

Համաբարբառների համակարգչային մշակման խնդիրները

Արդի գիտության նորագույն նվաճումներին հաղորդակից լինելը, տեխնիկական համապատասխան միջոցների գործնական կիրառությունը հայ լեզվաբանության մեջ կարող են զարգացման լայն հեռանկարներ բացել ավանդական այնպիսի բնագավառներում, ինչպիսին համաբարբառների կազմությունն է:

Համաբարբառները, ինչպես հայտնի է, բնագրային օրինակներով ներկայացնում են հեղինակի բառապաշարը՝ ըստ այբբենական հաջորդակա-նության, տալիս բառերի գործածության հաճախությունը: Վերջին տասնա-մյակներում կազմվել և լույս են ընծայվել ինչպես հինգերորդ դարի հայ մատենագրության, այնպես էլ հետագա շրջանի մի շարք հիմնականում գրաբարագիր հեղինակների երկերի համաբարբառներ:

Բհարկե, մեր նպատակներից դուրս է անդրադառնալ նախորդ շրջա-նում կազմված համաբարբառներին՝ ըստ ամենայնի արժևորելով դրանք, կամ խոսել դրանցում եղած բացթողումների ու նկատված թերությունների մասին: Պարզապես կարող ենք փաստել, որ, առանց տեխնիկական որևէ օժանդակության, կատարվել է հսկայական աշխատանք, սակայն ժամա-նակները փոխվել են, և տեխնիկայի աննախադեպ առաջընթացը հասցրել է նրան, որ անհրաժեշտություն է առաջացել վերանայելու համաբարբառ-ների կազմության նախկինում ընդունված ավանդական եղանակները: Որոնման սահմանափակ հնարավորություններ ունեցող նախկին մեքե-նագիր համաբարբառներին կարող են փոխարինել համակարգչային զա-նագան ծրագրերով ստեղծված, որոնողական ճկուն համակարգ ունեցող, համացանցում հասանելի էլեկտրոնային համաբարբառները:

Մեծամասամբ էլեկտրոնային եղանակով կազմվում են օտարալեզու համաբարբառները, որոնք պայմանականորեն կարելի է երկու խմբի բա-ժանել. համաբարբառներ, որոնք ներկայացնում են որևէ առանձին հեղի-նակի բառապաշարը (Ուիլիամ Շեքսպիր, Թոմաս Գրեյ, Ջոն Քիթս, Ջերարդ Հոփքինս, Ուիլիամ Բլեյք), և միաժամանակ մի քանի հեղինակների երկերի բառապաշարը ներկայացնող համաբարբառներ («Անգլիական պոեզիայի համաբարբառ»):

Ցավոք սրտի, հայերեն էլեկտրոնային համաբարբառների մասին խոսելը դեռ վաղ է, որովհետև հիմա էլ կազմվելուց հետո դրանք շարունակում են տպագրվել առանձին գրքով, կամ լավագույն դեպքում տպագիր համաբարբառը Microsoft Word փաստաթղթի տեսքով տեղադրվում է որևէ կայքում, օրինակ՝ այսպես է ներկայացվում Կիրակոս Գանձակեցու «Հայոց պատմութիւն» երկի համաբարբառը լեզվի ինստիտուտի կայքում⁸⁶:

Հայաստանի ամերիկյան համալսարանի թվանշային գրադարանում էլ փորձ է արվում մեքենական եղանակով կազմել էլեկտրոնային համաբարբառ, սակայն ներկայացված է միայն Եղիշեի պատմությունը, այն էլ ոչ թե համաբարբառը, այլ երկում գործածված բառերի և բառաձևերի՝ այբբենական հաջորդականությամբ դասավորված ցանկը:

Կարո դ էն սրանք էլեկտրոնային համաբարբառներ համարվել և օրինակ ծառայել հետագայում կազմվելիք համաբարբառների համար: Կարծում ենք՝ ո՛չ: Տպագիր համաբարբառի ներկայացման ձևաչափը փոխելը դեռևս չի նշանակում բովանդակության փոփոխություն, իսկ էլեկտրոնային համաբարբառը տպագրից տարբերվում է հենց բովանդակությամբ:

Էլեկտրոնային համաբարբառների կազմության համար պետք է.

