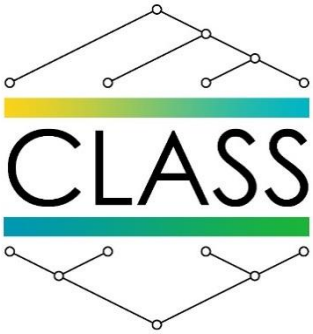
Co-funded by the Erasmus+ Programme of the European Union



Development of the interdisciplinary master program on Computational Linguistics at Central Asian universities

585845-EPP-1-2017-1-ES-EPPKA2-CBHE-JP

Analysis of international master programs by AMU with recommendations

Work Package 1 / Task 1.3 / Deliverable 1.2

March 2018

Content

Introduction

This report contains an analytical part of the outcomes of the Subtask 1.2 of the Work Package 1 (Preparation). The expected outcomes of this subtasks are defined as follows in the detailed description of the CLASS project:

„Analysis (of) international master programs content will be carried out by AMU, NUUz and ENU. The obtained outcome will be important for the developing the master program curriculum and qualification description on CL which is outcome 2.1 of WP2. Via this outcome the curriculum development instructions will be given by AMU and the main aspects will presented in order to integrate the master programs with good quality resources".

The document presents the analytical work done by the AMU team (Zygmunt Vetulani, Marek Kubis and Jacek Marciniak).

Scope of research

We analysed 9 examples of MA/MSc programs from Europe (7) and the USA (2). These programs have been elaborated/implemented in the following countries Australia, China, Czech Rep., France, Germany, Italy, Malta, Netherlands, Spain, Sweden, UK, USA. In 8 (out of 9) cases programs have been designed and implemented by the concerned universities. The only exception is the case of the "European Masters Program – Language and Communication Technologies" co-funded by the Erasmus+ Programme and involving a consortium of 9 universities from Australia, China, Czech Rep., France, Germany, Italy, Malta, Netherlands, Spain.

Sources

We explored the web sites addressed first of all to potential students.

https://lct-master.org/

http://www.brandeis.edu/departments/computer-science/comp-linguistics/

https://flov.gu.se/english/education/masters-second-cycle/mlt/programme-syllabus

http://www.uni-heidelberg.de/courses/prospective/academicprograms/computerling_ma_en.html

http://www.sas.rochester.edu/lin/graduate/MS.html

https://www.uni-stuttgart.de/en/study/study-programs/program/Computational-Linguistics-M.Sc./

http://www.sfs.uni-tuebingen.de/en/courses-of-study/courses-of-study-at-the-sfs/international-studies-in-computational-linguistics/international-ma-programme-iscl.html

http://www.uu.se/en/admissions/master/selma/program/?pKod=HSP2M

http://courses.wlv.ac.uk/course.asp?code=WL050P31UVD

Relevant aspects

We focussed on aspects that we consider crucial for the CLASS project.

These are:

- degree

- duration

- teaching language

- students' profile

- recruitment requirements

- curriculum structure (curses, projects, exams)

- mobility and flexibility issues

- tuition fees

Degree

Depending on the nature of the master program, respectively the MA or MSc degree is awarded as the final positive result of studies. As a rule, students are supposed to collect a proscribed amount of ECTS credit points (usually 120). A MA/MSc candidate is supposed to pass final exams and to present for acceptance the master thesis.

Duration

Standard duration is 2 years (4 semesters), however we have noticed non-standard solutions too (e.g. duration defined as 12 months divided in 3 semesters – in Wolverhampton, UK; 3 semesters at the University of Rochester, USA). Some universities offer two recruitment periods - autumn and spring (e.g. Uni. Stuttgart).

Teaching language

English is often used as teaching language, also in countries where it is not the local official language (e.g. Sweden). In some cases however the local language is used for teaching (e.g. German at the University of Heidelberg).

