Editor's note!

Welcome to our fifth CLASS Newsletter!
We realise that this has been a very trialing period for all of you.
Due to the COVID crisis, CLASS scheduled live meetings have been postponed indefinately. However, some CLASS tasks were carried on uninterruptedly and therefore in this isssue, we are pleased to present to you the major dissemination events and all partners' news related to CLASS prior to the pandemic, or after.
In addition, Pablo Gamallo from USC, CLASS project leader, describes the impact this crisis has had on the program and the overall progress of CLASS over the past months through a careful examination of all the academic activities implemented so far.

On behalf of CLASS we wish you all well.

Stay safe!

Message from the management

The COVID-19 pandemic restrictions inevitably affected the development of CLASS project. The consortium Annual Meeting planned for April 2020 on Al-Farabi Kazakh National Univertisty (KazNU) premises had to be cancelled. However, the partnership continued to work, and the activity was maintained. The Consortium is assessing the possibility to ask for an extension of the project to the EACEA.

The representatives of the National Erasmus+ Office in Kazakhstan carried out a monitoring visit in November 2019 at the Eurasian National University. The progress of activities was satisfactorily assessed with the local project partners, including representatives of the Al-Farabi Kazakh National University and the Kostanay State University. At the stage of the visit, the syllabi for the new Master programme in computational linguistics were being finalized and the content materials based on the blended learning approach were being developed.

In January 2020, the purchase of equipment was concluded. CLASS partner universities in Uzbekistan and Kazakhstan received computer hardware, worth €142.515,22, aiming at enabling the use of innovative ICT tools (both commercial and open source type) in the field of computational linguistics.

On behalf of USC, *Pablo Gamallo*

Academic activities: Report from the management

During the months of COVID-19 pandemic lockdown, the main activities of the CLASS project that have been carried out correspond to tasks included in WP3. Mainly, NLP tools and material for blended courses have been created.

Concerning the Uzbek language, **Urgench State University (UrSU)** ended up developing a morphological analyzer in the last few months. The analyzer, called Uzbek MORPhological Parser (UzMorpP), was written in Prolog. UzMorpP is freely available.

Due to COVID-19 quarantine, all courses in **KazNU** were done remotely. Professors of the educational program "Computational Linguistics" created new educational material to adapt to this complex situation. The following e-books were created for the "Language Resources" discipline:

•**Corpus Linguistics**: a glossary of basic terms and concepts (on kazakh language). – Almaty: Kazakh University, 2018. – 40 p. (Authors: Madieva G.B., Ismailova N.)

•**Corpus Linguistics**: Educational Dictionary of Key Terms and Concepts. Learner's Dictionary (on Russian and english languages)/ Compiled by G.B. Madieva, S. Bektemirova, N. Ismayilova. – Almaty: Kazakh University, 2019. – 52 p.

•**Innovative technologies and teaching methods. Dictionary of terms**. Learner's Dictionary/ Compiled by G. Madiyeva, S. Bektemirova, N. Ismayilova, D. Madiyeva – Almaty: Kazakh University, 2019. – 31 p.

Tools by KazNU	
Tools for Kazakh language	Bilingual tools (applied to Kazakh-English)
Tokenizer	Sentence aligner (tsv format)
Stopword removal	Tools to clean bitexts
Lemmatizer	Extractor from xml files
	Morfological segmentation

Table 1: NLP tools developed by KazNU in Python

It is importat to note that in the 2019-2020 academic years within the CLASS project, the Computer Linguistics master's degree was opened in UrSU. Due to the COVID-19 quarantine, during the last months several disciplines of the master have been taught remotely, namely: Introdcution to Programming for Natural Language Processing, Language Resources, Machine Translation Technology, Machine Learning in Natural Language Processing, and Natural Language Understanding. In addition, to promote the new master program, Samarkand **State Institute of Foreign Language (SamSIFL)** made public the following articles:

1.Implementation of Erasmus+ CLASS project at SamSIFL by O.Yusupov and F.Bakiev published by NEO in Uzbekistan in 2019 [http://www.erasmusplus.uz/images/shared/file/Erasmus+ publication 2019.pdf]

2.Роль и применение ИКТ в преподавании общей и специализированной терминологии будущим профессиональным переводчикам на базе языкового вуза by T.Nasrullayev published in the conference proceedings "MODERN LINGUISTIC RESEARCH: FOREIGN EXPERIENCE, ADVANCED RESEARCH AND INNOVATIVE METHODS OF TEACHING LANGUAGES" on May 15, 2020

3.Роль машинного перевода в изучении китайского языка by S.Bekniyazova published in the conference proceedings "MODERN LINGUISTIC RESEARCH: FOREIGN EXPERIENCE, ADVANCED RESEARCH AND INNOVATIVE METHODS OF TEACHING LANGUAGES" on May 15, 2020.

