THE PROGRAMS OF COMPUTATIONAL LINGUISTICS GRADUATE IN GERMAN AND US UNIVERSITIES

Irina N. Martynova
I.Ya. Yakovlev Chuvash State Pedagogical University, Cheboksary, Russia. E-mail: irinamartynova08@mail.ru
Universitaet Konstanz, Konstanz, Germany. E-mail: irinamartynova08@mail.ru

Lila A. Metelkova
I.Ya. Yakovlev Chuvash State Pedagogical University, Cheboksary, Russia. E-mail: lakhr76@gmail.com

Natalia G. Gordeeva
I.Ya. Yakovlev Chuvash State Pedagogical University, Cheboksary, Russia. E-mail: gord1307@list.ru

Larisa V. Nikitin ska
I.Ya. Yakovlev Chuvash State Pedagogical University, Cheboksary, Russia. E-mail: nikitin ska.larisa@yandex.ru

Margarita V. Emelianova
I.N. Ulyanov Chuvash State University, Cheboksary, Russia. E-mail: marg_emel@yahoo.com

Alena I. Trukova
I.N. Ulyanov Chuvash State University, Cheboksary, Russia. E-mail: alenatrukova@mail.ru

*corresponding author email: irinamartynova08@mail.ru

Abstract: The job of data innovations and computerized reasoning is always developing in the cutting edge world. Programmed language preparing with the assistance of PC programs is broadly utilized in different circles of human exercises. Preparing very qualified experts in this field is turning into a fundamental segment of the instructive procedure at present day colleges. The article goes for breaking down different practices of showing strategies for computational semantics at the alumni level. In the focal point of consideration are programs in this subject offered by chosen German and American colleges. The creators analyze diverse way to deal with developing prospectuses in computational semantics in Germany and the USA and the purposes behind picking specific courses and instructive systems. We utilize customary strategies for experimental investigation, for example, depiction and arrangement, content examination, correlation and unions. While still at its improvement organize showing PC semantics as a scholarly branch of knowledge lacks a very much grounded obvious structure and is affected, as it were, by instructive conventions of colleges that offer it. The aftereffects of the examination feature the fundamental focal points of the projects, these that can additionally be utilized to improve and systemize techniques to preparing in this cross-disciplinary territory. The investigation unmistakably shows the that at the present stage colleges are looking for the correct harmony between the semantics and the computational piece of the common language preparing schedules nearby with estimating the vital level of their hypothetical and reasonable segments. The accomplishments of the best German and American colleges in the field of preparing in computational etymology could be connected to the frameworks of advanced education in different nations.

Keywords: computational linguistics, graduate programs, teaching, German and US universities.

1. Introduction. Computer and internet technologies have already become an integral part of the modern society. Nowadays people can hardly imagine their lives without electronic devices and various software that literary embrace all the spheres of human activities. Sprawling of computer technologies goes alongside with the development of artificial intelligence that has already earned its keep in numerous applications used in everyday life almost worldwide. Automatic transactions machines, spelling correction and machine translation programs, speech recognition systems and grammar checking, online dictionaries, search engines – all these are present in our daily routines though quite often are practically invisible but equally powerful. All these tools combine resources of computer science and linguistics. This unity has already achieved great success and is ready for much more discoveries in the future. Studying language with the help of computers actually started with the emerging of the first computers. The idea had a solid basis of endless decoding tasks which the military and the scientists faced and solved during World War II. The middle of the XXth century was the time of unprecedented interest in machine translation. The general belief of that time was that it would be rather easy to create computer programs for this purpose. The initial euphoria vanished probably as quickly as it started. Scientists soon realized they had to go a long way before something worth attention appeared. Initial ideas have transformed greatly and expanded theoretically with the general development of hard- and software. It was further research conducted by linguists, mathematicians, engineers and IT specialists that brought about a new branch of science, computational linguistics. Computational linguistics (CL) is “an interdisciplinary field concerned with the processing of language by computers” (1). Combining knowledge from both linguistics and computer science CL is of a crucial importance in the XXI century as it is directly connected with the studies of artificial intelligence. CL and natural language processing have been boosting currently growing from a minor subfield into a powerful scientific discipline dealing with hi-profile projects. The recent research has documented the emerge of a totally new branch of science, digital humanities. This state-of-art field indicates how important computational approaches are for the humanities and social sciences nowadays (Bockwinkel, 2016; Videla, Carmen Gloria Burgos, et al. 2018 ). Qualified specialists in the sphere of CL or natural language processing are in high demand in many companies and corporations.
Thus, CL is gradually becoming a part of curricula in the growing number of colleges and universities all over the globe. The USA was among the first to start research and education in this field. The CL programs come under different names but their focus is evident: they are aimed at providing students with sufficient knowledge for creating life and science of the future. Natural language processing is taught at all level starting from undergraduate degree programs. CL courses can be included in the curricula of linguistics programs without mentioning it in the program title. Linguistics departments are introducing computer and speech labs to promote new research in the CL area. Various works have already been published on CL and its components and publications by the US authors are still used as guidebooks both by course leaders and students (Jurafsky & Martin, 2009). As the CL programs are only developing and there is still no unified idea of what they should look like except the fact that they include traditional linguistics courses such as phonetics, semantics, morphology and syntax, computer science and programming languages.

