan et al., 2021)
     - Non-autoregressive Decoding (Sui et al., 2021)

3. (1 hour) Robust KG Construction
   - Robustness Problem Discovery
     - Model Behavior Probing (Cao et al., 2021)
     - Robustness Evaluation (Wang et al., 2021)
• Data Correction
  – Data Denoising (Ma et al., 2021)
  – Data Bias Removal (Mehrabi et al., 2020)
• Robust Model Learning
  – Adversarial Training (Li and Qiu, 2021; Liu et al., 2022)
  – Robust Architecture Design (Zheng et al., 2022; Wang et al., 2022a)
  – Causal Inference (Zhang et al., 2021b)

4 Prerequisites

Anyone with a background in natural language processing can access this tutorial. Moreover, a basic understanding of neural networks, preferably with some knowledge of information extraction, knowledge graph, and pre-trained language models, is helpful.

5 Reading list

• “Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion”, (Dong et al., 2014)

• “Fonduer: Knowledge Base Construction from Richly Formatted Data”, (Wu et al., 2018)

• “A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios”, (Hedderich et al., 2021)

• “Few-Shot Named Entity Recognition: An Empirical Baseline Study”, (Huang et al., 2021)

• “Knowledge Extraction in Low-Resource Scenarios: Survey and Perspective”, (Deng et al., 2022)

• “Uncertainty-Aware Label Refinement for Sequence Labeling”, (Gui et al., 2020)

• “Reasoning with latent structure refinement for document-level relation extraction”, (Nan et al., 2020)

• “KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction”, (Chen et al., 2022d)

6 Presenters

Ningyu Zhang is an associate professor at Zhejiang University, leading the group about KG and NLP technologies. He is also a researcher at Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies (AZFT), Co-PI of the Alibaba Open Business Knowledge Graph1, which is devoted to benefiting e-commerce applications and to discovering socioeconomic values. He is a member of ACL, a member of the Youth Working Committee of the Chinese Information Processing Society of China, and the member of the Language and Knowledge Computing Professional Committee of the Chinese Information Processing Society of China. He has published many papers in top international academic conferences and journals such as ICLR, ACL, ENNLP, NAACL, and IEEE/ACM Transactions on Audio Speech and Language. He has served as PCs for NeurIPS, ICLR, KDD, ICML, AAAI, IJCAI and reviewer for ARR, TKDE, and TKDD. He has received the best paper award from the China Conference on Knowledge Graph and Semantic Computing and the best paper nominations from the International Joint Conference on Knowledge Graphs. He has won first place in the TREC Precision Medicine 2020 sponsored by the National Institute of Standards and Technology (NIST) and fourth place in the international semantic evaluation competition (SemEval 2021 Task4) sponsored by ACL. He has given multiple talks on information extraction and knowledge graph.

Email: zhangningyu@zju.edu.cn
Homepage: https://person.zju.edu.cn/en/ningyu

Tao Gui is an associate professor at the Institute of Modern Languages and Linguistics of Fudan University. He is the key member of the FudanNLP group2. He is a member of ACL, a member of the Youth Working Committee of the Chinese Information Processing Society of China, and the member of the Language and Knowledge Computing Professional Committee of the Chinese Information Processing Society of China. He has published more than 40 papers in top international academic conferences and journals such as ACL, ENNLP, AAAI, IJCAI, SIGIR, and so on. He has served as area chair or PCs for SIGIR, AAAI, IJCAI, TPAMI, and ARR. He has received the Outstanding Doctoral Dissertation Award of the Chinese Informa-

1https://kg.alibaba.com/
2https://nlp.fudan.edu.cn
tion Processing Society of China, the area chair favorite Award of COLING 2018, the outstanding Paper Award of NLPCC 2019, and a scholar of young talent promoting projects of CAST.

Email: tgui@fudan.edu.cn

Homepage: https://guitaowufeng.github.io

Guoshun Nan is a tenure-track professor in the School of Cyber Science and Engineering, Beijing University of Posts and Telecommunications (BUPT). He is a key member of the National Engineering Research Center of Mobile Network Security and a member of the Wireless Technology Innovation Institute of BUPT. Before starting his academic career, he also worked in Hewlett-Packard Company (China) for more than four years as an engineer. He is a member of ACL. He has a broad interest in information extraction, model robustness, multimodal retrieval, cyber security and the next generation of wireless networks. He has published more than ten papers in top-tier conferences such as ACL, CVPR, EMNLP, SIGIR, IJCAI, CKIM and Sigcomm. He served as a reviewer for ACL, EMNLP, AAAI, IJCAI, Neurocomputing and IEEE Transaction on Image Processing.

Email: nanguo2021@bupt.edu.cn

References


Recent Advances in Pre-trained Language Models: Why Do They Work and How Do They Work

Cheng-Han Chiang  
National Taiwan University  
dcml0714@gmail.com

Yung-Sung Chuang  
CSAIL, MIT  
yungsung@mit.edu

Hung-yi Lee  
National Taiwan University  
hungyilee@ntu.edu.tw

1 Brief Description

Deep learning-based natural language processing (NLP) has become mainstream research in recent years and has shown significant improvements over conventional methods. Among all deep learning methods, fine-tuning a self-supervisedly pre-trained language model (PLM) on downstream tasks of interest has become the standard pipeline in NLP tasks. Ever since ELMo (Peters et al., 2018) and BERT (Devlin et al., 2019) were proposed in 2018, models fine-tuned from PLMs have dominated numerous leader-boards in various tasks including question answering, natural language understanding, natural language inference, machine translation, and sentence similarity. Aside from applying PLMs on various downstream tasks, many have been delving into understanding the properties and characteristics of PLMs, including the linguistic knowledge encoded in the representations of PLMs, and the factual knowledge the PLMs acquire during pre-training. While it has been three years since PLMs were first proposed, there is no sign of decay in the research related to PLMs.

There were two tutorials focusing self-supervised learning/PLMs: a tutorial in NAACL 2019 (Ruder et al., 2019) and one in AACL 2020. However, given the ever-evolving nature of this realm, it is conceivable that there has been significant progress in the study of PLMs. Specifically, compared with PLMs back in 2019, when they are mostly held by tech giants and used in scientific research, the PLMs nowadays have become more widely adopted in various real-world scenarios by users with different hardware infrastructures and amount of data, and thus posing problems that have never arisen before. Substantial progress, including possible answers to the effectiveness of PLMs and new training paradigms, have been made to allow PLMs better deployed in more realistic settings. Hence, we see it necessary and timely to inform the NLP community about the recent advances in PLMs through a well-organized tutorial.

This tutorial is divided into two parts: why do PLMs work and how do PLMs work. Table 1 summarizes the content this tutorial will cover. This tutorial intends to facilitate researchers in the NLP community to have a more comprehensive view of the advances in PLMs during recent years, and apply these newly emerging techniques to their domain of interest. As self-supervised learning and PLMs are very popular in these days, we expect our tutorial to have at least 100 attendees.

Type of the tutorial  The type of this tutorial is Cutting-edge. We will cover the cutting-edge advances in PLMs which have been flourishing in the NLP community since 2020. No tutorial has systematically reviewed any topics that we aim to cover (as listed in Table 1) at ACL/EMNLP/NAACL/EACL/AACL/COLING.

2 Tutorial Structure and Content

Pre-trained language models are language models that are pre-trained on large-scaled corpora in a self-supervised fashion. Traditional self-supervised pre-training tasks mostly involve recovering a corrupted input sentence, or auto-regressive language modeling. After these PLMs are pre-trained, they can be fine-tuned on downstream tasks. Conventionally, these fine-tuning protocols include adding a linear layer on top of the PLMs and training the whole model on the downstream tasks, or formulating the downstream tasks as a sentence completion task and fine-tuning the downstream tasks in a seq2seq way. Fine-tuning PLMs on downstream tasks often yield exceptional performance gain, which is why PLMs have become so popular.

In the first part of the tutorial (estimated 40
mins), we will summarize some findings that partially explain why PLMs lead to exceptional downstream performance. Some of these results have helped researchers to design better pre-training and fine-tuning methods. In the second part (estimated 2 hrs 20 mins), we will introduce recent progress in how to pre-train and fine-tune PLMs; the new techniques covered in this part have been shown to bring significant efficiency in terms of hardware resource, training data, and model parameters while achieving superb performance.

Table 1: Works in the past three years (from 2020 to 2022) related to our tutorial, to list just a few.
2.1 Part I: Why Do PLMs Work

We will introduce several results that partially explain the effectiveness of PLMs from two aspects: empirical and theoretical.

2.1.1 Empirical Explanations

Many researchers have conducted empirical experiments to show what PLMs have learned during pre-training that aids downstream performance. They mostly construct a special pre-training dataset to examine the transferability of the PLM and draw connect the transferability of the PLM with the characteristic of the pre-training dataset. Block (A) in Table 1 lists the relevant works in recent years.

2.1.2 Theoretical Explanations

Some researchers aim to understand the effectiveness of PLMs by rigorous mathematics, as shown in block (B) in Table 1. Their results range from using statistical models to what PLMs are learning during pre-training, or bounding the generalization errors of the downstream tasks.

