

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

ՄԵՔԵՆԱՅԱԿԱՆ ԹԱՐԳՄԱՆՈՒԹՅԱՆ ՀԱՄԱԿԱՐԳԵՐՈՒՄ
ԳԻՏԵԼԻՔՆԵՐԻ ԲԱԶԱՆԵՐԻ ՁԵՎԱՎՈՐՄԱՆ ՄԿՋԲՈՒՆՔՆԵՐԸ

Առաջարկված է մեքենայական թարգմանության համակարգերի բաղադրիչ մաս հանդիսացող գիտելիքների բազայի ձևավորման ֆորմալ մի մեթոդ, որը թույլ է տալիս որոշել գիտելիքների բազայի հիմնական միավորների կազմը, այդ միավորների միջև հաստատվող իմաստաշարահյուսական հարաբերություններն արտահայտող կապերի կազմը և այդ կապերի ծածկագրման եղանակները:

Առանցքային բառեր. իմաստային թարգմանիչ, ինֆորմացիոն բազա, գիտելիքների բազա, կոնցեպտ, շարահյուսական հարաբերություն, իմաստային հարաբերություն, կոնցեպտների ստրուկտուրավորում, մուլտիպլետ, հարաբերական սխալ:

Ժամանակակից մեքենայական թարգմանության համակարգերում (ՄԹՀ) առավել արդյունավետ է համարվում դրանց աշխատանքի կազմակերպման հետևյալ սխեման.

1. Մուտքային տեքստը (տեքստի հատվածները) ձևափոխվում և բերվում են որոշակի մաթեմատիկական ֆորմալ տեքսի, որն այդ տեքստում առկա օբյեկտները և պրոցեսներն արտապատկերող փաստերի իմաստային նմանակը;
2. Ստացված նմանակի հիման վրա գեներացվում են ելքային տեքստի նախադասությունները՝ ելքային լեզվի բառարանների և քերականական կանոնների հիման վրա:

Նշված իմաստային նմանակների ներկայացման ընդունված ձև են համարվում n չափանի ռելացիաները: Ըստ [1]-ի և [3]-ի՝ ավելի հարմար է օգտագործել բինար (երկչափ $n = 2$) ռելացիաներ՝

$$G(L1, L2) \xrightarrow{Ph} R(k1, k2), \quad (1)$$

որտեղ $L1$ -ը և $L2$ -ը մուտքային նախադասության լեկսեմներն են, $k1$ -ը և $k2$ -ը $L1$ և $L2$ լեկսեմների իմաստային համարժեքները (որոնց այսուհետև կանվանենք կոնցեպտներ), $G(L1, L2)$ -ը (հետագայում $G12$) $L1$ և $L2$ լեկսեմների միջև հնարավոր շարահյուսական հարաբերությունն է, $R(k1, k2)$ -ը (հետագայում $R12$) այդ շարահյուսական հարաբերության իմաստային համարժեքը, իսկ Ph -ը (1) անցման կանոնը, որը երկչափ (բինար) հարաբերությունների դեպքում ներկայացվում է միարժեք համապատասխանության կանոններով: (1)-ում $L1$ -ի և $L2$ -ի միջև բինար հարաբերությունների գնահատման ժամանակ հարկավոր է ՄԹՀ-ի շարահյուսական

վերլուծության փուլը փոխարինել իմաստա-շարահյուսական վերլուծության փուլով: Մուտքային նախադասության իմաստային նմանակը (1) ձևափոխության դեպքում կստացվի ծառանման գրաֆի տեսքով, որի հանգույցները նախադասության Li անդամների իմաստային համարժեք հանդիսացող ki կոնցեպտներն են, իսկ դրանց միջև կապերը համապատասխանում են Li-երի միջև եղած շարահյուսական հարաբերություններին: Վերը նշված գրաֆը ստանալու համար ՄԹՀ-ն պետք է ունենա ինֆորմացիոն բազա, որի համար առաջարկվում է հետևյալ ենթաբազաներից կազմված կառուցվածքը.

