***Регистрационная форма участника:***

ФИО (полностью): *Sagimbayeva Nursulu Nurlankyzy / Сагимбаева Нұрсұлу Нұрланқызы*

Место учебы или работы: *КазНУ им.аль-Фараби*

Должность (для студентов - курс): *студентка 4 курса спец. Переводческое дело*

Контактные телефоны: 8 778 140 28 91; 8 707 809 69 77

Адрес электронной почты (E-mail): *nursulu200@mail.ru**;* *saule\_dus@mail.ru*

Почтовый адрес, по которому Вам можно выслать опубликованные материалы:

*nursulu200@mail.ru**;* *saule\_dus@mail.ru*

Секция конференции: *Филология, лингводидактика и перевоведение: актуальные проблемы и тенденции развития*

Устное сообщение или стендовый доклад: *устное сообщение*

Технические средства, необходимые для демонстрации доклада: *интерактивная доска для презентации*

Для студентов - Ф.И.О., уч. степень, ученое звание научного руководителя – *Сейденова С.Д., к.филол.наук, и.о.профессора*

**MACHINE TRANSLATION IN SOCIAL MEDIA: PECULIARITIES, CHALLENGES, AND OPPORTUNITIES**

*Student Sagimbayeva N.N.*

*under the supervision of Seidenova S.D.*

Al-Farabi Kazakh National University

Almaty, Kazakhstan

email: nursulu200@mail.ru

With the emergence and ubiquitous spread of technology, social media have become a major source of textual data. According to research data (S. Agarwal, S. Godbole, D. Punjani, and S. Roy, 2007; J. Carrera, O. Beregovaya, and A. Yanishevsky, 2009), user-generated content has specific traits that distinguish it from other texts:

1) it contains high levels of noise (such as errors, abbreviations, filler words like “um”, etc.);

2) it is not tied to a particular domain;

3) it is produced at a rapid pace and in large amounts;

4) it is focused on sharing information and knowledge, meaning that linguistic correctness or appropriate syntactic structure is usually not a priority. These and other peculiarities are thoroughly considered in the tasks related to natural language processing (NLP), text mining, hate speech detection, and many others.

Machine translation (MT) is another domain that works closely with user-generated content (UGC). There are many dimensions in which MT in social media can be useful. The examples include but are not limited to the following ones:

1. *socio-political:* during tensions in the society or in a global arena, a lot of content is produced by people with different views and from different linguistic communities. Proper MT enables access to the content from *different* sides of the conflict, thus giving the users the power to access original UGC, undistorted by the journalists or actors with political interests or bias towards the situation;
2. *economical:* internet-based shops and other businesses can encompass a much larger audience if their descriptions and captions are available in a multitude of languages. Social media influencers, charities, magazines, and other entities can benefit from this greatly, too;
3. *content inclusivity:* MT can translate captions for videos, movies, and other audiovisual content, giving the users access to more sources of information. These sources encompass domains of education, science, entertainment, and so on.

Despite the high demand for efficient MT, it still faces certain challenges. For example, the majority of the languages in the world don’t have rich enough corpora of their own, as well as rich *bilingual* parallel corpora to produce efficient automated translation into other languages. As Margaret Dumebi Okpor (2014) points out, building such corpora is very resource-consuming. Moreover, the situation is aggravated by different word orders in different languages (consider Japanese and English), the wide presence of idioms, slangs, neologisms, as well as linguistic practices like code-switching (a common phenomenon for the Kazakh society).

For these reasons, there is a need to discuss recent machine translation approaches and advancements, as well as challenges and perspectives in the field. Real-life examples from social media will contribute to a better understanding of the topic.