The work of linguists in Computational Linguistics

Irina Temnikova
GATE Institute, Bulgaria
irina.temnikova@gmail.com

1 Brief description of the tutorial content and its relevance to the computational linguistics community

This is an introductory tutorial, based on the presenter's experience as a linguist, translator, annotator, and researcher in the Computational Linguistics (CL) field. The tutorial is for linguists, beginning their careers in the Computational linguistics field, such as University students. It presents an overview of the tasks which are completed by linguists, assigned by more experienced researchers and companies, such as data collection, annotation, tools' evaluation, and post-editing of machine translation (including the use of language models). It provides definitions and explains some basic concepts, such as the main CL methods, the levels of text processing, Corpus Linguistics, Translation Technologies, and Natural Language Processing (NLP) applications. Some practical exercises will be given. The tutorial also provides practical tips for beginning linguists regarding where they can find related jobs in the field. The tutorial is compiled from previous University lectures of the presenter. This tutorial is of great value for linguists, especially those with no mathematical or programming skills, who begin their journey in the CL/NLP fields.

2 The type of tutorial

This tutorial is introductory.

3 Description of the target audience and any prerequisite background

3.1 Target Audience

The tutorial is designed to train linguists, beginning their careers in the Computational Linguistics/Natural Language Processing fields, as well as translators, planning to use Translation memories and Machine Translation.

Linguists with existing annotation, statistical and programming skills are also welcome.

More experienced researchers and annotators are welcome to join the discussion with their suggestions and personal experiences.

3.2 Prerequisite background

Linguistic knowledge or linguistic (University-level) education.

4 Outline of the tutorial structure and content

The tutorial is planned for 4 hours (four sessions of 45 minutes each). Below is an approximate outline (it is subject to change):

- Introductions and Basic Definitions
- Linguistic Research and Corpus Linguistics: Tasks and Tools
- Translation Technologies
- Main NLP Approaches, Levels of Text and Speech Pre-processing, and an Overview of Existing NLP Applications
- Linguists' Tasks in Computational Linguistics and Natural Language Processing (Research, Rules and Grammars, Data Collection and Generation, Annotation, Evaluation) – Methods and Tools
- Practical Exercises
- Names of Bulgarian (or international, depending on the Audience) CL Experts and Researchers; International Websites and Companies Hiring Linguists/Annotators; Translation Agencies

5 Explanation of how the tutorial incorporates previous research

The tutorial includes Irina Temnikova's experience as an Italian, French, English to Bulgarian translator and as a linguist with PhD in Computational linguistics. It also includes her practical experience as an annotator and in designing annotation and evaluation experiments with guidelines, and training human annotators and evaluators. While details about private clients' assignments are not disclosed, the presentations include bits from the successfully completed research project TRACES, emotion annotation in text, and running machine translation evaluation experiments for Bulgarian and Romanian within the BROD project.

6 Recommended reading, a bibliography, and if relevant, a list of programmes or applications

None - the tutorial is introductory.

7 One-paragraph summary of the instructor's research interests, expertise, and teaching experience

Dr. Irina Temnikova holds a B.A. degree in Italian linguistics, an M.A. and a PhD in Computational linguistics. She started her career as a translator and as a linguist with basic Perl and Python programming skills. During the first years of her career, she conducted pure linguistic and psychlinguistic research applied to the Natural Language Processing (NLP) field. During the last few years she self-taught herself and made a breakthrough progress in understanding, fine-tuning and training NLP models, as well in statistical calculations, while leading a research team and successfully completing the externally funded research project TRACES. During all her career, she worked as an annotator in freelance projects, designed and runned annotation and NLP systems evaluation experiments. Her research interests cover interdisciplinary and practical approaches to emotion detection, text readability and simplification, controlled languages and sublanguages, NLP for disaster computing, post-editing and evaluation of machine translation, extracting linguistic insights from interpreting transcriptions, corpus linguistics, and detecting deception and disinformation. Irina has extensive teaching experience of Italian language courses. Her University-level teaching experience includes single lectures on the topics of her research interests at undergraduate and masters level at the University of Wolverhampton, UK; a part of a course and a recently fully designed by her and completed whole course on the work of linguists in Computational linguistics at the undergraduate (B.A.) level in Sofia University, Bulgaria.

8 Estimate of the expected audience size, including information on previous iterations of the tutorial if applicable

5-25 linguists and translators, this is the first time this tutorial is given.

9 Description of any special requirements for technical equipment

A projector, Wi-Fi connection, and electricity; it is desirable for participants to bring their laptops

10 Statement of what materials will be available to the attendees

Slides, maybe some printed materials with exercises

11 Statement on whether the tutorial materials can be made publicly available on the CLIB 2026 website

Yes, in a shortened form.