

ՀՏԴ 519.765/768:681.51

ՀԱՇՎՈՂԱԿԱՆ ՏԵԽՆԻԿԱ ԵՎ
ԻՆՖՈՐՄԱՏԻԿԱ

Է.Ն. ՄԱՆՈՒԿՅԱՆ, Ա.Հ. ՇԱՐԱԲՅԱՆ, Ա.Ս. ՄԱՆՈՒԿՅԱՆ, Ա.Ա. ՄԵԼԻՔՅԱՆ

**ԹԱՐԳՄԱՆՉԱԿԱՆ ՀԱՄԱԿԱՐԳԵՐՈՒՄ ԿՈՆՏԵՔՍՏ-ԿԱԽՅԱԼ
ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ ԻՐԱԿԱՆԱՑՆՈՂ ՄԵՏԱԼԵԶՎԻ ՄՇԱԿՈՒՄԸ**

Հիմնավորված է մեքենայական թարգմանության համակարգերում հատուկ մասնագիտացված մետալեզվի օգտագործման անհրաժեշտությունը, տրված են այդ մետալեզվի օպերատորներին ներկայացվող պահանջները, ինչպես նաև՝ այդ օպերատորների և դրանցում օգտագործվող օպերանդների նկարագրությունները:

Առանցքային բառեր. իմաստային թարգմանիչ, գիտելիքների բազա, կոնցեպտ, մետալեզու, կոնտեքստ վերլուծություն, օպերատոր, օպերանդ:

Ըստ [1]-ի՝ ժամանակակից մեքենայական թարգմանության համակարգերում ընդունված մոտեցումներից մեկը միջին լեզվի օգտագործման միջոցով թարգմանության կազմակերպումն է: Այս մոտեցման դեպքում գիտելիքների բազայի միջոցով իրականացվում է տեքստի բազմափուլ վերլուծություն, ընդ որում, վերլուծության յուրաքանչյուր փուլում զգացվում է կոնտեքստ վերլուծության իրականացման անհրաժեշտություն: Վերջինիս իրականացման կանոնները տարբեր են՝ կախված թարգմանության մուտքային և ելքային լեզուներից, և կարող են տրվել միայն մասնագետ լեզվաբանի կողմից: Նմանօրինակ վերլուծության անհրաժեշտությունն ավելի հաճախ առաջանում է տարբեր լեզվախմբերին պատկանող լեզուներից մեկից մյուսը թարգմանություններ կատարելիս, օրինակ՝ անգլերեն-հայերեն թարգմանության դեպքում:

Դիտարկենք անգլերեն-հայերեն թարգմանություններին բնորոշ խնդիրներ, որոնց լուծման համար անհրաժեշտ է կատարել տեքստի կոնտեքստ վերլուծություն:

Այս լեզուներն ունեն բավականին տարբեր թե բառապաշար և թե քերականական կառուցվածքներ: Ըստ [2]-ի և [3]-ի՝ բառապաշարների տարբերությունների հետ կապված խնդիրները լուծվում են գիտելիքների բազայի կազմակերպչական միավորների՝ կոնցեպտների կազմի և դրանց միջև տրվող իմաստային կապերի ճիշտ ընտրության միջոցով: Սակայն լեզուների քերականական կառուցվածքների տարբերություններից բխող խնդիրը հնարավոր չէ լուծել նշված ճանապարհով: Ստորև ներկայացվում են անգլերենի և հայերենի քերականական կառուցվածքների որոշ տարբերություններ և դրանցից բխող թարգմանչական խնդիրները:

Ենթակա – ստորոգյալ – խնդիր կապակցությունների տարբերությունը երկու լեզուներում: Մեքենայական թարգմանությունը հիմնված է մուտքային նախադասության բառերի միջև շարահյուսական հարաբերությունների բացահայտման վրա:

Հայտնի է, որ անգլերենում նշված հարաբերություններ կազմող բառերի փոխդասավորություններն ավելի կոշտ են ու ամրակայված: Իսկ հայերենում այդ հարաբերություններում բառերի շարադասությունն անհամեմատ ավելի ճկուն է: Այս հանգամանքը թույլ չի տալիս, որ հայերեն-անգլերեն անցման ժամանակ կիրառվի բառացի թարգմանության սկզբունքը: Օրինակ, դիտարկենք հետևյալ նախադասությունը. <<Մարդը դրեց գիրքը սեղանին: – The man put the book on the table>>: Այս նախադասությունը հայերենում կարելի է գրել նաև հետևյալ շարադասությամբ. <<Մարդը գիրքը սեղանին դրեց>>: Սակայն եթե այս դեպքում փորձենք կատարել հայերենից անգլերեն բառ առ բառ թարգմանություն, ապա կստացվի՝ The man the book on the table put, որն անգլերենում բառերի անիմաստ շարան է:

Որոշիչ-որոշյալ կապակցությունների տարբերությունը երկու լեզուներում:

Հայերենում որոշիչ-որոշյալ կապակցության դեպքում որոշիչը դրվում է որոշյալից առաջ (բացառություն է միայն հետադաս որոշիչը, երբ որոշիչ-որոշյալն անջատվում են բուրժով): Անգլերենում որոշիչը որոշյալից առաջ կարող է լինել միայն այն դեպքում, եթե այն ոչ մի լրացում չունի (Օրինակ՝ գեղեցիկ աղջիկ – beautiful girl): Սակայն եթե որոշիչն արտահայտված է դերբայով և որևէ լրացումով (նախադաս դերբայական դարձված է), օրինակ՝ <<ջուր խմող մարդ>>, ապա այս դեպքում արդեն անգլերենում որոշիչն իր լրացման հետ միասին պետք է գրվի որոշյալից հետո: Հետևաբար, նշված օրինակի թարգմանությունը կլինի՝ the man drinking water (այլ ոչ թե՝ water drinking man):

Առաջին երկու խնդիրները լուծելու համար լեզվաբանը պետք է ունենա թարգմանության միջանկյալ փուլերում ստացվող նախադասության իմաստն արտահայտող ծառի ճյուղերի (կոնցեպտների) հետ մանիպուլյացիաներ կատարելու հնարավորություն, որոնց շնորհիվ նա, մասնավորապես, կկարողանա փոխել ճյուղերի դասավորությունը ծառի մեջ՝ ըստ ելքային լեզվի շարադասության կանոնների (ըստ [2]-ի, [4]-ի և [5]-ի):

Ենթակա – ստորոգյալ հարաբերությունների դեպքում բայի դեմքի և թվի հետ կապված խնդիրները: Անգլերենում հաճախ տարբեր դեմքեր կամ թվեր ունեցող բայերն արտահայտվում են նույն կազմությամբ: Օրինակ՝ Ես գնում եմ: – I go. Մենք գնում ենք: – We go. նախադասություններում <<գնալ>> բայի կազմությունը նույնն է՝ go. Դա կարող է հանգեցնել նրան, որ, օրինակ՝ <<I go home.–Ես գնում եմ տուն:>> նախադասությունը թարգմանվի՝ <<Ես գնում ենք տուն:>>: Այս խնդիրը լուծելու համար լեզվաբանը պետք է հնարավորություն ունենա բայի կազմության վերլուծության ընթացքում շարահյուսական կապեր ստեղծելիս այդ կապը ընդունելու համար անհրաժեշտ ստուգումներ կատարել (նշված օրինակում անհրաժեշտ է ստուգումով համոզվել, որ ենթակայի դեմքը և թիվը համընկնում են ստորոգյալի դեմքին ու թվին) և կախված դրանց արդյունքից՝ ընդունելի համարել կամ չհամարել տվյալ շարահյուսական կապը:

Անգլերեն նախադասություններում կրկնակի ժխտման բացառումը: Եթե նախադասության մեջ կա ժխտում արտահայտող բառ, ապա նման դեպքերում անգլերենում բայը չի կարող օգտագործվել ժխտական խոնարհմամբ: Օրինակ, <<Ես երբեք չեմ տեսել Փարիզը – I have never seen Paris. >>: Այս նախադասության մեջ << երբեք >> բառը (never) ժխտում է արտահայտում, ինչը, սակայն, չի խանգարում, որ հայերենում բայը ևս ժխտված լինի՝ <<չեմ տեսել>>: Իսկ անգլերենում բայը չի կարող ժխտական խոնարհմամբ լինել, այսինքն՝ I haven't never seen Paris. նախադասությունը սխալ է: Այս խնդիրը լուծելու համար լեզվաբանը պետք է հնարավորություն ունենա տեքստի վերլուծության վերջնական փուլում (երբ տեքստի վերլուծությունն արդեն ավարտվել է, ստացվել է դրա իմաստն արտահայտող ծառը, և մնացել է միայն այն վերածել ելքային լեզվով գրված տեքստի) փոխել նախադասության անդամների տարբեր պարամետրերը (նշված օրինակում պետք է փոխել բայի ժխտականության հայտանիշը):

Անգլերենում օգտագործվող բայական ֆրազները և դրանցից բխող խնդիրները: Անգլերենում գոյություն ունեն բազմաթիվ բայական ֆրազներ, որոնց մեջ բային կցվող օժանդակ մասնիկը (որը կարող է լինել ինչպես անմիջապես բայից հետո, այնպես էլ դրանից հեռացված) նախդիր կամ հետդիր է հանդիսանում, ընդ որում, այդ մասնիկը հիմնականում նախադասության մեջ առանձին իմաստային միավոր չի ներկայացնում: Թարգմանության ժամանակ անհրաժեշտություն է առաջանում գտնել նման անգլերեն բայական ֆրազների հայերեն համարժեքները: Ենթադրենք՝ ունենք հետևյալ նախադասությունը. “The car walked into the bus.”: Այս նախադասության մեջ “walk” բառը նշանակում է ոչ թե <<քայլել>>, այլ <<հարվածել>>, քանի որ այն “into” նախդրի հետ կազմում է բայական ֆրազ: Հետևաբար, այս նախադասության ճիշտ թարգմանությունը կլինի՝ <<Ավտոմեքենան հարվածեց ավտոբուսին>>, այլ ոչ թե՝ <<Ավտոմեքենան քայլեց ավտոբուսի մեջ>>, ինչպես կլիներ բառացի թարգմանության դեպքում: Բերենք ևս մեկ օրինակ. ”I switched the TV on.”: Այս նախադասության մեջ “switch” բառը նշանակում է ոչ թե՝ փոխանջատել (որը դրա հիմնական իմաստն է), այլ <<միացնել>>, քանի որ այն “on” նախդրի հետ կազմում է բայական ֆրազ այն տարբերությամբ, որ այս դեպքում արդեն նախդիրը հեռացված է բայից: Հետևաբար, այս նախադասության ճիշտ թարգմանությունը կլինի. <<Ես միացրեցի հեռուստացույցը>>, այլ ոչ թե՝ <<Ես փոխանջատեցի հեռուստացույցի վրա>>, ինչպես կլիներ բառացի թարգմանության դեպքում: Այս խնդիրը լուծելու համար թարգմանչական համակարգը պետք է կարողանա տեքստի վերլուծության սկզբնական փուլում ճանաչել բայական ֆրազները և փոխարինել դրանք իրենց համարժեք կոնցեպտներով, որպեսզի բայական ֆրազներ պարունակող տեքստերի վերլուծությունը (որը ներկայացնում է նախադասության բառերին համարժեք կոնցեպտների շարանի վերլուծությունը) կախված չլինի այդ տեքստերում ֆրազների առկայությունից:

Այսպես կոչված՝ <<ճկուն>> ֆրազները և դրանցից բխող խնդիրները: Անգլերեն քերականության մեջ գոյություն ունի բայական ֆրազների մի կառուցվածք, որի էությունը

յունը բացատրենք կոնկրետ օրինակով: Դիտարկենք հետևյալ անգլերեն նախադասությունը. “The man covered with dust.” - <<Մարդը փոշոտվեց>>: Սակայն եթե այս նախադասությունը թարգմանենք բառացի, ապա կստացվի. <<Մարդը ծածկվեց փոշով>>: Թվում է, թե այս տիպի ֆրազները ոչնչով չեն տարբերվում սովորական բայական ֆրազներից, սակայն դա այդպես չէ: Ի տարբերություն սովորական բայական ֆրազների՝ այս դեպքում ֆրազը ձևավորվում է՝ կախված ոչ թե բային կից բառից, այլ շարահյուսական այն կապից, որով տվյալ բառը կապվում է հիմնական բային: Օրինակ, եթե <<փոշի>> բառի փոխարեն գրենք <<ալյուր>> (flour)՝ “The man covered with flour.”, ապա այս նախադասությունը կթարգմանվի. <<Մարդն ալյուրոտվեց>>: Այսօրինակ ֆրազները մենք կանվանենք <<ճկուն>> ֆրազներ: Այս և նմանատիպ այլ օրինակների առկայությունը անգլերեն տեքստերում ստիպում է, որ դրանց դեպքում կատարվի հատուկ վերլուծություն՝ ֆրազը հասկանալու համար: Ընդ որում, այս ֆրազները պետք է ճանաչվեն ոչ թե տեքստի վերլուծության սկզբնական փուլում, ինչպես նախորդ դեպքում էր, այլ վերջում, երբ նախադասության ծառն արդեն ստացված է, քանի որ այս դեպքում ֆրազի ստեղծման հանգամանք է հանդիսանում միայն խիստ որոշակի շարահյուսական կապի առկայությունը: Այս խնդիրը լուծելու համար լեզվաբանը պետք է հնարավորություն ունենա կատարելու մանիպուլյացիաներ նախադասության իմաստն արտահայտող ծառի ճյուղերի (կոնցեպտների) հետ տեքստի վերլուծության վերջնական փուլում, ինչպես նաև, անհրաժեշտության դեպքում, ապահովել ծառի ճյուղի որոնումը (ըստ տրված որոնումային բանալիների), հեռացումը, ավելացումը կամ դրա պարամետրերի փոփոխումը:

Վերը նշված խնդիրները լուծելու համար առաջարկվում է մուտքային տեքստերի վերլուծությունն իրականացնող և կոշտ ալգորիթմով աշխատող ծրագրային միջոցներում ներդնել մոդուլներ, որոնց աշխատանքի ալգորիթմն արդեն կախված կլինի վերլուծվող նախադասության բառերը ներկայացնող կոնցեպտներից և դրանց միջև հաստատվող շարահյուսական կապերից: Այդ մոդուլները պետք է իրականացնեն կոնտեքստ-կախյալ քերականական վերլուծություն, իսկ, ինչպես նշվեց վերևում, վերլուծության ալգորիթմները խիստ բազմազան են և կախված են ընտրվող կոնցեպտի կամ շարահյուսական վերլուծության ծառի մեջ ստեղծվող կապի շրջապատից: Նշված վերլուծության կանոնները կարող են տրվել միայն լեզվաբանների կողմից: Հետևաբար, որպեսզի լեզվաբանը չծանրաբեռնվի ծրագրավորման հոգսով և պրոբլեմներով, անհրաժեշտ է թարգմանությունը կազմակերպելու համար օգտագործել պրոբլեմ-օրիանտացված մետալեզու, որով լեզվաբանը կարտահայտի կոնտեքստ-կախյալ վերլուծության իրեն հայտնի կանոնները: Այսպիսով, այս կանոնների իրականացումը տեխնիկապես պահանջում է վերը նշված մետալեզվով ներկայացված տեքստերի ինտերպրետացիա: Վերոհիշյալ նշված ճկուն ալգորիթմներով աշխատող մոդելները փաստորեն ներկայացնում են ունիվերսալ ինտերպրետատորի կանչ, իսկ մետալեզվով ներկայացված ինտերպրետացվող ծրագրերը, բնական է, պետք է կցվեն կոնցեպտ-

ներին և պահվեն արդեն գիտելիքների բազայում: Վերը նշված խնդիրները լուծելու համար անհրաժեշտ է, որ մետալեզվով գրված կոնտեքստ-կախյալ վերլուծություն իրականացնող ծրագրերը կանչվեն և ինտերպրետացվեն տեքստի անալիզի (երբ մուտքային տեքստը վերածվում է ծառի) կամ սինթեզի (երբ ծառը վերածվում է ելքային տեքստի) որոշակի փուլերի կատարման ընթացքում: Ընդ որում, նշված փուլերում մշակվող տեղեկույթը (մուտքային տեքստի ձևափոխված տեսքերը) ներկայացվում է հետևյալ երեք տեսքերով.

- որպես բառերի գծային շարան,
- որպես բառերի իմաստներին համապատասխանող կոնցեպտների շարան,
- տեքստի հնարավոր իմաստներն արտահայտող ծառանման գրաֆների բազմություն:

Տեքստի ներկայացման նշված ձևերից բխում են պահանջները մետալեզվի օպերատորների նկատմամբ: Մետալեզուն, բացի դասական համարվող վերագրման, անցման և ցիկլի կազմակերպման օպերատորներից, պետք է ունենա նաև այնպիսի օպերատորներ, որոնք կատարում են տեքստում բառի (կոնցեպտի) դիրքի ամրակայում, տեղաշարժեր բառերի կամ կոնցեպտների շարվածքում, վերընթաց և վայրընթաց շարժումներ տեքստի իմաստային ծառի ճյուղերով, ծառի ճյուղերի (ենթածառերի) ավելացում, հեռացում կամ տեղափոխում:

Տեքստի ներկայացման վերը նշված երեք ձևերը յուրահատուկ պահանջներ են առաջադրում նաև մետալեզվում օգտագործվող օպերանդներին: Լեզվաբանին մետալեզվով գրված ծրագրերի դաշտերի նկարագրման անհրաժեշտությունից ազատելու համար մետալեզվում օգտագործվում են ամրագրված անուններով և կառուցվածքով օպերանդներ, որոնք ցույց են տրված աղյուսակում:

Մետալեզվով գրված ծրագրերը կանչվում և ինտերպրետացվում են տեքստի անալիզի կամ սինթեզի կոնկրետ փուլերում: Մեքենայական թարգմանության համակարգի միջուկով այդ կանչերը կառավարելու համար անհրաժեշտ է, որ գիտելիքների բազայում տեղադրված ծրագրի վերնագիրն ունենա հետևյալ կառուցվածքը.

[LANGUAGE] PROGRAM [ANALYS] կամ {SYNTHESES} [(STAGE<փուլի_համար>)]:

Այստեղ [LANGUAGE] դաշտը ներկայացնում է թարգմանվող լեզվի սիմվոլային ծածկագիրը, [ANALYS] և {SYNTHESES} դաշտերի միջոցով նշվում է՝ ծրագիրը տեքստի՞ց ծառ, թե՞ ծառից տեքստ անցման ժամանակ պետք է կանչվի, իսկ [(STAGE<փուլի_համար>)] դաշտով նշվում է տեքստի վերլուծության փուլի համարը, որի ընթացքում պետք է ակտիվացվի տվյալ ծրագիրը:

Օրինակ, եթե որևէ կոնցեպտին կցված է ENG PROGRAM ANALYS (STAGE 3) վերնագիրն ունեցող ծրագիր, ապա այն պետք է ակտիվացվի անգլերեն տեքստի վերլուծության երրորդ փուլում այն դեպքերում, երբ տեքստի բառերի շարվածքում առկա է բառը, որին հավակնում է տվյալ կոնցեպտը:

Օպերանդի անվանումը	Նկարագրությունը
int <համար>	Այս օպերանդներն օգտագործվում են ամբողջ թվեր պահելու համար: Օրինակ՝ int, int1, int2 և այլն:
x <համար>	Այս օպերանդներն օգտագործվում են իրական թվեր պահելու համար: Օրինակ՝ x, x1, x2 և այլն:
pos <համար>	Այս օպերանդներն օգտագործվում են տեքստում բառի (կոնցեպտի) դիրքի համարը պահելու համար: Օրինակ՝ pos, pos1, pos2 և այլն:
pos.<շարահյուսական պարամետր>	Այս օպերանդներն օգտագործվում են տեքստում որևէ դիրքով բառի (կոնցեպտի) շարահյուսական տարբեր պարամետրերը պահելու համար: Օրինակ՝ pos.case (պահվում է pos դիրքում գտնվող բառի հոլովը), pos.person (պահվում է pos դիրքում գտնվող բայի կամ դերանվան դեմքը) և այլն:

Նկարագրված մետալեզոն իր ատրիբուտներով և դրանով գրված ծրագրերի կանչի կազմակերպման սխեման ստեղծվել և զարգացվել են անգլերեն - հայերեն մեքենայական թարգմանության համակարգի մշակման ընթացքում, և մետալեզվում առաջարկված լուծումների արդյունավետությունը գործնականորեն ապացուցվել է այդ համակարգի երկարամյա շահագործման ընթացքում:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. Манукян А.С., Казарян Р.Р. Лингвистические основы организации процесса смыслового машинного перевода // Вестник ГИУА (Политехник): Сборник научных и методических статей. - Ереван, 2010. - Том 2, № 1, – С. 270-276.
2. Մանուկյան Ա. Ս., Շարաբջյան Ա. Հ., Մելիքյան Ա. Ա., Մանուկյան Է. Ն. Մեքենայական թարգմանության համակարգերում գիտելիքների բազաների ձևավորման սկզբունքները // ՀՀ ԳԱԱ և ՀՊՃՀ Տեղեկագիր. Տեխն. գիտ. սերիա.- 2011.- Հ. 64, № 3.- էջ 257-265:
3. Շարաբջյան Ա.Հ., Մանուկյան Ա.Ս. Գիտելիքների բազաների ձևավորման մի մեթոդ մեքենայական թարգմանության համակարգերում // ՀՊՃՀ Լրաբեր.- Երևան, 2011.- Հ. 3, № 1.- էջ 223-227:
4. Манукян Э.Н., Манукян С.Н., Манукян А.С., Саядян А.С. Основы построения интеллектуальной системы машинного перевода текстов // Сборник материалов годичной научной конференции ГИУА. - Ереван, 2001.- Т. 1. – С. 306.
5. Arnold D., Balkan L., Meijer S. MACHINE TRANSLATION // An Introductory Guide, Linguistics, University of Essex, Colchester, CO4 3SQ, UK, doug@essex.ac.uk. - March 6, 2001. – P. 1-124.

ՀՊՃՀ (ՊՈԼԻՏԵԽՆԻԿ): Նյութը ներկայացվել է խմբագրություն 10.11.2012:

Э.Н. МАНУКЯН, А.А. ШАРАБЧЯН, А.С. МАНУКЯН, А.А. МЕЛИКЯН
ПРИМЕНЕНИЕ МЕТАЯЗЫКА В БАЗАХ ЗНАНИЙ СИСТЕМ МАШИННОГО
ПЕРЕВОДА

Обоснована необходимость применения особого специализированного метаязыка в системах машинного перевода. Заданы требования, которым должны соответствовать операторы этого метаязыка, а также приведены описания этих операторов и использованных в них операндов.

Ключевые слова: смысловой переводчик, база знаний, концепт, метаязык, контекстный анализ, оператор, операнд.

E.N. MANUKYAN, A.H. SHARABCHIAN, A.S. MANUKYAN, A.A. MELIKYAN
APPLYING A METALANGUAGE IN THE KNOWLEDGE BASES OF THE
MACHINE TRANSLATION SYSTEMS

The necessity of applying a specialized metalanguage in machine translation systems is substantiated, the requirements are set to the operators of this metalanguage, as well as descriptions of those operators and the operands used in them are introduced.

Keywords: semantic translator, knowledge base, concept, metalanguage, context analysis, operator, operand.