ҒТАМР 16.31.21 https://doi.org/10.26577/EJPh202220267

А.А. Карымхан, М.Қ. Мәмбетова*, Б. Нұрланғазықызы

Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
*e-mail: mmanshuk@gmail.com

МӘТІНДІ АВТОМАТТЫ ЖЕҢІЛДЕТУ: ЗЕРТТЕУЛЕР МЕН БАҒЫТТАР

Мақалада табиғи тілді өңдеу саласындағы дербес зерттеу бағыты ретінде автоматты мәтінді жеңілдетуге (ATS) жан-жақты шолу жасалады. Зерттеудің мақсаты – автоматты мәтінді жеңілдетудің дамуын, ағымдағы жағдайын және негізгі мәселелерін жүйелі және сыни тұрғыдан ұсыну. Scopus, Google Scholar және ACL Anthology дерекқорларындағы кең көлемді әдебиет шолуына сүйене отырып, 1998 жылдан 2025 жылға дейінгі зерттеулер талданады. Мақалада ATS дамуы ережеге негізделген тәсілдерден бастап, статистикалық және нейрондық модельдер арқылы үлкен тілдік модельдерді қолдануға дейінгі эволюциясы қарастырылады. Бұл үдеріс бірнеше тілде жоғары сапалы параллель корпустардың біртіндеп кеңеюімен қатар жүріп отырғаны көрсетіледі.

Күрделі сөздерді (1) анықтауды, сондай-ақ алмастыруды (2) іріктеу, (3) генерациялау және (4) дәрежелеуді қоса алғанда, лексикалық жеңілдету үдерісін талдауға ерекше назар аударылады. Зерттеу ережелерге, жиілікке немесе таза деректерге негізделген оқшауланған тәсілдер көбінесе өз шегіне жететінін және гибридті, лингвистикалық тұрғыдан хабардар шешімдер ең жақсы нәтиже беретінін көрсетеді. Негізгі мәселелер мағына мен үйлесімділіктің сақталуы, зерттеулерде ағылшын тілінің күшті басымдығы және қазақ тілі сияқты типологиялық тұрғыдан күрделі тілдер үшін ресурстардың жетіспеушілігі болып қала береді.

Мақалада мұндай тілдер үшін таза нейрондық тәсілдердің жеткіліксіз екені атап өтіледі. Оның орнына лингвистикалық тұрғыдан негізделген сенімді модельдеуге сүйенетін, сонымен қатар автоматтандырылған және нейрондық әдістермен толықтырылған кезең-кезеңмен жүзеге асатын тәсіл ұсынылады. Шолу автоматты мәтінді жеңілдетуді одан әрі дамыту үшін бағалау процедураларын жетілдірудің, сондай-ақ типологиялық тұрғыдан бағытталған, лингвистикалық негізделген зерттеулердің маңыздылығын көрсетеді.

Түйін сөздер: мәтін, жеңілдетілген мәтін, автоматты мәтінді жеңілдету, қазақ тілі.

A.A. Karymkhan, M.K. Mambetova*, B. Nurlangazykyzy

Al-Farabi Kazakh National University, Almaty, Kazakhstan
*e-mail: mmanshuk@gmail.com

Automatic Text Simplification: Studies and Directions

The article provides a comprehensive survey on automatic text simplification (ATS) as an independent research area in the field of natural language processing. The study aims to systematically and critically present the development, current state, and key challenges of ATS. Based on an extensive literature review in Scopus, Google Scholar and the ACL Anthology, research from 1998 to 2025 is analyzed. The article examines the evolution of ATS development from rule-based approaches to the use of large language models using statistical and neural models. It is shown that this process goes hand in hand with the gradual expansion of high-quality parallel corpora in several languages.

A particular focus is placed on the analysis of lexical simplification, including the (1) identification of complex words, as well as (2) selection, (3) generation and (4) ranking of substitutions. The study shows that isolated rule-based, frequency-based, or purely data-driven approaches often reach their limits and that hybrid, linguistically grounded solutions deliver the best results. Key challenges remain the preservation of meaning and coherence, the strong dominance of English in research, and the lack of resources for typologically complex languages like Kazakh.

The article notes that purely neural approaches to such languages are not enough. Instead, a step-by-step approach is proposed, based on linguistically sound reliable modeling, as well as complemented by automated and neural methods. The survey highlights the importance of improving evaluation procedures for the further development of automatic text simplification, as well as typologically oriented linguistically sound research.

А.А. Карымхан, М.К. Мамбетова*, Б. Нурлангазыкызы
Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
*e-mail: mmanshuk@gmail.com

Автоматическое упрощение текста: исследования и направления

В статье дается всесторонний обзор автоматического упрощения текста (ATS) как самостоятельного направления исследований в области обработки естественного языка. Цель исследования – систематически и критически представить развитие, текущее состояние и ключевые проблемы ATS. На основе обширного обзора литературы в Scopus, Google Scholar и ACL Anthology анализируются исследования с 1998 по 2025 год. В статье рассматривается эволюция развития ATS от подходов, основанных на правилах, до использования более крупных языковых моделей с помощью статистических и нейронных моделей. Показано, что этот процесс идет рука об руку с постепенным расширением высококачественных параллельных корпусов на нескольких языках.

Особое внимание уделяется анализу процесса лексического упрощения, включая определение (1) сложных слов, а также (2) отбор, (3) генерацию и (4) ранжирование замен. Исследование показывает, что изолированные подходы, основанные на правилах, частоте или исключительно на данных, часто достигают своих пределов, и что гибридные, лингвистически обоснованные решения дают наилучшие результаты. Ключевыми проблемами остаются сохранение смысла и связности, сильное доминирование английского языка в исследованиях и нехватка ресурсов для типологически сложных языков, таких как казахский.

В статье отмечается, что чисто нейронных подходов к таким языкам недостаточно. Вместо этого предлагается поэтапный подход, основанный на лингвистически обоснованном надежном моделировании, а также дополненный автоматизированными и нейронными методами. Обзор подчеркивает важность совершенствования процедур оценки для дальнейшего развития автоматического упрощения текста, а также типологически ориентированных лингвистически обоснованных исследований.

Ключевые слова: текст, упрощенный текст, автоматическое упрощение текста, казахский язык.

Кіріспе

Мәтінді автоматты түрде жеңілдету (ATS – automatic text simplification) – мәтіннің түсініктілігі мен оқылымдылығын оңтайландыру мақсатында оның ақпараттық бастапқы мазмұны мен мағынасын сақтай отырып, тілдік күрделілігін (лексикалық, синтаксистік және дискурстық деңгейде) автоматты оқытылған машиналық жүйелер арқылы төмендету үдерісі. Бұл үдерістің мақсаты – белгілі бір аудитория (балалар, шетелдіктер немесе ана тілінде тілдің сауаты төмен қолданушылар, екінші тілді үйренушілер мен оқытушылар, нейро-когнитивті бұзылулары бар немесе арнайы білімі жоқ адамдар) үшін күрделі құжаттардың мазмұнын оларды түсінуді жеңілдету үшін бейімдеу (Paetzold, 2016).

Тарихи түрғыдан мәтінді автоматты түрде жеңілдету басқа табиғи тілді өңдеу (NLP) тапсырмаларын жақсартудың алдын ала өңдеу кезеңі үшін қажет қадам ретінде дамыған. Табиғи тілді өңдеу (NLP) аясында мәтінді жеңілдету (text simplification, TS) парафраза жасау, мәтінді қысқаша мазмұндау және машиналық аударма сияқты бағыттармен тығыз байланысты. Осыған байланысты мәтінді жеңілдетуді бағалау тәсілдерінің едәуір бөлігі аталған салалардан алынған. Мысалы, синтаксистік жеңілдетуді жүзеге асыруда талдау (parsing) әдістері (Chandrasekar, 1996), сұрақ генерациялау (Heilman, 2010), ақпаратты шығару (Evans, 2011), фактілерді анықтау (Beigman Klebanov, 2004) және семантикалық рөлдерді белгілеу (Vickrey, 2008) сияқты тәсілдер қолданылып келеді.

Уақыт өте келе мәтінді жеңілдету тәсілдері қолмен аннотацияланған ережелерге негізделген жүйелерден автоматтандырылған модельдерге қарай дамыды. Бұл өзгеріс аталған бағыттың зерттеушілер арасында қызығушылық тудырып отырғанын және оның өзектілігін айқындайды. Осы шолу мақаласында автоматты мәтінді жеңілдету саласындағы негізгі зерттеулер қарастырылып, жүйеленеді. Жұмыстың мақсаты – осы бағыттағы ғылыми еңбектерге кешенді талдау жасап, бұрын қолданылған тәсілдерді сипаттау, сондай-ақ жеңілдетудің лексикалық, синтаксистік және семантикалық қырларын және қазіргі заманғы әдістерді талқылау. Жалпы алғанда, бұл бағыт әлемдік тілдер тәжірибесінде қарқынды дамып келеді.

Қазіргі таңда жасанды интеллект жүйелері мен ірі тілдік модельдер мәтін құруда жоғары

нәтижелер көрсетуде, бұл оларды мәтінді жеңілдету міндеттерінде тиімді құралға айналдырады. Соған қарамастан, мәтінді автоматты түрде жеңілдету NLP-дің қолданбалы тапсырмасы ретінде қарастырылатынын ескеру қажет. Үлкен тілдік модельдер техникалық түрғыдан маңызды болғанымен, олар толыққанды лингвистикалық және типологиялық талдаудың орнын баса алмайды.