2.2 Part II: How Do PLM Work

In this part, we will introduce some new techniques in pre-training and fine-tuning PLMs.

2.2.1 Pre-training

Improving Existing Pre-training Methods

Language model pre-training is a resource-hungry task when PLMs were first proposed, requiring a large amount of data, high-end hardware equipment, and lengthy pre-training time. To mitigate the above issues, some research aims to mitigate the above issues, as listed in block (C) in Table 1. Some of these works provide answers about the sufficient amount of data and time to pre-train a PLM that is good enough for downstream tasks, and others provide implementation optimization solutions to cut down the high-end requirement on hardware resources.

New Pre-training Methods

Aside from improving existing pre-training methods, there have also been new pre-training methods designed for specific downstream tasks. One of the important topics we aim to cover is applying contrastive learning on language model pre-training. Contrastive learning has been widely applied to pre-training models in computer vision, and we will introduce how contrastive learning has improved PLMs recently. Relevant works are listed in block (D) in Table 1.

2.2.2 Fine-tuning

In this part, we will go through several important fine-tuning protocols that have emerged recently. We categorize them based on the scenario in which the fine-tuning method is used.

Parameter-Efficient Fine-tuning

PLMs are enormous, often having millions or even billions of numbers of parameters. In the traditional fine-tuning method, fine-tuning each distinct downstream task produces a fine-tuned model that is are bulky as the original PLM. To reduce the number of parameters for fine-tuning PLMs on downstream tasks, there has been a surge of research on parameter-efficient fine-tuning in NLP, as listed in block (E) in Table 1.

Data-Efficient Fine-tuning

A large amount of labeled data is not always available for all downstream tasks, and it is thus important to find a way to apply the PLMs on downstream tasks with limited labeled data. These endeavors are included in block (F) in Table 1. We will discuss how to apply PLMs under different levels of labeled data scarcity.

In case we have a large amount of unlabeled data, semi-supervised learning fine-tuning protocols provide effective ways to utilize those unlabeled data and can boost the downstream performance. If those few labeled data are the only thing available, then we must harness the knowledge that the PLM possesses to aid the performance of few-shot learning. When we have no labeled data, zero-shot learning is still possible in certain cases, if you use the PLM correctly. We will discuss how to make a PLM able to perform well in the zero-shot setting.

Cross-Task Transfer

When we have a target task of interest, it is canonical to fine-tune the PLM on the target task. While transferring from PLMs leads to exceptional performance gain, sometimes we want more. This can be achieved by transferring from the PLMs and additional guidance from other auxiliary tasks in the form of intermediate task fine-tuning or multitask learning. Relevant works are listed in block (G) in Table 1. We will discuss how can cross-task transfer improve the downstream performance together with the power of PLMs.
3 Diversity

PLMs have shown promising results on different domains and have boosted the performance of low-resource languages on many tasks. The why part covered in this tutorial has the potential to help individuals of different groups to pre-train their own PLMs more efficiently. The how part covered in this tutorial specifically focuses on how to apply PLMs under different real-world scenarios with data scarcity and restricted model parameters, which will enable individuals of different groups to apply PLMs on the domains of interest in a more realistic setting. We see this tutorial to benefit diverse groups in the community.

The tutorial instructors are also diverse: Chuang is a PhD student in the USA, and Lee and Chiang are researchers in Taiwan. Also, Chuang and Chiang are currently Ph.D. students familiar with precise implementations, while Lee is a senior researcher with ten years of experience in human language processing research. This diversity in members enables our team to provide a thorough and detailed yet comprehensive and unified view on PLMs.

4 Prerequisites for Attendees

We expect the attendees to have basic machine learning concepts such as gradient descent and model optimization. The attendees will need to have basic knowledge in linear algebra and calculus to understand some contents in block (B) in Table 1. The attendees should also have minimal knowledge about PLMs and transformer models.

5 Reading List

We encourage attendees to read the following emblematic papers on PLMs and transformer model architectures:

- Transformer model: Vaswani et al. (2017)
- PLMs: Radford et al.; Devlin et al. (2019); Raffel et al. (2019)

6 Biographies of Presenters

Cheng-Han Chiang\footnote{\url{https://d223302.github.io/}} is a PhD student in National Taiwan University. His research focuses on natural language processing and self-supervised learning, and he has published several papers analyzing PLMs. He has experiences in giving lectures on machine learning topics: he gave a lecture on BERT in AI Summer School 2020\footnote{https://ai.ntu.edu.tw/?p=3534}, and his two lectures on graph neural network (in Mandarin) has received over 68k views on Youtube\footnote{https://www.youtube.com/watch?v=eybCCTKnwzA&ab_channel=Hung-yiLee}. He has also served as reviewers in EMNLP 2021, ICLR 2022, NeurIPS 2022, EMNLP 2022, and AAAI 2023.

Yung-Sung Chuang\footnote{https://people.csail.mit.edu/yungsung/} is a PhD student in Electrical Engineering and Computer Science at MIT CSAIL, where he works with Dr. James Glass. His research focuses on learning representations for natural language which helps downstream tasks such as natural language understanding, natural language generation, question answering. He has published several paper in this direction in EMNLP, ACL, NeurIPS, and NAACL. He also has served as reviewers in NeurIPS 2021, ICLR 2022, ICML 2022, NeurIPS 2022, EMNLP 2022, and AAAI 2023.

Hung-yi Lee\footnote{https://speech.ee.ntu.edu.tw/~hylee/index.php} is an associate professor of the Department of Electrical Engineering of National Taiwan University, with a joint appointment at the Department of Computer Science & Information Engineering of the university. His research focuses on deep learning, speech processing, and natural language processing. He owns a YouTube channel teaching deep learning (in Mandarin) with more than 8M views and 100k subscribers. He gave tutorials at ICASSP 2018\footnote{The tutorial has the most participants among the 14 tutorials in ICASSP 2018.}, APSIPA 2018, ISCSLP 2018, INTERSPEECH 2019\footnote{The tutorial also has the most participants among the 8 tutorials in INTERSPEECH 2019.}, SIPS 2019, INTERSPEECH 2020, ICASSP 2021, ACL 2021. He is the co-organizer of the special session on “New Trends in self-supervised speech processing” at INTERSPEECH (2020), the workshop on "Self-Supervised Learning for Speech and Audio Processing" at NeurIPS (2020), the workshop on "Meta Learning and Its Applications to Natural Language Processing" at ACL (2021), and the workshop on "Self-Supervised Learning for Speech and Audio Processing" at AAAI (2022). He will give the tutorial, "Self-supervised Representation..."
Learning for Speech Processing” with other researchers at ICASSP 2022 and NAACL 2022. He is the lead guest editor of IEEE JSTSP Special Issue on Self-Supervised Learning for Speech and Audio Processing, member of the Speech and Language Technical Committee (SLTC) of IEEE Signal Processing Society (SPS), SPS Education Center Editorial Board member, and Associate Editor for the SPS Open Journal of Signal Processing.

7 Open Access

We will allow our slides and video recording of the tutorial published in the ACL Anthology. All the slides and videos used in the tutorial, along with the reading lists related with the tutorial, will be updated at this tutorial website.

References


Shuxiao Chen, Koby Crammer, Hangfeng He, Dan Roth, and Weijie J Su. 2022. Weighted training for cross-task learning.


Yu Meng, Chenyan Xiong, Payal Bajaj, Paul Bennett, Jiawei Han, Xia Song, et al. 2021. Coco-lm: Correcting and contrasting text sequences for language model pretraining. Advances in Neural Information Processing Systems, 34.

Yu Meng, Chenyan Xiong, Payal Bajaj, Paul N Bennett, Jiawei Han, Xia Song, et al. 2022. Pretraining text encoders with adversarial mixture of training signal generators. In International Conference on Learning Representations.


Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2022. Finetuned language models are zero-shot learners.


When Cantonese NLP Meets Pre-training: Progress and Challenges

Rong Xiang\textsuperscript{1} Hanzhuo Tan\textsuperscript{1} Jing Li\textsuperscript{1} Mingyu Wan\textsuperscript{2} Kam-Fai Wong\textsuperscript{3}

\textsuperscript{1}Department of Computing, Hong Kong Polytechnic University, HKSAR, China
\textsuperscript{2}School of Continuing Education, Hong Kong Baptist University, HKSAR, China
\textsuperscript{3}Department of SEEM, The Chinese University of Hong Kong, HKSAR, China

\{rong.xiang, hanzhuo.tan\}@connect.polyu.hk
jing-amelia.li@polyu.edu.hk claramywan629@hkbu.edu.hk kfwong@se.cuhk.edu.hk

Abstract

Cantonese is a language from the Chinese family with over 73 million speakers in the world (García and Fishman, 2011; Yu, 2013). It is mostly used in colloquial scenarios (e.g., daily conversation and social media) and exhibits different vocabulary, grammar, and pronunciation compared to standard Chinese (SCN)\textsuperscript{1}, which is mainly designed for formal writing (Wong and Lee, 2018).

