Vítkovice, 15th Sep 2025

A teaching semester at UniCatt, Milan

Zdeněk Žabokrtský

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

How did it happen?

From: "**Passarotti Marco Carlo**" <marco.passarotti@unicatt.it>

To: "Zdenek Zabokrtsky" <zabokrtsky@ufal.mff.cuni.cz>

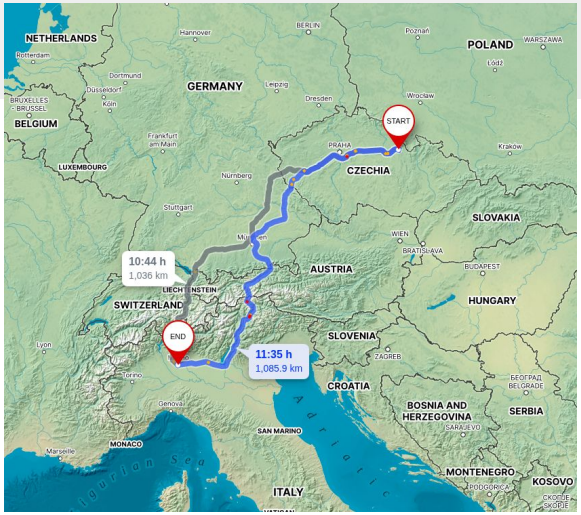
Sent: Tuesday, 16 April, 2024 15:11:12

Subject: Visiting Professor in Milan

Dear Zdeněk,

Milan

- 1000 km southwest from Jahodov
- in the Lombardy region in Italy
- 3M residents in the metropolitan city



About Università Cattolica del Sacro Cuore

- “Cattolica” for short
- an Italian private research university
- founded in 1921, but located partially in historical buildings, esp. Cloisters of Sant'Ambrogio
- 12 faculties (humanities)
- 40k students
- #331 in Best Global Universities, #442 in QS ranking

CIRCSE Department

- Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione (lit. Interdisciplinary Research Center for the Computerization of Expressive Signs)
- Flavio Massimiliano Cecchini, Chiara Colombo, Claudia Corbetta, Theodorus Fransen, Federica Iurescia, Roberta Grazia Leotta, Eleonora Litta Modignani Picozzi, Francesco Mambrini, Giovanni Moretti, Marco Passarotti, Matteo Pellegrini, Savina Raynaud, Paolo Ruffolo, Rachele Sprugnoli, Marinella Testori
- Recently an ERC-Consolidator Grant: LiLa - Linking Latin (2018-2023)

Master's degree study program "Linguistic Computing"

First year

- Computational Linguistics 1 (8 ECTS)
- Fundamentals of Computer Science for Linguistics (8 CFU)
- Project and Change Management (8 ECTS)
- English Language for Linguistic Computing (12 ECTS)
- Computational Linguistics 2 (8 ECTS)
- Formal Methods and Models for Computational Linguistics (8 ECTS)
- One elective course among: (8 ECTS)
  - Semantic Fundamentals for Natural Language Processing
  - Terminological Resources
- Theology Seminar

Second year

- Language I - Language and Phonology / Communicative Strategies (12 ECTS)*
- Curriculum in Humanities:
  - Natural Language Processing for Social Media (8 ECTS)
  - Digital Tools for the Humanities (8 ECTS)
- Curriculum in Business:
  - Artificial Intelligence and Natural Language Processing for Decision Making (8 ECTS)
  - Customer Behavior and Semantic Web (8 ECTS)
- One elective course among the following: (8 ECTS)
  - Linguistic Linked Open Data and Semantic Web
  - Theory and Method of Linguistic Annotation
  - Computational Philology and Digital Editing
- Internship, or one elective course among the following: (6 ECTS)
  - Issues of Theoretical Linguistics in Natural Language Processing and Computational Linguistics
  - Advanced Python for Linguistics
  - Data Structures and Database Systems Concepts
- Master Final dissertation (18 ECTS)

The course - what was given

- “Formal Methods and Models for Computational Linguistics “, 8 ECTS
- an obligatory course
- 2nd semester of the first year
- 6 teaching hours / week, 12 teaching weeks
- online teaching in certain weeks (either synchr. or asynchronous)
- 6 exam dates

Topics to be covered in the course

- partially inherited list of topics (statistics and probability, word&sentence representations, stringology, evaluation) ...
- ... but a strong demand for adding modern methods, esp. ANNs
- relative freedom as for the course's content
- so I decided to design the course from scratch