Мәтінді автоматты жеңілдету бағытындағы көптеген зерттеулер негізінен ағылшын тілінде жүргізілген. Бұл жағдай қолданылатын әдістер мен бағалау көрсеткіштерінің қалыптасуына тікелей әсер етті. Көп жағдайда мұндай тәсілдер тілге тәуелсіз ретінде ұсынылғанымен, оларды басқа типологиялық ерекшеліктері бар тілдерге бейімдеу мәселесі жеткілікті деңгейде зерттелмеген.

Бұл саладағы өзекті мәселелердің бірі – тілдік күрделілікті формалды түрде азайту мен мәтіннің оқырман үшін шынайы түсініктілігін арттыру арасындағы тепе-теңдік. Лексикалық немесе синтаксистік түрғыдан жеңілдету әрдайым мәтінді түсінуді жеңілдете бермейді. Керісінше, кейбір жағдайларда мағынаның бұрмалануына немесе мәтіндегі прагматикалық және дискурсивтік байланыстардың әлсіреуіне алып келуі мүмкін. Осыған байланысты қазіргі зерттеулерде мәтіннің мағынасын сақтау, когнитивтік жүктеме және коммуникативтік тиімділік мәселелеріне ерекше назар аударылуда.

Мәтінді автоматты түрде жеңілдету жүйелерін әзірлеу мен бағалауда лингвистикалық талдаудың рөлі де маңызды. Нейрондық модельдердің жоғары нәтижелеріне қарамастан, прагматикалық сәйкестік, мәтіннің тұтастығы және жанрлық ерекшеліктер сияқты аспектілер жиі ескерусіз қалып жатады. Сондықтан лингвист мамандардың қатысуы тек нәтижелерді интерпретациялау кезеңінде ғана емес, сонымен қатар оқыту корпустарын құрастыруда, жеңілдету критерийлерін белгілеуде және бағалау әдістерін әзірлеуде де қажет.

Сонымен қатар мәтінді жеңілдету жүйелерін әртүрлі тілдерге бейімдеу мәселесі ерекше өзектілікке ие. Морфологиясы күрделі, сөз тәртібі еркін немесе келісім жүйесі дамыған тілдер үшін жеңілдету стратегиялары аналитикалық тілдерге қарағанда өзгеше болуы тиіс. Бұл қолданыстағы модельдердің әмбебаптығына күмән туғызады және әр тілдің өзіндік ерекшеліктерін ескеретін зерттеулердің қажеттілігін көрсетеді.

Соңғы жылдары бұл саладағы жарияланымдар саны айтарлықтай артқанымен, зерттеулер әлі де жүйелі түрде біріктірілмеген. Көп жағдайда әртүрлі тәсілдер мен модельдер олардың тілдік алғышарттары мен шектеулерін салыстырмай, жеке-жеке қарастырылады. Нейрондық архитектуралар мен ірі тілдік модельдердің дамуы мәтінді жеңілдетудің жаңа әдістерін ұсынды, алайда олардың бұрынғы ережелік және статистикалық тәсілдермен сабақтастығы жеткілікті деңгейде талданбаған.

Демек тек қолданыстағы зерттеулерді жинақтап қана қоймай, мәтінді автоматты түрде жеңілдету әдістерінің эволюциясын сыни түрғыдан қарастыратын, шешілмеген мәселелерді айқындайтын және типологиялық әрі лингвистикалық аспектілерді ескеретін кешенді шолу жұмыстарына қажеттілік артып отыр.

Материалдар мен әдістер

Бұл жұмыс мәтінді автоматты түрде жеңілдету (ATS) саласында жарияланған зерттеулерді жинақтап, оларды жүйелі түрде талдауға арналған шолу сипатындағы зерттеу болып табылады. Негізгі мақсат – қазіргі қолданылып жүрген әдістер мен ресурстарды назарға ала отырып, автоматты мәтін жеңілдетудің заманауи тәсілдерін реттеп көрсету және оларды сыни түрғыдан бағалау.

Зерттеу әдебиеттерге кешенді талдау жасауға негізделгендіктен, шолу бөлімі мазмұнына қарай үш негізгі тараушаға бөлінді: біріншісінде ATS бағытының қалыптасу кезеңдері мен алғышарттары қарастырылады; екіншісінде қолданыстағы тәсілдердің шектеулері, соның ішінде қазақ тіліне қатысты мәселелер талданады; үшінші тараушада мәтінді жеңілдетудің негізгі әдістері сипатталады.

Дереккөздерді іріктеу барысында *Scopus, Google Scholar* және *ACL Anthology* сияқты жетекші академиялық платформалар пайдаланылды. Бұл ғылыми журналдарда жарияланған еңбектерді ғана емес, сонымен қатар табиғи тілді өңдеу саласындағы ірі халықаралық конференция материалдарын да қамтуға мүмкіндік берді.

Зерттеу ауқымына 1998 жылдан 2025 жылға дейінгі аралықта жарық көрген еңбектер енгізілді. Мұндай хронологиялық қамту ATS саласының даму динамикасын және әр кезеңдегі әдіснамалық өзгерістерді бақылауға жағдай жасайды. Іздеу үдерісі мәтінді автоматты түрде жеңілдетуге және соған байланысты ішкі міндеттерге

қатысты негізгі терминдер негізінде жүргізілді. Шолу шеңберін нақты сақтау үшін материалдар келесі өлшемдер бойынша таңдалды:

- біріншіден, мәтінді автоматты жеңілдетуге тікелей қатысты зерттеулер;
- екіншіден, әдістемелік немесе қолданбалы маңызы бар еңбектер;
- үшіншіден, лексикалық және/немесе синтаксистік жеңілдету тетіктері сипатталған жұмыстар.

Сонымен бірге, бұл шолудың белгілі бір шектеулері бар екенін атап өткен жөн. Ең алдымен, қамтылған әдебиеттер қолжетімді дерекқорлармен шектеледі. Бұған қоса, талданған зерттеулердің басым бөлігі ағылшын тілінде орындалған, бұл өзге тілдердегі тәжірибені толық қамтуға мүмкіндік бермейді. Сондай-ақ, автоматты мәтін жеңілдету саласындағы жарияланымдардың тілдік және тақырыптық әркелкілігі жеткіліксіз болғандықтан, кейбір тілдер мен қолдану салалары бұл зерттеу аясынан тыс қалып отыр.

Таңдап алынған еңбектер бірнеше негізгі параметрлер бойынша салыстырмалы түрде талданды. Атап айтқанда, мәтінді жеңілдету тәсілдері, қолданылатын операция түрлері және ресурстық базаның сипаты назарға алынды. Мұндай талдау ATS саласының даму бағыттарын айқындауға, сондай-ақ қазіргі қолданыстағы әдістердің әлсіз тұстарын, әсіресе ресурсы шектеулі тілдерге бейімдеу мәселелері тұрғысынан, анықтауға мүмкіндік берді.

Әдебиеттерге шолу

Мәтінді автоматты түрде жеңілдету үшін моно-тілдік параллель корпустарға шолу әр түрлі тілдер мен домендерде ATS-ке деген қызығушылықтың тұрақты және жүйелі өсуін көрсетеді. Зерттеулер біртіндеп 2015 жылы әзірленген Newsela (Л. Сюй және басқалары), 2020 ж. ASSET (Ф. Альва-Манчего және басқалары), 2020 ж. WikiAuto (Ч. Цзян және басқалары) сияқты ауқымды және жоғары сапалы ресурстар ағылшын тілінен бастап испан, итальян, француз, жапон, португал, неміс және басқа тілдерге ұласты (Ryan, 2023). Көптеген жағдайларда корпустар мақсатты аудиториямен – тілді шет тілі ретінде оқитын мектеп оқушылары, дислексиясы бар немесе оқуда қиындықтары бар адамдар үшін әзірленді, бұл ATS қолданбалы сипатын айқындады. Қолданылған деректерді жинау стратегиялары сарапшылар мен оқытушылар тарапынан қолмен орындалған жеңілдетуден бастап, краудсорсингке және Уикипедия негізінде автоматты түрде теңестірілген корпустарға дейін қамтиды, бұл ресурстардың сапасы мен ауқымдылығы арасындағы тепе-теңдікті іздеу үдерісін көрсетеді. Бұл жағдайлар жиынтықта ATS-тің NLP бағытындағы тәуелсіз және белсенді дамып келе жатқан бағыты ретінде қалыптасқанын, тілдік құрылымдағы, домендердегі және ресурстарды құру әдістеріндегі айырмашылықтар лингвистикалық және тапсырмаға тән тәсілдердің, әсіресе морфологиялық түрғыдан күрделі және аз ресурсты тілдер үшін қажеттілігін көрсетеді. Ескеретін жайт, орыс тіліндегі зерттеу дәстүрінде мәтінді автоматты түрде жеңілдетуге жақын жұмыстар көбінесе тар мағынада жеңілдету (simplification) емес, мәтінді бейімдеу (адаптация) түрғысынан сипатталады. Дегенмен қолданылатын принциптер мен операциялар ATS тәсілдерімен, соның ішінде лексикалық және синтаксистік жеңілдетумен, сөйлемдерді бөлумен және оқылымды басқарумен сәйкес келеді. *RuWikiLarge*, *RuSimpleSentEval* және *RuAdapt* сияқты корпустардың дамуы ресейлік зерттеушілердің осы тақырыпқа қызығушылығын көрсетеді (Sakhovskiy, 2021; Dmitrieva, 2021). Осылайша, терминологиялық айырмашылықтарға қарамастан, орыс мәтіндерін бейімдеу бойынша зерттеулер мәтінді автоматты түрде жеңілдетудің жалпы парадигмасына енеді.