1. շարահյուսական նկարագրի ենթաբազաներ,
2. գիտելիքների ենթաբազա (ԳԲ):

Նախադասության մեջ երկու բառերի միջև հարաբերությունը ունի երկու բաղադրիչ՝ շարահյուսական և իմաստային (ըստ [2]-ի): Բազմալեզու թարգմանիչներում նույն իմաստային հարաբերությանը կարող են համապատասխանել տեսքով տարբեր շարահյուսական հարաբերություններ: Այս հանգամանքը կարող է դժվարացնել թարգմանության ընթացքում մուտքային տեքստի բառերի միջև առկա շարահյուսական հարաբերությունների համարժեքի որոնումը ելքային տեքստում: Այս խնդիրը լուծելու համար առաջարկվում է յուրաքանչյուր R_{ij} հարաբերության իդենտիֆիկատորը (տվյալ դեպքում կապի կոդը) ներկայացնել երկու բաղադրիչով՝

$$R_{ij} = (R_0, R_1)_{ij}, \quad (2)$$

որտեղ R_0 -ն իմաստային բաղադրիչն է (արտահայտում է հարաբերությունների էությունը), իսկ R_1 -ը՝ դիտարկվող լեզուներում միևնույն R_0 իմաստային հարաբերությունն արտահայտող ձևաբանական տեսքերի հնարավոր կոմբինացիաների համարը: Վերը նշված բաղադրիչների ներմուծման անհրաժեշտությունը դիտարկենք բայ-գոյական հարաբերությունների օրինակով: Նշենք բայ-գոյական հարաբերությունների R_0 բաղադրիչի արտահայտման ձևերը: Այդ հարաբերությունները բաժանվում են երկու խմբի՝ հարաբերություններ, որոնք բնորոշ են որոշակի տիպի բայերին, և հարաբերություններ, որոնք կարող են լինել ցանկացած բայի մոտ: Առաջին խմբին պատկանող հարաբերությունները բխում են բայի էությունից, իսկ երկրորդ խմբին պատկանողները ցույց են տալիս տվյալ բայով ներկայացվող գործողության հետ կապված հատկանիշները (ձև, տեղ, ժամանակ և այլն): Բայ-գոյական հարաբերությունների համար առաջարկվում է R_0 բաղադրիչի հետևյալ կազմը՝ հայտնի ենթակայական ($R_0 = 1$) և ուղիղ խնդրի ($R_0 = 2$) հարաբերություններ; ուղիղ իմաստով հանգման ($R_0 = 3$), անջատման ($R_0 = 4$) և գործողության իրականացման միջոցի ($R_0 = 5$) հարաբերություններ; փոխաբերական իմաստով հանգման ($R_0 = 6$), անջատման ($R_0 = 7$) և գործողության իրականացման միջոցի ($R_0 = 8$) հարաբերություններ:

Օրինակ՝ <<Ես դրեցի գիրքը սեղանին: - I put the book on the table. >> կամ <<Ես գրկեցի մայրիկին: - I embraced the mother.>> նախադասությունների մեջ բայն ունի ուղիղ իմաստով հանգման հարաբերություն ($R_0 = 3$): Սակայն եթե գրենք՝ <<Ես դիմեցի նրան: - I applied to him.>>, ապա այս դեպքում բայն ունի արդեն

փոխաբերական իմաստով հանգման ($R_0=6$) հարաբերություն: Բերված օրինակներից երևում է նաև R_{ij} -ի R_i բաղադրիչի ներմուծման անհրաժեշտությունը, քանի որ դրանց մեջ նույն իմաստային R_0 հարաբերությունն արտահայտվում է տարբեր կերպ: Այս օրինակներում ակնհայտ է, որ հայերենից անգլերեն $R_0=3$ հարաբերության արտահայտման ձևերը երկուսն են՝

- $R_0 = 3, R_1 = 1$ – ուղիղ իմաստով հանգման հարաբերություն, հայերենում տրական հոլովով արտահայտված, իսկ անգլերենում օո նախդիրով կազմվող;
- $R_0 = 3, R_1 = 2$ – ուղիղ իմաստով հանգման հարաբերություն, հայերենում տրական հոլովով արտահայտված, իսկ անգլերենում առանց նախդիրի:

Իմաստաշարահյուսական վերլուծության համար պետք է, որ (3,1) և (3,2) հարաբերությունների ձևաբանական կողմն արտահայտված լինի հայերեն և անգլերեն լեզուների շարահյուսական նկարագրի ենթաբազաններից յուրաքանչյուրում, ինչպես նաև այդ հարաբերությունները տրված լինեն ԳԲ-ում:

