МУҲАНДИСЛИК КОРПУСИНИ ШАКЛЛАНТИРИШ МЕЗОНЛАРИ

Гўзал Хакимовна Нурматова
Навоий давлат кончилик институти доценти
g.nurmatova77@gmail.com

АННОТАЦИЯ

Мақолада муаллиф муҳандислик терминларни корпусдаги лингвистик таҳлили учун муҳандислик корпусини шакллантириш мезонлари тўғрисида ғарб ва шарқ тилшуносларнинг назарий фикрлари асосида муҳандислик корпусини шакллантириш борасида амалий ечимларини таклиф қилади.

Калит сўзлар: корпус, муҳандислик, репрезентативлик, корпус баланси.

CRITERIA FOR FORMATION OF ENGINEERING CORPUS

Guzal Khakimovna Nurmatova
Associate Professor of Navoi State Mining Institute
g.nurmatova77@gmail.com

ABSTRACT

In this article, the author proposes practical solutions to design an engineering corpus based on theoretical views of western and eastern linguists' criteria for designing an engineering corpus for linguistic analysis of engineering terms in the corpus.

Key words: corpus, engineering, representativeness, corpus balance, research article.

КИРИШ

Ҳар қандай корпусини шакллантиришда *(corpus design)* маълум бир мезонларга амал қилинади. Т. Мак Энери ва Р. Ксиаоларнинг фикрларига кўра қўйидагилар мезонлар корпусни шакллантиришда асосий ҳисобланади деб таъкидлайдилар:

1. Корпус репрезентативлиги *(Corpus representativeness).*
2. Корпус баланси *(Corpus balance).*
3. Корпусга намуналарни танлаш *(Corpus sampling).*
4. Корпусни салмоқлилиги *(*Corpus validity*).*

Ушбу мезонлар корпусни тузишда жуда аҳамиятли бўлиб, кейинги таҳлил босқичларинг сифатли олиб борилишини ва самарали натижаларни таъминлашга хизмат қилади.

Corpus representativeness мезони тўғрисида Т. Мак Энерининг Corpus-based language Studies номли китобда Лич қўйидагича фикр билдирган: “A Corpus is thought to be representative of the language varietyit is supposed to represent if the findings based on its contents can be generalized to the said language variety”.

Бу ерда Лич корпус репрезентативлиги унинг тилнинг ишлатилиш хилма-хиллигини *(language variety)* кўрсатувчи контентга асосланган натижалар тилнинг ишлатилиш ўша хилма-хиллигига кўра умумлаштира олинишидир деб таъкидлайди. Албатта бу ерда Дж.Лич корпус ичидаги маълумотнинг сифатини, яъни ундаги аниқланган тадқиқот натижалари корпусда тўпланган маълумотлар базаси тилнинг ишлатилиш хилма-хиллигига кўра қамраб олинишини назарда тутаяпти. Дейлик, муҳандислик корпусини шакллантириш учун муҳандислик терминлари қатнашган илмий мақолаларнинг танлаб олиниши, маълум муҳандислик соҳалари терминлари ўша мақолаларда фаол ишлатилиши ҳамда маълум жанрга (бизнинг ишимизда илмий жанр) тегишли бўлиши керак. Корпус репрезентативлигининг сифатига эришиш учун Д. Байбер қўйидагича фикр билдиради: “Representativeness refers to the extent to which a sample includes the full name range of variability in population” [Biber, 1993].

Репрезентативликнинг сифати корпус *намуналарни танлаш*нинг *(sampling)* танланган тил хилма-хиллигининг (бу ерда жанрининг) қанчалик кенг миқиёсда қамраб олининганлиги ва ўша жанрнинг кўрсата олинганлигидир. Бу эса ўз навбатида корпусни *танланган намуналар (Corpus sampling)* деган мезонни киритади. Ушбу танланган намуналар эса тадқиқотнинг долзарблигидан келиб чикиб *(Research question)*, унинг якуний натижалари қўйилган мақсаднинг мувофиқлигини кўрсатиши керак. Демак, ушбу мезон корпусни баланслантириш *(Corpus balance)* мезонини киритилишига ҳам туртки бўлиб, у корпус жанрнинг балансини, яъни унинг ҳажмини белгилайдиган омилдир. Умуман олганда корпусни репрезентативлигини унинг танланган намуналари ва баланси белгилаб беради.

Охирги мезон корпусни салмоқлилиги *(Corpus validity)* А. Лоренс ва М. Боувин томонидан таклиф қилиниб, у корпусни қанчалик қайишувчанлигини ва мослашувчанлигини кўрсатадиган ҳамда танланган намуналарнинг салмоғи ва тилнинг ташқи қийматини кўрсатадиган мезондир.

Демак, ҳар қандай корпусни шакллантириш учун ушбу мезонларга амал қилинган ҳолда амалга оширилиши зарур, чунки улар тадқиқот натижаларнинг

сифатини, яъни репрезентативлигини ва салмоғини кўрсатади. Корпусни шакллантиришда ушбу мезонлар уларнинг турига кўра ҳам ўлчанади ва мослашади. Бизнинг корпусимиз бир турдаги соҳа (муҳандислик) ва жанрга (илмий) тегишли бўлганлиги боис, у *ихтисослаштирилган корпус (Specialized Corpus)* турига тегишли бўлиб, махсус домейнли '*domain-specific'* ёки махсус жанрли '*genre-specific'* номли корпус ҳисобланади. Ҳар қандай корпус, шу жумладан ихтисослаштирилган корпус ҳам, корпуснинг репрезентативлик мезонини акс эттириш керак. Ихтисослаштирилган корпусда Мак Энери ва Уилсон фикрларига кўра репрезентативлик лексик даражанинг зичлиги *(closure)* билан ўлчанади [McEnery and Wilson 2001:166]. Корпусдаги сўзлар зичлигини ўлчаш учун корпус балансидаги токенлар[1] *(tokens)* сони бир хил сигментларга бўлиниши керак, шунда лексик бирликлар сони гипербола ўқида бир хил тақсимланади, яъни гиперболада лексик бирликларни *асимптотик тақсимоти* кузатилади.

Ихтисослаштирилган корпуснинг баланслантириш корпусни шаклланитиришдаги энг зарур мезонларан бири бўлсада, уни ўлчайдиган маълум бир илмий ўлчов мавжуд эмас. Баланслаштириш корпусга киритиладиган маълумотлар базасиннг тақсимланиши билан ўлчанади, яъни матн категориялари ва уларнинг пропорциялаш билан боғлиқдир. Шуни таъкидлаш жоизки, матнларни корпусга йиғиш учун учта омилга, яъни домейн *(domain)*, давр *(time)* ва маълумот ресурси *(medium)* боғлиик. Домейн омили танланган соҳа контекстига, давр эса матн нашр этилган йилга, маълумот ресурси эса матн манбаси бўлиб, улар корпус балансини беради. Мак Энери корпусдаги матнларни шакллантиришда, уларнинг ишлаб чиқариш *(production)* ва қабул қилиш *(reception)* омиллари муҳим аҳамиятга эга дея таъкидлайди. Матн ишлаб чиқариш деганда маълум бир турдаги маълумотларни корпусга танланган матнларни киритишдир. Матнни етказиш эса, танланган матнларни кенг кўламдаги қабул қилувчиларга ёки истеъмолчиларга етказиш учун ўша танланган матнлардир, яъни бу ерда ихтисослаштирилган корпус учун баланслаштириш мезони муҳим аҳамият касб этади. Ихтисослаштирилган коруснинг нафақат балансни, балки репрезентативлигини ҳам корпусни фреймлаш жараёни билан боғлиқдир (Мак Энери, А. Уилсон, 2001).

Агарда корпусни шакллантиришда ана шу 3 та асосий мезонларга амал қилинса, корпуснинг аҳамиятлилиги ва салмоғи ҳам ортади ва эришилган натижалар кенг кўламдаги истеъмолчиларга фойдалилиги аниқланади.

[1] Token - корпус лингвистикада лексик бирлик дегани ва у лексик бирликларнинг сонини билдиради.

Ҳинд тилшуноси Р. Кришнамурси ва турк тилшуноси И.Косем корпусни шакллантириш борасида айнан корпус жанр ва домейнининг таснифлаш муаммоларини ўрганиб чиқиб, турли корпусларда маълум бир домейн ёки жанр турлича номланишини таъкидлаб ўтган. Масалан, муҳандислик йўналиши бир корпусда технология ва муҳандислик (Technology and Engineering) домейнига тегишли бўлса, бошқа корпусда у физика фанлари ва муҳандислик (Physical Sciences and Engineering) ёки муҳандислик ҳар бир йўналиши алоҳида, математика ва компьютер фанлари (Mathematical and Computer Sciences), муҳандислик (Engineering), технология (Technology) каби номланиши мумкин. Бу эса ўз навбатида корпуснинг жанрига, яъни матн турига таъсир қилади. Айрим корпуслар битта жанрга (масалан илмий ҳисоботлар), айримлари эса турли жанрларга асосланган бўлиши мумкин (масалан диссертация ва дарсликлар). Корпусни шакллантиришдаги бундай ёндашув эса махсус корпуснинг репрезнтативлигини кўрсата олмайди.

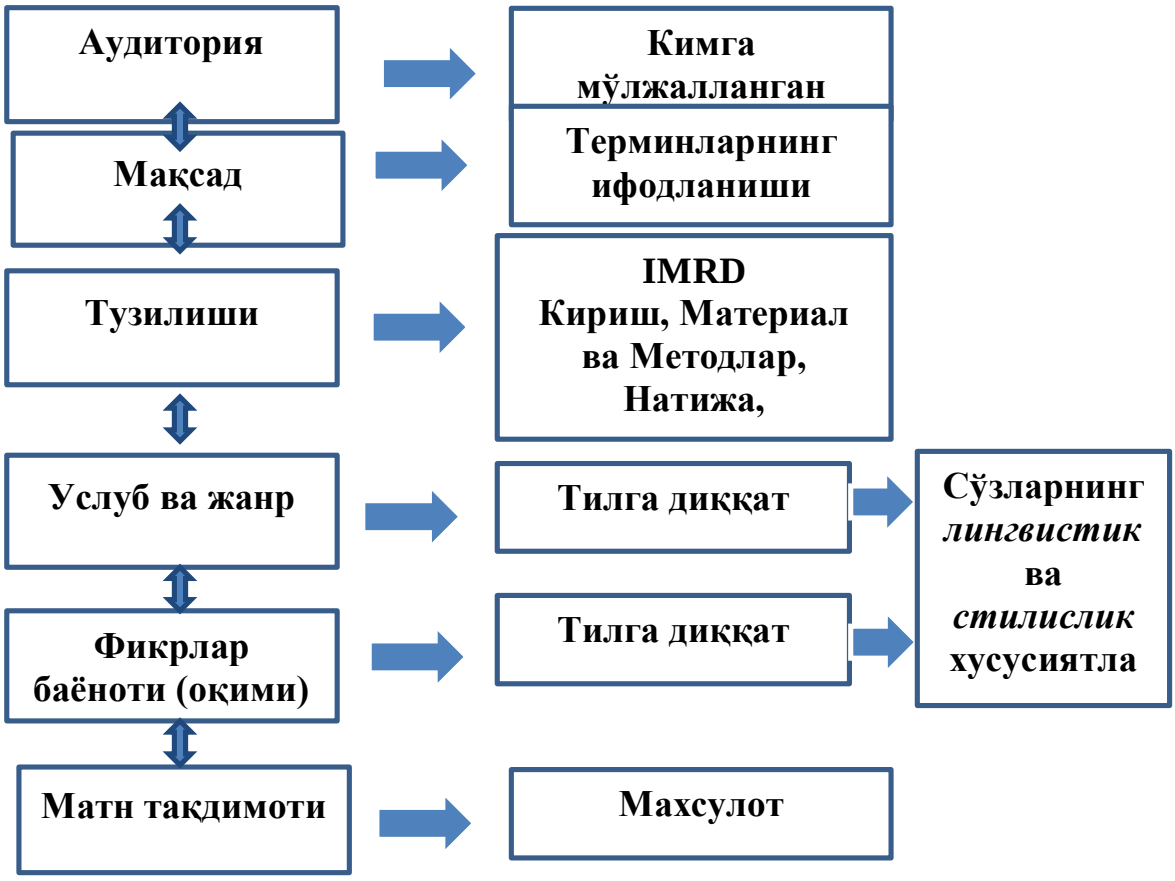
Демак, тадқиқотимиздаги махсус корпуснинг репрезентативлигини таъминлаш мақсадида, уни бир жанрли материалдан шакллантиришимиз лозим.

МАТЕРИАЛ ВА МЕТОДОЛОГИЯ

Муҳандислик терминларининг корпусини шакллантиришда бир жанрли ҳамда бир вақтда кўп домейнли махсус корпус йиғиш муҳимлигини ўрганиб чиқдик. Электрон базасидаги Plos One журнали кўзланган параметрларга тўғри келиши яъни муҳандислик соҳасининг кенг қамровли домейнга ва кенг қамроали ўрганувчиларги мўлжалланги боис ҳамда уларнинг бир жанрга тегишлиги айнан махсус корпуснинг тамойилларига тўғри келди.

Корпус материалининг шарҳи (изоҳи). Корпусдаги ҳам терминларни зичлигини ҳам корпус балансини таъминлаш мақсадида инглиз тилида аниқроғи жаҳон стандартларига кўра илмий мақолаларнинг ёзилишини ўрганиб чиқиб, терминлар мақолаларнинг айнан қайси қисмида фаол ишлатилишини аниқлаб олдик. Ж. Сволс ва К. Фик ўзларининг “Academic Writing for Graduate Students” номли китобида битирувчи ўрганувчилар учун илмий иш ёзиш кўникмаларини (бу ерда илмий иш деганда илмий мақола, илмий ҳисобот ва ҳоказалар ҳам тушинилади) ўргатар экан, илмий мақола ёзишда қабул қилинган **IMRD** (Introduction, Materials and Methods, Discussions, Results) яъни кириш, материал ва методлар, муҳокамалар ва натижалар номли бўлинмаларга амал қилиниш ва илмий мақола кимга **(Audience**-қандай ўрганувчиларга**)** йўналтирилганини, қандай услуб ва жанрда **(Style and Genre)** ёзилганлиги, мақсад **(Purpose),** термин ва тушунчаларнинг аудиторияга етказилишида **(Flow)** фикр ва

маълумотлар баёнининг кетма-кетлигга амал қилиниши, ҳамда матннинг **(Presentation)** ўзининг тақдим этилишини ҳисобга олиш кераклигини таъкидлаган. Уларнинг фикрларини қўйидаги схемада акс эттириш мумкин:



Мак-Энерининг фикрича, сўз зичлиги ва мўллиги *(closure and saturation)* корпуснинг балансини таъминлайди. Ж. Сволс ва К. Фикларнинг схемасига кўра Мак-Энери айтган ўша зичлик айнан илмий мақоланинг *услуб ва жанр* ҳамда *фикрлар баёнотида* кўришимиз мумкинлигини аниқладик. Бироқ сўзларнинг *лингвистик ва стилистик хусусиятлари* айнан IMRD нинг қайси қисмида кўпроқ ифодаланиши учун унинг ҳар бир қисмини кўриб чиқамиз, чунки корпусни шакллантиришда бизлар илмий мақоланинг қайси бўлимини ёки бўлимларини олишимиз кераклигини аниқлаб олишимиз керак. Ж. Сволс ва К. Фиклар ҳар қандай илмий мақоланинг кириш (Introduction) қисмида фикр ва маълумотларнинг кетма кетлигини ифодалайди ҳамда қўйидаги босқич (Move) ларга амал қилинишини аниқлаган:

1. Тадқиқот майдонини ҳосил қилиш *(Establishing a research territory).*

2. Ўрганилаётган масала ёки муаммони долзарблигини кўрсатиш мақсадида уни **асослаш зарурияти ва сабабларни кўрсатиш** *(Establishing a niche).*

3. Кўрсатилган асос ва сабаблар учун таклиф ва ечимларни кўрсатиш *(Occupying the niche).*

Ж. Сволс ва К. Фиклар ҳар қандай илмий мақоланинг *муҳокама ва хулоса* **(Discussion/Conclusion)** қисмида фикр ва маълумотларнинг кетма кетлигини қўйидаги босқич **(Move)** ларга амал қилинишини аниқлаган:

1. Олдинги маълумотни жамлаш *(Background Information.)*, яъни тадқиқот мақсади, теория ва методологияси тўғрисида *(*агар зарурият туғилса, аммо тадқиқотнинг айрим соҳаларида кўрсатилиш зарур*)* .

2. Асосий **натижаларни жамлаш** *(Summarizing and reporting key results)*ва баён қилиш (кўрсатилиш шарт);

3. Асосий **натижаларга изоҳ** (*Commenting on key result*) бериш (кўрсатилиш шарт);

4. Тадқиқотнинг чекловлари *(*Stating the limitations on the study*);*

5. Кейинги тадқиқотни амалга ошириш учун **тавсиялар** *(Making recommendation for future research and implementations)* бериш (агар ўринли бўлса).

НАТИЖА

Ушбу босқичлардан кўриниб турибдики, илмий мақоланинг муҳокама ва хулоса қисмидаги 1,2,5 пунктлари мақоланинг кириш қисмига ўхшаш 3 ва 4 пунктларида эса топилгпн натижаларга изоҳ ва тадқиқотнинг чекловларини кўрсатаяпти.

Илмий мақолаларнинг материал ва методлари (Materials and Methods) қисмида асосан тадқиқот материали ва усуллари тўғрисида сўз юритилганлиги боис, уларда кириш ёки натижа ва муҳокамалар қисмларидагидек сўзларнинг лингвистик ва стилистик хусусиятлари кузатилмайди. Илмий мақоланинг аннотация қисми ҳам алоҳида ўринга эга, чунки унда мақоланинг барча асосий қисмлари, яъни умумий маълумот, мақсад, метод, натижа ва хулосалардан иборат. Аммо унинг ҳажми қисқалиги боис ундаги сўзларнинг лингвистик ва стилистик хусусиятлари кам кузатилади.

Plos One журналидаги ёзиш талабларини ўрганиб чиқсак, у ерда IMRD стандартлариги қатъий амал қилинганлигини, қар бир қисмда Ж. Сволс ва К. Фикларнинг келтириб ўтган кетма кет босқичлардан иборатлигини Plos One журналининг сайтида берилган талабларида кўришимиз мумкин:

Инглиз тилидаги мақола кўпчилик томонидан ўқилишига жалб қиладиган унинг номи (Title) ва аннотацияси (Abstract) бўлиб, улар ҳам маълум принципларга кўра шаклланади. Масалан, мақоланинг номи мақоланинг бошида турсада, тадқиқот сўнгида шаклланади. У ўрганилаётган тадқиқлт объектининг номи, қамраб оладиган ҳажмни ва танланган соҳани ёрита оладиган бўлиши керак. Муҳандислик соҳаси ва унга турдош соҳалардаги мақолаларнинг аксарияти (90%) отли бирикма ва предлоглардан ёки –ing қўшимчаси қўшилган феъллардан иборатдир (Ж. Сволс ва К. Фик, 2012, б. 379).

Мақоланинг аннотация қисми тўғрисида гапирганимизда, у илмий мақолани бутун тақдирини, яъни қабул қилиишини ёки қабул қилинмаслигини баҳолаб берадиган қисм бўлиб, тадқиқотнинг рақобатбардошлигини белгилайди. Аннотацияларни ёзиш икки хил, тадқиқотнинг натижаларига қаратилган ва тадқиқотни жамлайдиган, усулларда ёзилиши мумкин, деб таъкидлайди Ж. Сволс ва К. Фик.

Кириш ҳамда натижа ва муҳокамалар бўлимларидаги талабларига эътибор берсак, айнан шу қисмларда лексик бирликлар фаол ишлатилишини кўришимиз мумкин. Мақоланинг айнан кириш қисмида сўзларнинг зичлигини ва шу боис уларнинг лингвистик ва стилистик хусусиятларни кўра олишимиздан, аммо мақоладаги натижа ва муҳокама қисмида худди шу нарсани кузатишимиз билан бирга яна шу сўзларнинг такрорланиб келиши эҳтимоллиги бўлганлиги боис, материал ва методлар қисмида эса тилнинг услуб ва жанри муҳим аҳамият касб этмаганлиги туфайли мақоланинг натижа ва муҳокамалар бўлимидагига кўра эса кириш қисмига қўйилган талабларнинг кўпоқ эканлигини ҳисобга олиб, у ерда сўларнинг лингвистик ҳусусиятлари кўпроқ акс этади деган фикрга келдик. Бироқ, Сволс ва Фиклар мақоланинг номи ва аннотациясида ҳам отли бирикмаларнинг учраши 90% ташкил қилиши деган фикрлари, терминларнинг учраш ҳолатини эҳтимолинини келтириб чиқаради.