The greatest choice of CL programs is provided by the US universities. Europe is also working hard on developing its IT business and consequently promoting CL at the universities. Analyzing the experience of these two countries in organizing CL degree programs would be a way to find out certain trends in this sphere, new possibilities for curriculum improvement and implementing the best practices at other countries. Throughout history German universities have been of a particular importance for Russia and have positively influenced the development of the Russian system of higher education. Nowadays using German models in CL education could help Russia create innovative programs and courses. We have chosen master’s programs for the analysis as in the case with CL they seem to be the most popular. Although provided on the bachelor’s level as well the interdisciplinary nature of CL demands of the students to have some background training in linguistics and/or computer science. (Nikolaeva, A. D., & Savvinov, V. M. 2016)

2. Literature Review. The research is based on the official information, documents and course descriptions published by the Universities of Konstanz (University of Konstanz, Faculty of Humanities., 2018), Heidelberg (Ruprecht-Karls-Universitaet Heidelberg, 2017), Munich (Ruprecht-Karls-Universitaet Heidelberg, 2017), Master-Studienabschluessene Washington (2017), University of Washington (2018), Colorado (University of Colorado Boulder, 2017) and Carnegie Mellon University in Pittsburg (Carnegie Mellon University, 2016). In problems arising during the process of teaching CL have already been in focus of attention of researchers. One of the issues addresses in the question of how CL should be taught to students with diverse academic background which is not a rare case in this interdisciplinary subject (Baldridge & Erk, 2008), R. Amaro (2016). The authors share their experience and propose ways of adjusting CL programs that as the result could suit well both computer science and linguistics students. For example, Jason Baldridge and Katrin Erk back in 2008 thought that the best solution for teaching CL was working out separate programs for students studying the humanities and the sciences. Similar concerns are expressed by E. Fossler-Lussier (2008) whose analysis includes the situations when the same CL courses are attended by undergraduate and graduate students. M. Koit (2008) in their article focused on the undergraduate programs in CL and changes in them which occurs due to the shift from the 4-year to the 3-year study pattern. Unfortunately, not so much research has been done yet to analyze this problem in detail.

