

Է.Ն. ՄԱՆՈՒԿՅԱՆ, Ա.Հ. ՇԱՐԱԲՉՅԱՆ, Է.Ս. ՄԱՆՈՒԿՅԱՆ

**ՄԵՔԵՆԱՅԱԿԱՆ ԹԱՐԳՄԱՆՈՒԹՅԱՆ ՀԱՄԱԿԱՐԳԵՐՈՒՄ ՑԱՆՑԱՅԻՆ
ԳԻՏԵԼԻՔՆԵՐԻ ԲԱԶԱՆԵՐԻ ԿԱԶՄԱԿԵՐՊՄԱՆ ԱՐԴՅՈՒՆԱՎԵՏՈՒԹՅԱՆ
ԲԱՐՁՐԱՑՄԱՆ ՄԵԹՈԴՆԵՐԸ**

Առաջարկված են մեքենայական թարգմանության համակարգերում ցանցային գիտելիքների բազաների կազմակերպման արդյունավետության բարձրացմանը նպաստող մեթոդներ, որոնք թույլ են տալիս զգալիորեն պակասեցնել այդ գիտելիքների բազաների հանգույցների միջև իրապես տրվող կապերի քանակը՝ առանց նախադասությունների վերլուծության որակի կորուստների:

Առանցքային բառեր. մեքենայական թարգմանության համակարգ, ցանցային գիտելիքների բազա, կոնցեպտ, իմաստաշարահյուսական հարաբերություն, մուլտիպլետ, մուլտիպլետի գազաթ:

Ժամանակակից մեքենայական թարգմանության համակարգերում (ՄԹՀ) առավել արդյունավետ է համարվում դրանց աշխատանքի կազմակերպման հետևյալ սխեման: Նախ՝ մուտքային տեքստի նախադասությունը ձևափոխվում և բերվում է մաթեմատիկական որոշակի ձևայնացված տեսքի, որը ծառանման գրաֆ է և, փաստորեն, ներկայացնում է մուտքային նախադասության իմաստային նմանակը: Այնուհետև՝ ստացված գրաֆի հիման վրա ելքային լեզվի բառարանների և քերականական կանոնների միջոցով գեներացվում է ելքային նախադասության տեքստը: ՄԹՀ-երի աշխատանքի կազմակերպման առաջին փուլը պայմանականորեն անվանում են տեքստի թարգմանության վերլուծության, իսկ երկրորդը՝ ելքային տեքստի համադրման (սինթեզի) փուլ: Վերլուծության փուլի ընթացքում ստացվում են նախադասության իմաստային նմանակները հանդիսացող մի քանի գրաֆներ [1]: Այդ պատճառով այս փուլի վերջում կատարվում է նաև ստացված գրաֆներից ամենահավանականի ընտրություն: Վերլուծության և համադրման փուլերն իրենց հերթին բաժանվում են տարբեր ենթափուլերի: Նշված ենթափուլերի ընթացքում մշակվող նախադասությունը կարող է ներկայացվել որպես բառերի զծային շարան, որպես բառերի իմաստներին հավակնող կոնցեպտների բազմությունների շարան կամ որպես նախադասության հնարավոր իմաստներն արտահայտող ծառանման գրաֆների բազմություն [1], [2]: Այսպիսով, տեքստի որակյալ թարգմանություն կատարելու համար անհրաժեշտ է ստանալ նախադասությունների հնարավորինս ճիշտ իմաստային նմանակները: Այժմ ավելի մանրամասն դիտարկենք դրանց ստացման գործընթացը և ներկայացնենք դրանց ձևավորման համար առաջարկվող մեթոդները:

Վերը նշված գրաֆը ստանալու համար ՄԹՀ-ն պետք է ունենա տեղեկատվական բազա, որի համար առաջարկվում է ենթաբազաների հետևյալ կառուցվածքը.