Students' profile

In most cases the concerned universities explicitly define the profile of candidates e.g. in terms of students' expectations concerning their professional career. In most cases, both computer science and linguistic background is presupposed. This prerequisite is obviously satisfied in cases where strong formal constraints are imposed (as e.g. previous acquisition of BA/BCs degree in computational linguistics - University of Tübingen (BA) or University of Stuttgart (BCs).

Recruitment requirements

The formal requirements, when applied, often impose possession of BA/BCs degree in Computational Linguistics or equivalent.

Certified proficiency in English is usually a formal inscription requirement.

Program principles

Within the 4-semester-based curriculum typically the first semester is dominated with the core, mandatory courses. The elective courses usually start in the second semester and typically dominate the third one. As a rule, preparation of the master thesis occupies the final semester, but sometimes starts earlier.

	University	Requirements	Project	Exams	Thesis/Title	
1	European Masters Program Consortium: - University of Saarlandes - University of Trento - University of Prague - University of Malta University of Lorraine - University of Groningen - University of the Basque Country - University of Melburne	BA, BSc or equivallent; 2 years;	obligatory project	120 ETCS	Master Thesis	double degree: study in two universities of the consortium (one year each); payable: 17.000/8.500 Euro; teaching language English;

	- Shanghai Jiao Tong University					
2	Brandeis University (USA)	2 years; BA/BSc;			Exit requirement: either an internship in computational linguistics or a master thesis	teaching language: English
3	University of Gotenburg (Sweeden)	2 years (one-year master program is proposed as well);	Elective project		YES	teaching language German; English required
4	University of Heidelberg (Germany)	BA, BSc		120 ETCS	MA Thesis	payable
5	University of Rochester (USA)	3 full-time semesters; BSc;	Cumulative (final) project		MSc	payable; teaching language: English
6	University of Stuttgart (Germany)	4 semesters; B.Sc. in Natural Language Processing;		120 ETCS		teaching language: English
7	University of Tübingen	2 years; BA in Computational Linguistics; teaching	Practical work as a student trainee	3 hours written exams (7 courses); obligtory master exam	obligatory MA thesis in Computational Linguistics	language: (generally) English
8	Uppsala University	4 semesters; Bachelor's degree equivalent to a Swedish Kandidatexamen, from an internationally recognised university (background in linguistics, language studies, and/or computer science); also 60 credits in language technology, computational linguistics, computer science, linguistics or a language subject; teaching language English (verified English	Project work	120 ETCS	MA	payable: application fee: SEK 900 tuition fee, total: SEK 240.000 total: 330.000 SEK (= 32 500Euro)

Table 2

Curriculum structure

The common characteristic observed for all master programs that we have analysed is the organisation of content into three parts.

- The part one consists of courses to be followed by all students in the predefined order.
- The part two is composed of elective courses to be followed by students in order to complete the required quota of courses (usually measured using the ECTS points system.
- The part three is composed of the obligatory items imposed by the concerned university in order to evaluate the master candidate's acquired skills and expertise; these are typically: master thesis, final exams and final project (all or some of them).

Typically the first semester is dominated by the part-one-activities. It is common to propose to students to listen some courses of the part two already in the second semester. The third semester is often dominated by the elective courses and activities (as projects and labs). The final semester (typically the fourth one) is dominated by subjects of the part three.

It is important to notice that the implementation of this scheme varies from one university to another.

Below, we present a number of examples for various universities.

Part one:

- methodologies, computational syntax and morphology, data structures, data organization and processing, logic, computability and complexity, formal languages and algorithms (European Master's Program)

- Programming for Linguistics, Fundamentals of Computational Linguistics, Programming in Java and C, Advanced Programming Techniques, Data Structures and the Fundamentals of Computing (required, but can be taken in the second year), Structure and Interpretation of Computer Programs (recommended), Mathematical Methods in Linguistics, Statistical Approaches to Natural Language Processing, Information Extraction, Phonological Theory (recommended), Syntactic Theory, Formal Semantics: Truth, Meaning, and Language, Architecture of Conversation: Discourse and Pragmatics (recommended) (Brandeis University)