Despite the substantial efforts in Chinese Natural Language Processing (NLP), most previous studies center around SCN, where limited work attempts to explore how to process Cantonese with the cutting-edge NLP techniques (Xiang et al., 2019; Lee et al., 2021). Modern NLP paradigms have been deeply revolutionized by large-scale pre-training models, e.g., BERT (Devlin et al., 2019) and GPT-3 (Brown et al., 2020), which have achieved SOTA performance on many NLP tasks via fine-tuning.

Although the general NLP field is thriving, Cantonese NLP has drawn limited attention so far, as demonstrated by the recent publications in the ACL Anthology — only 47 papers are related to “Cantonese”, compared to 7,018 papers for English, 2,355 for common Chinese, and 323 for Mandarin.

This tutorial will present a roadmap lining Cantonese NLP up to the SOTA practice based on pre-training. We will start with the previous progress made by linguistics and NLP researchers, followed by the major challenges caused by the language specificity, and end with the promising future directions to allow Cantonese and other low resource languages to benefit from the advanced NLP techniques. The details will be covered in PART II-V.

\begin{itemize}
\item **PART I: Cantonese NLP Overview (in 30 min).**
\item At the beginning, we will briefly introduce Cantonese language and its related research in NLP.
\item **PART II: Progress in Language Specificity, Resources, and Methodologies (in 40 min).** Cantonese (or Yue) is the second most popular dialect among all Chinese variants (Matthews and Yip, 2011). For-
\end{itemize}

\textsuperscript{1}Standard Chinese is known as Standard Northern Mandarin, which is emerged as the lingua franca among the speakers of various Mandarin and other varieties of Chinese (Hokkien, Cantonese, and beyond).
Cantonese Pre-training Linguistic Pre-training Challenges Prospect

Overview: Cantonese Pre-training
Motivation

Challenges
Prospect1
Overview:

Cantonese Pre-training
Multilingual

Summary
Future

PART III: Pre-training in SOTA NLP (in 40 min). The cutting-edge NLP takes advantages of the promising results achieved by the pre-training of language representations. A typical pre-trained and fine-tune scheme refers to pre-train a large model on massive unlabelled corpora by self-supervised objectives, and fine-tune the model on downstream tasks with task-specific loss. Such self-supervised objectives, e.g. Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) (Devlin et al., 2019) enable the model to gain generalized language representations without human supervision. During fine-tuning stage, the pre-trained representations can be further used to learn a specific Natural Language Understanding (NLU) task with small-scale annotations via incremental training.

Transformer (Devlin et al., 2019; Liu et al., 2019; Brown et al., 2020) is the most widely employed pre-trained architecture in NLP. The transformer encoder consumes the input text and project it into high-dimension vectors, which are feed into the transformer decoder to generate the output sequences. Inspired by the transformer architecture, researchers explore the transformer encoder for NLU tasks and transformer decoder for Natural Language Generation (NLG) tasks.

Since the transformer-based pre-training (Devlin et al., 2019; Liu et al., 2019) was introduced to the world, championing the leaderboards of many NLP benchmarks, the “pre-training and fine-tuning” paradigm has profoundly revolutionized the way we research NLP for most of the majority languages, such as English and Chinese. Nevertheless, the success of language pre-training is built upon the availability of rich language resources and large-scale textual corpora, hindering Cantonese and other low-resource languages from gaining the benefit of pre-training. The following presents the challenging low-resource issue in Cantonese.

PART IV: Challenges from Colloquialism and Multilingualism

Colloquialism
Multilingualism

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training

Overview: Cantonese Pre-training
Multilingualism (40 min). In Cantonese NLP, Colloquialism and Multilingualism jointly present the low-resource challenges. Cantonese is by nature colloquial, where using casual language would sparsify the context and require more data for contextual model training, making the low-resource issue serious. Like other low-resource language processing (Li et al., 2018), it is possible to gather large-scale data from social media. However, models might also compromise their performance using just noisy social media data. It not only requires effective data cleaning and augmentation, but also non-trivial model capabilities of capturing salient information in context with diverse quality.

Cantonese exhibits a code-switching convention with multiple languages. Substantially richer context is hence required for NLP models to gain the multilingual understanding capabilities, despite the limitation to make it happen in the low-resource scenarios. Although it is possible to transfer the knowledge gained in a similar language with rich resources (henceforth cross-lingual learning) (Friedrich and Gateva, 2017; Khalil et al., 2019; Zhang et al., 2019), Cantonese, as a vibrant language, absorbs the knowledge from numerous languages beyond SCN and English.

PART V: Future NLP Directions for Cantonese and other Low-Resource Languages (30 min). Data scarcity and limited methodology exploration are top issues for Cantonese in benefiting deep semantics and general NLP tasks. To mitigate low-resource problems, data augmentation is an alternative to scale up the Cantonese dataset for NLP model training. For example, we might employ heuristic rules (Ratner et al., 2017; Lison et al., 2020), machine learning (Şahin and Steedman, 2018) and information retrieval (Riedel et al., 2010; Hedderich et al., 2021) to automatically boost the data scales. In the augmentation process, we might need to learn how to distinguish SCN from Cantonese. Though both are encoded in the Chinese language system, the former dominates the Chinese resources while the latter is a minority (Wu and Dredze, 2020; Cui et al., 2021).

Cross-lingual learning might provide another be a promising alternative for the pre-training in low resource, which borrows knowledge from other languages (Wisniewski et al., 2014; Zhang et al., 2019; Khalil et al., 2019). We may take advantage of SOTA pre-trained transformers to capture the general and specific language features for transfer learning (Devlin et al., 2019; Clark et al., 2020). In addition, based on Cantonese’s phonological history, future work may consider injecting phonetic knowledge into language learning or developing multi-modal understanding across text and speech.

2 Type of the Tutorial

This is an introductory tutorial of Cantonese NLP, where we draw NLP community’s attention to look at the research of Cantonese — a language with over 73 million speakers in the world (García and Fishman, 2011; Yu, 2013) while only has 47 papers in ACL Anthology related to it. The tutorial will present a roadmap going through the essential issues regarding language specificity, data scarcity, research progress, and major challenges for Cantonese NLP to be benefited from the cutting-edge NLP paradigms based on language pre-training.

3 Target Audience

Our tutorial is designed for the attendees of premier computational linguistics conferences, who preferably have interests and working experience in the processing of Asian languages and low-resource languages. The audiences would better have the following prerequisites.

- **Language Representation Learning.** Familiar with the basic concepts and main ideas of language pre-training, e.g., word embeddings (Mikolov et al., 2013), BERT (Devlin et al., 2019), and how the learned representations are employed to train various NLP tasks.

- **Linguistics.** Have the basic knowledge of the fundamental linguistic concepts (Jurafsky, 2000), e.g., semantics, syntax, lexicography, morphology, phonetics, etc.

- **Machine Learning.** Understand the traditional machine learning paradigm using handcrafted features (Svensén and Bishop, 2007) and the trendy deep learning-based methods (Goodfellow et al., 2016) allowing automatic feature learning in neural architectures.

4 Tutorial Outline (3 hours)

- **PART I: Cantonese NLP overview (30 min).**
  - Background of Cantonese.
  - Brief review of Cantonese NLP.
  - Brief introduction of language pre-training.
  - Problem definition and motivation.
PART II: Progress in language specificity, resources, and methodologies (40 min).
- Brief history of Cantonese.
- Linguistic characteristics of Cantonese.
- Summary of Cantonese NLP resources.
- Summary of Cantonese NLP methodologies.

PART III: Pre-training in SOTA NLP (40 min).
- Language pre-training methods.
- Pre-training in low resource.
- Chinese pre-training.

PART IV: Challenges from colloquialism and multilingualism (40 min).
- How colloquialism challenges pre-training.
- How multilingualism challenges pre-training.

PART V: Future NLP directions for Cantonese and other low-resource languages (30 min).
- Summary of the tutorial.
- Future work for Cantonese NLP and beyond.

5 Reading list

For trainees interested in reading important studies before the tutorial, we recommend the following: Ouyang (1993); Snow (2004); Sachs and Li (2007); Vaswani et al. (2017); Devlin et al. (2019); Liu et al. (2019); Brown et al. (2020); Sun et al. (2019); Liu et al. (2019); Nguyen et al. (2020).

6 Tutorial Presenters

Our tutorial will contain 5 parts and here we introduce the presenter for each of them.

- Kam-Fai Wong (PART I). Kam-Fai Wong a full professor in the Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong (CUHK). His research interest focuses on Chinese natural language processing and database. He is the fellow of ACL and has published over 260 technical papers in different international journals, conferences, and books. Also, He was the founding Editor-In-Chief of ACM Transactions on Asian Language Processing (TALIP) and the president of Asian Federation of Natural Language Processing (AFNLP).

- Mingyu Wan (PART II). Mingyu Wan is a postdoctoral fellow at the Department of Chinese and Bilingual Studies of Hong Kong Polytechnic University. Her research interest includes Financial NLP, CSR Modelling, Misinformation Detection, Sentiment/Emotion Analysis, Machine Learning, Language Resource Construction etc. She has 6 journal publications and more than 10 international conference proceedings in the NLP venue. She organizes the first Computing Social Responsibility Workshop cohosted at LREC 2022 conference.