CL has long established itself as a field of great interest and importance for linguists, IT specialists and educators. It emerged as an independent branch of science in in the early 1960s when the term “computational linguistics” was created in order to widen the scope of studies in machine translation and suggest new ways of the possible use of computers in language processing. The CL constituent, corpus linguistics, has been and is used to build databases for various languages (Draxler, 2008). Modern science stresses the importance of the linguistics constituent in natural speech processing at the same time urging the need to consider mathematical patterns in order to fulfil linguistic tasks. CL is also widely employed in the psycholinguistic studies as it describes language processing the way it occurs in the brain. Thus, research in CL facilitates studies in speech production (Clark, 2013). Connection of CL with machine translation is highlighted in the works of E.A. Reiter (2018), I. Naim (2018). Reiter analyses the use of the BLEU in natural language processing systems while Naim, Riley and Gildea are trying to work out ways of eliminating the effects of orthographic similarities that hamper distinguishing between words in machine translation. S. Joty (2017), Guzman, Marquez and Nakov in their work deal with the new methods of machine translation evaluation. Issues of maximizing machine translation accuracy are raised in the article by G. Neubig and T. Watanabe (2016) who dwell on the algorithms used in the statistical machine translation. A number of papers is devoted to various aspects of applying machine translation to particular languages. D. Deng and N. Xue (2017) in their research investigate problems of translation from Chinese into English and do it with the help of a parallel treebank. One of the most important trends in CL nowadays is described by Christopher Ch. Manning (2015) which was published in the journal Computational Linguistics in December 2015. In his article Manning talks about Deep Learning and the possibilities which are predicted for this system in natural language understanding. These ideas of artificial intelligence are supported in the article by B. Grosz (2018) about creating computer systems that are able to hold conversations with humans. Researchers also address linguistics issues connected with CL. For example, M.A. Morey (2018) touch upon the subject of the dependences of discourse parsing on syntactic constituency parsing and to what extent this dependency can exist. The issues of morphological segmentation are discussed in the article by B. Can and S. Manandhar (2018). They touch upon the advantages of a probabilistic hierarchical clustering model as a new way format of segmentation in contrast to the traditionally used methods. CL is also used for making spelling correction programs. The work by A. Rozovskaya
(2017) is an attempt to find out the best variant with the help of the annotated learner data. C. Gardent and L. Perez-Beltrachini (2017) investigate different ways natural language generation and propose a hybrid approach analysis of which they present. Another linguistics problem is represented in the work by P. Tripodi and M. Pelillo (2017). These researchers are suggesting a new method of solving on of the most complicated problems in machine translation, that is choosing the only meaning of a word out of several. E. Shutova (2017) address the similar topic which poses even a greater difficulty for natural language processing, metaphor recognition and translation. The authors suggest the more extensive use of computer technologies. The scope of linguistics applications of CL is broad indeed and also includes text analysis. Computer models facilitate identification of various structures in a text, a method that can in its turn be applied to machine translation or text recognition (Stab & Gurevych, 2017; Tastan et al., 2018). A challenging approach to translation problems can be found in the work by A. Irvine, Ch. Callison-Burch (2017). They introduce a new method in which monolingual corpora is used for inducing variants of translation on the example of low frequency words. There is a possibility of using CL is research in the field of language dialects. It is based upon statistical methods and implies using corpora data (Nguyen & Eisenstein, 2017). The importance of the artificial intelligence is further emphasized by its growing use in medical applications. Nowadays CL based software can help people with the Alzheimer’s disease to compensate for breakdowns in communication and even avoid them (Chinaei, 2017). B. Roark and R. Sproat (2007) highlight the interconnection between linguistics and mathematics which is so characteristic of CL. In their work these scientists describe the process and results of applying algebraic methods to the formal semantic theory in order to analyze it.

3. Materials and Methods. In our research we focus on graduate CL degree programs in Germany and the USA of the two pointing out and discussing their particular features, academic nuances and innovative initiatives, which can be of advantage in the education systems of other countries. The methods used are those generally employed for empirical analysis like description, classification, data interpretation, content analysis, comparison and syntheses.

4. Results. One can take undergraduate and graduate programs in computational linguistics at several universities in Germany including the universities if Tuebingen, Saarland, Stuttgart, Konstanz, Munich and Heidelberg. Master programs from Konstanz, Heidelberg and Munich were taken for analysis as representing various approaches to teaching this complex discipline at universities with different background and the numbers of students enrolled.

The University of Konstanz offers an MA program dedicated to computational linguistics matters under the heading Speech and Language Processing. The program is run by the Department of Linguistics of the Faculty of Humanities. The University itself being a medium size one with about 11,000 students, it doesn’t have an undergraduate program in computational linguistics. Students who start the master’s program here are required to have some previous knowledge in linguistics while competence in computer science is not crucial. The syllabus indicates the two main areas of specification in which education is carried out. The first one deals with human language processing and consequently the fields where it can be used, namely psycholinguistics and neurolinguistics. The second area is dedicated to machine language processing that leads directly to applying computational models in digital sphere. The syllabus aims at providing the students with both theoretical and practical knowledge in language processing to enable them to be successful in the possible post graduate studies and/or at the job market. Students specializing in human language processing make a special emphasis on studying psycholinguistics and neurolinguistics, while those in the machine language processing subcategory deal mostly with symbolic and statistical methods. The principle that the University of Konstanz follows implies choosing between these two options after taking an obligatory module in foundational linguistics. Within its framework students are free to take any two classical linguistic courses though those who are further going to focus on machine language processing are advised to concentrate on Lexical Functional Grammar.

Further all the students need to study statistics and the basics of both human and machine language processing. Students specializing in human language processing are advised to focus their attention on experimental methods, while for machine language processing it is strongly advisable to learn at least one programming language. This MA program is also strongly research oriented with the point of providing the students with the necessary skills for conducting independent analysis in practical applications.