- Introduction to Programming, Introduction to formal linguistics, Natural Language Processing, Statistical Methods, Dialogue Systems, Computational Semantics, Computational Syntax (University of Gothenburg, Sweden)

- "Methods in Computational Linguistics" assessed by a written exam (University of Stuttgart, Germany)

- Advanced Parsing Techniques, Computational Semantics, Computational Morphology, Lexicon and Grammar Formalisms, Survey of NLP Applications, Symbolic Methods in Computational Linguistics, Machine Learning for Natural Language, Developing NLP resources for lesser-resourced languages, Corpus Annotation: Linguistic Foundations and Computational Linguistic Analysis, Computational Analysis of Discourse, Cognitive Models of Language Processing, Models of Natural Language in Computational Linguistics, NLP Applications: Methods, Resources, and Evaluation, Computational Linguistic Analysis in Learning and Education, NLP workflows, pipelines and toolkits, Machine Translation, Information Retrieval, Intelligent Computer-Assisted Language Learning, Current Topics in Statistical Machine Translation, Computational Approaches to Text Simplification, Corpus Annotation of Information Structure, Computational Approaches to Language Variation and Stylometrics, Integrated Models of Processing, NLP supporting Noticing and Awareness in Second Language Acquisition, Language Technology for Educational Assessment, Advanced Distributional Semantics, Scientific Visualization of Language Data, Current Topics in Machine Learning for Natural Language (University of Tübingen, Germany)

- Programming for language technologists I, Mathematics for language technologists, Natural language processing, Advanced programming for language technologists, Machine learning in natural language processing, Elective courses, Language technology: research and development (Uppsala University, Sweden)

- Python Programming, Computational Linguistics, Research Methods and Professional Skills, Corpus Linguistics with R, Machine Learning for NLP (University of Wolverhampton)

Part two:

- machine translation, information retrieval, question answering, speech recognition and generation, models of human language processing and understanding, psycholinguistics, multimodality, language resources, computational semantics, formal semantics, inference in NLP, artificial intelligence, knowledge representation, automated reasoning, semantic web, intelligent and multi-modal interfaces, cognitive modelling, computational psychology, neural networks, machine learning (European Master's Program, EU)

- Modal, Temporal and Spatial Logic for Language, Machine Translation, Information Retrieval, Computational Semantics, 136 Automated Speech Recognition, Information Extraction, Language Annotation for Machine Learning, Topics in Natural Language Processing, Natural Language Processing Systems (non-exhausted list) (Brandeis University, USA)

- Machine Learning, Dialogue Systems 2, Language Technology Project, Language Technology Resources, Embodied and Situated Language Processing (University of Gothenburg, Sweden)

- at least one of {Introduction to Language Sound Systems, Introduction to Grammatical Systems, Introduction to Semantic Analysis}; at least two of {Topics in Phonetics and Phonology, Data Science for Linguistics, Syntactic Theory, Phrase Structure Grammar, Topics in Experimental Syntax, Formal Semantics, Pragmatics, Computational Semantics, Methods

in Linguistic Research, Syntax, Graduate Semantics, Topics in Phonetics and Phonology, Formal Pragmatics}; two of {Introduction to Computational Linguistics, Natural Language Processing, Statistical Speech and Language Processing}; at least two of {Data Mining, Artificial Intelligence, Logical Foundations of Artificial Intelligence, Machine Learning} (University of Rochester, USA)

- Each student chooses two „Concentrations" (12 credits each) on a core area of Computational Linguistics. The choice is from {Computational Syntax and Semantics, Laboratory Phonology and Speech Processing, Statistical Natural Language Processing}. (University of Stuttgart, Germany)

- Advanced seminars offered by the Department of Linguistics or external seminars from another MA program; Practical work in the semester break or during semester as a student trainee (University of Tübingen, Germany)