ա. մշակել դրանց կազմության սկզբունքները՝ հաշվի առնելով տպագիր համաբարբառներում եղած սահմանափակումները և նպատակ ունենալով վերացնել դրանք,

բ. արդեն ստեղծված համաբարբառները դարձնել էլեկտրոնային,

գ. ընտրել այն հեղինակներին, որոնց երկերի համաբարբառների կազմությունը մյուսների համեմատությամբ ավելի կարևոր է,

դ. ստեղծել լեզվաբանի աշխատանքը մեծ չափով ավտոմատացնող ծրագրային ապահովման համակարգ:

Նախ տեսնենք, թե ինչ եղանակով և որտեղ կարող ենք ներակայացնել էլեկտրոնային համաբարբառները: Տրամաբանական է, որ ժամանակի ընթացքում ավելի մեծ թվով հեղինակների համաբարբառներ են կազմվելու, և տվյալների հենքը շարունակ մեծանալու է, ուստի նպատակահարմար է դրանք որևէ կայքում տեղադրելու փոխարեն ստեղծել հատուկ համաբարբառների համար նախատեսված նոր կայք: Այստեղ լրացուցիչ տեղեկություններ կտրվեն արդեն կազմված և մոտ ապագայում նոր կազմվելիք համաբարբառների, դրանց օգտագործման եղանակների և այլնի մասին:

⁸⁶ Տե՛ս <http://language.sci.am/hy/concordancias>:

Կայքի գլխավոր բաժիններից առաջինը կլինի գրադարանը, որտեղ թվային տարբերակով կներկայացվեն հեղինակների այն ստեղծագործությունները, որոնց համաբարբառները կազմվել են: Սա ընթերցողին այս կամ այն ստեղծագործությունն առցանց կարդալու հնարավորություն կտա, իսկ լեզվաբանը ազատ կարող է օգտվել այդ տեքստից հետազոտական աշխատանքներ կատարելիս:

Կայքի երկրորդ և կարևորագույն հատվածը կլինի համաբարբառների բաժինը: Սովորաբար տպագիր համաբարբառում բառերը ներկայացվում են այբբենական կարգով, էլեկտրոնայինում կարող ենք օգտվողին ընտրելու հնարավորություն տալ. բառերը այբբենական հաջորդականությամբ դասավորվեն, թե՛ ըստ գործածության հաճախության: Այնուհետև, նպատակահարմար է ներկայացնել հեղինակի բառապաշարը՝ սկսած առանձին ստեղծագործություններում եղած բառերից մինչև բոլոր ստեղծագործություններում եղած ամբողջական բառապաշարը՝ հիպերտեքստային հղումներով, գլխաբառերի տակ թեքվող խոսքի մասերի դեպքում դնել բառաձևերը, իսկ բառաձևերի տակ՝ բնագրային համապատասխան օրինակները: Ներկայումս եղած օտարալեզու էլեկտրոնային համաբարբառների մեծագույն մասում այբբենական կարգով դասավորված են ոչ թե բառերը, այլ բառաձևերը: Պատճառն այն է, որ ցանկացած տեքստում գործածված բառաձևերը հեշտությամբ կարելի է դասավորել այբբենական կարգով (իսկ այդ համաբարբառները կազմվում են մեքենական եղանակով), սակայն սա մեզ համար ընդունելի չէ: Ճիշտ է վերականգնել գլխաբառը և նոր միայն դրա տակ ներկայացնել համապատասխան բառաձևը:

Դրանից բացի՝ օտարալեզու համաբարբառների մեծագույն մասում, հիպերտեքստային հղումներից բացի, չկա բառը նարավորություն: Ենթադրենք՝ որևէ մեկին հետաքրքրում է **անրջել** բայի գործածությունը: Համապատասխան բառացանկերում կարդալու փոխարեն նա կարող է այդ բառը, ընտրել տեքստերի ծավալը, որոնցում ցանկանում է տեսնել դրա գործածության բոլոր օրինակները: Սա ևս անչափ կարևոր նոր հնարավորություն է, որը պետք է ունենան հայերեն էլեկտրոնային համաբարբառները:

Այժմ անդրադառնանք ձևային և բովանդակային այն տարբերություններին, որ հանդես կբերեն էլեկտրոնային համաբարբառները նախորդների համեմատությամբ: Այսպես՝ օգտագործման դյուրության տեսակետից էլեկտրոնայիններն այն առավելությունը կունենան, որ հարկ չի լինի թերթելով գտնել բառը: Ավելին՝ եթե մեզ անհրաժեշտ է ձևաբանական փոփոխության ենթարկված որևէ բառաձև, պետք է վերականգնել դրա ուղիղ

ձևը, ապա նոր միայն գլխաբառը գտնել տպագիր համաբարբառում: Մա բավական լուրջ խնդիր է հատկապես գրաբարի համաբարբառներում բառաձևեր որոնելիս, քանի որ գլխաբառը վերականգնելու համար պետք է շատ լավ իմանալ գրաբարի քերականությունը: Մինչդեռ էլեկտրոնային համաբարբառում կարելի է որոնել ցանկացած բառ ու բառաձև և որոնման արդյունքում գլխաբառն իր բառաձևերով կհայտնվի անմիջապես:

Մեքենագիր համաբարբառներում օրինակները բերվում են կիսատ՝ հիմնականում կարճ բառակապակցություններով: Սրա պատճառով հետագոտողը շատ հաճախ ստիպված է բառը փնտրել բնագրում՝ համատեքստում դրա ունեցած իմաստը ճշտելու համար: Էլեկտրոնային համաբարբառներում այս խնդիրը ևս հեշտությամբ լուծելի է: Իբրև ելակետ՝ բոլոր բառերը կարելի է ներկայացնել նախադասությամբ՝ օգտվողին հնարավորություն տալով ընդլայնելու նախադասությունը. ծավալային որևէ սահմանափակում չկա:

Քանի որ մեքենագիր համաբարբառները կազմվում են մարդկանց կողմից, նույնիսկ մեծագույն ուշադրության դեպքում հնարավոր է, որ դուրս մնան բնագրային որոշ օրինակներ կամ դրանք դիտավորյալ չներկայացվեն՝ բառահոդվածի ծավալը չմեծացնելու նպատակով: Էլեկտրոնային համաբարբառը այս խնդիրը ևս չի ունենա. բոլոր օրինակները տվյալների հենքից կընտրվեն մեքենական եղանակով և կպատկերվեն ամբողջական նախադասություններով:

Մեքենագիր համաբարբառները լրջորեն սահմանափակում են հետագոտողի հնարավորությունները լեզվական այս կամ այն երևույթը որևէ ժամանակաշրջանում ուսումնասիրելու առումով: Օրինակ՝ որևէ բառի գործածությունը հինգերորդ դարի հայ մատենագրության մեջ քննելու համար օգտվողը ստիպված է տվյալ բառը փնտրել բոլոր պատմիչների համաբարբառներում, գտնելուց հետո համեմատել դրանք, հաճախության ցանկերից գտնել հաճախության ցուցիչը, կարդալ օրինակները, երբեմն նաև՝ բնագրերի համապատասխան հատվածները: Այս ամենի փոխարեն էլեկտրոնային համաբարբառում կարելի է ընդամենը որոնել ցանկացած բառը՝ նախապես նշելով այն հեղինակներին, որոնց բնագրային օրինակները ցանկանում ենք տեսնել:

Ըստ էության՝ մեքենագիր համաբարբառները կարևոր են միայն բառագիտական մակարդակում, և դրանք՝ որևէ կերպ հնարավոր չէ օգտագործել քերականական իրողությունների ուսումնասիրության ժամանակ: Ենթադրենք՝ որևէ մեկին հետաքրքրում է անորոշ դերբայի կիրառությունը