Այժմ դիտարկենք կոնցեպտների կազմի ընտրման հարցը: Պարզ է, որ այս հարցի լուծման համար ելակետ են հանդիսանում թարգմանվող լեզվի բառերը: Եթե հայտնի են մուտքային A լեզվի բառերը՝ LA_i , ապա դրանցից յուրաքանչյուրի ներկայացման համար ԳԲ-ում պետք է լինի ki կոնցեպտ, որի անունը A լեզվում պետք է համընկնի LA_i -ի հետ: Եթե LA_i բառի թարգմանությունը B լեզվում LB_i -ն է, ապա այդ B լեզվում ki կոնցեպտի անունը պետք է լինի LB_i : Սակայն հաճախ նույն LA_i -ին B -ում կարող են համապատասխանել հոմանիշ չհանդիսացող տարբեր բառեր ($L_1B_i, L_2B_i, \dots, L_mB_i$): Այս դեպքում հնարավոր չի լինի կատարել A -ից B թարգմանությունը վերը նշված սխեմայով թարգմանության սինթեզի փուլում B լեզվում առաջացած անորոշության պատճառով: Խնդիրը լուծելու համար պետք է բազմացնել ki կոնցեպտը, այնպես որ տեղի ունենան իրար հետ չհամընկնող հետևյալ եղանակները (ըստ [3]-ի և [4]-ի).

$$LA_i \rightarrow k_{ij} \rightarrow L_i B_i, i = 1, \dots, n, \quad (3)$$

որտեղ k_{ji} –ով արտահայտված են LA_i բառի բոլոր իմաստները, $L_i B_i$ –ով՝ ելքային B լեզվում ամեն j -րդ իմաստն արտահայտող բառերը: Նման իրավիճակի առաջացումը վկայում է, որ LA_i -ն բազմիմաստ բառ է, և նրա իմաստներն արտահայտվում են տարբեր կոնցեպտներով: Ցանցային բազաներում կոնցեպտների հատկություններն արտահայտվում են իրենց կապերով: Հետևաբար, վերը նշված եղանակներում k_{ji} -երից յուրաքանչյուրը պետք է ունենա իր՝ մյուսներից տարբերվող կապերի կազմը: Այդ կապերը բնորոշվում են կապի տիպով և կապվող կոնցեպտով: Եթե խմբավորենք կապերն ըստ կապի իդենտիֆիկատորի տիպի, ապա այդ կապերի բազմությունը կարելի է ներկայացնել որպես՝

$$U = \bigcup_{k=1}^t R_k S_k, \quad (4)$$

որտեղ t -ն տվյալ կոնցեպտի մոտ տրվող տարբեր կապերի տիպերի քանակն է, իսկ (S_0 -ն բազայում առկա բոլոր կոնցեպտների բազմությունն է): Սովորաբար $t > 10$, իսկ կապերի տիպերը որոշվում են լեզվաբանի կողմից՝ նախադասության մեջ բառերի օգտագործման սխեմաների հիման վրա: Մյուս կողմից S_k -ի հզորությունը կախված է բազայի հզորությունից և կարող է ընդունել շատ մեծ արժեքներ: Բազայի ձևավորման ընթացքում կոնցեպտների ցանկացած $\{i, j\}$ զույգի համար պետք է, որպեսզի i -րդ կոնցեպտի U_i կապերի բազմությունը հավասար չլինի j -րդ կոնցեպտի U_j կապերի բազմությանը, ինչպես նաև այդ բազմությունները ընդգրկված չլինեն իրար մեջ, այսինքն՝

$$U_i / U_j \neq 0, U_j / U_i \neq 0, (i = 1, \dots, n, j = 1, \dots, n, i \neq j), \quad (5)$$