1. ATS саласының даму тарихы мен алғышарттары

Базалық лексиканы қамтитын сөздік қор және шектеулі грамматикалық ережелерге негізделген, халықаралық қарым-қатынас үшін қолжетімді, қарапайым мәтіндерді қалай жазуға болатындығы туралы алғашқы ұсыныстар 1937 ж Ч.К. Огденнің «Негізгі ағылшын» (ағ. Basic English) және 1987 жылғы Д. Кристалдың «Қарапайым ағылшын бастамасы» (ағ. Plain English initiative) болды. 1990 жылдардың аяғынан бастап (1) интеллектуалдық мүмкіндіктері шектеулі адамдарға мәтіндерді қалай жазу керек (Freyhoff, 1998), (2) жалпы ақпаратты кең аудиторияға және (3) веб-мазмұнды қалай қол жетімді етуге болады және т.б. сияқты зерттеулер түсінікті мәтіндерді қалай жазу керектігі туралы хабардарлықты арттыруға бағытталған көптеген бастамалар пайда болды. Оқуға оңай жаңалықтар мәтіндерді ұсынатын мамандандырылған веб-сайттар қазір көптеген елдерде жиі кездеседі (мысалы, ағылшын тілі үшін Newsela, жапон тілі үшін News Easy Japan және т.б.).

Ағылшын тіліндегі ATS зерттеулерін шартты түрде үш кезеңге бөлуге болады:

1) нақты анықталған түрлендірулерге бағытталған ережелерге негізделген жүйелер (2010 жылға дейін);

2) параллель TS деректерінде оқытылатын бақыланатын машиналық оқыту жүйелері (2010-2014/2016);

3) мәтінді жеңілдететін нейрондық жүйелер (2015/2017 жылдан бастап осы күнге дейін). Бұл кезеңге Deep Learning және трансформациялық модельдер тән.

4) Үлкен тілдік модельдер (LLM) негізінде жеңілдету (қазіргі уақыт).

Осы бағыттың көрнекті өкілдері болып М. Шардлоу, А. Сидхартан, Г.Х. Пэтцольд, Л. Специя, Р. Чандрасекар, Д. Каучак, Р. Эванс, Т. Кадзивара, С. Штайнер, В. Сю, Ф. Альва-Манчего және т.б. саналады.

Мәтінді автоматты түрде жеңілдету (Automatic Text Simplification, ATS) қатаң белгіленген ережелерге негізделген алғашқы жүйелерден қазіргі заманғы нейрондық және трансформаторлық модельдерге дейін және соңғы уақытта үлкен тілдік модельдерге (LLM) негізделген шешімдерге дейін ұзақ жолдан өтті. Бұл даму жолы мәтіндерді әр түрлі оқырмандар тобына, соның ішінде сауаттылығы төмен, когнитивті немесе тілдік қиындықтары бар адамдарға және тілді жетік білмейтіндерге қол жетімді және түсінікті ету секілді практикалық тапсырмаларға байланысты болды (Espinosa-Zaragoza, 2023). Барлық технологиялық өзгерістермен ATS-тің негізгі мақсаты мәтіннің бастапқы мағынасы мен ақпараттық мазмұнын мүмкіндігінше сақтай отырып, оның тілдік күрделілігін төмендету арқылы өзгеріссіз қалады.

1.1 Ережелерге негізделген тәсілдер (Rule-Based Approaches). Алғашқы автоматты жеңілдету жүйелері қолмен жасалған лингвистикалық ережелерге сүйенді. Сарапшылар оларды күрделі сөздер мен синтаксистік құрылымдарды анықтап, оларды қарапайым аналогтармен алмастыратындай етіп тұжырымдады (Hijazi, 2022). Мысалы, ереже сирек кездесетін немесе абстрактілі сөзді жиілік синонимімен ауыстыруды немесе ұзын, шамадан тыс жүктелген сөйлемді бірнеше қысқа және құрылымдық қарапайым сөздерге бөлуі мүмкін. Мұндай жүйелер формальды грамматикалық түрлендірулерге негізделген лексикалық алмастыру мен синтаксистік парафразаны белсенді қолданды.

Алайда, *rule-based* тәсілдері масштабтың нашар болуы, ережелерді әзірлеу мен қолдау үшін айтарлықтай күш-жігерді қажет етуі, тілдің өзгергіштігі мен анық көрсетілмеген түсініксіздігімен күресу қиындығы сияқты шектеулер көптеген кедергілер тудырды.

1.2 Статистикалық машиналық әдістердің пайда болуы *data-driven* тәсілдеріне маңызды қадам болды. Ережелерді қолмен кодтаудың орнына, мұндай жүйелер жеңілдету заңдылықтарын автоматты түрде анықтау үшін күрделі және жеңілдетілген мәтіндердің параллель корпустарын қолдана бастады. Статистикалық машиналық аударма (SMT) әдістері ATS тапсырмаларына бейімделді, мұнда жеңілдету «күрделі» тілден «қарапайым» тілге аударма ретінде қарастырылды (Romero, 2022). Лексикалық жеңілдету корпуста есептелген сөздерді ауыстыру ықтималдығына, ал синтаксистік жеңілдету сөйлемдерді қайта құрудың типтік схемаларына негізделген. Бұл тәсілдер деректерді жақсырақ қорытындылауға және ережелерді қолмен әзірлеуге тәуелділікті азайтуға мүмкіндік берді, бірақ олардың тиімділігі көбінесе шектеулі болатын параллель корпустардың көлемі мен сапасына байланысты болды. Сонымен қатар SMT-ге негізделген тәсілдер кейде жеңілдетілген нәтижеде еркін сөйлеу мен грамматикалық дұрыстықты сақтаумен күресті.

1.3 Deep learning зерттеуінің дамуы мәтінді автоматты түрде жеңілдетудегі бетбұрыс болды, бұл нейрондық желілер мен трансформаторлық модельдердің пайда болуымен айтарлықтай прогреске әкелді (Corti, 2023). Бұрынғы тәсілдерден айырмашылығы, бұл модельдер алдын ала анықталған мүмкіндіктерге немесе статистикалық туралауға сүйенбейді. Олар күрделі тілдік көріністерді деректерден тікелей автоматты түрде алады. Көбінесе назар аудару механизмдерімен (attention mechanism) толықтырылған кодтау және декодтау архитектуралары кеңінен қолданыла бастады. Бұл тәсіл негізінен күрделі бастапқы мәтінді үздіксіз векторлық көрініске айналдырады, содан кейін осы көрініс негізінде жеңілдетілген нұсқа жасалады.

Трансформаторлық модельдер мәтінді автоматты түрде жеңілдетуді дамытуда ерекше маңызды рөл атқарды, мәтін ішіндегі ұзақ мерзімді қатынастарды ескерудің және тізбектерді тиімді өңдеудің жоғары мүмкіндігін көрсетті. Олар қолданатын өзіне назар аудару механизмдері модельге өңдеу кезінде сөйлемдегі әрбір сөздің маңыздылығын анықтауға мүмкіндік береді

(Bakker, 2024). Бұл лексикалық деңгейде дәлірек жеңілдетуге және синтаксистік қарапайымдылықтың тұрақтылығына әкеледі. Нәтижесінде, жеңілдетілген мәтіндері әдетте үйлесімді, еркін және грамматикалық тұрғыдан дұрыс болады. Нейрондық модельдер, әсіресе трансформаторлар, қазіргі уақытта сөйлем деңгейіндегі қарапайымдылықта ең жақсы нәтижелерді көрсетеді. Олар күрделі сөздік қорын сәтті анықтап, ауыстырады, синтаксисті қайта құрылымдайды, сөйлемдерді бөледі және қайта жазуды жүзеге асырады.

ATS-тегі ең соңғы жетістік – кең ауқымды *тілдік модельдерді* пайдалану. Мұндай модельдер, соның ішінде GPT-нің әртүрлі нұсқалары, мәтіндік деректердің үлкен көлемі бойынша алдын ала оқытылады, бұл оларға тілдің семантикасын, синтаксисін және прагматикасын терең түсінуге мүмкіндік береді. Олардың негізгі артықшылығы – мәтінді икемді түрде түсіндіру және жасау, сондай-ақ нұсқаулар немесе дәл баптау арқылы жеңілдету тапсырмасына бейімделу қабілеті (Färber, 2025).

Кең ауқымды тілдік модельдерді қолдану арқылы жеңілдету олардың тіл мен әлем туралы кең біліміне негізделген. Бұл грамматикалық тұрғыдан дұрыс, табиғи болып қалатын және бастапқы мағынаны қажетті контексте дәл жеткізетін мәтіндерді жасауға мүмкіндік береді. Алдыңғы тәсілдерден айырмашылығы, мұндай модельдер тек параллель корпустар мен алдын ала анықталған ережелерге сүйенбейді. Олар лексикалық, синтаксистік және семантикалық жеңілдетуді бір процесте орындай алады және нәтижелерде жоғары тұрақтылықты көрсете алады. Бұл мәтіннің күрделілік деңгейін әртүрлі мақсатты аудиториялардың қажеттіліктеріне икемді түрде реттеуге болатын жекелендірілген жеңілдетуге жол ашады.

Автоматты мәтінді жеңілдету жүйелерін әзірлеу көбінесе мақсатты аудиторияға байланысты. Қарапайымдылық ұғымы әртүрлі болуы мүмкін. Шет тілін үйренушілер үшін жеңілдетілген сөздік қоры мен сөйлем құрылымы маңызды. Когнитивті ерекшелігі бар адамдарға идеялардың айқын көрсетілуі және логикалық тұрғыдан ұйымдастырылған мәтін қажет. Сондықтан, жеңілдету деңгейін нақты пайдаланушыға икемді түрде реттеу мүмкіндігі бар заманауи жүйелер барған сайын дамып келеді.

2. Қолданыстағы тәсілдердің шектеулері және қазақ тілі

Жоғары сапалы деректердің жетіспеушілігі, әсіресе ағылшын тілінен басқа тілдер үшін (біздің жағдайда қазақ тілі), күрделі мәселе болып қала береді. Сенімді модельдерді оқыту барлық тілдер үшін қолжетімді емес күрделі және қарапайым мәтіндердің үлкен жиынтығын қажет етеді. Сонымен қатар әр тілдің өзіндік ерекшеліктері бар. Бұл жеңілдету міндетін күрделендіреді және мамандандырылған шешімдерді қажет етеді. Осыған байланысты зерттеушілер үлкен корпустарға тәуелділікті азайту жолдарын іздестіруде және модельдерді оқытуға әмбебап тәсілдерді қолдануда.

Бұрын жеңілдету негізінен сөйлем деңгейінде қарастырылған. Бұл тәсіл мәтіннің тұтастай логикасы мен үйлесімділігін әрқашан сақтай бермейді. Қазіргі заманғы зерттеулер идеялардың үйлесімділігі мен жалпы түсінікті сақтау үшін құжат бойынша жеңілдетуге қарай жылжуда. Сонымен қатар жеңілдету сапасын бағалау қиын болып қала береді, себебі автоматтандырылған көрсеткіштер әрқашан адамның қабылдауын көрсете бермейді. Сондықтан, пайдаланушылардың бағалаулары мен мамандандырылған критерийлердің рөлі, әсіресе қолданбалы салаларда артып келеді.

Автоматты мәтін жеңілдетуде (ATS) басым болып отырған тәсіл деректерге *(data-driven)* негізделген, яғни модельдер күрделі және жеңілдетілген сөйлемдер жұптары арқылы жеңілдету операциялары үйретіледі (Alva-Manchego, 2020). Бұл қолданылатын ресурстардың қолжетімділігі мен сипаты ATS жүйелерінің сапасын айқындауда шешуші рөл атқарады. Қазіргі таңда мұндай ресурстар жеткілікті деңгейде негізінен ағылшын тілі үшін ғана қолжетімді, себебі осы саладағы зерттеулер тарихи түрғыда көбіне ағылшын тіліне бағытталған. Сонымен қатар көптеген жұмыстар шектеулі бағалау метрикаларына сүйенеді, бұл модельдердің әртүрлі домендерде, тілдерде және түрлі мақсатты аудиториялардың қажеттіліктері түрғысынан қалай жұмыс істейтіндігі туралы сұрақтардың ашық күйінде қалуына әкеледі.

Дәстүрлі түрде сенімділігі мен басқарылу мүмкіндігіне байланысты қолмен жасалған ресурстарға басымдық берілді (Ryan, 2023). Алайда мұндай ресурстарды құрудың жоғары құны мен еңбек сыйымдылығы олардың ауқымын кеңейтуге, домендік қамтуды арттыруға және тілдік әртүрлілікті қамтамасыз етуге елеулі түрде кедергі келтіреді. Осы шектеулерді еңсеру мақсатында зерттеушілер ресурстарды құрудың бақылаусыз әдістеріне жүгінді. Олардың қатарына, ең алдымен, Уикипедия мен Қарапайым Уикипедия (Simple Wikipedia) негізіндегі ке-

лісілген корпустардан сөйлем жүптарын автоматты түрде іріктеу (Kauchak, 2013), сондай-ақ краудсорсингке негізделген тәсілдер жатады (Pellow, 2014).

Сонымен қатар әдебиетте Уикипедияны ATS міндеттері үшін әмбебап ресурс ретінде пайдаланудың дұрыстығына қатысты күмәндар да айтылуда (Xu, 2015). Әсіресе ресурсы шектеулі тілдер үшін өзекті болып отырған параллель деректердің тапшылығы мәселесін шешу мақсатында, перефразалау генерациясына негізделген (Brunato, 2016; Martin, 2022) немесе машиналық аударма тәсілдерін қолданатын параллель корпустарды автоматты түрде құру әдістері ұсынылды (Miliani, 2023). Соңғы жылдары Үлкен тілдік модельдер (LLM) ATS саласына жаңа парадигма енгізіп, синтетикалық ресурстарды жасау мүмкіндігін ашты, алайда мұндай деректердің сапасы әлі де мұқият әрі жүйелі бағалауды қажет етеді (Ryan, 2023).

Агглютинативті және морфологиялық түрғыдан күрделі болып табылатын қазақ тілі ресурсы шектеулі тілдерге тән деректер тапшылығы мәселесін айқын көрсетеді. Осы тұрғыдан алғанда, дәл осындай тілдер үшін автоматты, синтетикалық және LLM-ге негізделген шешімдер ұсынылуда. Дегенмен, LLM-дерді лингвистикалық талдаудың баламасы ретінде емес, құрал ретінде қарастыру қажет. Қазақ тілі үшін мәтінді жеңілдетудің нейрондық жүйелерін тиімді үйрету мақсатында алдымен қолмен аннотацияланған және лингвистикалық тұрғыдан тексерілген жеңілдету моделін әзірлеу қажет, ол бастапқы модель ретінде қызмет ете алады. Тарихи тұрғыда ресурсы мол тілдер үшін мәтінді жеңілдету жүйелерінің дамуы дәл осындай жүйелі лингвистикалық жұмыстардан басталған.

Осылайша, ATS саласындағы ғылыми олқылықтардың бірі – қолданыстағы зерттеулердің басым бөлігі ағылшын тіліне бағытталып, бағалау метрикаларының шектеулі жиынтығына сүйенуі, ал басқа тілдерге, домендерге және мақсатты аудиторияларға қолдану мүмкіндіктерінің жеткілікті деңгейде зерттелмеуі болып табылады. Бұны түрік тілі мәтіндерін жеңілдету бойынша жүргізілген А. Явуздың зерттеуі де дәлелдейді. Түрік тілі – түркі тілдер тобына жататын, агглютинативті және морфологиялық түрғыдан бай тіл, ол мәтінді жеңілдету әдістері бастапқыда жасалған тілдерден, әсіресе ағылшын тілінен, елеулі түрде ерекшеленеді (Yavuz, 2022). Түрік тіліне қатысты зерттеулер де әртүрлі міндеттер мен домендер үшін параллель корпустардың болмауы data-driven және нейрондық тәсілдерді қолдануды айтарлықтай шектейтінін көрсетеді.

Қазақ тілі үшін мәтінді жеңілдету жүйелерін дамытуға қатысты жекелеген алғышарттар аз. Соның бірі – көптілді ірі тілдік модельдерге (LLM) негізделген, нүсқаулыққа бейімделген KazSIM жобасы. Қателерді талдау нәтижелері негізгі қиындықтардың стильдің өзгеруі, шақ формаларының ауысуы және семантикалық ауытқулармен байланысты екенін көрсетеді. Бұл құбылыстар қазақ тілінің агглютинативті морфологиясы мен икемді синтаксисімен тығыз байланысты (Toleu, 2025). Аталған байқаулар қазақ тілі үшін лингвистикалық тұрғыдан негізделген, кешенді мәтін жеңілдету жүйесін әзірлеудің қажеттілігін қосымша дәлелдейді.