Պարույր Սևակի «Մարդը ասիի մեջ» ժողովածուում: Տպագիր համաբարբառը որևէ կերպ օգտակար լինել չի կարող, իսկ ահա էլեկտրոնային համաբարբառում բավական է որոնել անորոշ դերբայի

-ել, -ալ վերջավորություններն առանձին և համապատասխան հոլովածներով, և դրա գործածության բնագրային բոլոր օրինակները կպատկերվեն անմիջապես: Այսպես կարելի է ուսումնասիրել քերականական շատ իրողություններ. սա մեկն է այն նոր հնարավորություններից, որ կունենան էլեկտրոնային համաբարբառները:

Այսպիսով՝ հայերեն էլեկտրոնային համաբարբառներում, ինչպես տեսանք, բառերը ներկայացվելու են գլխաբառերով, յուրաքանչյուր գլխաբառի տակ տրվելու են համապատասխան բառաձևերը, բառաձևերի տակ՝ բնագրային օրինակները: Գլխաբառերը դասավորվելու են այբբենական կարգով կամ ըստ հաճախության, տրվելու է բառի որոնման հնարավորություն: Ակնհայտ է, որ տպագիրների համեմատությամբ էլեկտրոնային համաբարբառները հանդես կբերեն զգալի տարբերություններ և կունենան մի շարք առավելություններ:

Նոր համաբարբառներ կազմելուց առաջ նախնառաջ անչափ կարևոր է արդեն եղած տպագիր համաբարբառները (հասոկապես մատենագիրների երկերի) դարձնել էլեկտրոնային ու տեղադրել այդ կայքում՝ ըստ անհրաժեշտության շտկելով դրանցում եղած վրիպակները և, որ ամենակարևորն է, ընդլայնելով բնագրային օրինակների ծավալը:

Ցանկացած դեպքում համաբարբառի կազմությունը բարդ է և աշխատատար, ուստի անհրաժեշտ է ճշտել՝ ի՞նչ չափով է կարևոր տվյալ համաբարբառի կազմությունը: Այն դեպքում, երբ մենք չունենք Հովհ. Թումանյանի, Վ. Տերյանի, Ե. Չարենցի և հայ պոեզիայի մյուս ականավոր ներկայացուցիչների համաբարբառները, որոնք շատ կարևոր են այդ հեղինակների լեզվի և ոճի ուսումնասիրության, բառագործածության, պատկերավորման համակարգի քննության համար, շարունակում են կազմվել գրաբարագիր այնպիսի հեղինակների երկերի համաբարբառներ, որոնք, վերոնշյալների համեմատությամբ, պակաս կարևոր են:

Կարծում ենք՝ ճիշտ կլիներ ստեղծել հայ դասական պոեզիայի կամ արձակի էլեկտրոնային համաբարբառը, որտեղ ներկայացված կլինեն գոնե դասական հեղինակները: Ավելին՝ ընդլայնելով ընդգրկման սահմանը՝ ցանկալի է ներառել նաև ժամանակակից հեղինակներին՝ պատկերացում կազմելու համար արդի հայ պոեզիայի զարգացման միտումների մասին գոնե լեզվագործածության առումով: Այդ կերպ կարելի է տեսնել, թե խոսքի

պատկերավորման համար ժամանակակից հեղինակների գործերում որ բառերն են առավել հաճախադեպ, կարելի է նույնիսկ համեմատել նախորդ և ժամանակակից այլ հեղինակների՝ նույն թեմայով գրված ստեղծագործությունների բառապաշարի հետ:

Օրինակ՝ Վիոլետ Գրիգորյանի «Մեր» վերնագրով պոեմի ամենահաճախադեպ գոյականներն են *մարմին (19) սեր (7), պաշիկ, տոտիկ (6)* բառերը, ամենահաճախադեպ բայերը՝ *արտաշնչել (5), ճոճել (4), համբուրել (4), ճնկվել (3)* բառերը: Վ. Տերյանի բանաստեղծական չորս շարքերում միասին վերցրած («Մթնշաղի անուրջներ», «Փշե պսակ», «Գիշեր և հուշեր», «Ոսկի հեքիաթ») *մարմին* բառը գործածվել է ընդամենը տասը անգամ, իսկ ամենահաճախ գործածված գոյականները *սիրտ, կյանք, հոգի* բառերն են, բայերից ամենաշատը գործածվում են *կանչել, սիրել, երգել* բառերը: Այսպես կարելի է հակադրել սիրտ տերյանական ընկալումը հայ պոեզիայի էրոտիկ թևի ներկայացուցիչ բանաստեղծուհու՝ նյութական իմաստ արտահայտող բառերով սեր զգացմունքի մարմնական դրսևորմանը:

Քանի որ հատկապես արձակագիրների մեծածավալ ստեղծագործությունների համաբարբառների կազմությունը ենթադրում է հսկայական աշխատանք, անհրաժեշտ է պարզել, թե ուրիշ լեզուների համար արդեն ստեղծված համակարգչային ծրագրերը ի՞նչ չափով օգտակար կլինեն մեզ համար, և ի՞նչ նոր ծրագրեր մենք կարող ենք ստեղծել: Օրինակ՝ կան շատ կայքեր, որոնք հայերեն տեքստը հաջողությամբ վերածում են առանձին բառերի, դասավորում այբբենական կարգով, շատ ծրագրերով էլ կարելի է անմիջապես տեսնել որոնված բառի բնագրային բոլոր օրինակները: Ինչ խոսք, այս կերպ մեծապես հեշտանում է համաբարբառ կազմող լեզվաբանի աշխատանքը:

Անշուշտ, կարելի է հաջողությամբ ստեղծել նաև հատուկ ծրագրեր հենց հայերեն համաբարբառների կազմության համար (լեմատիզացիոն ծրագիր), սակայն քանի որ դրանք համաբարբառների կազմության լեզվաբանական կողմին չեն առնչվում, այժմ դրանց չենք անդրադառնա:

Այնուամենայնիվ, բոլոր դեպքերում հնարավոր չէ ամբողջապես մեքենայացնել համաբարբառների կազմությունը, քանի որ համակարգչային ծրագիրը գործնականում չի կարող անթերի կատարել լեզվաբանի աշխատանքը. որևէ ծրագիր չի կարող առանձնացնել հարադրավոր բայերն ու դարձվածքները: Ուստի բոլոր տեսակի եղած և նորաստեղծ համակարգչային ծրագրերը պետք է ընդամենը հեշտացնեն լեզվաբանի աշխատանքը,

իսկ համաբարբառները պետք է շարունակեն կազմվել բացառապես լեզվաբանների կողմից:

Վաղուց արդեն ժամանակն է, որ մեզանում համաբարբառների կազմությունը թեևակոխի որակական զարգացման միանգամայն նոր փուլ. նոր մեթոդներով և տեխնիկական նորագույն միջոցներով ստեղծվեն նոր համաբարբառներ, որոնք կապահովեն փաստական հսկայական նյութ լեզվական տարբեր իրողությունների քննության համար:

Акопян Фрида - Проблемы компьютерной обработки конкордансов. - В статье представлены принципы создания армянских электронных конкордансов в соответствии с которым новый веб-сайт будет создан, где не только печатные конкордансы будут доступны в электронном виде, но и новые конкордансы для разных авторов будут собраны. Важность создания новых лингвистических инструментов подчеркивается, потому что с помощью компьютерных программ можно создать конкордансы, которые будут способствовать изучению языковых явлений.

Hakobyan Frida - The problems of computational processing of concordances. - The article presents principles of creating Armenian electronic concordances according to which a new web site will be created where not only printed concordances will be available in an electronic form, but also new concordances for different authors will be compiled. The importance of creating new linguistic tools is underlined because with the help of software programs it is possible to create concordances which will contribute to investigation of linguistic phenomena.