4.ERASMUS+ PROJECT EXPERIENCES ON THE COMPUTATIONAL LINGUISTICS MASTER DEGREE IN THE CENTRAL ASIAN UNIVERSITIES by N. Abdurakhmonova, M. Aripov, A. Duarte, K. Georgouli, C. GómezRodríguez, Y. Kandalina, B. Maia, A. Sharipbay, U. Tukeyev, G. Urazboev, Z. Vetulani, O. Yusupov published in INTED 2020.

SamSIFL has also developed blended courses for the disciplines "Language analysis" and "Language resources ". Currently they are preparing teaching materials such as lecture presentations, seminar worksheets, tasks, exercises, tests as well as two video rollers. They are planning to finish these tasks till August, 2020 and make them online.

To enroll master students they have been discussing this issue continuously with the Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan. Finally, according to the Resolution of the President of the Republic of Uzbekistan No. PP-4749, 6 entry quotas for MA Computational Linguistics at SamSIFL have been confirmed, two of which are state scholarships.

Tools by UrSU

Urgench State University project members developed automatic morphological parsing tools for the Uzbek language (UzMorph) under supervision of professor ZygmuntVetulani (AMU) as NLP Tools development subtask (WP3.3) of CLASS project. This research was partially developed by the postdoc grant of the Gayrat Matlatipov within his Erasmus Mundus Fellowship at the Adam Mickiewicz University from 2008.
A morphological parser UzMorpP implemented in Prolog. The program uses lexicons for roots (1000 entries) and suffixes (108).

To use UzMorpP system please take the following steps:
Input: uzb.tst - the tool accepts as input a list of words from an external input text file.
Output: the tool will return a list of all possible morphological segmentations of a word with corresponding descriptions.
The predicate do/0 is used to initiate the UzMorpP system. This predicate reads inputs from the text file (here 'uzb.tst'). Here every word should be separated with a comma or space and must contain only Uzbek characters. The Uzbek alphabet has two compound characters (o' and g') where a latin letter is followed by the apostrophe (ASCII code 39). In order to correctly process Uzbek words using these characters the UzMorpP tool changes the apostrophe to (ASCII code 95).

An operational prototype of UzMorpP is installed for free inspection at:
https://drive.google.com/open?id=1TK-uleV7ftpWIfKPboVR4Gu-Fxe6Vknv

Tools and course by KazNU

Monolingual NLP tools.
1. Tokenization tool. The tokenization algorithm is implemented using regular expressions. The tokenizer receives one sentence at the input, and breaks it into tokens.
2. Tool for removing stop words. Dictionary of stop words is compiled. In the Kazakh language, stop words include ododay (interjection), shylau (particles, unions), elikteu (imitative words), esimdik (pronoun), modal szzder (modal words).
3. Lemmatization tool. Based on the Porter algorithm, according to the grammatical rules of Kazakh, affixes are sequentially cut off from the end of the word: that is how stemming is performed.

Bilingual corpora development tools.
1. Tool for align parallel bilingual corpora. Aligns sentences from file pairs in a given folder based on their names. Uses hunaligntool.
2. Tool for clean alphabet. Checks for character substitutions between characters from Kazakh and English alphabets. Corrects cases where one Kazakh letter is found between two English letters or one English letter is found between two Kazakh letters.
3. Tool for clean text. Cleans texts in Kazakh and English:
- removes zero-width characters, heading and trailing spaces, substitutes various space characters with a plain single space;
- cleans empty lines;
- substitutes various quotation mark characters with plain quotation marks
- substitutes various hyphen characters with a plain hyphen.
4. Tool for combine texts into one. Combines files from a given directory into one file based on the file name ("eng" or "kaz").
5. Tool for extracts titles and texts from XML files. Extracts titles and texts from XML files. Saves them into text files.
6. Tool for normalization and tokenization. Performs normalization and tokenization on a given file. Uses sacremoses tool.
7. Tool for morphological segmentation. Performs morphological segmentation of Kazakh text.

Blended learning course development in the discipline "Machine Translation Technologies."
The people responsible for Blended learning course development in the discipline "Machine Translation Technologies" are Zhandos Zhumanov and Aliya Turganbaeva. The lecture materials and materials of practical lessons were prepared and were download into Moodle system.
(https://dl.kaznu.kz/course/view.php?id=101303)
Due of coronavirus quarantine in KAZNU all educational process from March 2 onwards moved to study remotely. Therefore, courses "Formal Grammar" and "Understanding of Natural Languages" were moved to MOODLE system of KAZNU with lecture and practical materials.

The latest CLASS achievement!

Message from Ulugbek Khodiev, Head of Department, Ministry of Higher and Secondary Special Education of the Republic of Uzbekistan.

With a resolution of the President of the Republic of Uzbekistan, quotas for admission to higher educational institutions were approved. According to the decree, from the 2020/2021 academic year, the preparation of masters in the specialty "Computer Linguistics" begins in the following universities: National University of Uzbekistan; Tashkent State University of Uzbek Language and Literature; Samarkand State Institute of Foreign Languages. Congratulations to all the participants in the project!
Sincerely, *Ulugbek Khodiev, Head of Department, Ministry of Higher and Secondary Special Education of the Republic of Uzbekistan.*

CLASS@INTED 2020: "Sharing the Passion for Learning"

CLASS Project was pesented at the 14th International Technology, Education and Development Conference, in Valencia, Spain, between March 2-4, 2020.

The presentation was carried out by Prof. Katerina Georgouli (UNIWA, Athens) and the paper was included in the INTED2020 Proceedings.

PAPER AUTHORS

N. Abdurakhmonova,
M. Aripov,
A. Duarte,
K. Georgouli,
C. Gómez-Rodríguez,
Y. Kandalina,
B. Maia,
A. Sharipbay,
U. Tukeyev,
G. Urazboev,
Z. Vetulani,
O. Yusupov

The paper presented the experience of implementing the ERASMUS+ CLASS project. It discussed the relevance of training in computational linguistics in connection with the active development of artificial intelligence and its use in various industries as one of the top priorities at present. The paper explained the development and teaching of relevant disciplines as: Machine Learning in NLP, Language Resources, Language Analysis, Speech Processing, Machine Translation Technology, Understanding Natural Language, Ontologies and Semantic Technologies, Statistical Methods in NLP, Formal Grammar, and Python Programming. It featured the modern pedagogical technologies that were implemented, such as blended learning, and project based learning as well as an analysis of the needs of the labor market where more than 50 companies were interviewed.

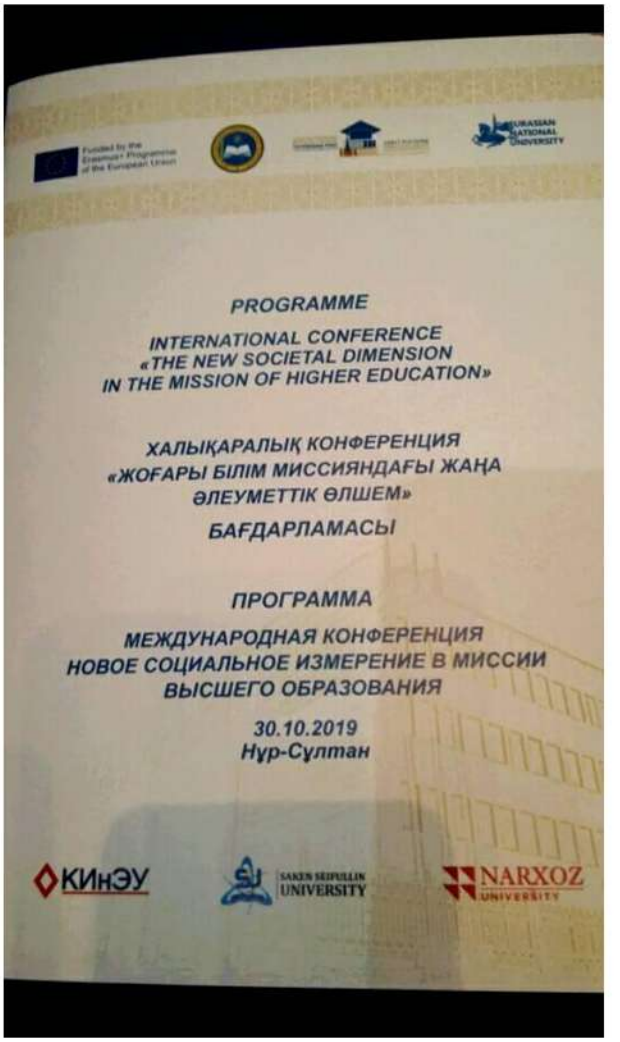
The paper also examined the tools that have been developed for processing the Kazakh and Uzbek languages that will be used in organizing and conducting the educational process for this program. The impact of this educational program both at the individual level, at the level of teachers involved in the implementation of this educational program, and at the institutional level was found to be very significant.

CLASS Project @ UNESCO International Conference, Paris, France

Two members of CLASS project professor Zygmunt Vetulani (Adam Mickiewicz University) and associate professor Nilufar Abdurakhmonova (Tashkent State university of the Uzbek language and literature) participated in International Conference "Language Technologies for All (LT4All): Enabling Linguistic Diversity and Multilingualism Worldwide" within the framework of the 2019 International Year of Indigenous Languages, a three-days (4-6 December 2019) International Conference organized at UNESCO Headquarters, Paris, France. As CLASS project fruitful collaboration their investigation in the field of language technology will be published in the proceedings of the conference.

CLASS Project @ International Conference in Nur-Sultan, October 2019

CLASS Project was presented at the International Conference entitled: "The New Societal Dimension in the Mission of Higher Education." The event was held in Nur-Sultan last October



Kazakh Language in computer linguistics Round Table

A round table was held on December 9 at the Art Lane gallery in Almaty organized by KazNU and LLP "Translators Group." The discussion was moderated by Ospan Berdaly (into Kazakh and Russian languages).

The purpose of the round table was to discuss current issues related to computer linguistics in Kazakhstan, taking into account the transition to Kazakh and Latin alphabets, NLP and linguistic technologies. The audience consisted of researchers, linguists, translators, students, business and government agencies, media.

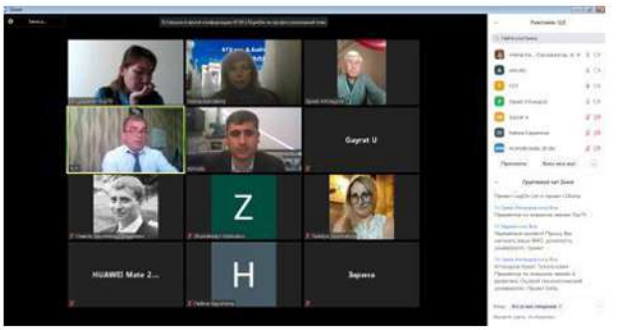
The discussion revolved around the following issues:
-create corpora of Kazakh language;
-use the neural network in machine translation and in translation from the Kazakh language;
-creation of glossaries with terminological dictionaries for the Kazakh language and
-training of personnel for work in the field of computer linguistics.

Staff week@USC

Last November, the USC hosted the fourth edition of Staff Week, a meeting that aims to promote the exchange of synergies that revert to optimizing the management of mobility programs. Fourty Universities participated in the event, twenty of which came form non-EU countries. CLASS member Gayrat Urazboev was also there representing UrSU University. The event was covered by "La Voz de Galicia."

New Challenges for the Implementation of ERASMUS +

CLASS was presented at a round table entitled "New Challenges for the Implementation of Erasmus+" that was organized by Baitursynov Kostanay State Univesity. Participants came from universities in Spain, Kyrgyzstan, Russia, Uzbekistan, Tajikistan, Kazakhstan.



SCOPUS Award 2019 granted to Dr. Gayrat Matlatipov

Dr. Gayrat Matlatipov was awarded the SCOPUS AWARD for his research works done in studying the Uzbek language

Stay Safe!