

Արեւմտահայերէնը նորարար թեքնաբանութեան կիզակէտին

Գալուստ Կիւլպէնկեան Հիմնարկութեան Հայկական Բաժանմունքը ուրախ է յայտարարելու որ *Արեւմտահայերէնի Ծառադարանը* (Treebank) թողարկուեցաւ 15 մայիս 2021-ին եւ այժմ առկայ է Universal Dependencies Consortium-ին մէջ:

Ծառադարանները անհրաժեշտ գործիքներ են լեզուի մը կենսունակութիւնը ապահովելու համար. անոնք կ'երաշխաւորեն որ լեզուաբանական եւ թեքնաբանական գիտութիւններու ոլորտներուն մէջ լեզուն քայլ պահէ արդի զարգացումներուն հետ: Ծառադարաններու շնորհիւ, համակարգչային ծրագիրներ կը «մարզուին» որպէսզի ճանչնան նախադասութեան մը քերականական եւ լեզուական տարբեր տարրերը. այսինքն՝ ծրագիրը կը ճանչնայ բայը, գոյականը, առոգանութեան եւ կետադրութեան նշանները եւ աւելին: Ծառադարանները քերականական մանրամասն ծանօթագրութիւններով օժտուած նախադասութիւններու շտեմարաններ են: Անոնց շնորհիւ յեղափոխութիւն մը տեղի ունեցաւ ինստիտուտական թուականներուն միջազգային *Բնական Լեզուի Մշակում*-ի ոլորտին մէջ:

Ծառադարանները կենսական դեր կը խաղան լեզուաբանական ժամանակակից մշակումի համակարգերու մէջ. օրինակ՝ համակարգչային մեքենայական թարգմանութիւններ, իմաստաբանական վերլուծիչներ, լեզուին բաղադրիչ մասնիկներու ծանօթագրութիւններ եւ այլն: Բաժանմունքի տնօրէն Ռազմիկ Փանոսեան կը բացատրէ՝ «որպէսզի լեզու մը ուղղագրութիւն եւ քերականութիւն սրբագրելու գործիքներ ունենայ, կամ մեքենայական թարգմանութեան հնարաւորութիւնը իրականացնէ, այդ լեզուն պետք է Ծառադարան մը ունենայ: Ուրախ ենք որ Հիմնարկութիւնը կենսական դեր խաղաց այս ժամանակակից գործիքը ի սպաս դնելու անոնց որոնք կը փափաքին լեզուի եւ արհեստագիտութեան ոլորտին մէջ աշխատիլ»:

Universal Dependencies ծրագիր մըն է որ տարբեր լեզուներու համար միջ-լեզուական հասարակաց Ծառադարաններ կը մշակէ, որուն այժմ մաս կը կազմեն թէ՛ արեւելահայերէնը, եւ թէ՛ արեւմտահայերէնը: Այս ծրագիրը այլազան լեզուներու հասարակաց շարահիւսութիւնները կը ճանչնայ եւ կը ծանօթագրէ, արտօնելով նաեւ լեզուական իւրայատկութիւններու ծանօթագրութիւնները:

Այս նոր Ծառադարանը մշակած է ArmTDP խումբը, համադեկավարութեամբ՝ Մարատ Եաւրումեանին (Երեւանի Պետական Համալսարան) եւ Հրանդ Խաչատրեանին («Երեւանէն YerevaNN» հետազօտական կեդրոն), եւ հիմնուած է «Հայերէնի շարահիւսական ծառերու UD դարան»-ի արեւմտահայերէնի բաժնին վրայ: Արեւմտահայերէնի Ծառադարանը կ'աւելնայ 114 տարբեր լեզուներով 202 ծառադարաններուն շարքին:

Ճառադարանը ամբողջութեամբ նոր պատրաստուած է, եւ ուրեմն *Բնական Լեզուի Մշակում*-ի ոլորտին մէջ արեւմտահայերէնի համար ստեղծուող որեւէ գործիքի հիմքը կրնայ հանդիսանալ: 2021-ի աւարտին, պիտի թողարկուի Ճառադարանին թարմացուած եւ ընդլայնուած տարբերակը:

Արեւմտահայերէնի Ճառադարանը այժմ կը պարունակէ 1780 նախադասութիւն եւ 7.5 միլիոն բառ. հիմնուած է 50 հեղինակներու 110 գործերուն վրայ, գրուած 1895-էն 2010-ի միջեւ: Չանագան գրական աշխատութիւններ ընդգրկուած են՝ վիպագրութիւն, նամակագրութիւն, ճամփորդական օրագրութիւն, քաղաքական եւ գրական դասախօսութիւններ ու ճառեր, յուշագրութիւն եւ ակելին: Ան կ'օգտուի *Հայաստանի Ամերիկեան համալսարանի Հայ մատենագրութեան թուանշային գրադարանի* շտեմարանէն:

Արեւմտահայերէնի Ճառադարանը, եւ իր շնորհիւ մշակուած բոլոր *Բնական Լեզուի Մշակում*-ի գործիքները անհրաժեշտ են որպէսզի լեզուն օժտուի նորարար միջոցներով, եւ քայլ պահէ թեքնաբանական առաջընթացի արդի մակարդակին հետ: