

ՏՎՅԱԼՆԵՐԸ՝  
ՈՐՊԵՍ ԲՆԱԿԱՆ ՊԱՇԱՐ  
ԵՎ ՀԶՈՐՈՒԹՅԱՆ ԱՂԲՅՈՒՐ

*Արա Մարջանյան՝*

*Կարեն Վարդանյանի (1987-2016թթ.) հիշատակին*

Արդի աշխարհի թերևս ամենաբնութագրական միտումներից են հետևյալ երկուսը. տվյալները (*data*), հատկապես՝ թվայնացված տվյալները, դառնում են համընդհանուր և ունիվերսալ արժեք՝ հին աշխարհի ոսկու նման, իսկ երկրների կամ կորպորացիաների տվյալների շտեմարաններն այլևս գիտակցվում են որպես կարևորագույն «բնական պաշար»՝ նավթի կամ նույն ոսկու նման: Ավելին, երկրների հզորության (*might*) գնահատականներում, բնական ու մարդկային պաշարներ, տեխնոլոգիական ունակություններ կարևորագույն ցուցիչների կողքին, այսօր բերվում են նաև երկրում առկա տվյալների մեծ կենտրոնների (*SU*Կ, *Big Data Centers*) քանակն ու դրանց որակական դասը (*Tier*)<sup>1</sup>:

Մյուս կարևոր միտումը լեզվական (թվայնացված՝ գրավոր և բանավոր) հենքի վրա տվյալների մեծ շտեմարանների ստեղծման (լեզվական ազգային համակազմներ՝ կորպուսներ, օրական լրահոսի, սոցցանցերի շտեմարաններ, մշակութաբանական և պատմական կորպուսներ և այլն) և այդ տվյալներն այս կամ այն նպատակով մշակելու արդի մեթոդների (*Big Data Mining, Big Data Processing*) ստեղծումն է: Առաջընթացն այստեղ պարզապես աննախադեպ է, թե՛ կիրառական բնույթի տարատեսակ ձեռնարկումներում, թե՛ հիմնարար նշանակության ուսումնասիրություններում (մշակութաբանություն, սոցիալական և լեզվական հոգեբանություն, կոր-

<sup>1</sup> «Նորվանք» ԳԿՀ Քաղաքական հետազոտությունների կենտրոնի ավագ փորձագետ, տ.գ.թ., ա.գ.ա.:

<sup>1</sup> *Մարջանյան Ա.Հ.*, Բառերի տիեզերք II. Կուլտուրոմիկա, Տվյալների մեծ շտեմարանները և մենք, «21-րդ ԴԱԸ», պատրաստվում է տպագրության:

պուսային լեզվաբանություն և այլն), թե՛ գիտելիքների և տվյալների մեծ շտեմարանների ստեղծման առումով<sup>2</sup>: Եվ նորից, երկրների հզորության գնահատականներում լեզվական տարատեսակ կորպուսների առկայությունը ներկայում նույնքան կարևոր դերակատարում ունի, որքան բանակային կորպուսների առկայությունը:

Բերենք կիրառական բնույթի մի քանի օրինակ: Սոցցանցերի ԹԻՀ մշակման արդի մեթոդները թույլ են տալիս իրականացնել տվյալ երկրի հանրության տարբեր խմբերի *բիհեյվորիստական* մոդելավորումը (*Behavioral modeling*)<sup>3</sup> և վարքի կանխատեսումը: Օրական լրահոսի և սոցցանցերի գրառումների «զգացմունքների լեռնահանման» (*text sentiment mining*, 2004) և տեքստերի տեղավայրային կոդավորման (*full-text geocoding*, 2007) միջոցով հնարավոր եղավ մոդելավորել «Արաբական գարնան» իրադարձությունները, դրանց ուժգնությունը՝ Թունիսում և Եգիպտոսում (ռեժիմների փոփոխության) և թուլությունը՝ Մաուդյան Արաբիայում (իշխանափոխության ցածր հավանականություն), տոնայնա-տեղորոշման (*geocoded tone*) գլոբալ գրաֆի կառուցման միջոցով տեղորոշել առանձին անհատների, մասնավորապես Օ. բին Լադենի գտնվելու վայրը<sup>4</sup>: Իսկ օգտատերերի պրոֆիլավորման արդի մեթոդներն էականորեն օժանդակեցին ԱՄՆ-ում նախագահական ընտրություններում Դ.Թրամփի հաղթանակին<sup>5</sup>:

### **Հայ Վիքիպեդիան՝ որպես կորպուս. քանակական ասպեկտ**

Ցավոք, նշված ուղղություններով Հայաստանում ուսումնասիրությունները գտնվում են դեռ սաղմնային վիճակում. թերևս, մեկ բացառությամբ՝ խոսքը հայերեն *Wikipedia*-ի մասին է:

<sup>2</sup> Մարջանյան Ս.Հ., Բառերի տիեզերք. Հաճախակիության բառարանները և Google N-grams գործիքը, «21-րդ ԴԱԸ», N4 (74), 2017:

<sup>3</sup> Eagle N., Pentland A.S., Eigenbehaviors: identifying structure in routine. *Behav. Ecol. Sociobiol.*, 63, 2009, pp. 1057-1066.

<sup>4</sup> Leetaru K.H., Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone In Time And Space". *First Monday*, No 16 (9), 5 September 2011.