3. Мәтінді жеңілдетудің негізгі әдістері

3.1 Лексикалық жеңілдету

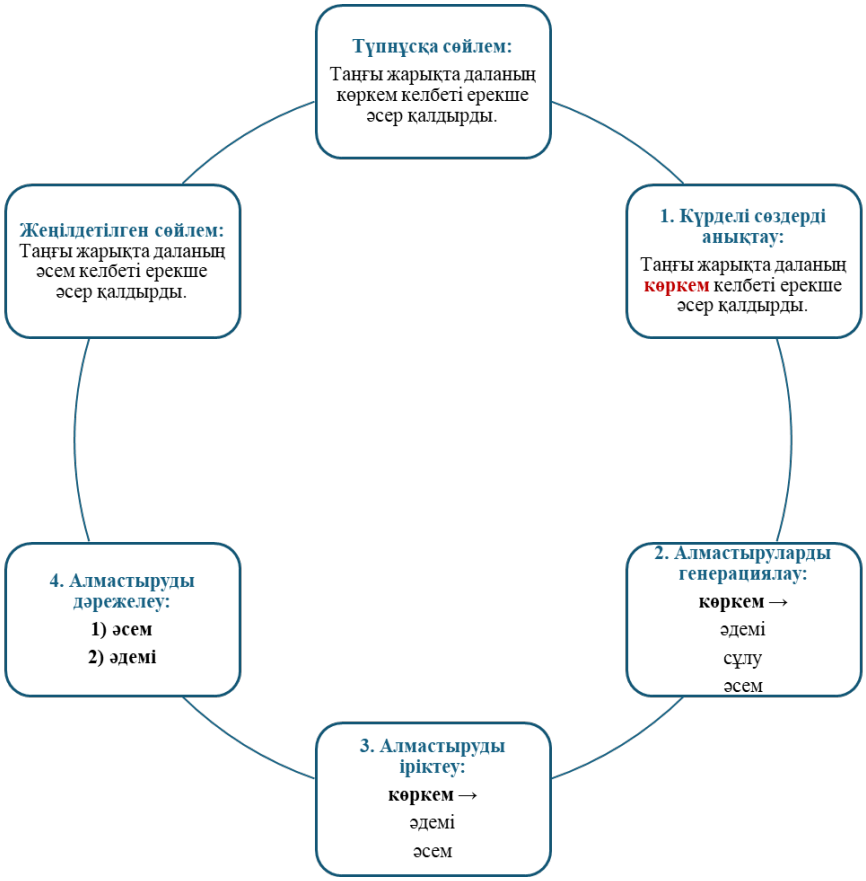
Лексикалық жеңілдету (Lexical simplification) дәстүрлі түрде мәтіндегі күрделі сөздерді қарапайым синонимдері/баламалармен алмастыру міндеті ретінде қарастырылады. Қолданыстағы жүйелердің көпшілігі бұл тапсырманы 1) күрделі сөздерді анықтау (*CWI – Complex Word Identification)*; 2) мүмкін болатын алмастыруларды генерациялау *(SG – Substitution Generation);* 3) алмастыруды іріктеу *(SS – Substitution Selection),* яғни оларды контекстік сүзуден өткізу және 4) алмастыруды дәрежелеу *(SR – Substitution Ranking)* (қарапайымдылық дәрежесі бойынша) тізбеуді қамтитын көп сатылы процесс ретінде жүзеге асырады (Shardlow, 2014a) (1-сызба).

Лексикалық жеңілдетудің алғашқы тәсілдері негізінен WordNet сияқты семантикалық ресурстарға және сөздердің жиілік көрсеткіштеріне, атап айтқанда Кучер-Фрэнсис жиіліктеріне сүйенді, бұл кейінгі зерттеулердің негізін қалады (Biran, 2011). Алайда, соңғы зерттеулерде лексикалық түсініксіздік жағдайында шектеулі корпустарға негізделген жиілік критерийлерін қолдану жеңілдетудің сапасының төмендеуіне әкелетінін көрсетті. Осыған байланысты, кейінгі зерттеулерде тілдік модельдерді, контекстік векторларды және кеңейтілген жиілік ресурстарын қолдана отырып, сөз мағынасының түсініксіздігін жоюға және контекстті есепке алуға баса назар аударылды (Bott, 2012). Дамудың қосымша бағыты фразалық лексикалық жеңілдету болды, бұл мүмкін түрлендірулердің ауқымын кеңейтуге және жеке сөздерді ауыстырумен салыстырғанда мағынаның бұрмалану қаупін азайтуға мүмкіндік берді. Лексикалық жеңілдету саласындағы маңызды кезеңдердің бірі – SemEval

(2012) лексикалық алмастыру тапсырмасы (Specia, 2012), ол синонимдерді қарапайымдылық деңгейі бойынша реттеу мәселесін жеке қарастырып, әртүрлі тәсілдерді салыстыруға мүмкіндік беретін стандартталған бағалау негізін ұсынды.

1-сызба

Лексикалық жеңілдету тізбегі



Ескерту: Сызба мақала авторлары тарапынан құрастырылды

Лексикалық жеңілдету (LS) зерттеулері шартты түрде екі негізгі бағытқа бөлінеді: 1) ережелерге негізделген тәсілдер және 2) деректерге негізделген тәсілдер. Тарихи түрғыда ережелерге негізделген әдістер бірінші дамыды, үлкен параллель корпустар жоқ тілдер үшін қолданыла берді және ең алдымен, деректерге негізделген модельдерді оқыту үшін қажет болды.

Алғашқы лексикалық жеңілдету жүйелерінің бірі 1998 жылы ұсынылған афазиясы бар оқырмандар үшін ағылшын газеттерінің мәтіндерін жеңілдетуге бағытталған (Carroll, 1998). Жүйенің архитектурасына екі негізгі модуль кірді – *анализатор* және *жеңілдеткіш:* бірінші мәтінге лингвистикалық талдау жасады, ал екіншісі алынған құрылымдарды қабылдауға жеңіл түрге айналдырды. Анализатор лексикалық белгілеу, морфологиялық және синтаксистік талдау компоненттерінен түрды. Осы кезең аяқталғаннан кейін лексикалық және синтаксистік қосалқы компоненттерді қамтитын жеңілдету модулі қолданылды. Лексикалық компонент әр сөз үшін WordNet көмегімен синонимдер жиынтығын жасады және Оксфорд психолингвистикалық базасының деректері негізінде ең жиі нұсқаны таңдады, содан кейін ол жеңілдетілген мәтінде де қолданылды. Болашақта ережелерге негізделген лексикалық жеңілдету тәсілдерді бірнеше ішкі санаттарға бөлуге болады.

3.2 Синтаксистік жеңілдету

Синтаксистік жеңілдету (Syntactic simplification) мәтіннің бастапқы ақпараттық мазмұны мен жалпы мағынасын сақтай отырып, күрделі синтаксистік құрылымдарды түрлендіруге бағытталған. Синтаксистік түрғыдан күрделі құрылымдардың қатарына әдетте пассивті құрылымдарды қоса алғанда, бағыныңқы байланыстар, күрделі сөйлемдер жатады. Көптеген зерттеулер синтаксистік жеңілдетуді үш негізгі кезеңді қамтитын көп сатылы процесс ретінде қарастырады (Shardlow, 2014b).

Бірінші кезеңде мәтіннің синтаксистік құрылымын анықтау және талдау ағашын құру мақсатында анализ жасалады. Осы кезеңде сөздер мен сөз тіркестері негізгі сөйлемнің үзінділерін білдіретін «супертегтерге» топтастырылған. Құрылымдық презентация негізінде сөйлемнің синтаксистік күрделілік дәрежесі анықталады және оны жеңілдету қажеттілігі туралы шешім қабылданады. Бұл кезең салыстыру ережелерінің көмегімен де, екілік жіктеуіштерді, мысалы, SVM (Syntactic Vector Machine) көмегімен де жүзеге асырылуы мүмкін.

Синтаксистік ағашқа түрлендіру кезеңінде күрделі сөйлемдерді бөлу, элементтерді қайта құру немесе жеке құрылымдарды алып тастау сияқты жеңілдету операцияларын жүзеге асыратын қайта жазу ережелерінің жиынтығына сәйкес өзгертулер енгізіледі. Соңғы кезең нәтиженің үйлесімділігін, өзектілігін және жалпы оқылуын арттыруға бағытталған мәтінді қосымша өңдеуді қамтиды.

Синтаксистік жеңілдетуді де зерттеуде екі негізгі тәсіл ерекшеленеді: ережелерге негізделген және деректерге негізделген, қолданыстағы жүйелердің көпшілігі ережелерге сүйенеді. Мұндай жүйелердің тиімділігі көбінесе лингвистикалық сипаттаманың тереңдігіне және талдау құралдарының, соның ішінде талдаушылар мен теггерлердің сапасына байланысты.

Синтаксистік жеңілдетудің алғашқы rule-based тәсілдерінің бірі 1996 жылы ұсынылды және сөйлемдерді жеңілдетуді талдау үшін алдын-ала өңдеу кезеңі ретінде қарастырды (Chandrasekar, 1996). Авторлар жеңілдету процесін сөйлем құрылымын талдаудың дәйектілігі және кейіннен түрлендіру ережелерін қолдану деп сипаттады. Негізгі назар салыстырмалы сөйлемдер, аппозитивтер және салалас және сабақтас сияқты құрылымдарға аударылды. Ережелердің қолмен түжырымдалуы салыстырмалы құрылымдарды бірнеше қарапайым сөйлемдерге айналдыруға мүмкіндік береді, осылайша мәтіннің синтаксистік күрделілігі төмендейді.

К. Вудсенд және М. Лапата түпнұсқа және жеңілдетілген сөйлемдердің синтаксистік ағаштары арасында сәйкестік орнатуға мүмкіндік беретін квази-синхронды грамматикаға негізделген мәтінді автоматты түрде жеңілдету моделін ұсынды. Модель параллель корпуста оқытылады, мұнда сөйлемдердің екі нұсқасы да алдынала талдаудан өтеді және жеңілдету ережелерін автоматты түрде шығарады (Woodsend, 2011). Жүйе шеңберінде ережелердің үш түрі қалыптасады: синтаксистік жеңілдету, лексикалық жеңілдету және сөйлемдерді бөлу. Бөлу ережелері күрделі конструкцияларға, соның ішінде құрастырылған және бағыныңқы сөйлемдерге, салыстырмалы конструкцияларға және қосылатын элементтерге қолданылады. Бірнеше ықтимал жеңілдетулер болған кезде грамматиканы, сөйлем ұзындығын және оқылымды ескере отырып, ең тиімді нұсқаны таңдау үшін бүтін сызықтық бағдарламалау қолданылады. Эксперименттік нәтижелер ұсынылған модель қарапайымдылық, дұрыстық және мағынаны сақтау көрсеткіштері бойынша негізгі жүйелерден асып түсетінін көрсетті.