The same research focus model is shared by the Institute of Computational Linguistics (ICL) of the Ruprecht-Karls-University of Heidelberg. The oldest and one of the largest universities in Germany, University of Heidelberg offers this master’s program only to those who have a bachelor degree in the same field. Their graduates of the Computer Linguistics Master’s program are supposed to be able to work on new projects using the team skills acquired while studying. Strong research orientation is reinforced by the demand to solve practical problems in the field of computational linguistics. The range of research studies in Heidelberg is rather wide including both obligatory and optional courses. All the students in this program must take seminars in research and project planning, carrying out a project and scientific writing. They can also choose from conference organizing or participation, research internship, a textbook compilation and taking part in software release. Unlike Konstanz, ICL in Heidelberg stresses the division line between specializations in formal and applied linguistics vs theoretical and applied computational linguistics. The Heidelberg syllabus suggests a broad scope of training which involves seminars in general linguistics, computer science and mathematics. Linguistics and computational linguistics modules are obligatory for both specializations. Courses in theoretical and practical computational linguistics include Automata and Graph Theories, Machine Learning, Formal Languages, Methods of Statistical and Algorithmic Speech Processing, Information Retrieval, Machine Translation,
Artificial Intelligence, Speech Recognition and Synthesis and others. Studying formal and applied linguistics involves courses dealing with linguistic grammatical theories, syntax, semantics, pragmatics, morphology and phonology, language learning systems, cognitive, contrastive and corpus linguistics.

Another German institution of higher education, Ludwig-Maximilians-University in Munich accounting for more than 50,000 students enrolled offers both bachelor’s and master’s programs in Computational Linguistics. The studies are provided by the Center for Information and Speech Processing (CIS) that was founded in 1987. Candidates seeking for a master’s degree in Munich are expected to have studied computational linguistics or similar areas like linguistics, programming or mathematics at the bachelor level. Special attention is given to programming, it is a must for future graduate students to have so expertise in it. In contrast to the programs offered by the universities in Konstanz and Heidelberg that one from Munich is a Master of Science program. Its linguistic share is definitely smaller than at the previously mentioned universities. The ratio of linguistic and computational constituents in CL is still an issue that deserves careful consideration and has already been discussed by researchers (Roark & Sproat, 2007). The program falls into three main areas, including computer science, linguistic and algorithmic methods and foundations of computer linguistics. During first three semesters graduate students need to take master seminars in lexicon, syntax and semantics, information systems and algorithmic and formal foundations of computer linguistics. The seminars are obligatory though students are free to choose in which semester to take them. Alongside these seminars the program includes a broad and well-structured system of training in the field computer linguistics proper. It embraces three semesters and comprises multilevel (basic, advanced and proficiency) theoretical courses plus special contact hours for practical exercises where students are encouraged to further deepen their knowledge of the subject. During the third semester students can choose to take a course in experimental work and tools in computer linguistics.

The number of universities offering computational linguistics programs in the USA is rather large together with the variety of courses they offer. The most important one for many years has been and still remains the Stanford University. It hosts the oldest CL programs in the USA and possesses a number of labs and workshops. Besides Stanford is the home of some of the best-known CL researchers and text books authors. There is a variety of other high-quality institutions in the USA where CL is taught. One of them is the University of Washington (UW) where this degree can be obtained both through full-time studies and online. Implementation of distance learning in this sphere set the UW apart not only from German universities, but from many educational institutions in the USA and other countries. If in Germany one needs four semesters to earn the degree, in Seattle it can take from one up to three years. To get the degree students need to take nine courses and complete their master’s project. Of the nine course six are required and the other three are elective. Among the required courses only two are on linguistics while the rest deal with natural language processing which reflects the scientific emphasis of the program. All these courses require the students to be familiar with both basics of linguistics and computer programming. The computer languages needed are Perl, C, C++, Java or Python. Linguistics courses are also structured with an emphasis on language processing, for example, Introduction to Syntax for Computational Linguistics. A lot of attention is paid to various techniques in natural language processing with three courses devoted to this topic. The elective courses are also mostly from the field of computational linguistics. Another feature that differentiates the UW program from those taught in Germany is what the choice of options concerning what a master’s project can be like. Writing a thesis at the end of graduate studies is a long-lasting academic tradition which is many centuries old. Indeed, the idea of a final paper is simple and logic: students should be able to carry out an independent research using the knowledge and skills acquired during their studies at the university. At the UW they go further and make it possible for the students to choose between a master’s thesis and a six- to ten-week internship. Though this idea might seem to be breaking away from academic traditions, for sure it stresses the computer science character of the UW graduate program and highlights the huge demand of language processing in the modern world.

