

ԼԻԼԻԹ ՓԱՐԵՄՈՒԶՅԱՆ

ՀԱՄԸՆԴՀԱՆՈՒՐ ՑԱՆՑԱՅԻՆ ԼԵԶՈՒ

ՀՀ Գիտությունների ազգային ակադեմիայի Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի, Հր. Աճառյանի անվան լեզվի ինստիտուտի, Երևանի պետական համալսարանի գիտնականների կողմից զգալի աշխատանքներ են կատարվել լեզվաբանության և ինֆորմատիկայի կիրառական ուղղությունների ձեռքբերումները հայագիտության մեջ կիրառելու նպատակով: Խոսքերը ակտիվորեն մասնակցել է UNL համակարգի (Համընդհանուր ցանցային լեզու)¹ հայկական մոդուլի ստեղծմանը: Աշխատանքներ են կատարվել հայերենի ձևային նկարագրության, տեքստի ավտոմատ մշակման, ավտոմատ բառարանների և համակարգչային տվյալների հենքերի ստեղծման ուղղությամբ:

Նախապես մշակված ու ներդրված լինելով գոկտոր Հ. Ուշիդայի և գործընկերների կողմից (1996 թ.), UNL -ն իր շարունակական զարգացումն ապրեց մի քանի գիտական ինստիտուտների, շահագրգիռ կազմակերպությունների կողմից՝ Տոկիոյի (Ճապոնիա) Միավորված ազգերի կազմակերպության համալսարանի Հեռանկարային հետազոտությունների ինստիտուտի նախաձեռնությամբ:

2001 թ. հունվարին ՄԱԿ-ի համալսարանը ստեղծեց առանձին կազմակերպություն՝ UNDL հիմնադրամը², որը պատասխանատու է UNL ծրագրի զարգացման և կառավարման համար:

UNL-ն իբրև համակարգչային համընդհանուր, արհեստական ցանցային լեզու, նպատակաուղղված է համաշխարհային սարդոստայնի ինֆորմացիայի նկարագրմանը, պահպանմանն ու տարածմանը՝ որևէ բնական լեզվից՝ համակարգչային միջավայրից անկախ տեսքով (տե՛ս ներդիր, նկ. 6): Ի տարբերություն մեքենական թարգմանության համակարգերի, որոնք նախատեսված են մի բնական լեզվից մյուսը տեքստերի թարգմանության համար, UNL համակարգը նախատեսվում է ավտոմատացված եղանակով, նաև՝ մարդու միջամտությամբ, բնական լեզվով ներկայացված տեքստը UNL -ի փոխարկելու, այնուհետև նրանից պահանջվող տեքստը բնական լեզվով սինթեզելու համար: Պարզ է, որ համակարգը կարող է սպասարկել տվյալ բնական լեզուն, եթե ներկայացվի հատուկ ծրագրային և լեզվաբանական միջավայր:

¹ www.unlweb.znet.

² <http://www.undlfoundation.org>.

Այսպիսի միջավայր է ստեղծվել ՄԱԿ-ի հիմնական լեզուների՝ արաբերենի, չինարենի, անգլերենի, ֆրանսերենի, ռուսերենի ու իսպաներենի համար: Բացի վերոնշյալ լեզուներից, ստեղծվում են միջավայրեր նաև հինգիլի, ինդոնեզերենի, իտալերենի, պորտուգալերենի, թայերենի, հայերենի համար:

Հայերենը ներգրավվեց UNL ծրագրի աշխատանքներին 2001թ.: Մինչև 2014թ. Հայ մասնագետների կողմից ստեղծվեց UNL-ի աշխատունակ տարբերակ, որը ներառում էր UNL-հայերենբառարան (մոտ 37 000 բառ ու բառային միավոր) հայերեն- UNL և UNL-հայերեն անցման կանոններով, UNL հայկական սերվերով, UNL-ի տեքստային գրաֆի կառուցիչով և այլն: UNL լեզուն նախատեսված է մուտքային բնական լեզվի բառերի, բառակապակցությունների և նախադասությունների կամ ամբողջ տեքստերի իմաստների ներկայացման համար:

Գիտական հոդվածներ են ներկայացվել 2003, 2005, 2007, 2009, 2011, 2013 թթ. CSIT (International Conference on Computer Science and Information Technologies) և այլ միջազգային գիտաժողովներում: 2005 թ. Քունիսում կայացած ինֆորմացիոն և կոմունիկացիոն տեխնոլոգիաների համաշխարհային գագաթնաժողովին՝ ՄԱԿ-ի տաղավարում ներկայացվել են կատարված աշխատանքները: 2006 թ. հայկական խմբի ձեռքբերումները ներկայացվեցին UNL ծրագրի ղեկավարների հետ Հայաստանում տեղի ունեցած աշխատանքային հաշվետու հանդիպման ժամանակ:

UNL լեզվի հիմնական բաղադրիչը բառարանն է, որն ամփոփում է հայերենի բառերը, զբաղեցնելով UNL համարժեքները, վերջիններիս համապատասխանեցված համընդհանուր բառերը (UW), որոնք կոնցեպտուալ տարրեր կամ հասկացություններ են և յուրաքանչյուր բառի ձևաբանական ու շարահյուսական բնութագրերը: UW բառերը պարունակում են իմաստային և կառուցվածքային սահմանափակումներ, որոնք հնարավորություն են ընձեռում որոշ չափով վերացնել բնական լեզուներին բնորոշ բազմիմաստությունը: Բառարանում ներկայացվում են բառերի ձևաբանական բնութագրերը: Մյուս կարևոր բաղադրիչներն են՝ գիտելիքների հենքը (KB), որում թվարկվում են տարբեր իմաստների հնարավոր հարաբերությունները, UW բառերի իմաստը նկարագրող հատուկ բնութագրերը (բնութագրեր՝ օբյեկտ, գործողության ազնետ, ժամանակ, տեղ և այլն) և իմաստաշարահյուսական երկանգամ հարաբերությունները: KB-ն մշտապես համալրվում է հասկացությունների ավելացմանը զուգընթաց: Հետևաբար, UNL լեզվի տեքստը UW բառերի՝ ատրիբուտներով օժտված և իմաստաշարահյուսական հարաբերություններով փոխկապակցված ներկայացում է (տե՛ս ներդիր, նկ. 7):

UNL համակարգի կարևոր մոդուլներն են EUGENE (dEep-to-sUrface

natural language GENERator)³ և IAN (Interactive ANalyser)⁴ փոխարկման (էնկոնվերտոր) և վերլուծման (դեկոնվերտոր) ծրագրային կիրառական մոդուլները: Առաջինը իրականացնում է բնական լեզվով տրված տեքստի ձևափոխում (վերլուծություն և փոխարկում) UNL գրաֆի, երկրորդը՝ UNL գրաֆից համապատասխան բնական լեզվի տեքստի սինթեզ:

Այս մոդուլների համար կարևոր մաս են կազմում անալիզի և սինթեզի կանոնները, որոնք և հիմնականում որոշում են տրված բնական լեզվի ներկայացման որակը UNL համակարգում: Հայերեն մոդուլի համար կազմվել է մոդուլի լեզվաբանական ապահովում: Էնկոնվերտոր և դեկոնվերտոր ծրագրերը համընդհանուր են բոլոր լեզուների համար, և սա է բոլոր բնական լեզուներին մատչելի լինելու գլխավոր գրավականը: Հայկական մոդուլի ստեղծման համար համաձայնագրի միջոցով UNDL հիմնադրամից խումբը ստացել է UNL համակարգի անհրաժեշտ ծրագրային մոդուլները, ինչն էլ ունեցավ իր անգնահատելի դերը հայերեն մոդուլի մշակման գործում: Լեզվաբանական ապահովումից բացի, հայ մասնագետները մշակել են նաև ծրագրային ապահովում, որի օգնությամբ իրականացվել են հայերենի հետ կապված բոլոր բառարանները, կանոնները, դրանց վերլուծությունը և ստուգումը:

UNL -հայերեն հայկական մոդուլն այժմ ունի հետևյալ կառուցվածքը.