որտեղ n -ը բազայում առկա կոնցեպտների քանակն է: Այս դեպքում արդեն հնարավոր է դառնում լուծել LA բառի բազմիմաստության խնդիրը: Դրա համար թարգմանվող նախադասության մեջ որոշվում են LA -ի շարահյուսական կապերը նախադասության հարևան բառերի հետ: Այդ կապերի կազմի համեմատումը LA -ի հավակնորդ կոնցեպտների հարաբերությունների հետ հնարավորություն է տալիս որոշել LA -ի ճիշտ հավակնորդ կոնցեպտը նախադասության մեջ՝ B լեզվում ընտրել LA -ի ճիշտ իմաստը: ՄԹՀ-երում ԳԲ-ն ունենում է որոշիչ դեր, սակայն նրա ձևավորումը պահանջում է ծավալուն ինֆորմացիայի մշակում: Բացի այդ, ԳԲ-ն որպես կապակցված կոնցեպտների ամբողջություն դիտարկելով՝ կարելի է պնդել, որ վերջինս հանդիսանում է այդ կոնցեպտներով ներկայացվող առարկայական աշխարհի սխեմատիկ մոդելը: Իրական աշխարհի հետ կապված տեքստերի վերլուծությունը ցույց է տալիս, որ այդ տեքստերի բառերի միջև եղած շարահյուսաիմաստային հարաբերությունները (տես (2) բանաձևը) ունենում են խիստ անհավասարաչափ խտությամբ հանդիպման հաճախություններ: Այսպիսով, տեքստերի վերլուծության բարձր ճշտություն ապահովելու համար անհրաժեշտ է, որ ԳԲ-ի կոնցեպտների միջև տրվող հարաբերություններին կցվի ևս մի պարամետր՝ f , որը ցույց կտա նախադասությունների մեջ այդ հարաբերության հանդիպման հաճախությունը: Այսպիսով (2)-ը կընդունի արդեն հետևյալ տեսքը՝

$$R_{ij} = (R_0, R_1, f)_{ij}: \quad (6)$$

ԳԲ-ի ստեղծումը պահանջում է, որ դրա սկզբնական ձևավորումը կատարվի լեզվաբանի կողմից, ով պետք է լուծի բառերի բազմիմաստության խնդիրը և որոշի ԳԲ-ի կոնցեպտների կազմը: ԳԲ-ի սկզբնական ձևավորումից հետո միայն ծրագրային փաթեթների օգնությամբ կարելի է վերահսկել (3) և (5) պայմանները և առաջարկել ճշգրտումներ (6)-ում f բաղադրիչի արժեքի վերաբերյալ: Սակայն (4)-ի մեջ t -ի և S_k հզորության մեծ արժեքները կարող են հանգեցնել հետևյալ խնդիրներին.

1. Բազայի ձևավորումը կդառնա խիստ աշխատատար՝ կմեծանա լեզվաբանի կողմից տրվող կապերի քանակը, կդժվարանա (3) և (5) պայմանների վերահսկման գործընթացը:
2. Կմեծանան ԷՀՄ-ի հիշողության մեջ ԳԲ-ի կոնցեպտները ներկայացնող կազմակերպչական աղյուսակների ծավալները, որի պատճառով կմեծանա

տեքստերի թարգմանության ընթացքում այդ աղյուսակների հետ որոնումային գործընթացների տևողությունը:

Նշված պրոբլեմների լուծման նպատակով այս աշխատանքում առաջարկվում է թարգմանչական համակարգերի ԳԲ-ների ձևավորման ժամանակ օգտագործել հետևյալ մոտեցումները.

1. R_{ij} հարաբերությունների բազմությունը բաժանել երկու չհատվող ենթախմբերի՝ բազային հարաբերություններ և դրանցից ածանցվող հարաբերություններ:
2. Կատարել կոնցեպտների ստրուկտուրավորում և դրա օգնությամբ ստեղծել սեմանտիկ բրգաձև ցանցեր, որոնցում, ի հաշիվ ժառանգելիության մեխանիզմների կիրառման, կտրուկ կփոքրանան S_k բազմությունների հզորությունները:

Առաջին կետի ներդրման դեպքում բազային հարաբերությունները ֆիքսվում են ԳԲ-ում լեզվաբանի կողմից, իսկ երկրորդ խմբի հարաբերություններն ածանցվում են դրանցից որոշակի ալգորիթմներով, որոնց իրականացման մոդուլները ներդրվում են թարգմանիչի միջուկի կազմում: Դա թույլ է տալիս նվազեցնել բազայում ուղղակիորեն տրվող հարաբերությունների քանակը: Յուրաքանչյուր այս մոտեցման հնարավորությունները օրինակով: Ենթադրենք պահանջվում է թարգմանել հետևյալ ֆրագմենտը՝ «<Բանտարկյալի փախուստը բանտից ձախողվեց:>> Այստեղ <<բանտարկյալ>> և <<փախուստ>> բառերի միջև առկա է հատկացուցչային կապ: Սակայն այդ կապը կարող է դասվել ածանցյալ հարաբերությունների դասին, այսինքն այն կարող է ստացվել (ժառանգվել) բազային հանդիսացող <<ենթակաստորոգյալ>> կապից, որն արտահայտում է <<Բանտարկյալը փախչում է>> իմաստային հարաբերությունը, և պետք է նշվում է ԳԲ-ում <<փախչել>> կոնցեպտի մոտ: Նշված ածանցյալ հարաբերության ստացումն ապահովելու համար պետք է ԳԲ-ում <<փախուստ>> գոյականը հատուկ կապի միջոցով կապել <<փախչել>> բայից, որից նա ծրագրային միջուկում ներդրված մոդուլների պետք է ժառանգի իր հատկացուցչային կապերը: Նկատենք, որ նման ածանցյալ հարաբերությունների քանակը բավականին մեծ է, և դրանց ցանկի և ստացման ալգորիթմների հարցը առանձին լեզվաբանական խնդիր է և դուրս է այս աշխատանքի քննարկման սահմաններից:

Այժմ դիտարկենք կոնցեպտների ստրուկտուրավորման խնդիրը: Նախ սահմանենք <<մուլտիպլետի>> գաղափարը: Մուլտիպլետը բազայի կոնցեպտներից կազմված միաստիճան հիերարխիկ կառուցվածք է: Մուլտիպլետի գագաթը վերացական կոնցեպտ է, որը բազա է մտցվում մուլտիպլետի ձևավորմանը զուգընթաց: Նշենք, որ այդ կոնցեպտները մուլտիպլետի ձևավորման փուլում չեն իդենտիֆիկացվում թարգմանվող լեզվի բառերով: Մուլտիպլետի ստորադաս կոնցեպտների քանակը պետք է մեծ լինի մեկից: Մուլտիպլետի ստորադաս կոնցեպտների կապը գագաթի հետ հաստատվում է հատուկ կապով, որով ստորադասին թույլ է տրվում ժառանգել իր վերադասի՝ R_{ij} կապերով արտահայտվող բոլոր հատկությունները: Այս հատկության շնորհիվ մուլտիպլետի ձևավորման փուլում այն R_{ij}-երը, որոնք առկա են մուլտիպլետի բոլոր ստորադասների մոտ, տեղափոխվում են և տրվում

մուլտիպլետի գազաթ հանդիսացող կոնցեպտի մոտ: Այսպիսով, մուլտիպլետի ներդրումը բազայի մեջ թույլ կտա (4)-ում փոքրացնել Sk-ի հզորությունը՝ լուծելով վերնում նշված բազայի ձևավորման ժամանակ առաջացող խնդիրները: Sk-ի հզորությունն ավելի կփոքրանա, եթե մուլտիպլետի ձևավորման փուլում դրա ստորադասների ցանկում կարողանան ընդգրկվել նաև այլ մուլտիպլետների գազաթներ հանդիսացող կոնցեպտներ, որոնք այդ պահին արդեն առկա կլինեն ԳԲ-ում: Մուլտիպլետի ձևավորման ընթացքում դրա գազաթին վերագրվում է մակարդակի համար: Վերջինս մեկով ավելին է, քան տվյալ մուլտիպլետի ամենաբարձր մակարդակն ունեցող ստորադասի մակարդակի համարը (այդ կոնցեպտը կարող է լինել նաև մեկ այլ մուլտիպլետի գազաթ): Մինչ ստրուկտուրավորումը բոլոր կոնցեպտներին վերագրվում է զրոյական մակարդակ: Այսպիսի ստրուկտուրավորման ժամանակ ԳԲ-ն դառնում է բրգաձև համակարգերի ամբողջություն, որոնցից յուրաքանչյուրի գազաթը հանդիսացող մուլտիպլետը (մուլտիպլետի գազաթը) լեզվում առկա խոսքի մասերից որևէ մեկն է (գոյական, ածական, բայ և այլն): Նման բրգաձև համակարգ ստեղծելու ընթացքում նվազագույն հզորությամբ Sk-ներ ստանալու համար այս աշխատանքում առաջարկվում է օգտագործել հետևյալ երկու սկզբունքները.