- Elective courses (no detailed information) (Uppsala University, Sweden)

- Machine Translation and Other NLP Applications, Translation Tools for Professional Translators (University of Wolverhampton)

Part three:

- project and master thesis (European Master's Program, EU)

- either an internship in computational linguistics or a master thesis (Brandeis University, USA)

- Master's Project (University of Gothenburg, Sweden)

- Master's Thesis (University of Heidelberg)

- no info provided (University of Rochester, USA)

- Research Module (Research Seminar + Research Colloquium), CL Team Laboratory, Master's Thesis (University of Stuttgart, Germany)

- Final oral exam and composition of a master's thesis (University of Tübingen, Germany)

- Master's thesis in language technology (Uppsala University, Sweden)

- 15,000-word dissertation (University of Wolverhampton)

Mobility and flexibility issues

Mobility and flexibility are two principles generally respected in all cases we have studied.

As a rule the generally applied (in Europe) Bologna model strongly promotes students' mobility. In most cases master studies in EU are open for students from other universities, in particular for international students.

In some (rare) cases recruitment to the master studies is however practically impossible for students coming from external universities because of very rigid principle of continuation with respect to previous BA studies supposed to be completed at the same university – as this is the case at the University of Rochester, USA.

In some cases, student mobility is considered as an integral element of the curriculum. This is the case of the Erasmus project "European Master's Program – Language and Communication Technologies" based on the principle of „double degree" according which every student studies one year each at two different European (or non-European) consortium partner universities.

Flexibility of curricula is implemented via the ETCS credit point system (or equivalent in non-EU universities) and a substantial offer of elective courses.

Tuition fees

In most cases studies are payable and fees vary substantially from one university to another. For example, at the universities belonging to the European Master's Program Consortium studies cost 17.000 Euro (8.500 for EU residents), whereas at the Uppsala University students are supposed to pay 32.000 Euro (application+tuition fees).

Content of international educational master program in Computer Linguistics (recommendations on the ground of the analysed curricula of international master programs)

This proposal takes into accounts specific requirements of the CLASS project (on the basis of the project proposal document and individual consultations at the Kick-off-meeting in Santiago de Compostela)

Objective: formation of well-trained professional staff apt to undertake a collective work in creation of language technologies (both software and language resources) within professional teams of language technology developers (supervised by NLP and HLTs experts);

preparation for professional career in the sector of language industries. Alumni will constitute the core of future highly competent human resource to launch language industry for the concerned languages.

Teaching language: English and/or local language /depending on needs and possibilities/.

Implementation for: Kazakhstan and Uzbekistan

Recruitment requirements: bachelor degree in applied computer science or equivalent. It is assumed that all candidates are familiar with computer science and linguistics at the basic level.

- computer science: practical experience with programming in at least one widely used modern programming language, familiarity with one of the main operational systems (Windows, Unix/Linux, Android), familiarity with data processing and storage

- linguistics: good knowledge of the grammatical system of one's mother tongue, good knowledge of traditional English grammar, fluency in standard English (in both speaking and writing); recommended familiarity in a second foreign language

Duration: 4 semesters (including preparation of master thesis)

Evaluation: 120 ECTS credit points collected for the courses and other activities strictly relevant to CL, final exam, final project, master thesis. Recommendation: the program may be partially realised at external universities using the ECTS credit point system, preferably within the ERASMUS+ framework (co-tutoring may be accepted in case of long duration external internship).

Main tasks for the first year

First semester /preparatory/:

- compensatory lectures for completion of linguistic background for non-linguists and enhancement of computer science competence of linguists (e.g. Introduction to Computational Linguistics, Computational Morphology, Formal Grammars)

- subject matters of general importance for language technologies (subjects mandatory for all students) (e.g. Introduction to programming for NLP (incl. Python, R, Prolog,...)