- Hanzhuo Tan (PART III). Hanzhuo Tan is a Ph.D. student at the Department of Computing of Hong Kong Polytechnic University. His research interest includes self-supervised pre-training, NLP for social media, etc. He has 2 journal paper published in IEEE Transactions. He did six-month internship at Baidu PaddleNLP group on pre-training social transformer.

- Rong Xiang (PART IV). Rong Xiang is a postdoctoral fellow at the Department of Computing, Hong Kong Polytechnic University (PolyU). His research interests are acquisition and the application of human intelligence into machine learning networks. He has done substantial work in sentiment analysis, social media analysis and lexical semantics. He has published over 20 research papers in premier NLP venues. He co-organized CogALex 2020 and PACLIC 33.

- Jing Li (PART V). Jing Li is an assistant professor at the Department of Computing, Hong Kong Polytechnic University (PolyU). Before joining PolyU, she was a senior researcher in Tencent AI Lab. Her research interests are topic modeling, language representation learning, and NLP for colloquial and social media languages. She has published over 30 research papers in the top NLP venues and was invited to serve as the action editor for ACL rolling review (ARR) and the area chair for ACL 2021.

7 Other Information

Inclusion of Others’ Work. This tutorial will survey the progress of Cantonese NLP and language pretraining, which substantially contain others’ work.

Divergency considerations. Audiences who cannot speak Cantonese or Chinese will also be able to understand our tutorial. It will be conducted in English, where Cantonese cases will be presented with their English translations. Background knowledge will be provided to lower prerequisites (only those in Section 3 are needed). In the tutorial, we will discuss how the findings from Cantonese NLP can be generalized to other low-resource languages to benefit audiences in diverse streams.

Estimation of audience size. 100-200.
References


1 Tutorial Description

As natural language technology becomes ever-present in everyday life, people will expect artificial agents to understand language use as humans do. Nevertheless, most advanced neural AI systems fail at some types of interactions that are trivial for humans (e.g., ask a smart system “What am I pointing at?”). One critical aspect of human language understanding is situated reasoning, where inferences make reference to the local context, perceptual surroundings, and contextual groundings from the interaction. In this cutting-edge tutorial, we bring to the NLP/CL community a synthesis of multimodal grounding and meaning representation techniques with formal and computational models of embodied reasoning. We will discuss existing approaches to multimodal language grounding and meaning representations, discuss the kind of information each method captures and their relative suitability to situated reasoning tasks, and demonstrate how to construct agents that conduct situated reasoning by embodying a simulated environment. In doing so, these agents also represent their human interlocutor(s) within the simulation, and are represented through their virtual embodiment in the real world, enabling true bidirectional communication with a computer using multiple modalities.

“Grounding” in much of the NLP literature involves linking linguistic expressions to information expressed in another modality, often images or video (Yatskar et al., 2016; Li et al., 2019). Examples include linking semantic roles to entities in an image, or joint linguistic-visual attention between a caption and an image or video. Efforts have also focused on creating common meaning representation formalisms for linguistic data that are known to be relatively expressive, easy to annotate, and extensible to accommodate linguistic diversity, scale, and support inference, e.g., Copestake et al. (2005); Banarescu et al. (2013); Cooper and Ginzburg (2015); Pustejovsky et al. (2019); Lai et al. (2021).

Robust human-computer interactions and human-robot interactions will require representations with all these features, that encode the different modalities in use in such an interaction and ground them to the shared environment, enabling bidirectional, symmetric communication, and shared reference. Central to such situated meaning is the recognition and interpretation of gesture in the common ground (Holler and Wilkin, 2009; Alahverdzhieva et al., 2018).

Certain problems in human-to-human communication cannot be solved without situated reasoning, meaning they cannot be adequately addressed with ungrounded meaning representation or cross-modal linking of instances alone. Examples include grounding an object and then reasoning with it (“Pick up this box. Put it there.”), referring to a previously-established concept or instance that was never explicitly introduced into the dialogue, underspecification of deixis, and in general, dynamic updating of context through perceptual, linguistic, action, or self-announcement. Without both a representation framework and mechanism for grounding references and inferences to the environment, such problems may well remain out of reach for NLP.

An appropriate representation should accommodate both the structure and content of different modalities, as well as facilitate alignment and binding across them. However, it must also distinguish between alignment across channels in a multimodal dialogue (language, gesture, gaze), and the situated grounding of an expression to the local environment, be it objects in a situated context, an image, or a formal registration in a database. Therefore, such a meaning representation should also have the basic facility for situated grounding; i.e., explicit mention of object and situational state in context.

To date there has been interest in creating meaning representations that capture multimodality, and in multimodal corpora that capture language use in a situated environment (e.g., Chen et al.
yet the two have been largely distinct. We will demonstrate how to bring these together into grounded meaning representations that capture language, gesture, object, and event semantics that can be used to not only represent situated meaning, but drive situated reasoning in embodied agents that occupy a three-dimensional environment.

There have been recent *ACL tutorials on meaning representations (Lopez and Gilroy, 2018; Koller et al., 2019), on common-sense reasoning (Sap et al., 2020), and on common ground and multimodality (Alahverdzhieva and Lascarides, 2011). To our knowledge this is the first time these three areas have been brought together with situated, grounded reasoning for an NLP/CL audience.

This tutorial will cover the most pressing problems in situated reasoning: namely, those requiring both multimodal grounding of expressions, as well as contextual reasoning with this information. Three example areas we will cover are:

**“Make Me Another”** Grounding an underspecified item or concept to previous elements of a dialogue requires an understanding of both what is salient in context, and of what elements of that item or concept are relevant to the situation inhabited by the interlocutors (Schlangen and Skantze, 2011). For example, if someone is cooking a stack of pancakes for someone else, and the diner says “make me another,” a human would likely infer a reference to a single pancake, not the whole stack. The computational mechanisms for representing the elements of the environment and making this inference are richly involved. Addressing this problem and similar ones is an important part of building agents that respond to queries and requests in ways that are situationally appropriate.

**Underspecification of Deixis** The referent of a deixis may be ambiguous, though it naturally grounds to an object if one is available (Alahverdzhieva and Lascarides, 2011). Adding demonstratives like “this” or “there” naturally selects for objects vs. locations, and we will present models to capture these joint gesture-language semantics (Alahverdzhieva and Lascarides, 2010). Even coupling gesture and language may be insufficient for reasoning. “Pick up this one and put it there,” plus deixis, singles out a liftable object in the embedding space, and a possible ambiguity. If *there* refers to a location, then the command can be fully grounded in space. But, if the referent of the deixis is an object, additional reasoning must be conducted vis-à-vis what part of the object accommodates both the action *put* and the denotatum of *this one*. The many possible interpretations lead to rich reasoning strategies in situated space.

**Dynamic Updating of Context Through Announcement** When a participant in a dialogue sees, says, does, or realizes something new, the external and/or internal world changes for the participants, along with the capabilities for reasoning over the situation. For example, someone can verbally or gesturally announce an intent or provide information; perceptually demonstrate that something is present or absent; visibly act on a request or command; and personally realize something based on the current context. Each of these requires situated grounding and reasoning within those worlds.

### 1.1 Outline

This tutorial comprises 4 45-minute parts. 1) We will first present existing approaches to multimodal grounding, in the form of cross-modal linking (Yatskar et al., 2016; Yang et al., 2016; Sadhu et al., 2021) or linguistic-visual attention (Antol et al., 2015; Shih et al., 2016; Zhu et al., 2018; Sood et al., 2020) along with datasets that exist for this purpose (e.g., Kontogiorgos et al., 2018; Chen et al., 2019; Shridhar et al., 2020), and 2) common approaches to structured meaning representation (Copestake et al., 2005; Banarescu et al., 2013; Cooper and Ginzburg, 2015). 3) We will describe the formulation of common ground as a data structure of the information associated with a state in a dialogue or discourse (Clark et al., 1983; Stalnaker, 2002), and how it can be used to ground elements like gestures and situations to meaning representations (Lascarides and Stone, 2009; Alahverdzhieva et al., 2018). Each section will focus the material with regard how the discussed frameworks treat the grounding and reasoning questions from Sec. 1.

4) Finally we will present some of own work, including the grounded modeling language VoxML (Pustejovsky and Krishnaswamy, 2016) and a demonstration of building agents capable of situated reasoning in VoxWorld (Krishnaswamy and Pustejovsky, 2016; Pustejovsky and Krishnaswamy, 2021), a platform built on VoxML for developing embodied agent behaviors. We will provide a starter scene with an agent who can act upon the world, and discuss the computational and modeling considerations that go into developing
distinct types of agents, such as virtual collaborative assistants (Krishnaswamy et al., 2017), mobile robots (Krajovic et al., 2020; Tellex et al., 2020), and self-guided exploratory agents (Tan et al., 2019; Pustejovsky and Krishnaswamy, 2022), comparing our own framework to others’.

**Technical requirements** We have no special hardware requirements for this tutorial except for a projector or display screen.