ХУЛОСА

Демак, корпусни шакллантириш борасида уни ташкил қилувчи маълумотлар баъзаси, яъни материалнинг танланиши бўйича мавжуд бўлган муаммолар кўриб чиқилиди. Корпус материали унинг репрезентативлигини белгилайдиган асосий кўрсаткич бўлиб, бу борада турли қарашларни ўрганиб чиқдик ва корпусни бир жанрли (илмий) махсус корпус сифатида шакллантиришга қарор қилдик. Корпус материалини эса муҳандислик соҳасининг кенг қамровига эга бўлган (44 та йўналиш) PLOS ONE электрон журналини танладик. Корпусдаги домейнларни аниқлаб олиш учун

Ўзбекистоннинг муҳандислик йўналишлари мавжуд бўлган ОТМлар ва Plos ONE журналидаги 12 та муҳандислик йўналиш танлаб олинди. Корпусдаги танланган намуналар сони ҳар бир домейнда ҳар хил бўлсада, уларнинг сонини ўзгартирмасликка қарор қилдик, биринчидан ушбу домейнларнинг ҳар бирини алоҳида корпус қилиб анализ қилиб чиқишимиз мумкин, иккинчидан, ушбу домейнларнинг спецификасидан келиб чиқсак, масалан дейлик ўша механика муҳандислиги, муҳандисликнинг энг қадимги ва қолган муҳандислик йўналишларнинг асоси ҳисобланганлиги боис, табиийки ушбу домейннинг ҳажми ҳам кенг бўлиши, учинчидан корпусимиз бир жанрли ва махсус корпус турига кирганлиги боис махсус корпуснинг ҳажми қанчалик катта бўлса шунчалик унинг репрезентативлигини таъминлайди (Т. Мак-Энери, 2006) деган фикрга амал қилдик. Мақолаларнинг бошқа қисмларига нисбатан кириш қисмида муҳандислик лексик бирликларнинг мақоланинг кириш қисмига қўйилган талаблардан ва уни ёзиш учун амал қилиш керак бўлган принципларга кўра, лексик бирликларнинг фаол ишлатилиши натижасида терминларининг лингвистик ҳусусиятлари кўпроқ намоён бўлиши эҳтимоли кўпроқ эканлигини ҳисобга олиб, корпус асосидаги муҳандислик терминларининг лингвистик таҳлилни муҳандисликка оид мақолаларнинг кирш қисмидан олиш мақсадга мувофиқ деган фикрга келдик.

ФОЙДАЛАНИЛГАН АДАБИЁТЛАР

1. Н. Атабоев. The main theoretical problems of corpus linguistics. www.fledu.uz Илмий методик журнал. 2019. Б. 47.
2. McEnery T., Xiao R., Tono Y. Corpus-based language Studies. –NY: *Routledge Applied Linguistics*. 2006. Pp. 386. P. 13
3. Biber D., 1993. “representativeness in corpus design”. Literary and Linguistic Computing 8/4: 243-257.
4. K. Aijmer and B. Altenberg (eds) English Corpus Linguistics, pp. 8-29. London: Longman.
5. A. Laurence, M. Bowen. The Asian ESP Journal Special Edition, Volume 9, Issue 2, 2013, p. 2-25.
6. Krishnamurthy R.,Kosem I. (2007). Issues in creating a corpus for EAP Pedagogy and Research. *Journal of English for academic purposes*, 6 , 356-373. www.elsevier.com/locate/jeap
7. Hui-Hsien Feng. Designing, implementing, and evaluating an automated writing evaluation tool for improving EFL graduate students’ abstract writing: A case in

Taiwan: Dissertation for degree of Doctor of Philosophy (PhD) in Applied Linguistics and Technology. – Iowa State University, 2015. Pp. 194.

REFERENCES

1. H. Атабоев. The main theoretical problems of corpus linguistics. www.fledu.uz Илмий методик журнал. 2019. Б. 47.
2. McEnery T., Xiao R., Tono Y. Corpus-based language Studies. –NY: *Routledge Applied Linguistics*. 2006. Pp. 386. P. 13
3. Biber D., 1993. “representativeness in corpus design”. Literary and Linguistic Computing 8/4: 243-257.
4. K. Aijmer and B. Altenberg (eds) English Corpus Linguistics, pp. 8-29. London: Longman.
5. A. Laurence, M. Bowen. The Asian ESP Journal Special Edition, Volume 9, Issue 2, 2013, p. 2-25.
6. Krishnamurthy R.,Kosem I. (2007). Issues in creating a corpus for EAP Pedagogy and Research. *Journal of English for academic purposes*, 6 , 356-373. www.elsevier.com/locate/jeap
7. Hui-Hsien Feng. Designing, implementing, and evaluating an automated writing evaluation tool for improving EFL graduate students’ abstract writing: A case in Taiwan: Dissertation for degree of Doctor of Philosophy (PhD) in Applied Linguistics and Technology. – Iowa State University, 2015. Pp. 194.