Brief information about the project

Name of the project	AP15473441 "Development of word-formation markings in the corpus of texts of the kazakh language"
Relevance	The most relevant today is machine translation, automatic recognition of language elements and synthesis of spoken speech, spelling and grammar checking tools, modern frequency dictionaries, etc. To strengthen the activities of the National Corpus of the Kazakh language, an urgent issue is to increase the types of linguistic markings introduced into it, prepare its linguistic base, and consider software capabilities.
Purpose	The development of a language base and the development of software for the input of word-formation markings in the Kazakh language into the corpus.
Objectives	- to characterize the concepts of "derivative word" and word-formation markings in the corpus; - determination of the involvement of affixes in word formation and form formation; - analysis of the development of the introduction of word-formation markings in the National Corpus of the Turkish language; - analysis of the development of the introduction of word-formation markings in the National Corpus of the Russian language; - compilation of a list of paired words, a list and dictionary of merged words, a table of word-formation affixes for each class of words as linguistic instructions that serve as the basis for building an algorithm for automatic recognition of words in the corpus; - preparation of the development of automatic recognition in the corpus of paired words, merged words and derived words that are word-formation markings.
Expected and achieved results	- the description of the concept of "derivative word" and word-formation methods in the corpus, types of markings is given; - the ratio of affixes to word formation and form formation is determined; - analysis of the development of the introduction of word-formation markings in the National Corpus of the Turkish language; - analysis of the development of the introduction of word-formation markings in the National Corpus of the Russian language; - as linguistic instructions that serve as the basis for the formation of an algorithm for automatic recognition of words in the corpus, a list of paired words, a list and dictionary of merged words, a table of word-formation affixes for each class of words will be compiled; - the development of automatic recognition in the corpus of paired words, merged words and derived words, which are word-formation markings, will be developed.

	- implementation of publications: the results of the research are published in the form of articles in domestic scientific publications and foreign scientific journals and conferences recommended by the Higher Attestation Commission: - At least 2 (two) articles in scientific publications indexed in the Arts and Humanities Citation Index or Social Science Citation Index in the Web of Science database or containing at least 35 (thirty-five) percentiles of CiteScore in the Scopus database Or: 1 (one) monograph with the share of a postdoctoral fellow who has been reviewed and submitted for publication by holders of at least two academic titles of professor and/or a PhD degree of at least 6 p. s.; - at least 5 (five) articles and (or) reviews in peer-reviewed foreign and (or) domestic publications recommended by the CCONVO.
Research team members with their identifiers (Scopus Author ID, Researcher ID, ORCID, if available) and links to relevant profiles	Pirmanova Kunsulu Kambarbekkyzy, PhD postdoctoral student, https://orcid.org/0000-0003-3783-3199
List of publications with links to them	1. Pirmanova K.H., Akar A., Myrzakhanova G.K. Ways of transmitting partially word-formation affixes in the national corpus of the Kazakh language. TILTANYM No. 1 (89) 2023 https://www.tiltanym.kz/jour/article/view/799/493 2. Pirmanova K.K., Torekhanova A., Ashimova N. Theory and methodology of the national linguistic corps of the world. Abylai Khan atyndagi Kazakh halykaralyk katynastar zhane alem tilderi University KHABARSHYSY No. 2 (69), 2023 https://bulletin-philology.ablaikhan.kz/index.php/j1/article/view/713 3. Pirmanova K.H., Akar A., Kaymerdenova N.J., Halabaeva I.M. Ways of transmitting complete word-formation affixes in the national corpus of the Kazakh language. TILTANYM No. 3, 2023 https://www.tiltanym.kz/jour/article/view/1151 4. Pirmanova K.K., Karbozova B. On the software of the database of the electronic corpus of Kazakh texts. Publishing house.Agency: Proceedings of the 1st International Scientific Conference "Fundamentals and Trends in Research" (December 1-2, 2022). Copenhagen, Denmark, 2022. 233 p. ISBN 978-2-684-08428-6. DOI 10.5281/zenodo.7396104 https://ojs.publisher.agency/index.php/FTR/article/view/420 5. Pirmanova K.K., Zaurbekova G., Tokhmyrzaev D.O. Derivatives and methods of word formation (structural aspect of word-formation designations). POLISH SCIENTIFIC JOURNAL. №67 (2023), ISSN 3353-2389 6. Pirmanova K.K., Zaurbekova G., Ayazbekova K. The concept of a sign and types of signs in the case. Scientific heritage (Budapest, Hungary). ISSN 9215 — 0365 № 123 (123) (2023)

	7. Pirmanova K.K., Tatieva S., Kazezova A. Development of a program of word-formation markings in the corpus of texts of the Kazakh language. The journal "Znanie misel" is registered and published in Slovenia. ISSN 3124-1123, №83/2023
Patents	-