1. դիտարկվող լեզուների շարահյուսությունների նկարագրությունները պարունակող ենթաբազաներ,

2. գիտելիքների ենթաբազա (ԳԲ):

Նախադասության մեջ ցանկացած հարևան երկու բառեր կարող են ունենալ կապ-հարաբերություն: Այդ հարաբերությունն ունի երկու բաղադրիչ՝ շարահյուսական և իմաստային [2], [3]: Ընդ որում, նույն իմաստային հարաբերությանը կարող են համապատասխանել տեսքով տարբեր շարահյուսական հարաբերություններ: Այս հանգամանքը կարող է դժվարացնել թարգմանության ընթացքում մուտքային տեքստի բառերի միջև առկա իմաստաշարահյուսական հարաբերությունների համարժեքի որոնումը ելքային տեքստում: Այս խնդիրը լուծելու համար առաջարկվում է յուրաքանչյուր R_{ij} հարաբերության իդենտիֆիկատորը (տվյալ դեպքում կապի կոդը) ներկայացնել երկու բաղադրիչով՝

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

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 դեպքում: 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; գործողության հանգման ($R_0 = 3$), անջատման ($R_0 = 4$) և միջոցի ($R_0 = 5$) անուղղակի խնդիրների հարաբերություններ; փոխաբերական իմաստով գործողության հանգման ($R_0 = 6$), անջատման ($R_0 = 7$) և միջոցի ($R_0 = 8$) հարաբերություններ: Օրինակ՝ <<Ես դրեցի գիրքը սեղանին: - I put the book on the table.>> կամ <<Ես հարվածեցի դարպասին: - I hit the goal.>> նախադասությունների մեջ բայն ունի հանգման անուղղակի խնդրի հարաբերություն ($R_0 = 3$): Բերված օրինակներից երևում է նաև R_{ij} -ի R_1 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՝

- $R_0 = 3, R_1 = 1$ – հանգման անուղղակի խնդրի հարաբերություն, հայերենում տրական հոլովով արտահայտված, իսկ անգլերենում *on* նախդիրով կազմվող;

- $R_0 = 3, R_1 = 2$ – հանգման անուղղակի խնդրի հարաբերություն, հայերենում տրական հոլովով արտահայտված, իսկ անգլերենում՝ առանց նախդիրի:

Այսպիսով, նախադասության իմաստաշարահյուսական վերլուծության համար հարկավոր է, որ (3,1) և (3,2) հարաբերությունների ձևաբանական կողմն արտահայտված լինի հայերեն և անգլերեն լեզուների շարահյուսական նկարագրերի ենթաբազաներից յուրաքանչյուրում, ինչպես նաև այդ հարաբերությունները տրված լինեն ԳԲ-ում համապատասխան կոնցեպտների մոտ:

Մուտքային տեքստերի հոսքում միևնույն բառը և դրա իմաստային համարժեք կոնցեպտը կարող են հանդես գալ նախադասություններում տարբեր բառերի հարևանությամբ և տարբեր դերերում: Այստեղից հետևում է, որ ԳԲ-ում նշված բառին համապատասխան կոնցեպտը պետք է ունենա տեքստերից բխող բոլոր հարաբերությունները ԳԲ-ի մյուս կոնցեպտների հետ: Ակնհայտ է որ ունիվերսալ թարգմանիչներում ԳԲ-ն կստացվի կապերով խիստ ուռճացված: Նման պայմաններում առաջանում են մեծ պրոբլեմներ թե այդ ԳԲ-ի վարման և թե պահպանման հետ կապված: Այդ պրոբլեմների լուծման ուղղություններից մեկը միաստիճան հիերարխիկ կառուցվածքների՝ մուլտիպլետների գաղափարի ներմուծումն է: Մուլտիպլետները ստեղծվում են ԳԲ-ի կոնցեպտների մոտ դեպի մուլտիպլետի գազաթը հատուկ ծառայողական կապերի տրման միջոցով (IR – inheritable relation): Այս կապերի միջոցով ապահովվում են մուլտիպլետի գազաթի մոտ տրված իմաստային կապերի ժառանգելիության մեխանիզմները դրա ստորադասների կողմից: Մուլտիպլետների հիերարխիկ համակարգի ստեղծումը թույլ է տալիս զգալիորեն պակասեցնել ԳԲ-ում կոնցեպտների մոտ տրվող իմաստային կապերի քանակը, քանի որ այդ հիերարխիկ ստեղծման պրոցեսում մուլտիպլետի ստորադասների մոտ տրված հարաբերությունները դրա ստեղծման պրոցեսում տեղափոխվում են և տրվում այդ մուլտիպլետի գազաթի մոտ: Մուլտիպլետների մասին առավել մանրամասն խոսվում է [3]-ում:

Հոդվածում առաջարկվում է ԳԲ մտցնել նոր՝ ժառանգելիության հատկություններ ունեցող ծառայողական կապեր, որոնք թույլ կտան փոքրացնել ՄԹՀ-երի գիտելիքների բազաներում փաստացի տրված իմաստաշարահյուսական հարաբերությունների քանակը: Առաջարկվում են այդ ծառայողական կապերի օգտագործման վրա հիմնված մեթոդներ, որոնք կարող են կիրառվել ԳԲ-ից տեղեկույթն ընթերցող օպերատորներում և թույլ են տալիս բացահայտել սեմանտիկ ցանցի (ցանցաձև գիտելիքների բազայի) ցանկացած երկու հանգույցների (k_1 և k_2) միջև տրված իմաստային հարաբերության (R) առկայության կամ բացակայության փաստը, նույնիսկ եթե այդ իմաստային հարաբերությունը անմիջապես տրված չէ այդ կոնցեպտների միջև: Առաջարկվող մեթոդների հիմքում ընկած է հետևյալ մոտեցումը.

▪ Ցանցի k_1 կետից փնտրվում է $k_1 \rightarrow k_2$ հարաբերության՝ ցանցի այլ կետերով անցնող այլընտրանքային ճանապարհ: Վերլուծելով այդ ճանապարհի վրա ընկած հարաբերությունների կազմը՝ որոշվում է ցանցի k_1 և k_2 հանգույցների միջև առկա հարաբերությունը:

Այժմ դիտարկենք այդ մեթոդները:

Մեթոդ 1: Սեմանտիկ ցանցում փնտրվում է ճանապարհի k_1 -ից դեպի k_2 -ը, որն ունի հետևյալ կառուցվածքը.

$$k_1 \xrightarrow{IR^m} K_{1Parent} \xrightarrow{R_{12}} K_{2Parent} \xrightarrow{IR_n} k_2,$$

որտեղ $K_{1Parent}$ -ը k_1 կոնցեպտի վերադաս հանդիսացող մուլտիպլետներից մեկի գազաթը ներկայացնող կոնցեպտն է, IR^m -ը՝ m հատ ծառայողական կապերի հաջորդականությունը, որոնց տրման շնորհիվ k_1 կոնցեպտն ընդգրկվել է $K_{1Parent}$ գազաթով ներկայացվող մուլտիպլետի ստորադասների կազմում, $K_{2Parent}$ -ը՝ k_2 կոնցեպտի վերադաս հանդիսացող մուլտիպլետներից մեկի գազաթը ներկայացնող կոնցեպտը, IR_n -ը՝ n հատ ծառայողական կապերի հաջորդականությունը, որոնց հաջորդական տրման շնորհիվ k_2 կոնցեպտն ընդգրկվել է $K_{2Parent}$ գազաթով ներկայացվող մուլտիպլետի ստորադասների կազմում, իսկ R_{12} -ը՝ $K_{1Parent}$ -ի և $K_{2Parent}$ -ի միջև եղած իմաստաշարահյուսական հարաբերությունը: Ընդ որում, m -ը և n -ը կարող են ընդունել զրոյական արժեքներ:

Նման կառուցվածքով ճանապարհի առկայության դեպքում այս աշխատանքում պնդվում է, որ այդ R_{12} հարաբերությունն առկա է նաև k_1 -ի և k_2 -ի միջև, քանի որ $K_{1Parent} \xrightarrow{R_{12}}$ հարաբերությունը IR^m ծառայողական կապերի շնորհիվ ժառանգվում է, և ստանում ենք $k_1 \xrightarrow{R_{12}}$ հարաբերությունը: Նույն կերպ՝ $K_{2Parent}$ -ից ստացվում է՝ $k_2 \xrightarrow{R_{12}}$ հարաբերությունը:

Մեթոդ 2: Սեմանտիկ ցանցում փնտրվում է ճանապարհի k_1 -ից (k_1 -ը գոյական է) դեպի k_2 -ը, որն ունի հետևյալ կառուցվածքը.

