

ԼԻԼԻԹ ՓԱՐԵՄՈՒՉՅԱՆ

Բանասիրական գիտությունների թեկնածու

ՀՀ Գիտությունների ազգային ակադեմիա

lili@sci.am

ՀԱՄԸՆԴՀԱՆՈՒՐ ՑԱՆՑԱՅԻՆ ԼԵՉՈՒ. ՀԱՅԿԱԿԱՆ ՄՈԴՈՒԼ

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: Խոսմբն 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 դոկտոր Հ. Ուշիդայի և գործընկերների կողմից (1996 թ.), UNL-ն իրշարունակական զարգացումն ապրեց մի քանի գիտական ինստիտուտների, շահագրգիռ կազմակերպությունների կողմից՝ Տոկիոյի (Ճապոնիա) Միավորված ազգերի 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 (նկ. 1): 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

¹ <http://www.unlweb.net/unlweb/>

² <http://www.undlfoundation.org>

պահանջվող տեքստը բնական լեզվով սինթեզելու համար: Պարզ է, որ համակարգը կարող է սպասարկել տվյալ բնական լեզուն, եթե ներկայացվի հատուկ ծրագրային և լեզվաբանական միջավայր:

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

Հայերենը ներգրավվեց UNL ծրագրի աշխատանքներին 2001թ.: Մինչև 2014 թ. հայ մասնագետների կողմից ստեղծվեց UNL-ի աշխատունակ տարբերակ, որը ներառում էր UNL-հայերեն բառարան (մոտ 37000 բառ ու բառային միավոր)՝ հայերեն-UNL և UNL-հայերեն անցման կանոններով, UNL հայկականսերվերով, UNL-ի տեքստային գրաֆիկառուցիչով 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. Թուրինում 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, դրանց 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 ազենտ, ժամանակ, տեղ և այլն) և իմաստաշարահյուսական երկանդամ հարաբերությունները: KB-ն մշտապես համարվում է՝ հասկացությունների ավելացմանը զուգընթաց: Հետևաբար, UNL լեզվի տեքստը UW բառերի՝ ատրիբուտներով օժտված և

իմաստաշարահյուսական հարաբերություններով փոխկապակցված ներկայացում է (նկ. 2):

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 մատչելիլիներով լիավոր 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 բոլորը առարանները, 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 գրաֆիկառուցիչ

Ձևաբանական վերլուծիչ և համադրիչ

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 (<http://www.unlweb.net>):

Աշխատանքների կատարման տարբեր փուլերում խումբը ստացել է ֆինանսական օժանդակություն ՀՀ Կառավարությունից («ՀՀ պետական գիտական հաշվողական համակարգի ստեղծում» նպատակային ծրագիր, 2003–2006

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

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

թթ.), դրամաշնորհ «Գալուստ Գյուլբենկյան հիմնադրամ» կազմակերպությունից («UNL-ի հայերենի մոդուլի մշակում» նախագիծ, 2003–2006 թթ.), UNDP հայկական մասնաճյուղից («Ինֆորմացիոն հասարակություն և դեմոկրատական կառավարում» նախագծի շրջանակներում, 2001–2003 թթ.) և UNDL հիմնադրամից LPP և MIR A-1, 1, 2, AESPO-1, BRUNO-1, NADIA-A1, UGO-A1 ծրագրերի բառարանները (NL-հայերեն և հայերեն-NL) կազմելու համար (2009–2011թթ.):

Այժմ խումբն աշխատում է ՀՀ Սփյուռքի նախարարության և ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի միջև կնքված պայմանագրի համաձայն, որի արդյունքում մշակվել է էլեկտրոնային բառարան հայերենի (և՛ արևելահայերեն, և՛ արևմտահայերեն) համար՝ օգտագործելով UNL միջավայրը՝ արդեն իսկ ստեղծված ծրագրային միջոցներով ու տվյալների հենքով:

Այսպիսով, UNL-NL բառարանի տարածական տարբերակների և NL բառարանի քերականական բնութագրերի դաշտում ներգրավվեց արևմտահայերենը՝ քերականական բնութագրով:

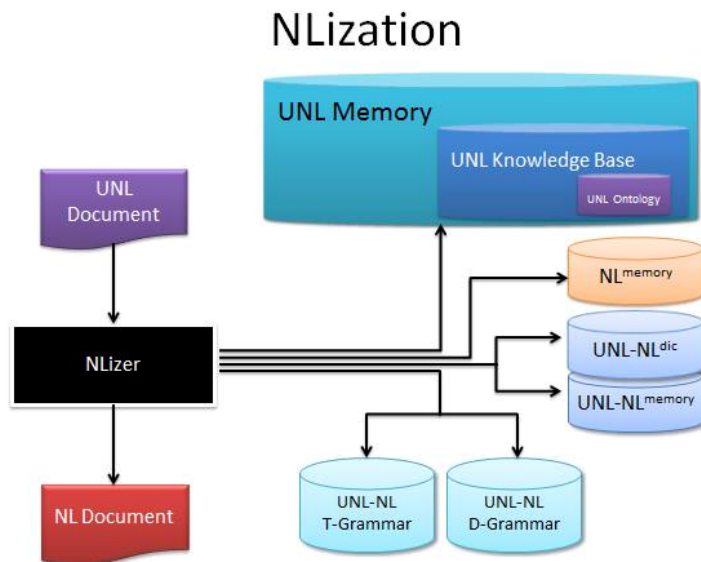
Ստեղծված UNL-հայերեն երկլեզու բառարանի բառերի ու բառային միավորների իմաստները նկարագրված են UNL-ով ու մշակվել է *բառապաշարային տվյալների հենք*⁵ WordNet-ի օրինակով: Բառարանը կարող է ընդգրկվել բազմալեզու բառարանային համակարգում, ինչպիսին է MUHIT համակարգը՝ սրանով իսկ փորձելով ամրագրել անչափ կարևոր ու կենսական մեր լեզուն՝ արևմտահայերենը պահել-պահպանելու պահանջը:

Ավելացնելով *նախորդ փուլում արդեն իսկ կազմված* արևմտահայերեն (արևելահայերեն) -UNL բառարանի բառային միավորների թիվը, տարվում են աշխատանքներ ցանցային որոնողական համակարգ ստեղծելու ուղղությամբ, որը հնարավորություն կտա օգտագործողին կայքի միջոցով տեսնել բառարանի պարունակությունը, կատարել բառերի փնտրում և թարգմանության հետ մեկտեղ ստանալ բառերի ձևաբանական և իմաստաբանական նկարագրություն:

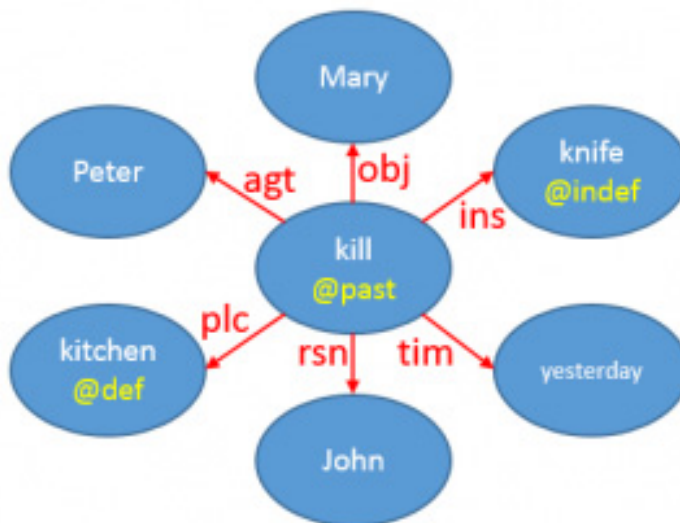
Օգտվելով UNL ցանցից՝ համակարգը հնարավորություն կտա ստանալ հայերեն բառերի բազմաթիվ լեզուներով թարգմանություններ: Օգտագործողը, նշված բառը չգտնելու կամ թերություններով գտնելու դեպքում, կկարողանա առաջարկել թարգմանության և նկարագրության իր տարբերակները, որոնք կդիտարկվեն մասնագետի կողմից՝ հետագայում համակարգին ավելացնելու նպատակով:

Գործարկելով որոնողական բազան՝ հնարավորություն կստեղծվի միավորել ու համախմբել աշխարհասփյուռ հայությունը UNL համակարգի շուրջ:

⁵ <http://www.unlweb.net/muhit/>



Նկ. 1 UNL համակարգ



Նկ. 2 "Peter", "kill", "Mary", "yesterday", "knife", "kitchen" և "John" – համընդհանուր բառերը (Universal Words)⁶, "agt" (agent), "obj" (patient), "tim" (time), "ins" (instrument), "plc" (place) և "rsn" (reason) – համընդհանուր հարաբերություններ (Universal Relations)⁷, "@past", "@def" and "@indef" – համընդհանուր բնութագրեր (Universal Attributes)⁸

⁶ http://www.unlweb.net/wiki/Universal_Words

⁷ http://www.unlweb.net/wiki/Universal_Relations

⁸ http://www.unlweb.net/wiki/Universal_Attributes

LILIT PAREMUZYAN

PHD in Philology

NAS of RA

lili@sci.am

A UNIVERSAL NETWORKING LANGUAGE: THE ARMENIAN MODULE

SUMMARY

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.

ЛИЛИТ ПАРЕМУЗЯН

Кандидат филологических наук

НАН РА

lili@sci.am

УНИВЕРСАЛЬНЫЙ СЕТЕВОЙ ЯЗЫК: АРМЯНСКИЙ МОДУЛЬ

РЕЗЮМЕ

UNL – вычислительный универсальный искусственный сетевой язык, предназначенный для описания, сохранения и распространения глобальной информации от любого естественного языка, независимо от типа компьютерной среды. Работа над армянским модулем началась в 2001 г. Были выполнены работы по созданию формального описания армянского языка, автоматической обработке текста, созданию автоматических словарей и компьютерных баз данных.

В настоящее время группа занята выполнением совместного проекта с Министерством Диаспоры. Результаты проекта будут включать создание UNL-армянского двуязычного словаря, в котором значения будут описаны на UNL, а разработка словарной базы данных – по примеру WordNet. Словарь может включаться в многоязыковую словарную систему, например, система MUNIT.