1. ԳԲ-ում առկա ցանկացած կոնցեպտ (թե թարգմանվող լեզվի բառերով արտահայտվող, և թե բազա ներմուծված մուլտիպլետների գազաթները ներկայացնող) պետք է ընդգրկված լինի որևէ մուլտիպլետի ստորադասների կազմում (բացի խոսքի մասերը ներկայացնող կոնցեպտներից):
2. Բուրգի կառուցման ժամանակ ընթացիկ մուլտիպլետի ստորադասների կազմին հավակնորդ կոնցեպտների ընտրման պահին այդ մուլտիպլետի որակի գնահատականի (Q) որոշման համար օգտագործել հետևյալ ֆունկցիան.

$$Q = n \frac{m_0}{m_0 + \sum_{i=1}^n m_i}, \quad (7)$$

որտեղ n -ը կառուցվող մուլտիպլետի ստորադասների ընթացիկ կազմի հավակնորդ կոնցեպտների քանակն է, m_0 -ն՝ մուլտիպլետի k_0 գազաթի մոտ տեղափոխված կապերի քանակը, իսկ m_i -ն i -րդ ստորադասի մոտ եղած կապերի քանակը: Եթե բուրգի ձևավորման յուրաքանչյուր տակտում ապահովվեն Q-ի առավելագույն արժեքներ, ապա դինամիկ ծրագրավորման սկզբունքի համաձայն՝ ստացվող սեմանտիկ բրգաձև ցանցի որակը մոտ կլինի իր օպտիմալ արժեքին: Նկատենք, որ այս ֆունկցիան փորձում է համատեղել մուլտիպլետի ձևավորման ընթացքում երկու իրար հակասող պահանջները՝ ստորադասների կազմի քանակի (n) ընդլայնումը և մուլտիպլետի k_0 գազաթի մոտ տրվող կապերի քանակը (m_0): Քանի որ n -ի մեծացումը հանգեցնում է m_0 -ի փոքրացմանը, m_0 -ի փոքրացումն էլ բերում է m_i -ի մեծացմանը, ապա այդ դեպքում n -ի մեծացումը հանգեցնում է (7)-ում բաղադրիչի փոքրացմանը: Այսպիսով, Q-ի առավելագույն արժեք ապահովելու համար անհրաժեշտ է գտնել n -ի օպտիմալ արժեքը: Սակայն իրականում

մուլտիպլետի ձևավորման պրոցեսում պետք է հաշվի առնել նաև (6)-ում առկա f պարամետրը: Երբ մուլտիպլետի ստորադաս կոնցեպտների մոտից նույն ($R0, R1$) կապը տեղափոխում ենք դեպի $k0$ գագաթ, ապա վերջինիս մոտ այդ կապի f պարամետրի արժեքը պետք է լինի բոլոր տեղափոխված կապերի f պարամետրերի թվաքանական միջինը: Վերջին հանգամանքը հաշվի առնելով՝ Q -ի արտահայտության ճշգրտված տեսքը կլինի հետևյալը.

$$Q = n \frac{\sum_{i=1}^{m_0} f_{i\bar{m}g}}{\sum_{i=1}^{m_0} f_{i\bar{m}g} + \sum_{i=1}^n \sum_{j=1}^{m_i} f_{ji}}, \quad (8)$$

որտեղ $f_{i\bar{m}g}$ -ը գագաթ տեղափոխված i -րդ կապի f պարամետրի միջին արժեքն է, իսկ f_{ji} -ն՝ մուլտիպլետի i -րդ ստորադասի j -րդ կապի f պարամետրի արժեքը: Բնական է, որ գագաթ տեղափոխվող կապերի մոտ f_i -երի փոխարինումը $f_{\bar{m}g}$ -ով կարող է հանգեցնել սխալների: Այդ պատճառով մուլտիպլետի ընտրումը (n -ի օպտիմալ արժեքի և ստորադասների կազմի ընտրումը) պետք է կատարվի ոչ միայն Q -ի օպտիմալ արժեքին հասնելու պայմանով, այլ նաև հետևյալ սահմանափակումը հաշվի առնելով.