$$k_1 \xrightarrow{IR_{V-N}} K_{Verb} \xrightarrow{IR^m} K_{VerbParent} \xrightarrow{R_{VerbRel}} K_{2Parent} \xrightarrow{IR_n} k_2,$$

որտեղ IR_{V-N} -ը մեր կողմից առաջարկվող ծառայողական հարաբերություն է k_1 -ից դեպի K_{Verb} բայը, $K_{VerbParent}$ -ը՝ K_{Verb} կոնցեպտի վերադաս հանդիսացող մուլտիպլետներից մեկի գազաթը ներկայացնող կոնցեպտը, իսկ $R_{VerbRel}$ -ը՝ բայի և $K_{2Parent}$ կոնցեպտի միջև իմաստաշարահյուսական հարաբերությունը:

Նման կառուցվածքով ճանապարհի առկայության դեպքում այս աշխատանքում պնդվում է, որ R_{12} հարաբերությունն առկա է k_1 -ի և k_2 -ի միջև, քանի որ $K_{VerbParent} \xrightarrow{R_{VerbRel}}$ հարաբերությունը IR^m և IR_{V-N} ծառայողական կապերի շնորհիվ ժառանգվում և փոխակերպվում է $k_1 \xrightarrow{R_{12}}$ հարաբերությանը: Նույն կերպ՝

$\xrightarrow{R_{VerbRel}} K_{2Parent}$ -ից ստացվում է $\xrightarrow{R_{12}} k_2$ հարաբերությունը: $R_{VerbRel} \rightarrow R_{12}$ համապատասխանությունները բխում են կիրառական լեզվաբանության կանոններից, իսկ դրանց ներկայացման աղյուսակները կցվում են IR_{V-N} ծառայողական կապերին:

Նշված մեթոդը պարզաբանենք կոնկրետ օրինակով: Վերցնենք <<փախչել>> բայը և <<փախուստ>> գոյականը: Ակնհայտ է, որ չնայած այս հասկացությունները տարբեր խոսքի մասեր են, սակայն <<փախչել>> բայի շատ իմաստաշարահյուսական կապեր կարող ժառանգվել <<փախուստ>> գոյականի մոտ և փոխարինվել դրանց համապատասխան այդ գոյականին բնորոշ հարաբերություններով: Օրինակ՝ <<բանտարկյալը փախավ>> ենթակայական հարաբերությունից ($R_{VerbRel}$) կարող է ժառանգվել <<բանտարկյալի փախուստը>> հատկացուցչային հարաբերությունը (R_{12}):

Մեթոդ 3: Սեմանտիկ ցանցում փնտրվում է ճանապարհ k_1 -ից (k_1 -ն ածական է) դեպի k_2 -ը, որն ունի հետևյալ կառուցվածքը.

$$k_1 \xrightarrow{IR_{V-A}} K_{Verb} \xrightarrow{IR^m} K_{VerbParent} \xrightarrow{R_{VerbRel}} K_{2Parent} \xrightarrow{IR_n} k_2$$

որտեղ՝ IR_{V-A} -ն մեր կողմից առաջարկվող ծառայողական հարաբերություն է k_1 -ից դեպի K_{Verb} բայը, $K_{VerbParent}$ -ը K_{Verb} կոնցեպտի վերադաս հանդիսացող մուլտիպլետներից մեկի գազաթը ներկայացնող կոնցեպտն է, իսկ $R_{VerbRel}$ -ը բայի և $K_{2Parent}$ կոնցեպտի միջև իմաստաշարահյուսական հարաբերությունը:

Նման կառուցվածքով ճանապարհի առկայության դեպքում այս աշխատանքում պնդվում է, որ R_{12} հարաբերությունն առկա է k_1 -ի և k_2 -ի միջև, քանի որ $K_{VerbParent} \xrightarrow{R_{VerbRel}}$ հարաբերությունը IR^m և IR_{V-A} ծառայողական կապերի շնորհիվ ժառանգվում և փոխակերպվում է $k_1 \xrightarrow{R_{12}}$ հարաբերությանը: Նույն կերպ՝ $\xrightarrow{R_{VerbRel}} K_{2Parent}$ -ից ստացվում է՝ $\xrightarrow{R_{12}} k_2$ հարաբերությունը: $R_{VerbRel} \rightarrow R_{12}$ համապատասխանությունները բխում են կիրառական լեզվաբանության կանոններից, իսկ դրանց ներկայացման աղյուսակները կցվում են IR_{V-A} ծառայողական կապերին:

Նշված մեթոդը պարզաբանենք կոնկրետ օրինակով: Վերցնենք <<գեղեցկանալ>> բայը և <<գեղեցիկ>> ածականը: Ակնհայտ է, որ չնայած այս հասկացությունները տարբեր խոսքի մասեր են, սակայն <<գեղեցկանալ>> բային առնչվող որոշակի կապեր ծրագրային եղանակով կարող են ժառանգվել <<գեղեցիկ>> ածականը ներկայացնող կոնցեպտի մոտ և փոխարինվել դրանց համապատասխան այդ ածականին բնորոշ հարաբերություններով: Օրինակ՝ <<Աղջիկը գեղեցկացավ>> կոնցեպտից ($R_{VerbRel}$) կարող է ժառանգվել <<գեղեցիկ աղջիկ>> որոշիչ-որոշյալ հարաբերությունը (R_{12}):

Այսպիսով, բերված մեթոդների պրակտիկ իրականացման փորձը ISMA մեքենայական թարգմանիչում ցույց տվեց, որ դրանք հեշտությամբ ենթարկվում են ալգորիթմացման և թույլ են տալիս զգալիորեն մեծացնել ցանցային գիտելիքների բազայի արդյունավետությունը:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Շարաբչյան Ա.Հ., Մանուկյան Ա.Ս.** Գիտելիքների բազաների ձևավորման մի մեթոդ մեքենայական թարգմանության համակարգերում // ՀՊՃՀ Լրաբեր. – 2011. – Հատոր 3, N1.- էջ 223-227:
2. **Շարաբչյան Ա.Հ., Մանուկյան Է.Ն., Մանուկյան Ա.Ս., Մելիքյան Ա.Ա.** Մեքենայական թարգմանության համակարգերում գիտելիքների բազաների ձևավորման սկզբունքները // ՀՀ ԳԱԱ և ՀՊՃՀ Տեղեկագիր. – 2011. – Հատոր 64, N3. էջ 257-264:
3. **Մանուկյան Է.Ն., Շարաբչյան Ա.Հ., Մանուկյան Ա.Ս.** Տեքստային վերլուծիչների գիտելիքների բազաների կազմակերպման մեթոդ-մեքենայական համակարգերի մշակումը // ՀՊՃՀ Լրաբեր. – 2014. – Մաս 1.- էջ 116-121:
4. **Manukyan E., Manukyan A., Manukyan S., Manukyan K.** Constructive Approach to the Selection of Metalanguage for Control of the Linguistic Process of Machine Translation // Proc. Of CSIT-2003 Conference. -Yerevan, Armenia, 2003. -P. 251-253.

Э.Н. МАНУКЯН, А.А. ШАРАБЧЯН, Э.С. МАНУКЯН

МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ОРГАНИЗАЦИИ СЕТЕВЫХ БАЗ ЗНАНИЙ В СИСТЕМАХ МАШИННОГО ПЕРЕВОДА

Предложены методы, повышающие эффективность организации сетевых баз знаний в системах машинного перевода, которые позволяют значительно уменьшить количество реально устанавливаемых отношений между узлами этих баз знаний, при этом не ухудшая качество анализа и перевода предложений.

Ключевые слова: система машинного перевода, сетевая база знаний, концепт, семантико-синтаксическое отношение, мультиплет, вершина мультиплета.

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

METHODS FOR INCREASING THE EFFICIENCY OF NETWORK KNOWLEDGE BASES ORGANISATION IN THE MACHINE TRANSLATION SYSTEMS

Methods for increasing the efficiency of network knowledge bases in the machine translation systems are proposed allowing to significantly reduce the quantity of the really established relations between the nodes of these knowledge bases, without deteriorating the quality of the analysis and the translation of sentences.

Keywords: machine translation system, network knowledge base, concept, semantic-syntactical relation, multiplet, multiplet top.