Нәтижелер мен талқылау

Автоматты лексикалық жеңілдету саласындағы зерттеулерге жүргізілген шолу әдістердің дамуындағы елеулі прогреске қарамастан, осы салада бірқатар тұрақты проблемалар сақталатынын көрсетеді. Әдебиеттерді талдау LS жүйелерінің көпшілігінің негізгі шектеулері күрделі сөздерді анықтау кезеңімен және мүмкін болатын ауыстыруларды таңдаумен байланысты екенін көрсетеді. Бұл әсіресе жиілік пен корпус критерийлерін анықтау кезінде көрінеді. Контекстті есепке алу және лексикалық түсініксіздікті жою әдістері жеңілдетудің сапасын ішінара жақсартуға мүмкіндік бергенімен, олар мәселені толығымен шешпейді және жиі ауыстырудан бас тартуға әкеледі. Бұл классикалық rule-based және data-driven тәсілдері белгілі бір тиімділік шегіне жететінін және неғұрлым икемді, белгілі бір тілге бағытталған талдауды қажет ететінін көрсетеді.

Күрделі сөздерді анықтауға арналған зерттеулерді талдау (Complex Word Identification) бұл кезең лексикалық жеңілдетудің ең маңызды және осал компоненттерінің бірі екенін көрсетеді. CWI-дің болмауы немесе оны *бәрін жеңілде-*

my (simplify everything) стратегиясы немесе қатаң жиілік шектері сияқты жеңілдетілген критерийлер негізінде жүзеге асыру жүйелі түрде қажетсіз ауыстыруларға, мағынаның бұрмалануына және мәтіннің грамматикалық байланысының бұзылуына әкеледі. Шекті және лексикаға бағытталған тәсілдер жеке домендерде белгілі бір тиімділікті көрсетсе де, олардың қолданылуы ресурстарға және мақсатты аудиторияға тәуелділікпен шектеледі. Заманауи зерттеулер жиілік белгілерін, лексикалық ресурстарды және машиналық оқыту әдістерін біріктіретін гибридті шешімдерді қолдану арқылы ең жақсы нәтижелерге қол жеткізілетінін көрсетеді. Қазақ тілі сияқты ресурстары шектеулі тілдер үшін бұл тұжырымдар өзекті, өйткені мамандандырылған корпустар мен лексикондардың болмауы қолданыстағы CWI стратегияларын тікелей қарызға алуды қиындатады. Бұл мәтінді автоматты түрде жеңілдетудің келесі кезеңдері үшін тұрақты негіз бола алатын CWI-ге лингвистикалық және лингвистикалық негізделген тәсілдерді әзірлеу қажеттілігін көрсетеді.

Алмастыруды генерациялау кезеңіне (Substitution Generation) арналған зерттеулерді талдау лексикалық жеңілдетудің бұл компоненті толықтық пен дәлдік арасындағы қатынасқа сөзсіз байланысты екенін көрсетеді. Бір жағынан, тиімді SG мүмкін болатын жеңілдетулердің кең ауқымын қамтамасыз етуі керек, бірақ іс жүзінде бұл көбінесе маңызды емес кандидаттардың көп санының пайда болуына әкеліп соғады, бұл кейінгі іріктеу кезеңдерін қиындатады. Тезаурустар мен мамандандырылған мәліметтер базасы сияқты лингвистикалық ресурстар сенімді ауыстыруды қамтамасыз етеді, бірақ олардың шектеулі қамтылуы және жоғары құру құны олардың әмбебаптығын төмендетеді. Параллель корпустар мен сөздердің таратылған көріністеріне негізделген автоматты тәсілдер генерация процесін масштабтауға мүмкіндік береді, бірақ лексикалық түсініксіздік пен шу мәселесін күшейтеді. Заманауи зерттеулер лингвистикалық ресурстарды автоматты генерациямен және контекстік сүзумен біріктіретін гибридті шешімдердің артықшылығын көрсетеді. Қазақ тілі үшін бұл тұжырымдар ерекше маңызға ие, өйткені параллель корпустар мен лексикалық ресурстардың шектелуі бақыланбайтын алмастырулар генерациясын аса қауіпті етеді және тілге тән және лингвистикалық түрғыдан негізделген SG стратегияларының қажеттілігін атап көрсетеді.

Алмастыруды іріктеу кезеңіне арналған зерттеулерді талдау (Substitution Selection) оның жеңілдетілген мәтіннің мағынасы мен грамматикалық дұрыстығын сақтаудағы шешуші рөлін растайды. SS стратегиясының болмауы айтарлықтай семантикалық бұрмалануларға әкеледі, бұл осы кезеңді LS жүйелерінің сапасы үшін маңызды етеді. Қолданыстағы тәсілдер сөздікке бағытталған дизамбигациядан автоматты статистикалық және нейрондық әдістерге дейін, алайда біріншісі ресурстармен шектеледі, ал екіншісі күрделі орнатуды қажет етеді. *POS (Part-of-Speech)*, яғни *сөз таптары* тегтері немесе семантикалық жақындық сияқты жеке белгілерді пайдалану контекстке тәуелді сөздерді өңдеу үшін жеткіліксіз болып шығады. Ресурстары шектеулі тілдер, соның ішінде қазақ тілі үшін бұл біріктірілген және лингвистикалық негізделген SS стратегияларды талап етеді.

Алмастыруды дәрежелеу немесе *бағалау* кезеңіне (Substitution Ranking) бағытталған зерттеулерді талдау дәл осы кезең жеңілдетілген мәтіннің қарапайымдылығы мен сәйкестігінің соңғы деңгейін анықтайтынын көрсетеді. Жиілік көрсеткіштері, олардың таралуына қарамастан, шектеулі болып шығады, өйткені олар сөздерді қолдану контекстін ескермейді. Неғұрлым тұрақты нәтижелер қарапайымдылық жиілік, лексикалық және семантикалық сипаттамаларды қоса алғанда, бірнеше белгілердің тіркесімі ретінде модельденетін тәсілдерді көрсетеді. Машиналық оқыту әдістері рейтинг сапасын одан әрі жақсартуға мүмкіндік береді, бірақ олардың тиімділігі оқу деректерінің көлемі мен сапасына тікелей байланысты. Қазақ тілі сияқты ресурстары шектеулі тілдер үшін бұл тек жиілікке ғана емес, сонымен қатар тілге тән ерекшеліктерге бағытталған SR-стратегияларды мұқият таңдау және бейімдеудің маңыздылығын көрсетеді.

Бұл түрғыда ресурстары шектеулі және типологиялық құрылымы күрделі тілдерге қолданыстағы тәсілдерді қолдану мәселесі аса маңызды. Жоғары морфологиялық вариациямен және икемді синтаксиспен сипатталатын агглютинативті тілдер үшін негізінен ағылшын тіліне арналған әдістерді тікелей беру қиынға соғады. Қазақ тілі – ірі параллель корпустар мен мамандандырылған лексикалық ресурстардың болмауы қолданыстағы LS-жүйелердің шектеулерін күшейтетін жағдайдың дәлелі болып табылады. Осы шолудың мақсаты – мәтінді автоматты түрде оңайлатуды дамытуды талдау негізінде қазір-

гі заманғы тәсілдердің негізгі әдіснамалық және ресурстық шектеулерін анықтау және қазақ тілі үшін мәтінді жеңілдету жүйелерін әзірлеу үшін неғұрлым өзекті бағыттарды белгілеу.

TS жүйелерінің көпшілігі сөйлем деңгейінде жеңілдету мен бағалау жүргізу үшін адам сарапшыларын пайдаланады. Бұл әдетте үш аспектке негізделеді: *(1) қарапайымдылық, (2) грамматикалық дұрыстық (немесе тілдік бірізділік)* және *(3) мағынаның сақталуы.* Қарапайымдылық жеңілдетілген сөйлемнің қаншалықты қарапайым екенін, ал грамматика жеңілдетілген сөйлемнің грамматикалық дұрыстығын анықтайды. Мәнді сақтау жеңілдетілгеннен кейін бастапқы мағынаның қаншалықты жақсы сақталатынын анықтайды.