$$\frac{\max_{i=1 \dots n} |t_{\bar{m}g} - t_i|}{f_{\bar{m}g}} \leq \alpha, \quad (9)$$

որտեղ α -ն հարաբերական սխալի թույլատրելի արժեքն է:

ԳԲ-ի ձևավորման, (3) և (5) պայմանների հսկման, ԳԲ-ում կոնցեպտների կիսման և ստանձման, մուլտիպլետների ընտրման և գնահատման գործընթացը պահանջում է մեծածավալ ինֆորմացիայի մշակում: Նշվածի իրականացումը պահանջում է մարդ-մեքենայական համակարգի առկայություն: Այսպիսով մուլտիպլետի ստեղծման պրոցեսում մարդը պետք է ունենա որոշակի դերակատարություն: Որպեսզի նա կարողանա տարբեր մանիպուլյացիաներ կատարել մուլտիպլետների հետ, անհրաժեշտ է, որ մուլտիպլետի գագաթները ներկայացնող վերացական կոնցեպտները մարդու կողմից ստանան բառային անուններ, որոնք պետք է տրվեն լեզվաբանի կողմից՝ ելնելով դրանց ստորադասների կազմից և ներկայացնեն այդ ստորադասների ընդհանրացումը: Այդ անվանումները չպետք է հանդիսանան թարգմանվող լեզվի տարբեր: Այսպիսով նշված մարդ-մեքենայական համակարգերը պետք է ընդունեն հատուկ խմբագրիչների տեսք, որոնք բացի ԳԲ-ի կազմակերպչական միավորների որոնման, ստեղծման կամ հեռացման օպերացիաներից, թույլ կտան նաև գնահատել սեմանտիկ, բրգաձև համակարգերի որակը, իսկ նրա բաղադրիչների՝ մուլտիպլետների ստեղծման պրոցեսում ԳԲ-ն մշակող մասնագետին կտան համապատասխան խորհրդատվություններ:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Манукян А., Казарян Р.** Лингвистические основы организации процесса смыслового машинного перевода // Вестник ГИУА (Политехник): Сборник научных и методических статей. - Ереван, 2010.-Т. 1. – С. 270
2. **Манукян Э., Манукян С., Манукян А., Саядян А.** Основы построения интеллектуальной системы машинного перевода текстов // Сборник материалов годичной научной конференции ГИУА. -Ереван, 2001.-Т. 1. – С. 306.
3. **Manukyan E., Manukyan A., Manukyan S., Manukyan K.** Constructive Approach to the Selection of Metalanguage for Control of the Linguistic Process of Machine Translation // In Proc. Of CSIT-2003 Conference. -Yerevan, Armenia. - P. 251-253.
4. **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(Պ): Նյութը ներկայացվել է խմբագրություն 29.02.2011:

Ա.Տ. ՄԱՆՈՒԿՅԱՆ, Ա.Ա. ՄԱՐԱԲՇՅԱՆ, Ա.Ա. ՄԵԼԻԿՅԱՆ, Յ.Ն. ՄԱՆՈՒԿՅԱՆ

ПРИНЦИПЫ ОБРАЗОВАНИЯ БАЗЫ ЗНАНИЙ В СИСТЕМАХ МАШИННОГО ПЕРЕВОДА

Предложен формальный метод образования одной из составляющих в системах машинного перевода - базы знаний, который позволяет определить состав единиц базы знаний, состав связей, выражающих семантико-синтаксические отношения между ними, и методы кодирования этих связей.

Ключевые слова: смысловой переводчик, информационная база, база знаний, концепт, синтаксическое отношение, семантическое отношение, структурирование концептов, мультиплет, относительная ошибка.

A.S. MANUKYAN, A.H. SHARABCHIAN, A.A. MELIKYAN, E.N. MANUKYAN

THE PRINCIPLES OF KNOWLEDGE BASE FORMATION IN MACHINE TRANSLATION SYSTEMS

The formal method of forming the knowledge base as one of the components of machine translation systems, which allows to define the unit structure of the knowledge base, the structure of connections expressing semantico-syntactical relations between them and methods of these connections coding is proposed.

Keywords: semantic translator, information base, knowledge base, concept, syntactical relations, semantic relations, structuring of concepts, multiplet, relative error.