- subjects required by the higher education Authority of concerned countries as mandatory in order to obtain the Master degree (if necessary continued during further semesters) (should not exceed 15% of educational content) (obligatory, but not ECTS scored)

- foreign languages (English and rudiments of other languages representing different language families /French, German, Hungarian,.../)

Second semester:

- lectures of general importance for language technologies (subjects mandatory for all students) /continuation and new matters/ (e.g. Language Resources and Databases, Machine learning for NLP - laboratory of language processing tools and language resources (team

project) – practical initiation to the development of software and language resources (work in mini-teams)

- Master Seminar (collective introduction to master thesis; general methodological issues)

- specialistic advanced courses (based on already acquired knowledge in Computer Linguistics) (elective courses)

- foreign languages (continuation)

Main tasks for the second year

Third semester:

- lectures of general importance for language technologies (subjects mandatory for all students) /continuation and new matters/ (e.g. Computational Semantics, Dialogue Systems)

- master seminar (individual)

- elective courses (including those related to the master thesis) - laboratory of language processing tools and language resources – practical initiation to the development of software and language resources (work in mini-teams) (continuation)

- foreign languages (continuation)

Fourth semester:

- elective courses (including those related to the master thesis) - laboratory of language processing tools and language resources – practical initiation to the development of software and language resources (work in mini-teams) - project termination /beta testing, final documentation/

- foreign languages

- master seminar (individual)

- final redaction of the master thesis

- master exams

Curriculum scheme (draft)

Subject	Working hours					ECTS
	Lecture	Classes	Laboratory	Personal work[1]	Σ	
Semester 1						30
Introduction to Programming for NLP (Python, R, Prolog,...)	60	0	60	180	300	10
Introduction to Computationa Linguistics	30	30	0	120	180	6
Computational Morphology	30	30	0	120	180	6
Formal Grammars	30		30	120	180	6
Foreign Languages	0	30	0	30	60	2
Semester 2						26
Parsing	30	0	30	120	180	6
Language Resources and Databases	30	0	30	120	180	6
Machine Learning for NLP	30	0	30	120	180	6
Elective course 1	30	0	0	120	180	X
Team Project (laboratory of language processing tools and language resources)	0	0	75	75	150	4
Master Project Seminar (collective)	0	30	0	90	120	2
Foreign Languages	0	30	0	30	60	2
Semester 3						22
Computational Semantics	30		30	120	180	6
Dialogue Systems	30		30	120	180	6
Elective course 2	30	0	30	120	180	X
Elective course 3	0	0	30	120	180	X
Team Project (laboratory of language processing tools and language resources)	0	0	75	75		4
Master Project Seminar (individual)	0	15	0	15		2
Foreign languages	0	30	0	30	60	4
Semester 4						18
Elective course 4	30	0	30	120	180	X
Elective course 5	30	0	30	120	180	X
Elective course 6	30	0	30	120	180	X
Team project (laboratory of language processing tools and language resources)	0	0	75	75	150	4
Master's project seminar (individual)	0	15	0	15	30	2
Master's project and thesis	0	0	60	240	300	12
All elective courses (1 - 6)	180		180	720	1080	24
Total	**450**	**210**	**705**	**2535**	**3900**	**120**

Table 2. Computational Linguistic Master – curriculum draft

1 The figures presented in the column „personal work" should be interpreted as recommendations.

2 Elective courses, as well as other program units may be carried out at external universities, e.g. within Erasmus+ student mobility or summer internship (by courtesy of the home university) (recommended).

Acknowledgements
Thanks are due to Mr. Sanatbek Matlatipov (NUUz) and other consulted colleagues for their remarks.

Report done on 20th March 2018 by:

Zygmunt Vetulani (AMU CLASS coordinator)

Marek Kubis

Jacek Marciniak

Adam Mickiewicz University in Poznań, Faculty of Mathematics and Computer Science

Dept. of Computer Linguistics and Artificial Intelligence

Contact: AMU CLASS Coordinator, Prof. Dr. Zygmunt Vetulani, +48 6-1777296, vetulani@amu.edu.pl