

**Հակոբյան Ֆրիդա** (Հայաստան,  
ԳԱԱ լեզվի ինստիտուտ)

**ՀԱՅԵՐԵՆԻ ԸՆԴՀԱՆՐԱԿԱՆ ԿՈՐՊՈՒՄԻ  
ԿԱԶՄՈՒԹՅԱՆ ՄԿՋԲՈՒՆՔՆԵՐԸ**

Քսաներորդ դարի երկրորդ կեսին ստեղծված լեզաբանական ուղղությունների թվում առանձնանում է *կորպուսալին լեզվաբանությունը*<sup>1</sup>՝ որպես խոսքային օրինակների միջոցով լեզվական իրողությունների ուսումնասիրության հետաքրքիր և ինքնատիպ եղանակ:

Կորպուսային լեզվաբանության բնագավառում ձեռք բերված մեծ հաջողությունները հիմնականում պայմանավորված են լեզվաբանության մեջ տեսական ու կիրառական տարբեր նպատակներով համակարգչային տեխնիկայի լայն գործածությամբ և այնպիսի ծրագրերի ստեղծմամբ, որոնց միջոցով ուսումնասիրվում ու ներկայացվում է լեզուն:

Ընդհանուր առմամբ՝ կորպուսային լեզվաբանությունը զբաղվում է լեզվական կորպուսների՝ մեծ, ներկայացուցչական, սահմանափակ ծավալ ունեցող և էլեկտրոնային եղանակով հասանելի հենքերի կազմությամբ: Լեզվական կորպուսների միջոցով փորձ է արվում ուսումնասիրել լեզուն՝ իր տարբեր դրսևորումներով և բովանդակային հարստությամբ:

Հայ լեզվաբանության մեջ լեզվական կորպուս ստեղծելու ուղղությամբ տարվող աշխատանքները սահմանափակվել են միայն արևելահայերենի ազգային կորպուսի՝ ԱՐԵՎԱԿ-ի<sup>1</sup> ստեղծմամբ, որը բավարար չէ թե՛ որակական առումով, թե՛ ընդգրկման ծավալով:

Քանի որ լեզվական ցանկացած իրողության համաժամանակյա և տարածամանակյա լիարժեք քննությունը

<sup>1</sup> [http://www.eanc.net/EANC/search/?interface\\_language=am](http://www.eanc.net/EANC/search/?interface_language=am):

հնարավոր է կատարել միայն համապատասխան կորպուսի առակայության դեպքում, էսպես կարևորվում է հայերենի նոր՝ *ընդհանրական կորպուսի* կազմությունը, որի հիմքում ընկած կլինեն համաշխարհային կորպուսային լեզվաբանության արդի փորձն ու հայերենի լեզվական առաձնահատկությունները: Մեր նպատակներից դուրս համարելով կորպուսի կազմության սկզբունքների լեզվաբանական և ծրագրային կողմի հանգամանակից ներկայացումը՝ կխոսենք միայն դրա կառուցման ուրվագծի մասին՝ առանձնացնելով կիրառության հիմնական ոլորտները:

Եթե լեզվական կորպուս ասելով հասկանում ենք տարբեր բնույթի տեքստերի ամբողջական հավաքածու, ապա պայմանականորեն մեկից ավելի տեքստերի ցանկացած ժողովածու կարող է կորպուս համարվել: Մակայն կան ընդհանրական մի շարք հատկանիշներ, որոնցով բնութագրվում են տեքստային հենքերն ընդհանրապես: Այդ հատկանիշներից ամենահիմնականն ու թերևս գլխավորն այն է, որ տեքստային հենքում եղած տեքստերի միջոցով պետք է ներկայացվի լեզուն: Հետևաբար, կորպուսը պետք է կազմել այնպես, որ, դրանում եղած տեքստերն ուսումնասիրելով, հետազոտողը ստանա լեզվական ցանկացած իրողության ամբողջական պատկերը: Հենքում եղած տեքստերը սկզբունքորեն կարելի է ուսումնասիրել երկու եղանակով: Առաջին դեպքում կարող ենք քննել բոլոր տեքստերը մեկ առ մեկ: Միանգամից նկատենք, որ այս սկզբունքը գործնականում կիրառելի չէ, որովհետև մեծ ցանկության դեպքում անգամ կարելի է միայն մեռած լեզուների տեքստային բոլոր օրինակները քննել, իսկ կենդանի լեզվում տեքստերն անթիվ են: Նաև որևէ կերպ հնարավոր չէ կազմել տեքստային ամբողջական հենք. մի կողմից՝ կարող են դուրս մնալ հազվադեպ գործածվող օրինակները, մյուս կողմից՝ ուշադրությունը բևեռելով միայն սակավ գործածական օրինակների վրա՝ կարող են չընդգրկվել նույնիսկ հաճախադեպ լեզվական օրինակները:

Ստացվում է այնպես, որ մենք ամեն դեպքում կարող ենք քննել միայն սահմանափակ թվով տեքստային օրինակներ, ուստի նոր կորպուսում տեքստերի ընդգրկման տեսակետից պետք է կարևորել ոչ թե դրանց քանակական շատությունը, այլ տարաբնույթ լինելն ու լեզվական հնարավորինս մեծ թվով տարբեր օրինակներ ներկայացնելը: Հետևաբար, ընդհանրական կորպուսի ամենաբնութագրական ու կարևոր հատկանիշը պետք է դարձնել ներկայացուցչականությունը:

Մյուս կարևոր հատկանիշը, որով լեզվական կորպուսը տարբերվում է այլ բնույթի տեքստային հենքերից, պիտակավորումն է, որը էական նշանակություն ունի յուրաքանչյուր կորպուսի օգտակարության հարցում: Այն համապատասխան պիտակների կցումն է բառային յուրաքանչյուր միավորի: Ընդհանրապես լեզվի կառուցվածքային տարբեր միավորների համար, ըստ լեզվական ուսումնասիրությունների նպատակի, կան պիտակավորման տարբեր տեսակներ՝ հնչյունական, բառագիտական, քերականական (ձևաբանական և շարահյուսական): Կարծում ենք՝ հայերենի ընդհանրական կորպուսի պիտակավորման համակարգը լիքաժեք դարձնելու համար նպատակահարմար է օգտագործել ինչպես հնչյունական, բառագիտական, այնպես էլ քերականական պիտակավորումը: ԱՐԵՎԱԿ-ում, օրինակ, կիրառվում են միայն ձևաբանական պիտակներ, որոնք, ըստ կորպուսի հեղինակների, արտահայտում են հայերենի խոսքի մասերի քերականական բոլոր կարգերը: Հարկ է ընդգծել, սակայն, որ առանձնացվող քերականական կարգերը երբեմն չեն համապատասխանում նորմատիվ քերականությանը: Օրինակ՝ առանձին խոսքի մաս է համարվում նախդիրը, առանձնացվում է վեց հոլով. սեռականն ու տրականը համարվում են առանձին հոլովներ, իսկ հայցականը ուղղականից չի առանձնացվում: Ածականի համեմատության աստիճանները համարվում են քերականական կարգ, ընդ որում՝ առանձնացվում է միայն գերադրական աստիճանը, բաղդատակա-

նը՝ ոչ: Բայի եղանակներից ներկայացվում են միայն երեքը՝ ըղձական, ենթակայական (sic), հրամայական: Ենթակայական եղանակն առանձնացնելիս հեղինակները հավանաբար նկատի են ունեցել ենթադրականը: Իբրև քերականական կարգ է առանձնացվել գոյականացումը՝ իր գոյականացված, անորոշ դերբայ, հարաբերական գոյական տեսակներով՝ նկատի ունենալով այլ խոսքի մասերի գոյականաբար գործածությունը<sup>2</sup>:

Ի տարբերություն ԱՐԵՎԱԿ-ի, որն աստիճանաբար պատմական կորպուս է դառնում, քանզի դրանում ներկայացված են XIX դարի երկրորդ կեսից մինչև 2009թ. կազմված տեքստերը, նպատակահարմար է, որ ընդհանրական կորպուսն արտացոլի առավել ընդգրկուն ժամանակաշրջան՝ ներառելով հայերենի պատմական զարգացման տարբեր շրջաններում գրաբարով, միջին հայերենով, ինչպես նաև արևելահայերենով ու արևմտահայերենով ստեղծված տեքստեր: Անշուշտ, սա էլ իր բարդությունները կունենա, քանի որ գործնականում գուցե հնարավոր չլինի այդ ծավալի ծրագրային բազա ստեղծել: Դրա համար կարելի է կորպուսում ներառվելիք տեքստերը կատարել ընտրողաբար՝ ըստ նախօրոք մշակված սկզբունքների՝ տեքստերի բազմաքանակության փոխարեն առավել կարևորելով լեզվական հնարավորինս բազմաբնույթ օրինակներ ներկայացնելը:

Ըստ այդմ՝ այն կբաժանվի ենթակորպուսների՝ ըստ հայերենի պատմական զարգացման տարբեր շրջաններն ընդգրկող տեքստերի: Կունենանք գրաբարի, միջին հայերենի, արևելահայերենի ու արևմտահայերենի, ինչպես նաև խոսակցական լեզվի ենթակորպուսներ: Տարբերվելով օտարալեզու մյուս կորպուսներից՝ այն իրավամբ կարող է համարվել միաժամանակ թե՛ պատմական և թե՛ զուգադրական կոր-

---

<sup>2</sup>[http://www.eanc.net/EANC/search/frame\\_parts/gramsel.php?interface\\_language=am&search\\_language=armenian1&contexts\\_output\\_language=armenian1:](http://www.eanc.net/EANC/search/frame_parts/gramsel.php?interface_language=am&search_language=armenian1&contexts_output_language=armenian1)

պուս: Լեզվում կատարվող ամենավերջին փոփոխություններին հաղորդակից դառնալու համար կարևոր է կորպուսի շարունակական հարստացումը նորանոր տեքստերով՝ հասնելով նրան, որ այն վերածվի այսպես կոչված մոնիտորային կորպուսի, այսինքն՝ մեքենական եղանակով հարստացվի համացանցային զանազան աղբյուրներից: Ի դեպ՝ մոնիտորային կորպուսների ստեղծման նախադեպեր կան: Օրինակ՝ Բիրմինգհեմի համալսարանում Ջոն Մինքլերը ձեռնամուխ է եղել COBUILD կոչվող մոնիտորային կորպուսի ստեղծմանը: Սա, նրա կարծիքով, տեքստերի անվերջ բազմություն է, որովհետև կարելի է շարունակ մեծ թվով տեքստեր ավելացնել՝ մեծացնելով տեքստային հենքի ծավալը:

Հայերենի ընդհանրական կորպուսի կազմությունը բավական բարդ և աշխատատար գործընթաց է, որը պահանջում է լեզվաբանների և ծրագրավորողների սերտ համագործակցություն, սակայն այն ինքնանպատակ չէ, քանի որ կունենա ինչպես տեսական-լեզվաբանական, այնպես էլ գործնական-կիրառական հսկայական նշանակություն՝ փաստական հարուստ նյութ ապահովելով լեզվաբանների համար:

Կորպուսի օգտակարության համար կարևոր է հետազոտողների առաջ բացել այն հեռանկարները, որ կունենա այս կորպուսը, որովհետև նույնիսկ հիմա, երբ տեքստային հենքի և որոնողական ծրագրի միջոցով կարելի է հարյուրավոր օրինակներ գտնել հաշված վայրկյանների ընթացքում, շատ հետազոտողներ չեն հրաժարվում օրինակները քարտագրելուց, համակարգիչ մուտքագրելուց:

Ընդհանրական այսպիսի կորպուսը նախևառաջ հրաշալի հնարավորություն կնձեռի լեզվի պատմության մասնագետներին *պատմական զարգացման ընթացքում հայերենի կրած փոփոխություններն ուսումնասիրելու համար*: Օրինակ՝ Հելսինկիի կորպուսում նախդիրների գործածության հաճախականությունը անզլերենի զարգացման տարբեր շրջաններում ուսումնասիրելիս պարզվել է, որ ամենահա-

ճախ գործածվող նախադիրները by-ը և of-ն են, որոնք հավասար գործածություն են ունեցել մինչև տասնհինգերորդ դարը. դրանից հետո by նախադիրը երեք, իսկ մինչև 1640 թվականը՝ ութ անգամ ավելի հաճախ է գործածվել, քան of-ը<sup>3</sup>:

Կորպուսը օգտակար փաստական նյութ կապահովի նաև գրական արևելահայերենի ու արևմտահայերենի հարցերով հետաքրքրվողների համար: Եթե շարունակաբար հարստացվի նաև արևմտահայերենի ենթակորպուսը, այդ կերպ հնարավոր կլինի կատարել լեզվական երևույթների զուգադրական քննություն տեքստային մեծաթիվ օրինակներով: Զուգադրական այդպիսի կորպուսները հաջողությամբ կիրառվում են, օրինակ, ամերիկյան և բրիտանական անգլերենների ուսումնասիրության ժամանակ: Այսպես՝ օգտագործելով ամերիկյան անգլերենի Brown և բրիտանական անգլերենի LOB տեքստային հենքերը՝ փորձ է արվել ուսումնասիրել արական և իգական սեռերի ներկայացուցիչներին տրվող անուններին փոխարինող դերանունների գործածությունը: Պարզվել է, որ արական սեռի ներկայացուցիչներին անվանող բառերն ավելի հաճախ են գործածվում լեզվում, քան իգական սեռի ներկայացուցիչներին տրվող անունները, ընդ որում՝ իգական սեռի անունները գործածության տեսակետից ավելի հաճախադեպ են բրիտանական, քան ամերիկյան անգլերենում:

Զուգադրական կորպուսներում կատարված ուսումնասիրություններով փորձ է արվել նաև բացահայտել քերականական մի շարք իրողություններ: Օրինակ՝ քանի որ կանայք, տղամադկանց համեմատությամբ, ավելի շատ հանդես են գալիս որպես գործողության կրող, իսկ տղամարդիկ՝ գործող անձ, արական սեռի ներկայացուցիչների անունները նախադասության մեջ ավելի շատ հանդես են գալիս որպես

---

<sup>3</sup> <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/>:

ենթակա, իգական սեռի ներկայացուցիչներին՝ որպես ուղիղ խնդիր:

Արևելահայերեն-արևմտահայերեն գուգադրական կորպուսի ստեղծման կարևորությունն ավելի է ընդծվում հայերենի այս երկու ճյուղերի մերձեցման ուղիների որոնման համատեքստում:

Պետք է կարևորել նաև խոսակցական ենթակորպուսի ստեղծումն ու հետագա հարստացումը տեսաձայնագրված հոլովակներով: Հատկապես հիմա, երբ հնարավոր է տեսա-հոլովակում հնչած բանավոր խոսքը վերածել տեքստի, հեշտությամբ կարելի է տեսնել, թե ինչ տարբերություններ կան գրական և խոսակցական լեզուների միջև, հասարակության տարբեր անդամները, ըստ իրավիճակի, լեզվական ինչ միջոցներ են ընտրում հաղորդակցվելու համար, և առհասարակ խոսակցական լեզուն զարգացման ինչ միտումներ ունի:

Հայերենի ընդհանրական կորպուսը՝ իր տեքստային սահմանափակ օրինակներով, պետք է ոչ միայն ներկայացնի լեզվական կառուցվածքը, հասանելի լինի էլեկտրոնային եղանակով, այլև ունենա ծրագրային ապահովման ամբողջական համակարգ, քանզի կորպուսի օգտակարությունն առավելապես պայմանավորված է համակազային համապատասխան ծրագրերի առկայությամբ: Շնորհիվ կորպուսում համակարգչային ծրագրերի կիառության՝ լեզվական այն աշխատանքները, որոնց կատարման համար նախկինում երկար ժամանակ էր պահանջվում, հնարավոր կլինի իրականացնել շատ արագ: Այդ ծրագրերը կկիրառվեն տեքստային հենքում եղած լեզվական օրինակներն ամենայն ճշգրտությամբ և մեծ արագությամբ որոնելու, միայնաց հետ համեմատելու, ըստ տարբեր հատկանիշների՝ դասակարգելու համար:

Ի վերջո, պետք է հասնել տեքստի ավտոմատ մշակմանն էապես նպաստող այնպիսի ծրագրերի ստեղծմանը, որոնց միջոցով կորպուսում եղած յուրաքանչյուր տեքստ

կվերածվի նախադասությունների, նախադասությունները՝ բառակապակցությունների, բառակապակցությունները՝ բառաձևերի, բառաձևերը՝ բառերի, վերջիններիս կազմում էլ կառանձնացվեն բառի հիմնական և երկրորդական ձևայինները: Մեքենական եղանակով տեքստի՝ որպես լեզվական խոշոր միավորի աստիճանական տրոհման միջոցով կհանգենք լեզվի իմաստային նվազագույն միավորին՝ ձևային: Տեքստի ավտոմատ մշակման այսպիսի համակարգը կարող է հաջողությամբ ծառայել լեզվի մակարդակային բոլոր միավորների քննությանը՝ ավտոմատացնելով կիրառական այնպիսի բնագավառներում տարվող աշխատանքներ, ինչպիսիք են բնագրային համաբարբառների կազմումը, բառարանները նոր բառերով համալրելը և այլն:

ԱՐԵՎԱԿ-ի համեմատությամբ ունենալով ձևային և բովանդակային նկատելի տարբերություններ և շահեկանորեն տարբերվելով օտարալեզու հեղինակավոր կորպուսներից՝ հայերենի ընդհանրական կորպուսը որակական նոր մակարդակի կհասցնի հայոց լեզվի լիակատար և առավել ամբողջական ուսումնասիրությունը:

Hakobyan Frida – The Principles of Creating an Armenian General Corpus.– The purpose of the article is to present a brief description of the linguistic and programming aspects of the principles of the Armenian general corpus. Special attention has been paid to features such as representativeness, tagging and software. The spheres where the newly compiled corpus can be successfully applied are also highlighted. It is emphasised that having differences in comparison with EANC and beneficially distinguishing from authoritative foreign-language corpuses the general corpus will lead the new and complete study of Armenian into a qualitatively new level.

Акопян Фрида – Принципы составления всеобщего корпуса армянского языка.– Цель статьи представить краткое описание языковых и программных аспектов принципов армян-



ского общего корпуса. Особое внимание было уделено таким особенностям, как репрезентативность, маркировка и программное обеспечение. Выделены также сферы, в которых можно с успехом применить новый корпус. Подчеркивается, что отличие корпуса от ВАНК и авторитетных иноязычных корпусов общий корпус армянского языка будет способствовать его более полному изучению на качественно новом уровне.