Қолмен бағалаудың да бірқатар шектеулері бар. Шығыс сөйлемінің (жеңілдетілген сөйлем нұсқасы) үш аспектісін бағалау үшін тілді білу қажет. Сонымен қатар адамдардың «қарапайым» деңгей жайлы түсініктері сәйкес келмейді және бір-бірінен ерекшеленеді. Бұл әсіресе бірнеше адамдар қатысқан кезде TS жүйелерін салыстыруды дәл емес етеді. Мәтінді қолмен жеңілдету баяу және қымбат болумен қатар (білікті редакторларды қажет етеді), интернетте жарияланған жаңа ақпаратты үнемі жаңартып отыруға, әртүрлі жазбаша мазмұнды үсынуға немесе оны жеке деңгейде бейімдеуге мүмкіндік бермейді. Бұл автоматты немесе кем дегенде жартылай автоматты (қолмен өңдеуден кейінгі кезеңмен) TS жүйелеріне қажеттілік туғызды. Қазақ тілі үшін де осы тәсіл тиімді, себебі қазақ тілінде материалдардың автоматты алгоритм немесе дерекке негізделген (data-driven) жеңілдету жүйелерін құрастыру үшін, алдымен осы жүйелерді дұрыс оқытуға қажет лингвист мамандар қолмен, сауатты, тілдің ережелеріне сай қазақ тіліндегі мәтіндерді автоматты жеңілдетуге негіз болатын лингвистикалық моделін анықтау керек.

Қорытынды

Мәтінді автоматты жеңілдету саласындағы айтарлықтай жетістіктерге қарамастан, бұл – әлі де әмбебап және толығымен сенімді шешімнен алыс, қосымша зерттеулер жүргізу, бағалаудың тұрақты әдістерін әзірлеу және сапалы жағдайларды жасау қажеттілігін көрсететін тақырып. Э. Дюпу атап өткендей, қазіргі заманғы нейрондық модельдердің көптеген жетістіктері көбінесе лингвистикалық теорияға нақты сүйенбестен үлкен көлемдегі деректерді ауқымды оқыту арқылы қол жеткізіледі (Dupoux, 2018). Алайда, қазақ тілі сияқты инфокеңістікте ресурстары шектеулі тілдер үшін бұл тәсіл жеткіліксіз, өйткені деректердің қарапайым өсуі тілдің типологиялық және морфологиялық күрделілігін өтемейді. Бұл түрғыда мәтінді жеңілдетудің негізгі принциптерін құрылымдауға және кейінгі автоматты шешімдер үшін сенімді негіз орнатуға қабілетті қолмен және әдіснамалық түрғыдан дұрыс құрылған лингвистикалық модель ерекше мәнге ие болады. Қазақ тілі үшін мүндай модель түрлендірулердің бақылануын қамтамасыз ететін және семантикалық және грамматикалық бұрмалану қаупін төмендететін автоматты жеңілдету жүйелерін қалыптастырудың бастапқы нүктесі бола алады. Осылайша, қазақ тілі үшін ATS-ті дамыту лингвистикалық түрғыдан негізделген база нейрондық және деректерге негізделген (data-driven) тәсілдерді қолданудан бұрын қалыптасып, оларды толықтыратын кезең-кезеңімен жүзеге асатын үдеріс ретінде қарастырылғаны орынды.

Сонымен қатар TS-ті бағалаудың лингвистикалық қыры да қосымша қиындықтар туындатады, атап айтқанда, жеңілдету деңгейін өлшеудегі дәлсіздіктер мен тілдің субъективті сипаты. Дегенмен, бұл салаға деген ғылыми қызығушылықтың бұрын-соңды болмаған деңгейге жетуін ескере отырып, алдағы уақытта түрақты ілгерілеу күтіледі. Осы түрғыдан алғанда, бұл мақала TS саласымен танысқысы келетін зерттеушілер үшін негіз қалаушы шолу бола алады деп үміттенеміз.

Мүдделер қақтығысы

Авторлар мүдделер қақтығысының жоқ екенін мәлімдейді.

Мақала авторларының үлесі

А.А. Карымхан зерттеудің ғылыми-теориялық аппаратын, әдіснамалық негізін қалыптастырды және әдістер, нәтижелер мен талқылау бөлімдеріне үлес қосты. М.Қ. Мәмбетова зерттеу идеясын бастапқыдан соңғы кезеңге дейін қадағалады, мақаланың жалпы тұжырымдамасын әзірлеуге және жүзеге асыруға жауапты болды. Б. Нұрланғазықызы шетелдік зерттеулерді сұрыптады, салыстырмалы анализ арқылы нәтижелерді талдауды жүргізді және талқылау бөліміне үлес қосты. Барлық авторлар қолжазбаның соңғы нұсқасын қарап шықты, мақұлдады және жариялауға келісім берді.

Әдебиеттер

Paetzold, G., & Specia, L. (2016). Unsupervised lexical simplification for non-native speakers. *Proceedings of the AAAI Conference on Artificial Intelligence, 30*(1). https://doi.org/10.1609/aaai.v30i1.9885

Chandrasekar, R., Doran, C., & Srinivas, B. (1996). Motivations and methods for text simplification. In *Proceedings of the 16th Conference on Computational Linguistics (COLING'96)* (pp. 1041–1044). https://aclanthology.org/C96-2183/

Heilman, M., & Smith, N. A. (2010). Good question! Statistical ranking for question generation. In *Proceedings of the NAACL-HLT Conference* (pp. 609–617). https://aclanthology.org/N10-1086/

Evans, R. (2011). Comparing methods for the syntactic simplification of sentences in information extraction. *Literary and Linguistic Computing, 26*(4), 371–388. https://doi.org/10.1093/llc/fqr034

Beigman Klebanov, B., Knight, K., & Marcu, D. (2004). Text simplification for information-seeking applications. In *Proceedings of the OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 735–747). https://www.sciweavers.org/publications/text-simplification-information-seeking-applications

Vickrey, D., & Koller, D. (2008). Sentence simplification for semantic role labeling. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics* (pp. 344–352). https://aclanthology.org/P08-1040/

Ryan, M. J., Naous, T., & Xu, W. (2023). Revisiting non-English text simplification: A unified multilingual benchmark. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 4898–4927). https://doi.org/10.18653/v1/2023.acl-long.269

Sakhovskiy, A., Izhevskaya, A., Pestova, A., Tutubalina, E., Malykh, V., Smurov, I., & Artemova, E. (2021). RuSimpleSentEval-2021 shared task: Evaluating sentence simplification for Russian. In *Proceedings of the International Conference "Dialogue"* (pp. 607–617). https://doi.org/10.28995/2075-7182-2021-20-607-617

Dmitrieva, A., & Tiedemann, J. (2021). Creating an aligned Russian text simplification dataset from language learner data. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing* (pp. 73–79). Association for Computational Linguistics. https://aclanthology.org/2021.bsnlp-1.8/

Freyhoff, G., Hess, G., Kerr, L., Tronbacke, B., & Van der Veken, K. (1998). *Make it simple: European guidelines for the production of easy-to-read information for people with learning disability*. ILSMH European Association.

Espinosa-Zaragoza, I., Abreu-Salas, J., Lloret, E., Moreda, P., & Palomar, M. (2023). A review of research-based automatic text simplification tools. In *Proceedings of Recent Advances in Natural Language Processing – Large Language Models* (pp. 321–330). https://doi.org/10.26615/978-954-452-092-2_036

Hijazi, R., Espinasse, B., & Gala, N. (2022). GRASS: A syntactic text simplification system based on semantic representations. In *Proceedings of the 11th International Conference on Natural Language Processing*.

Romero, M., Calderón-Ramírez, S., Solís, M., Pérez-Rojas, N., Chacón-Rivas, M., & Saggion, H. (2022). Towards text simplification in Spanish: A brief overview of deep learning approaches. In *Proceedings of the IEEE 4th International Conference on BioInspired Processing* (pp. 1–7).

Corti, L., & Yang, J. (2023). ARTIST: ARTificial intelligence for simplified text [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2308.13458

Bakker, J., & Kamps, J. (2024). Beyond sentence-level text simplification: Reproducibility study of context-aware document simplification. In *Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context @ LREC-COLING 2024* (pp. 27–38). ELRA and ICCL. https://aclanthology.org/2024.determit-1.3/

Färber, M., Aghdam, P., Im, K., Tawfelis, M., & Ghoshal, H. (2025). SimplifyMyText: An LLM-based system for inclusive plain language text simplification [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2504.14223

Alva-Manchego, F., Scarton, C., & Specia, L. (2020). Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics, 46*(1), 135–187. https://doi.org/10.1162/coli_a_00370

Kauchak, D. (2013). Improving text simplification language modeling using unsimplified text data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1537–1546). https://aclanthology.org/P13-1151/

Pellow, D., & Eskenazi, M. (2014). An open corpus of everyday documents for simplification tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 84–93). https://doi.org/10.3115/v1/W14-1210

Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics, 3*, 283–297. https://doi.org/10.1162/tacl_a_00139

Brunato, D., Cimino, A., Dell'Orletta, F., & Venturi, G. (2016). PaCCSS-IT: A parallel corpus of complex-simple sentences for automatic text simplification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 351–361). https://doi.org/10.18653/v1/D16-1034

Martin, L., Fan, A., de la Clergerie, É., Bordes, A., & Sagot, B. (2022). MUSS: Multilingual unsupervised sentence simplification by mining paraphrases. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 1651–1664). https://doi.org/10.48550/arXiv.2005.00352

Miliani, M., Alva-Manchego, F., & Lenci, A. (2023). Simplifying administrative texts for Italian L2 readers with controllable transformer models. In *Proceedings of CLiC-it 2023* (pp. 303–315). https://aclanthology.org/2023.clicit-1.37/

Yavuz, A. (2022). Automatic lexical simplification for Turkish [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2201.05878

Toleu, A., Tolegen, G., & Ualiyeva, I. (2025). Fine-tuning large language models for Kazakh text simplification. *Applied Sciences, 15*(15), 8344. https://doi.org/10.3390/app15158344

Shardlow, M. (2014a). A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications, 4*(1), 58–70. https://doi.org/10.14569/SpecialIssue.2014.040109

Biran, O., Brody, S., & Elhadad, N. (2011). Putting it simply: A context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Short Papers)* (pp. 496–501).