Students

- around 35 students, quite international
- bc in humanities, no univ-level maths whatsoever
- some intro to bash in the 1st semester, Python only simultaneously
- most of the students highly motivated

My implementation of the course

- Six modules (usually two weeks per module)
  - Module A: Course Intro; **Intro to Probability and Information Theory**
  - Module B: **Graph Theory; Formal Languages**
  - Module C: **About Words** - Formalizing the Lexical Space
  - Module D: **Machine Learning**
  - Module E: **Deep Learning**
  - Module F: **Evaluation in NLP**
- 400+ slides (85% new)
- 150 exercises

Final written test

- Cattolica's standard grading scale: 0-30 pts
- exam questions sampled from a previously published pool of 120 questions
- an example:

Question 1: What is a uniform distribution? [1 point]

Question 2: Explain the notion of conditional probability. Give an example. [2 points]

Question 3: What is smoothing (in the context of language modeling)? Why do we need it? [2 points]

Question 4: How many edges are there in a complete (undirected) graph with N nodes? [1 point]

Question 5: If a Hidden Markov Model is applied on POS tagging, transition probabilities correspond to what? [1 point]

Question 6: Can you describe a regular expression over the alphabet of input symbols {a,b} which accepts strings containing exactly three a's? [2 points]

Question 7: Give examples of derivation, compounding and borrowing, two for each, in Italian *(or any other language, but please translate it into English then)*. [2 points]

Question 8: Explain what is hapax legomena ratio and what it is used for. [2 points]

Question 9: Give three examples of pairs of words whose Levenshtein distance is 3. [1 point]

Question 10: How can you compute cosine similarity of two vectors (formula needed)? [1 point]

Question 11: Select any two classification methods, describe them and compare them. [2 points]

Question 12: What is activation function (in the context of artificial neural networks). Give an example. [1 point]

Question 13: Why are word embeddings used in the modern NLP? What are typical properties of word embeddings? [2 points]

Question 14: What is k-fold cross-validation? [1 point]

Question 15: Give examples of baseline solutions for at least three distinct NLP tasks, one for each (you can choose any). [2 points]

The biggest challenges

- “Linguistic computing” without proper background in mathematics - not easy
- No reliable knowledge of programming at the moment
- No HWs allowed

Results

- Feedback from the students
  - an official questionnaire, to be filled before registration for an exam date
  - criteria such as explanation clarity, quality of teaching materials, clear examination format, motivation, ...
  - average values of most evaluated criteria between 7 and 9 (on a 0-9 scale), probably
  - a repeated complaint: more time is needed for such a course
- However, eventually very good scores at the final exam
  - >80% of students >80% points

Differences from teaching experience at MFF UK

- both positive and negative **consequences of the fact that Cattolica is private**
- technical **support personnel** everywhere
- a **Blackboard license** - very comfortable, in my opinion more efficient than wiki/Moodle/LangTech solutions
- **outsourced IT solutions**, supersafe and modern-looking, but
  - Not a single “SIS”, rather a collection of different online systems
  - OTP mobile phone confirmation needed even for email login
  - an electronic signature needed for filling exam grades
  - univ-level shared printers => very difficult to print anything
- crowded mensa, no kitchenettes allowed (perhaps safety)

Differences from teaching experience at MFF UK, cont.

- More comfort for students, slightly less comfort for teachers
  - **6 exam dates** for each course specified by the faculty, beyond teacher's control
  - **no homework** assignments possible
  - some dorm facilities available only for student inhabitants, not for professor inhabitants
  - **evaluation from students seems to be considered crucial**
- Interleaved in-room and **online lectures** (obligatory in the whole program)
  - **not really appreciated** by the students
- Milan center extremely expensive, many teachers as well as researchers travel from substantial distances (Trento, Berlin...)
- At MFF, we often complain about SIS, but it could be much worse :)

Lessons learned

- NLP intro without background in maths - a kind of an **extreme experience**
- Slide production: switched from LaTeX to **Google Slides**
- **ChatGPT** is a fantastic assistant for rapid slide development (just that...)
- A **Learning Management System** can be efficient for creating a course web
- **Google colab** - very useful for quick in-class exercises, both bash and Python
- ÚČNK's **KonText** - extremely useful for real-time tours across languages
- A few **stereotypes broken** (about Italy, about private universities...)

Some sights

Museo Nazionale
Scienza e Tecnologia
Leonardo da Vinci

Parco naturale della Valle del Ticino

Museo Civico di Storia Naturale di Milano

A trip to Switzerland

Thank you.