**Distribution of materials** We plan to make all tutorial materials fully available to the community.

2 Target and Expected Audience

This tutorial will be of interest to both researchers in meaning representation, and in multimodal NLP and grounding, particularly those interested in both theoretical and data-driven approaches to language grounding and those interested in treating automated reasoning as more than just a pure machine learning problem. The diverse approaches to linguistic grounding of situated meaning have also provoked significant interest from the robotics community. Given the increased interest in interactive agents and grounding for robotics at the *ACL community, as witnessed by the recent creation of Language Grounding to Vision, Robotics, and Beyond tracks at most *ACL venues, this tutorial, that synthesizes various approaches to situated conversation and interaction will be a timely way to bring these two communities closer. We expect this tutorial will draw an audience of roughly 30-45.

2.1 Requisite Background

This tutorial will be self-contained. However, to get the most out of this tutorial, attendees will want to be familiar with both theoretical and machine-learning approaches to semantics. Familiarity with common meaning representation frameworks, such as abstract meaning representation (Banarescu et al., 2013) or minimal recursion semantics (Copestake et al., 2005), is desirable, as is familiarity with multimodal language and vision techniques, such as VQA or image captioning (Antol et al., 2015; Shih et al., 2016). Participants will be invited to “code along” for the last part of the tutorial if they so desire, for which knowledge of C# and the Unity game engine will be advantageous but not prerequisite.

3 Breadth and Reading List

This tutorial draws on a wealth of both theory and applied research in multimodal semantics, includ-
4 Instructors

Nikhil Krishnaswamy is Assistant Professor of Computer Science at Colorado State University and director of the Situated Grounding and Natural Language Lab (www.signallab.ai). He received his Ph.D. from Brandeis University in 2017. His primary research is in situated grounding and natural language semantics, using computational, formal, and simulation methods to study how language works and how humans use it. He is the co-creator of VoxML. He has taught courses on machine learning and NLP, previously taught at EACL 2017 (with J. Pustejovsky), and he will be co-teaching (also with J. Pustejovsky) at ESSLLI 2022 on multimodal semantics of affordances and actions. He has routinely received positive feedback as an instructor, including “always willing to engage in in-depth discussions regarding class material,” “his understanding of the subject matter is phenomenal,” “my favorite course this semester,” and “he clearly spends a lot of time making his lectures engaging.” He has served on the PC for ACL, EACL, NAACL, EMNLP, AAAI, AACL, etc. Email: nkrishna@colostate.edu, Website: https://www.nikhilkrishnaswamy.com.

James Pustejovsky is the TJX Feldberg Chair in Computer Science at Brandeis University, where he is also Chair of the Linguistics Program, Chair of the Computational Linguistics M.S. Program, and Director of the Lab for Linguistics and Computation. He received his B.S. from MIT and his Ph.D. from UMass Amherst. He has worked on computational and lexical semantics for 25 years and is chief developer of Generative Lexicon Theory; the TARSQI platform for temporal reasoning in language; TimeML and ISO-TimeML, a recently adopted ISO standard for temporal information in language; the recently adopted standard ISO-Space, a specification for spatial information in language; and the co-creator of the VoxML modeling framework for linguistic expressions and interactions as multimodal simulations VoxML (co-created with N. Krishnaswamy), enables real-time communication between humans and computers or robots for joint tasks, utilizing speech, gesture, gaze, and action. He is currently working with robotics researchers in HRI to allow the VoxML platform to act as both a dialogue management system as well as a simulation environment that reveals real-time epistemic state and perceptual input to a computational agent. Email: jamesp@brandeis.edu, Website: https://www.pusto.com.

5 Diversity

Situated reasoning and grounding inherently cross language boundaries. Language grounding in English can be compared to language grounding a low-resourced language by way of a situated model. From a research perspective these are important questions to answer, to explore how different languages represent the same environment or situation. Therefore situated reasoning is an important potential way to broaden the linguistic diversity of NLP, and we hope the meaning representation component of this tutorial may inspire broadening meaning representations to more languages yet.

The instructors are junior and senior faculty, respectively, established in the NLP community. We actively recruit women and underrepresented minorities to our respective research groups, and plan to promote this tutorial to an international and diverse audience. We are experienced instructors in a hybrid format, and we will accommodate and promote remote attendance to broaden participation.

6 Ethics Statement

Computational agents that reason situationally necessarily require sight and hearing, and come with concomitant ethical issues regarding computer vision and speech recognition. In the course of this tutorial, we will discuss many of the considerations surrounding user privacy and storing user data (or, in the case of our own research, explicitly not doing that (Wang et al., 2017)). We will also discuss adapting speech recognition models to user diversity as part of the multimodal grounding section (Krishnaswamy and Alalyani, 2021).

Real-time, situated reasoning requires smaller, lightweight models. While we use large models where necessary, our use of meaning representations to guide search within multimodal grounding tasks provides a way to accomplish this task with less computational overhead and resource use.
References


The Battlefront of Combating Misinformation and Coping with Media Bias

Yi R. Fung\(^1\), Kung-Hsiang Huang\(^1\), Preslav Nakov\(^2\), Heng Ji\(^1\)
\(^1\)University of Illinois Urbana-Champaign
\(^2\)Mohamed bin Zayed University of Artificial Intelligence
\{yifung2, khhuang3, hengji\}@illinois.edu
preslav.nakov@mbzuai.ac.ae

Abstract

Misinformation is a pressing issue in modern society. It arouses a mixture of anger, distrust, confusion, and anxiety that cause damage on our daily life judgments and public policy decisions. While recent studies have explored various fake news detection and media bias detection techniques in attempts to tackle the problem, there remain many ongoing challenges yet to be addressed, as can be witnessed from the plethora of untrue and harmful content present during the COVID-19 pandemic and the international crises of late. In this tutorial, we provide researchers and practitioners with a systematic overview of the frontier in fighting misinformation. Specifically, we dive into the important research questions of how to (i) develop a robust fake news detection system, which not only fact-check information pieces provable by background knowledge but also reason about the consistency and the reliability of subtle details for emerging events; (ii) uncover the bias and agenda of news sources to better characterize misinformation; as well as (iii) correct false information and mitigate news bias, while allowing diverse opinions to be expressed. Moreover, we discuss the remaining challenges, future research directions, and exciting opportunities to help make this world a better place, with safer and more harmonic information sharing.

1 Introduction

The growth of online platforms has greatly facilitated the way people communicate with each other and stay informed about trending events. However, it has also spawned unprecedented levels of inaccurate or misleading information, as traditional journalism gate-keeping fails to keep up with the pace of media dissemination. These undesirable phenomena have caused societies to be torn over irrational beliefs, money lost from impulsive stock market moves, and deaths occurred that could have been avoided during the COVID-19 pandemic, due to the infodemic that came forth with it, etc. (Allcott and Gentzkow, 2017; Rapoza; Solomon et al., 2020). Even people who do not believe the misinformation may still be plagued by the pollution of unhealthy content surrounding them, an unpleasant situation known as information disorder (Wardle et al., 2018). Thus, it is of pertinent interest for our community to better understand, and to develop effective mechanisms for remedying misinformation and biased reporting.

The emerging nature of news events, which also span diverse domains (e.g., economy, military, health, sports, etc.) and reporting style (e.g., long text vs. short text, realistic image vs. artistic image, etc.), makes misinformation detection and characterization challenging. Combating fake news and biased reports involve an interdisciplinary research area of reasoning on the semantics, style, cross-media contextualization, background knowledge, and propagation patterns, among others (Saquete et al., 2020; Pennycook and Rand, 2021; Collins et al., 2021). Moreover, the recent trends towards a more comprehensive understanding of the source stance, reporting intent, target audience, and propaganda technique behind a problematic piece of news (Zhang and Ghorbani, 2020) require greater socio-cultural norm and common sense awareness.

In this half-day tutorial, we aim to present a systematic overview of technological advancement in tackling interconnected tasks related to misinformation, media bias, malicious intent monitoring, and corrective actions. First, we will review prevailing paradigms and data resources for misinformation detection and characterization. Moreover, we will discuss the latest approaches to automatically explain why a news piece is inaccurate or misleading, and perform rectification of biased reporting. The participants will learn about trends and emerging challenges, representative deep neural network models, ready-to-use training resources, as well as how state-of-the-art language (and multimedia)
techniques can help build applications for the social good.

2 Outline of Tutorial Content

2.1 Background and Motivation [20min]

We begin motivating the tutorial topic with a selection of real-world examples of fake news and their harmful impacts to society, followed by a pedagogical exercise of how humans tend to approach the problem of fake news detection, characterization, and correction. We will point out conceptual distinctions amongst various types of fake news, including serious fabrication in news journalism about misattributed or nonexisting events, oversensationalized clickbaits, hoaxes which are false with the intention to be picked up by traditional news websites and satire which mimic genuine news but contain irony and absurdity (Rubin et al., 2015). For example, in general, news articles more likely involve serious fabrications, while social media posts involve more humour such as satire and hoaxes. We will also describe the cognitive, social and affective factors that lead people to form or endorse misinformed views (e.g., intuitive thinking, illusory truths, source cues, emotions, etc.), and the psychological barriers to knowledge revision after misinformation has been corrected, including correction not integrated, selective retrieval, and continued influence theories (Ecker et al., 2022).