Bott, S., Rello, L., Drndarević, B., & Saggion, H. (2012). Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In *Proceedings of COLING 2012* (pp. 357–374). https://aclanthology.org/C12-1023/

Specia, L., Jauhar, S. K., & Mihalcea, R. (2012). SemEval-2012 task 1: English lexical simplification. In *Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 347–355). https://aclanthology.org/S12-1046/

Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998). Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI'98 Workshop on Integrating Artificial Intelligence and Assistive Technology* (pp. 7–10). https://www.researchgate.net/publication/2740075_Practical_Simplification_of_English_Newspaper_Text_to_Assist_Aphasic_Readers

Shardlow, M. (2014b). Out in the open: Finding and categorising errors in the lexical simplification pipeline. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 1583–1590). http://www.lrec-conf.org/proceedings/lrec2014/pdf/479_Paper.pdf

Woodsend, K., & Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 409–420). https://aclanthology.org/D11-1038/

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition, 173*, 43–59. https://doi.org/10.48550/arXiv.1607.08723

References

Alva-Manchego, F., Scarton, C., & Specia, L. (2020). Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics, 46*(1), 135–187. https://doi.org/10.1162/coli_a_00370

Bakker, J., & Kamps, J. (2024). Beyond sentence-level text simplification: Reproducibility study of context-aware document simplification. In *Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context @ LREC-COLING 2024* (pp. 27–38). ELRA and ICCL. https://aclanthology.org/2024.determit-1.3/

Beigman Klebanov, B., Knight, K., & Marcu, D. (2004). Text simplification for information-seeking applications. In *Proceedings of the OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 735–747). https://www.sciweavers.org/publications/text-simplification-information-seeking-applications

Biran, O., Brody, S., & Elhadad, N. (2011). Putting it simply: A context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Short Papers)* (pp. 496–501).

Bott, S., Rello, L., Drndarević, B., & Saggion, H. (2012). Can Spanish be simpler? LexSiS: Lexical simplification for Spanish. In *Proceedings of COLING 2012* (pp. 357–374). https://aclanthology.org/C12-1023/

Brunato, D., Cimino, A., Dell'Orletta, F., & Venturi, G. (2016). PaCCSS-IT: A parallel corpus of complex-simple sentences for automatic text simplification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 351–361). https://doi.org/10.18653/v1/D16-1034

Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998). Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI'98 Workshop on Integrating Artificial Intelligence and Assistive Technology* (pp. 7–10). https://www.researchgate.net/publication/2740075_Practical_Simplification_of_English_Newspaper_Text_to_Assist_Aphasic_Readers

Chandrasekar, R., Doran, C., & Srinivas, B. (1996). Motivations and methods for text simplification. In *Proceedings of the 16th Conference on Computational Linguistics (COLING'96)* (pp. 1041–1044). https://aclanthology.org/C96-2183/

Corti, L., & Yang, J. (2023). ARTIST: ARTificial intelligence for simplified text [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2308.13458

Dmitrieva, A., & Tiedemann, J. (2021). Creating an aligned Russian text simplification dataset from language learner data. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing* (pp. 73–79). Association for Computational Linguistics. https://aclanthology.org/2021.bsnlp-1.8/

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition, 173*, 43–59. https://doi.org/10.48550/arXiv.1607.08723

Espinosa-Zaragoza, I., Abreu-Salas, J., Lloret, E., Moreda, P., & Palomar, M. (2023). A review of research-based automatic text simplification tools. In *Proceedings of Recent Advances in Natural Language Processing – Large Language Models* (pp. 321–330). https://doi.org/10.26615/978-954-452-092-2_036

Evans, R. (2011). Comparing methods for the syntactic simplification of sentences in information extraction. *Literary and Linguistic Computing, 26*(4), 371–388. https://doi.org/10.1093/llc/fqr034

Färber, M., Aghdam, P., Im, K., Tawfelis, M., & Ghoshal, H. (2025). SimplifyMyText: An LLM-based system for inclusive plain language text simplification [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2504.14223

Freyhoff, G., Hess, G., Kerr, L., Tronbacke, B., & Van der Veken, K. (1998). *Make it simple: European guidelines for the production of easy-to-read information for people with learning disability*. ILSMH European Association.

Heilman, M., & Smith, N. A. (2010). Good question! Statistical ranking for question generation. In *Proceedings of the NAACL-HLT Conference* (pp. 609–617). https://aclanthology.org/N10-1086/

Hijazi, R., Espinasse, B., & Gala, N. (2022). GRASS: A syntactic text simplification system based on semantic representations. In *Proceedings of the 11th International Conference on Natural Language Processing*.

Kauchak, D. (2013). Improving text simplification language modeling using unsimplified text data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1537–1546). https://aclanthology.org/P13-1151/

Martin, L., Fan, A., de la Clergerie, É., Bordes, A., & Sagot, B. (2022). MUSS: Multilingual unsupervised sentence simplification by mining paraphrases. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 1651–1664). https://doi.org/10.48550/arXiv.2005.00352

Miliani, M., Alva-Manchego, F., & Lenci, A. (2023). Simplifying administrative texts for Italian L2 readers with controllable transformer models. In *Proceedings of CLiC-it 2023* (pp. 303–315). https://aclanthology.org/2023.clicit-1.37/

Paetzold, G., & Specia, L. (2016). Unsupervised lexical simplification for non-native speakers. *Proceedings of the AAAI Conference on Artificial Intelligence, 30*(1). https://doi.org/10.1609/aaai.v30i1.9885

Pellow, D., & Eskenazi, M. (2014). An open corpus of everyday documents for simplification tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 84–93). https://doi.org/10.3115/v1/W14-1210

Romero, M., Calderón-Ramírez, S., Solís, M., Pérez-Rojas, N., Chacón-Rivas, M., & Saggion, H. (2022). Towards text simplification in Spanish: A brief overview of deep learning approaches. In *Proceedings of the IEEE 4th International Conference on BioInspired Processing* (pp. 1–7).

Ryan, M. J., Naous, T., & Xu, W. (2023). Revisiting non-English text simplification: A unified multilingual benchmark. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 4898–4927). https://doi.org/10.18653/v1/2023.acl-long.269

Sakhovskiy, A., Izhevskaya, A., Pestova, A., Tutubalina, E., Malykh, V., Smurov, I., & Artemova, E. (2021). RuSimpleSentEval-2021 shared task: Evaluating sentence simplification for Russian. In *Proceedings of the International Conference "Dialogue"* (pp. 607–617). https://doi.org/10.28995/2075-7182-2021-20-607-617

Shardlow, M. (2014a). A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications, 4*(1), 58–70. https://doi.org/10.14569/SpecialIssue.2014.040109

Shardlow, M. (2014b). Out in the open: Finding and categorising errors in the lexical simplification pipeline. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 1583–1590). http://www.lrec-conf.org/proceedings/lrec2014/pdf/479_Paper.pdf

Specia, L., Jauhar, S. K., & Mihalcea, R. (2012). SemEval-2012 task 1: English lexical simplification. In *Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 347–355). https://aclanthology.org/S12-1046/

Toleu, A., Tolegen, G., & Ualiyeva, I. (2025). Fine-tuning large language models for Kazakh text simplification. *Applied Sciences, 15*(15), 8344. https://doi.org/10.3390/app15158344

Vickrey, D., & Koller, D. (2008). Sentence simplification for semantic role labeling. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics* (pp. 344–352). https://aclanthology.org/P08-1040/

Woodsend, K., & Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 409–420). https://aclanthology.org/D11-1038/

Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics, 3*, 283–297. https://doi.org/10.1162/tacl_a_00139

Yavuz, A. (2022). Automatic lexical simplification for Turkish [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2201.05878

Авторлар туралы мәлімет:

Карымхан Ақмарал Адилқызы – PhD докторант, Әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, e-mail: k.akmaral2309@gmail.com).

Мәмбетова Мәншүк Құдайбергенқызы – филология ғылымдарының кандидаты, қауымдастырылған профессор, Әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, e-mail: mmanshuk@gmail.com).

Нұрлангазықызы Балнұр – оқытушы, Әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, e-mail: bbaitileuova@gmail.com).

Information about the authors:

Karymkhan Akmaral Adilkyzy – PhD Student, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: k.akmaral2309@gmail.com).

Mambetova Manshuk Kudaibergenovna – Candidate of Philological Sciences, Associate Professor, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: mmanshuk@gmail.com).

Nurlangazykyzy Balnur – lecturer, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: bbaitileuova@gmail.com).

Сведения об авторах:

Карымхан Акмарал Адилкызы – PhD докторант, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, e-mail: k.akmaral2309@gmail.com).

Мамбетова Маншук Кудайбергеновна – кандидат филологических наук, ассоциированный профессор, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, e-mail: mmanshuk@gmail.com).

Нурлангазыкызы Балнур – преподаватель, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, e-mail: bbaitileuova@gmail.com).

Келіп түсті: 1 сәуір 2026 жыл
Қабылданды: 3 маусым 2026 жыл