Carnegie Mellon University in Pittsburg, Pennsylvania, is well-known not only in the USA, but worldwide for its renowned programs in computer science and natural language processing. The Language Technology Institute of this university host a number of bachelor’s, master’s and postgraduate programs including a CL related one. It is a Master of Language Technologies which is a Master of Science degree. Courses constituting the program are related to speech processing, information extraction, machine translation, natural language processing, machine learning, and computational biology. Most of these are provided by the Language Technology Institute and in addition to it one third is taught by the School of Computer Science. The program is highly research oriented and also stresses the importance of developing public speaking skills. In addition to standard requirements Carnegie Mellon obliges its graduate students to devote summers while in the program to scientific research. By the end of the second year of their studies all the students must give oral presentations of their work which are assessed by the faculty members and can be attended by other students. At the same time a masters’ thesis is suggested as necessary only to the students who wish to pursue a research career in the future and apply for the PhD program. It might seem to be a controversy to the general research-oriented paradigm of the whole institute. However, this fact just clearly indicates that the whole program is highly practically oriented. Like at the University of Washington a thesis can be substituted by a research project for those who choose completing it. One can assume that Carnegie Mellon are deeply concerned about their graduates’ abilities to implement the knowledge they get at the university in practical applications. The thesis-free model of a Master’s program is also employed by some other US universities offering degree studies in Linguistics. A noteworthy blend of
computer science and linguistics can be observed in the CLASIC, a graduate program of the University of Colorado. Run by two departments simultaneously the Master of Science program in CL, Analytics, Search and Informatics was first presented in 2016 and probably marks a new trend in interdisciplinary education. The goal of the program is to make students competent both in theoretical linguistics and computer science with a special emphasis on such languages as English, Chinese, Arabic and Urdu. The computer science predominance of the program becomes obvious through the study of the syllabus where there are twice less courses in linguistics than in natural language processing, artificial intelligence, machine translation and learning. The CLASIC program is noticeably practically oriented. Instead of a Masters’ thesis students are expected to complete a research project which should be connected with and apparently worked out during their internships. These projects are evaluated not only by the departments professors, but also by employers or industry supervisors.

5. Discussion and Conclusions. Nowadays the world witnesses an unprecedented boom of computer and internet technologies. Artificial intelligence has firmly established in the lives of billions of people around the globe. Studying and learning how CL works has become essential for the successful economic development. At present master’s degree programs seem to be the most compact and popular way of teaching CL. More and more universities in various countries evince interest in CL programs. Germany being the leading economy in Europe is a place where CL is taught at all levels of higher education. Universities with CL programs are situated mostly in Bavaria and Baden-Wuertemberg. Graduate programs they offer are of two type: Master of Arts and Master of Science. Such a diversity is no way illogical. It just clearly highlights the dual nature of CL and the fact that modern system of education is still seeking for the most appropriate methods of teaching this subject. CL combines in itself linguistics and computer science. Master of Arts programs seem to be more suitable for students with no previous experience in information technologies while Master of Science ones are better for science bachelors with mathematics and information technologies courses prevailing. The other models involve differentiation between theoretical and applied CL. The advantage of this approach in evident: it helps students to make decisions about future career at business or in research organizations.

The idea of practically oriented CL programs dominates at American universities. Being CL founders, the USA have already advanced far ahead of other countries and offer a wide variety of CL courses at their universities. Usually the best CL programs can be found at the universities which are considered top institutions for teaching computer science. Like CL programs in Germany, those in the USA tend to be mostly computer science oriented. The special innovative feature of the US universities is new approach to the concept of a master’s thesis. In the USA a thesis is becoming optional for CL graduate students which sounds revolutionary for the field of linguistics. It most likely happens so due to the desire to make the programs more practically oriented. Instead of writing a thesis, which is still possible for the students who are opting for PhD degrees in the future, US universities stress the necessity of working on CL related projects. Students are supposed to work on the during their internships as well together with the internship supervisors who later can take part in the projects’ evaluation. Another valuable feature is that students are expected to present the results of their research to other students, and faculty members. German and US programs in CL could be used to create similar programs in other countries or to upgrade the already existing ones. The research results could be well applied to the Russian system of higher education where CL has just been introduced at a few universities. We believe that further studies of international experience in teaching CL can contribute to the improvement and unification of these programs. Best practices could be transferred to different countries and help to promote the development of modern information technologies. Undergraduate and post-graduate programs in CL are also well worth analyzing and systemizing in order to try of elaborating some universal guidelines for CL courses.

6. Acknowledgements. The publication is the outcome of the assignment №34.12808.2018/12.2 for carrying out a scientific project on the topic «Making scientific research in the framework of international scientific and educational cooperation upon the program «Immanuel Kant».

References