2.2 Fake News Detection [60min]

Bearing these properties in mind, we introduce:

- **stylistic** approaches that focus on lexical features, readability, and syntactic clues (Pérez-Rosas et al., 2018; Rashkin et al., 2017; Choshen et al., 2019)

- **fact-checking** approaches that compare check-worthy content with background knowledge, such as external knowledge bases (FreeBase, WikiBase, etc) and previously fact-checked claims (Baly et al., 2018; Shaar et al., 2020; Hu et al., 2021; Liu et al., 2021; Guo et al., 2022)

- **semantic-consistency** approaches that extract features related to single-document discourse-level coherence (Karimi and Tang, 2019) and cross-document event-centric coherence (Wu et al., 2022) in text. Extending to cross-media domain, the common strategy is to check text–image consistency (Tan et al., 2020; Huang et al., 2022; Aneja et al., 2021) and text–video consistency (Wang et al., 2022).

- **propagation patterns** that capture confounding factors from the dynamics of how a news topic spreads and the social network interactions (Lu and Li, 2020; Shu et al., 2020; Cheng et al., 2021).

We will discuss the merits and the limitations of these different lines of fake news detection approaches. For example, fact-checking approaches may not fare well for early rumours or breaking news not yet groundable to an established background knowledge (Zhou et al., 2019; Guo et al., 2022), in which case, the credibility of the news source can offer complementary assistance (Cheng et al., 2021). Stylistic approaches may be simple but yet effective for detecting low-quality human-written fake news, but not so good for machine-generated misinformation, which is stylistically consistent regardless of the underlying motives (Schuster et al., 2020). We then cover recent approaches (Lee et al., 2021b; Fung et al., 2021) that leverage a combination of these elements for greater representation power and robustness. Importantly, we also cover works that explore the diachronic bias of fake news detection and portability across data in different time and language settings (Murayama et al., 2021; Gereme et al., 2021).

Special Note on Neural Fake News Generation & Detection:

Advancements in natural language generation spawn the rise of news generation models which represent a double-edged sword (Zellers et al., 2019). On one hand, malicious actors may irresponsibly take advantage of the technology to influence opinions and gain revenue. But, on the other hand, it can also be used as a source of machine-synthesized training data for detector models to overcome data scarcity since real-world fake news tends to be eventually removed by platforms, as well as a tool for threat modeling to develop proactive defenses against potential threats. We review how popular detectors perform on fake news created from large-scale language and vision generator model (Zellers et al., 2019; Güera and Delp, 2018; Agarwal et al., 2019). We also review progress in
generating fake news that better aligns with the key
topic and facts (Mosallanezhad et al., 2021; Shu
et al., 2021; Fung et al., 2021), and work towards
applying topic/fact-constrained fake news gener-
tion to construct silver-standard data annotations
for finer-grained fake news detection (Fung et al.,
2021).

2.3 Fake News Characterization [30min]
To better understand and fight fake news, we next
address some fundamental questions of character-
izing fake news based on underlying source bias,
reporting agenda, propaganda techniques, and tar-
get audience (Buchanan, 2020). First, we intro-
duce modeling approaches for detecting political
and socio-cultural biases in news articles (Kulka-
mini et al., 2018; Fan et al., 2019; Baly et al., 2020;
Forbes et al., 2020). Next, we introduce the recent
EMU benchmark that require models to answer
open-ended questions capturing the intent and the
implications of a media edit (Da et al., 2021). We
cover methodologies for identifying the specific
propaganda techniques used, e.g., smears, glitter-
ing generalities, association transfer, etc. (Dimi-
trov et al., 2021). We also discuss the latest explo-
rations in predicting the intended target of harmful
media content, e.g., the person, the organization,
the community, or the society level (Pramanick
et al., 2021).

2.4 Corrective Actions for Misinformation
and Biased News Reporting [30min]
After misinformation has been detected and cate-
gorized based on its various characteristics, there
is naturally follow-up interest in corrective expla-
nations on why a piece of information is fake or
misleading, and how to report less biased and more
comprehensive news in general. Hence, we cover
frameworks for explaining why a given piece of
news is actually fake news through the leverage
of reader comments, as well as appropriate strate-
gies for placing the corrective explanations based
on user studies (Shu et al., 2019; Brashier et al.,
2021). We also cover research on mitigating me-
dia bias, such as through neutral article generation
(Lee et al., 2021a).

Industry Initiatives: We further point out recent
actions by tech companies with media-hosting plat-
forms for fighting fake news. With urges from the
government, they experiment with removing eco-
nomic incentives for traffickers of misinformation,
promoting media literacy, suspending improper
posts and accounts, and adding colored labels, with
corrections constructed from a community-based
point system similar to Wikipedia, directly beneath
misinformation posted by public figures.

2.5 Concluding Remarks & Future Directions
[30min]
Finally, we summarize the major remaining chal-
gen in this space, including the detection of sub-
tle inconsistencies, enforcing schema or logical
constraints in the detection, identifying semanti-
cally consistent but misattributed cross-media pair-
ings, and greater precision in fine-grained explana-
tions for the detected misinformation.

3 Specification of the Tutorial
The proposed tutorial is a cutting-edge tutorial
that introduces new frontiers in research on bat-
tling misinformation and news bias. The pre-
SENTED topic has not been covered by previous
ACL/NAACL/AACL tutorials in the past four
years. While there has been an EMNLP’20 tutorial
on “Fact-Checking, Fake News, Propaganda, and
Media Bias: Truth Seeking in the Post-Truth Era”
(Nakov and Da San Martino, 2020) and a COL-
ING’20 tutorial on “Detection and Resolution of
Rumors and Misinformation with NLP” (Derczyn-
ski and Zubiaga, 2020), fake news is a continuously
evolving and extremely important societal problem.
In our tutorial, we place particular emphasis on the
latest lines of development, including an empha-
sis on multimedia contextualization, sociocultural
awareness in characterization, and corrective ac-
tions. We estimate at least 75% of the work we
reference has not been covered in the two previous
aforementioned tutorials. We further estimate that
at least 75% of the research covered in this tutorial
is by researchers other than the instructors.

Audience and Prerequisites Based on the level of
interest in this topic, we expect around 100 partici-
pants. While no specific background knowledge is
assumed of the audience, it would be best for the
attendees to know basic deep learning, pre-trained
word embeddings (e.g., Word2Vec) and language
models (e.g., BERT).

Reading List We recommend the literature cited
in this paper, particularly: the rising threats of neu-
ral fake news (Zellers et al., 2019; Chawla, 2019),
knowledge-driven misinformation detection (Hu et al., 2021; Fung et al., 2021; Guo et al., 2022), intent characterization (Buchanan, 2020; Da et al., 2021), and study of fake news impact from a psychological point of view (Ecker et al., 2022).

**Desired Venue** The most desired venue for this tutorial would be AACL-IJCNLP’2022. The majority of our tutorial speakers have educational experience in Asia. At the same time, we also represent a global diversity in our research work.

**Open Access** We agree to allow the publication of the tutorial materials and presentation in the ACL Anthology. All the materials will be openly available at the UIUC Blender Lab website.

### 4 Tutorial Instructors

Below, we give the biographies of the speakers.

**Yi R. Fung** is a second-year Ph.D. student at the Computer Science Department of UIUC, with research interests in knowledge reasoning, misinformation detection, and computation for the social good. Her recent works include the INFO-SURGEON fake news detection framework, and multiview news summarization. Yi is a recipient of the NAAACL’21 Best Demo Paper, the UIUC Lauslen and Andrew fellowship, and the National Association of Asian American Professionals Future Leaders award. She has also been previously selected for invited talk (1 hour presentation) at the Harvard Medical School Bioinformatics Seminar. Additional information is available at [https://yrf1.github.io](https://yrf1.github.io).

**Kung-Hsiang Huang** is a first-year Ph.D. student at the Computer Science Department of UIUC. His research focuses on fact-checking and fake news detection. Prior to joining UIUC, he obtained his B.Eng. in Computer Science from the Hong Kong University of Science and Technology, and his M.S. in Computer Science is from USC. He is also a co-founder of an AI startup, Rosetta.ai. Additional information is available at [https://khuangaf.github.io](https://khuangaf.github.io).

**Preslav Nakov** is a Principal Scientist at the Qatar Computing Research Institute (QCRI), HBKU, who received his PhD degree from the University of California at Berkeley (supported by a Fulbright grant). Dr. Nakov is President of ACL SIGLEX, Secretary of ACL SIGSLAV, a member of the EACL advisory board, as well as a member of the editorial board of Computational Linguistics, TAACL, CS&L, IEEE TAC, NLE, AI Communications, and Frontiers in AI. His research on fake news was featured by over 100 news outlets, including Forbes, Boston Globe, Aljazeera, MIT Technology Review, Science Daily, Popular Science, The Register, WIRED, and Engadget, among others. He has driven relevant tutorials such as:


Additional information is available at [https://en.wikipedia.org/wiki/Preslav_Nakov](https://en.wikipedia.org/wiki/Preslav_Nakov).