- UNL-հայերեն բառարան՝ 37 000 բառ ու բառային միավոր
- Հայերեն էնկոնվերտոր
- Հայերեն դեկոնվերտոր
- UNL-հայերեն սերվեր (www.unl.am)
- UNL տեքստերի գրաֆի կառուցիչ
- Զևարանական վերլուծիչ և համագրիչ
- Գիտական հաշվետվություններ և հոդվածներ

Հայկական մոդուլի աշխատանքները խթանելու համար 2007 թ. ստեղծվեց UNL-հայկական կենտրոնը: Հայկական խումբը ներգրավվեց UNL-ի EOLSS և LIS ծրագրերի աշխատանքներում, 2009-2014 թթ.՝ LPP, MIR A-1, 1, 2, AESPO-1, BRUNO-1, NADIA-A1, UGO-A1 ծրագրերին (**Ошибки! Недопустимый объект гиперссылки.**):

Աշխատանքների կատարման տարբեր փուլերում խումբը ստացել է ֆինանսական օժանդակություն ՀՀ Կառավարությունից («ՀՀ պետական գիտական հաշվողական համակարգի ստեղծում» նպատակային ծրագիր, 2003-2006 թթ.), գրամաշնորհ «Գալուստ Գյուլբենկյան» հիմնարկությունից («UNL-ի հայերենի մոդուլի մշակում» նախագիծ, 2003-2006 թթ.), UNDP Հայաստանի գրասենյակից («Ինֆորմացիոն հասարակություն և ղեմնկրատական կառավարում» նախագիծի շրջանակներում, 2001-2003 թթ.) և UNDL հիմնադրամից LPP և MIR A-1, 1, 2, AESPO-1, BRUNO-1, NADIA-A1, UGO-A1 ծրագրերին:

³ <http://dev.undlfoundation.org/generation/index.jsp>.

⁴ <http://dev.undlfoundation.org/analysis/index.jsp>.

գրերի բառարանները (NL-հայերեն և հայերեն-NL) կազմելու համար (2009-2011 թթ.):

Այժմ խումբն աշխատում է ՀՀ Սփյուռքի նախարարության և ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի միջև կնքված պայմանագրի համաձայն «Համընդհանուր ցանցային լեզվի օգտագործմամբ հայերենի բառապաշարային տվյալների հենքի մշակում» թեմայով նախագծի իրականացման ուղղությամբ: Նախագծի նպատակն է մշակել էլեկտրոնային բառարան հայերենի (և՛ արևելահայերեն, և՛ արևմտահայերեն) համար՝ օգտագործելով UNL միջավայրը՝ ՀՀ Գիտությունների ազգային ակադեմիայի Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ու Հր. Աճառյանի անվան լեզվի ինստիտուտների կողմից արդեն իսկ ստեղծված ծրագրային միջոցներով ու տվյալների հենքով:

Առաջարկվում է UNL-NL բառարանի տարածական տարբերակների և NL բառարանի քերականական բնութագրերի դաշտում ներդրավել արևմտահայերենը՝ քերականական բնութագրով:

Նախագծի իրականացման արդյունքում կստեղծվի UNL-հայերեն երկլեզու բառարան, որի իմաստները նկարագրված կլինեն UNL-ով ու կմշակվի բառապաշարային տվյալների հենք՝ WordNet-ի օրինակով: Բառարանը կարող է ընդգրկվել բազմալեզու բառարանային համակարգում, ինչպիսին է MUHIT համակարգը⁵ սրանով իսկ փորձելով ամրագրել մեր լեզվի անշափ կարևոր ու կենսական տարբերակը՝ արևմտահայերենը պահել-պահպանելու պահանջը՝ հայ ազգի վերապրումն ապահովելու նպատակով:

Lilit Paremuzyan **Universal Networking Language**

UNL- is a computational universal artificial network language which aims at description, preservation and dissemination of global information, from any natural language, regardless of the form of the computer environment. The works of the Armenian module began in 2001. There have been done works on a creation of the formal description of Armenian language, the automatic processing of text, the creation of automatic dictionaries and computer data bases.

Now the group is working on the implementation of a joint project with the Ministry of Diaspora. The result of the project will be the creation of the UNL- Armenian bilingual dictionary, the meanings of which will be described in UNL and the development of vocabulary data basis on the example of WordNet. The dictionary can be included in the multilingual dictionary system, such as the MUHIT system is.

⁵ <http://www.unlweb.net/muhit/>.