**Heng Ji** is a Professor at the Computer Science Department of the University of Illinois Urbana-Champaign, and an Amazon Scholar. Her research interests focus on NLP, especially on Multimedia Multilingual Information Extraction, Knowledge Base Population and Knowledge-driven Generation. She was selected as “Young Scientist” and a member of the Global Future Council on the Future of Computing by the World Economic Forum. The awards she received include “AI’s 10 to Watch” Award, NSF CAREER award, Google Research Award, IBM Watson Faculty Award, Bosch Research Award, Amazon AWS Award, ACL2020 Best Demo Paper Award, and NAACL2021 Best Demo Paper Award. She has given a large number of keynotes and 20 tutorials on Information Extraction, Natural Language Understanding, and Knowledge Base Construction in many conferences including but not limited to ACL, EMNLP, NAACL, NeurIPS, AAAI, SIGIR, WWW, IJCAI, COLING and KDD. A selected handful of her recent tutorials include:

- EMNLP’21: Knowledge-Enriched Natural Language Generation.
- ACL’21: Event-Centric Natural Language Processing.

Additional information is available at [https://blender.cs.illinois.edu/hengji.html](https://blender.cs.illinois.edu/hengji.html).
Ethical Considerations

Technological innovations often face the dual usage dilemma, in which the same advance may offer potential benefits and harms. For the news probing methodologies introduced in this tutorial, the distinction between beneficial use and harmful use depends mainly on the data and intention. Proper use of the technology requires that input corpora be legally and ethically obtained, with the target goal to fight misinformation and mal-intents. Besides, training and assessment data may be biased in ways that limit the system performance on less well-represented populations and in new domains—causing performance discrepancy for different ethnic, gender, and other sub-populations. Thus, questions concerning generalizability and fairness need to be carefully considered when applying news analysis techniques to specific settings. A general approach to ensure proper application of dual-use technology should incorporate ethical considerations as the first-order principles in every step of the system design, maintain transparency and interpretability of the data, algorithms, and models, and explore counter-measures to protect vulnerable groups.

References


Xueqing Wu, Kung-Hsiao Huang, Yi Fung, and Heng Ji. 2022. Cross-document misinformation detection based on event graph reasoning. In NAACL 2022, Seattle, WA, USA.


A Tour of Explicit Multilingual Semantics: Word Sense Disambiguation, Semantic Role Labeling and Semantic Parsing

Roberto Navigli, Edoardo Barba, Simone Conia
Sapienza NLP Group
Sapienza University of Rome
first.lastname@uniroma1.it

Rexhina Blloshmi∗
Amazon Alexa AI
Berlin, Germany
blloshmi@amazon.de

Abstract

The recent advent of modern pretrained language models has sparked a revolution in Natural Language Processing (NLP), especially in multilingual and cross-lingual applications. Today, such language models have become the de facto standard for providing rich input representations to neural systems, achieving unprecedented results in an increasing range of benchmarks. However, questions that often arise are: firstly, whether current language models are, indeed, able to capture explicit, symbolic meaning; secondly, if they are, to what extent; thirdly, and perhaps more importantly, whether current approaches are capable of scaling across languages.

In this cutting-edge tutorial, we will review recent efforts that have aimed at shedding light on meaning in NLP, with a focus on three key open problems in lexical and sentence-level semantics: Word Sense Disambiguation, Semantic Role Labeling, and Semantic Parsing. After a brief introduction, we will spotlight how state-of-the-art models tackle these tasks in multiple languages, showing where they excel and where they fail. We hope that this tutorial will broaden the audience interested in multilingual semantics and inspire researchers to further advance the field.

1 Tutorial Description and Relevance

Over the past few years, the field of Natural Language Processing (NLP) has witnessed tremendous growth, mainly thanks to the increasingly wide availability of modern pretrained language models, such as ELMo (Peters et al., 2018), BERT (Devlin et al., 2019), and BART (Lewis et al., 2020), which have enabled unprecedented results in a broad range of tasks, from Neural Machine Translation to Question Answering, Information Retrieval and Text Summarization, inter alia. However, important questions that naturally arise when looking at the recent impressive gains in the field are whether such powerful language models learn to encode semantic knowledge and, if they are, to what extent. More importantly, the escalating interest in multilingual NLP demands approaches that are able to identify and transfer semantics across a multitude of languages, especially those for which there is a scarce amount of data available.

In this tutorial, we will review recent studies in lexical and sentence semantics, paying special attention to state-of-the-art approaches and how they tackle multilinguality in three fundamental tasks for Natural Language Understanding (NLU): Word Sense Disambiguation (WSD), Semantic Role Labeling (SRL) and Semantic Parsing (SP). In addition to an introduction to multilingual NLU, for each task we will provide, i) a gentle introduction, ii) an overview of the inventories and resources most commonly adopted, iii) an outline of current approaches with a particular focus on multilinguality and cross-linguality in order to understand their strengths and shortcomings, and also pointing to promising directions for future work. Although there have been previous tutorials on Semantics in NLP, especially on SP (Lopez and Gilroy, 2018; Gardner et al., 2018; Koller et al., 2019), our tutorial will, instead, focus on the challenges of multilinguality and cross-linguality and how recent approaches based on pretrained language models tackle them.

Despite the increasing performance of huge language models in NLU tasks, recent studies have demonstrated that the integration of explicit semantics into deep learning techniques is beneficial not only in terms of performances (Levine et al., 2020), but also interpretability (Wiedemann et al., 2019) and cross-lingual transfer (Blloshmi et al., 2020).

2 Tutorial Structure and Contents

The tutorial will be structured in a bottom-up fashion: participants will be introduced to multilingual
semantics, first at the lexical level with WSD, and then at the sentence level with SRL and SP, highlighting the most effective approaches to date, but also their weaknesses and future directions to address these.

2.1 Word Sense Disambiguation (WSD)

The tutorial will start with WSD as the lowest level of semantic abstraction. Its objective is to assign the most appropriate sense to a word in context from a finite set of possible choices (Navigli, 2009), which usually come from predefined sense inventories. Although, at a first glance, WSD may seem a simple task to a human, it has proven to be extremely challenging for machines. Indeed, depending on the sense inventory of choice, different linguistic phenomena may make the task difficult to tackle with standard classification techniques. Nonetheless, being able to link raw text to knowledge bases is fundamental in NLP (McCoy et al., 2019; Bender and Koller, 2020), bringing benefits in several fields, such as Machine Translation (Liu et al., 2018; Pu et al., 2018; Campolungo et al., 2022), Information Extraction (Delli Bovi et al., 2015), and Information Retrieval (Blloshmi et al., 2021b). We will start with an introduction to the task, presenting its most common formulation along with the challenges it poses. Then, we will describe state-of-the-art systems, highlighting their core contributions. Finally, we will conclude by presenting open challenges in multilingual WSD.

Resources for WSD. We will first present the standard resources currently in use for WSD, starting with WordNet (Miller et al., 1990), i.e., the most widely used sense inventory, and Open Multilingual Wordnet (Bond, 2011) and BabelNet (Navigli and Ponzetto, 2012; Navigli et al., 2021), two multilingual extensions of WordNet.

Current approaches in WSD. After the initial success of purely data-driven neural models in WSD (Yuan et al., 2016), subsequent approaches started to leverage information coming from knowledge bases in addition to standard training datasets (Huang et al., 2019; Bevilacqua and Navigli, 2020). We will put a special focus on state-of-the-art systems that rely on relational knowledge (Bevilacqua and Navigli, 2020) and sense definitions as additional knowledge (Blevins and Zettlemoyer, 2020; Barba et al., 2021a,b). We will explain how these approaches are data-efficient and why they are important, especially for low-resource languages.

2.2 Semantic Role Labeling (SRL)

While WSD is concerned with lexical-level meaning, SRL (Gildea and Jurafsky, 2000) investigates sentence-level semantics and is usually described informally as the task of automatically answering the question “Who did What to Whom, Where, When, and How?” (Márquez et al., 2008). More precisely, its objective is to extract the predicate-argument structure of a sentence and, therefore, it is considered by some as a form of shallow Semantic Parsing. Over the years, SRL has been proven to be beneficial in several tasks, such as Question Answering (Shen and Lapata, 2007), Machine Translation (Marcheggiani et al., 2018), Video Understanding (Sadhu et al., 2021), and data augmentation (Ross et al., 2022). Following a general introduction to SRL, the tutorial will highlight some key details about the most popular predicate-argument structure inventories for SRL, the salient characteristics of current state-of-the-art systems, and why everything becomes more complex when trying to tackle multilingual and cross-lingual SRL.

Inventories for SRL. The tutorial will overview the main challenges that current predicate-argument structure inventories pose for multilingual and cross-lingual SRL, with particular focus on PropBank-style inventories (Palmer et al., 2005; Xue, 2008; Jindal et al., 2022), FrameNet (Baker et al., 1998) and VerbAtlas (Di Fabio et al., 2019).

Current approaches in SRL. Given its close ties with syntax, over the years one of the main distinctions between proposed approaches is whether they have chosen to rely on syntactic features (He et al., 2019; Marcheggiani and Titov, 2020; Conia and Navigli, 2020), or not (Marcheggiani et al., 2017; Cai et al., 2018). The tutorial will briefly cover the advantages and disadvantages of relying on syntax in multilingual SRL, but also highlight annotation projection techniques for cross-lingual SRL (Akbik et al., 2015; Daza and Frank, 2020), and how recent trends in multi-task learning (Conia et al., 2021) and generation (Blloshmi et al., 2021a; Paolini et al., 2021; Conia et al., 2022) are going beyond traditional approaches, hinting at new directions in SRL.

2.3 Semantic Parsing (SP)

Finally, the tutorial will bring participants to a higher level of semantic abstraction: SP, indeed, may be seen as “the task of mapping natural lan-
guage sentences into complete formal meaning representations which a computer can execute” (Kate and Wong, 2010). Here we focus on formalisms that aim at encoding text in an abstract form that captures aspects of meaning – as opposed to executable formalisms for SP – that can be reusable in various scenarios, thus being domain independent. Indeed, SP formalisms have been successfully integrated into numerous downstream applications, such as Machine Translation (Song et al., 2019), Text Summarization (Hardy and Vlachos, 2018), Human-Robot Interaction (Bonial et al., 2020) and Question Answering (Kapanipathi et al., 2021). Nevertheless, research in SP has mainly focused on English, with only a handful of attempts in other languages.

**Formalisms for SP.** Over the years, various different formalisms have been proposed to encode semantic structures. We will first overview the most popular formalisms, such as Elementary Dependency Structures (Oepen and Lønning, 2006, EDS), Prague Tectogrammatical Graphs (Hajič et al., 2012, PTG), Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013, UCCA), Universal Decompositional Semantics (White et al., 2016, UDS), with a main focus on Abstract Meaning Representation (Banarescu et al., 2013) and BabelNet Meaning Representation, its fully-semantic extension (Martínez Lorenzo et al., 2022).

**Current approaches in SP.** SP is receiving ever growing attention that has led to numerous approaches of different flavors. Indeed, the advantages and disadvantages of parser types are variable across different formalisms. We will focus on two categories of approaches: graph-based ones (Zhang et al., 2019; Cai and Lam, 2020), that consist of transducing natural utterances into graphs, and sequence-to-sequence ones, that produce linearized graph structures for a given input text (Ge et al., 2019; Bevilacqua et al., 2021a). Due to the recent development of encoder-decoder pretrained architectures, sequence-to-sequence approaches to SP are emerging as the best-performing methods, not only in English (Bevilacqua et al., 2021a), but also in other languages (Procopio et al., 2021b).

### 3 Type, Prerequisites and Audience

This is a cutting-edge tutorial. State-of-the-art approaches for three key areas of multilingual lexical and sentence semantics will be presented, and some of them will be discussed in detail. We expect **80-120 attendees** from different fields as the barriers to entry will be low:

- **Math prerequisites:** Linear algebra, e.g., matrix operations, linear/non-linear functions.
- **Machine Learning prerequisites:** General concepts of classification, e.g., token classification, sequence labeling, sequence-to-sequence.
- **NLP prerequisites:** High-level notions about pretrained language models.

### 4 Reading List

Recommended work to read before the tutorial:

- Bevilacqua et al. (2021b): a survey on recent trends in WSD;
- Blevins and Zettlemoyer (2020) and Barba et al. (2021a): two recent WSD systems that take advantage of sense definitions;
- Márquez et al. (2008) and Hajic et al. (2009): an introduction to SRL and the largest gold benchmark for multilingual SRL;
- He et al. (2019) and Conia et al. (2021): two recent approaches to multilingual SRL, a syntax-aware and a syntax-agnostic one;
- Koller et al. (2019) and Oepen et al. (2020): tutorial on recent work and shared task on SP;
- Banarescu et al. (2013) and Bevilacqua et al. (2021a): the introduction to the AMR formalism for SP and a state-of-the-art system for AMR parsing and generation;

### 5 Tutorial Outline (3h)

#### Part 0: Introduction (10 minutes)

Introduction, motivation, goals, how the tutorial is organized.

#### Part 1: WSD (40 minutes)

- Introduction to WSD, formulation, examples;
- Sense inventories for WSD: WordNet, Open Multilingual WordNet and BabelNet;
- Current approaches in multilingual WSD: purely data-driven vs. knowledge-enhanced supervision; going beyond sense inventories.
QA & Break (10 minutes)

Part 2: SRL (40 minutes)

• Introduction to SRL, formulation, examples;
• Predicate-argument structure inventories: the case of multilingual and cross-lingual SRL;
• Current approaches in multilingual and cross-lingual SRL: syntax-aware vs syntax-agnostic systems, annotation projection techniques, and novel directions.

QA & Break (10 minutes)

Part 3: SP (40 minutes)

• Introduction to SP, formulation, examples;
• Main formalisms for SP;
• Current approaches in cross-lingual SP: annotation projection, data augmentation via translation, generation.

QA & Break (10 minutes)

Part 4: Conclusion (20 minutes). Where to go from here, general considerations, a look to the future of explicit lexical and sentence semantics.

6 Pedagogical Material

Part 1 (WSD), Part 2 (SRL) and Part 3 (SP) will include brief hands-on sessions. These will be supported by interactive demos and Jupyter/iPython/Colab notebooks to invite participants to play with high-performance pretrained systems for WSD, SRL and SP. All material (slides, notebooks, pretrained models) will be freely available online to let discussions continue beyond the tutorial and for teaching purposes.

7 Presenters

Roberto Navigli is a Full Professor in the Department of Computer, Control and Management Engineering (DIAG) of Sapienza University of Rome, from which he also obtained his Ph.D. in Computer Science in 2007. At Sapienza he has taught courses for 4 Master’s programmes (CS, CS Engineering, AI & Robotics and Data Science), including NLP. He has been a keynote speaker at more than 30 conferences and workshops, including IJCNLP, IJCAI-ECAI (early career spotlight), AMLD, SwissText+KONVENS, CLNLP, RANLP, TALN, eLex.

In 2014, he co-presented a (pre-neural) tutorial on “Multilingual WSD and Entity Linking” at COLING. In 2016, he co-presented a tutorial on “Semantic Representations of Word Senses and Concepts” at ACL. He has worked and published with around 200 researchers from all over the world in more than 200 papers in the area of NLP with a particular focus on Natural Language Understanding and multilinguality, attracting 18,000+ citations.

Rexhina Blloshmi is a Machine Learning Scientist at Amazon Alexa AI in Berlin. Her PhD focused on Semantic Parsing. She contributed in this field with several publications in AI and NLP conferences (3 EMNLP, 1 IJCAI and 3 AAAI), mainly on English and Cross-Lingual Abstract Meaning Representation and Semantic Role Labeling, but also on novel formalisms such as BabelNet Meaning Representation.

Edoardo Barba is a third-year PhD Student in NLP at Sapienza University of Rome. His research is mostly focused on Word Sense Disambiguation. He contributed to several articles regarding both state-of-the-art and data efficient systems for WSD (Barba et al., 2021a) as well as Data Augmentation techniques for Multilingual WSD (Barba et al., 2020; Procopio et al., 2021a). Teaching Assistant in 2020 and 2021 for the NLP course at Sapienza (taught in English).

Simone Conia is a third-year PhD Student in NLP at Sapienza University of Rome. His research revolves around multilingual and cross-lingual semantics, with numerous papers on WSD and SRL published at *ACL and other top-tier conferences. Simone is recipient of an Outstanding Paper Award at NAACL-2021 for his work on cross-lingual SRL. Teaching Assistant in 2020 and 2021 for the NLP course at Sapienza (taught in English).

8 Ethics & Diversity Statement

We do not foresee any major ethical issue for the topics covered in this tutorial. We acknowledge that pretrained language models may show biases towards some stereotypes, cultures, ethnic and/or social groups: perpetrating such biases is not in our intentions. We will cover a variety of languages, including Arabic, Chinese, English, French, German, Italian, Spanish: we hope that our effort can promote new studies aimed at making lexical and sentence semantics increasingly more inclusive of lower-resource languages.
References


Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge,


Claudio Delli Bovi, Luca Teleseca, and Roberto Navigli. 2015. Large-scale information extraction from textual definitions through deep syntactic and semantic analysis. Transactions of the Association for Computational Linguistics, 3:529–543.


Hardy Hardy and Andreas Vlachos. 2018. Guided neural language generation for abstractive summariza-


Abelardo Carlos Martínez Lorenzo, Marco Maru, and Roberto Navigli. 2022. Fully-Semantic Parsing and
Generation: the BabelNet Meaning Representation.


Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger,


Author Index

Barba, Edoardo, 35
Blloshmi, Rexhina, 35

Chiang, Cheng-Han, 8
Chuang, Yung-Sung, 8
Conia, Simone, 35

Fung, Yi, 28
Gui, Tao, 1

Huang, Kung-Hsiang, 28
Ji, Heng, 28

Krishnaswamy, Nikhil, 22

Lee, Hung-yi, 8
Li, Jing, 16

Nakov, Preslav, 28
Nan, Guoshun, 1
Navigli, Roberto, 35

Pustejovsky, James, 22
Tan, Hanzhuo, 16

Wan, Mingyu, 16
Wong, Kam-Fai, 16

Xiang, Rong, 16
Zhang, Ningyu, 1