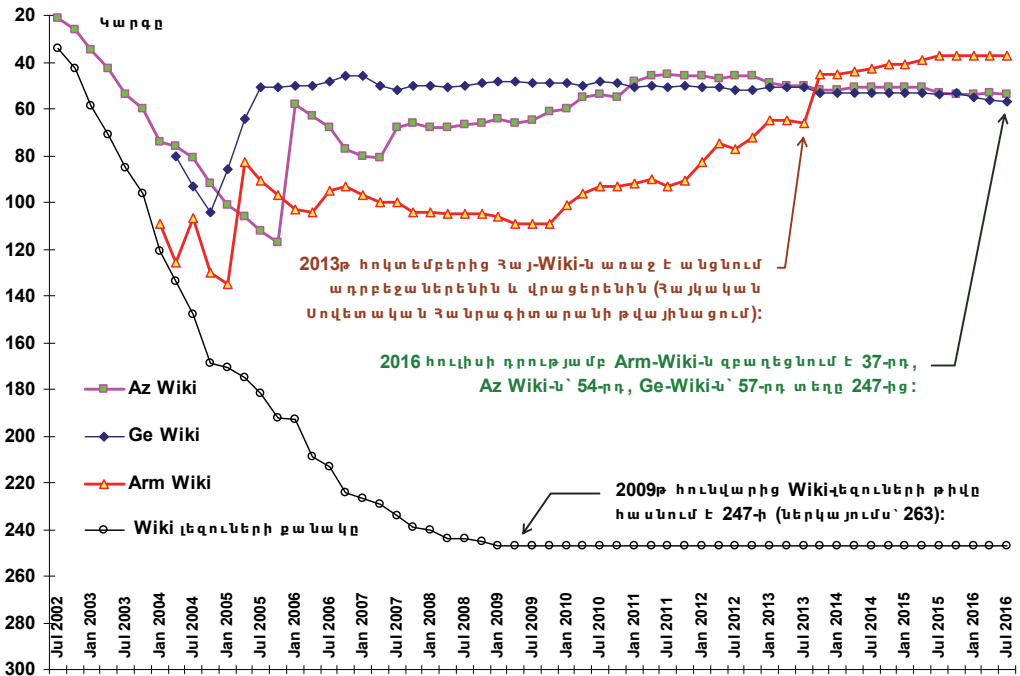
<sup>5</sup> Youyou Wu, Kosinski, Michal Stillwell David, Computer-based personality judgments are more accurate than those made by humans. *PNAS*, vol. 112, no. 4. January 27, 2015; Трамп победил на выборах благодаря ученым и Big Data. *Das Magazin*, 8 декабря 2016,

<https://hi-tech.mail.ru/news/how-trump-won/?frommail=1&fromnews=1>

Ընդհանրապես, *Wikipedia*-ն թողարկվել է 2001թ. հունվարի 15-ին, միայն անգլերեն: 2001թ. ավարտին առկա էին արդեն 18 լեզուներով 20.000 հոդվածներ: Լեզուների թիվը հասավ 26-ի՝ 2002-ի վերջին, 46-ի՝ 2003-ին և 161-ի՝ 2004-ի ավարտին: 2009-ի հունվարին լեզուների թիվը հասավ 247-ի, իսկ ներկայում այն կազմում է 263: Հոդվածների թվի առումով 2007թ. սեպտեմբերին անգլերեն *Wiki* հանրագիտարանը (ավելի քան 2 մլն հոդված) առաջին անգամ գերազանցեց դեռ Մին դինաստիայի օրոք (1408թ.) կազմված և անցած 6 դարերի ռեկորդակիր չինական Յուն Չժի (*Yon gle*) հանրագիտարանին: 2005-ին հեղինակավոր *Nature* ամսագրի տվյալներով՝ *Wikipedia*-ն (անգլերեն կորպուսը) իր ճշգրտությամբ մոտեցավ հռչակավոր *Encyclopædia Britannica* հանրագիտարանին<sup>6</sup>:

Գծապատկեր 1

Հայերեն, վրացերեն և ադրբեջաներեն Wiki-կորպուսների կարգաչիսն տեղը 2017թ. հունվարի դրությամբ



Ադրբեջաներեն *Wikipedia*-ն մեկնարկեց 2002թ. հուլիսին (Գծապատկեր 1): 2002-ին իր հիմնական քանակական ցուցանիշով՝ *Wiki* հոդվածների թվով (A) այն զբաղեցնում էր 21-րդ տեղը (կարգը) 34-ից: Մինչ 2004-ի հոկտեմբերը

<sup>6</sup> Internet encyclopaedias go head to head with *Encyclopaedia Britannica*. *Nature*, 438, 900-901; 2005.

ադրբեջաներեն *Wiki*-կորպուսը Անդրկովկասի առաջատարն էր: Այն զիջում է վրացերենին միայն 2011թ. հունվարին: Հայալեզու *Wiki*-ն սկսեց բարելավել իր կարգային տեղը՝ սկսած 2009-ից, երբ հոկտեմբերին գրանցած 109-րդ կարգից հասավ 66-ի 2013թ. հուլիսին: Վճռական թռիչքը տեղի ունեցավ 2013-ի հոկտեմբերին, երբ հայալեզու *Wiki*-ն իր շուրջ 47 հազ. հոդվածներով հայտնվեց առաջին 50-յակում՝ գրավելով 45-րդ կարգը: 2016թ. հոկտեմբերին հայերեն *Wiki*-ն իր մոտ 210 հազ. հոդվածներով զբաղեցնում էր 37-րդ կարգը՝ շոշափելիորեն գերազանցելով ադրբեջաներեն (111 հազ. հոդված, 54-րդ կարգ) և վրացերեն (109 հազ. հոդված, 57-րդ կարգ) *Wiki*-կորպուսներին: Հայերեն *Wiki*-ի լեզվական կորպուսի ծավալը, եթե այն գոյություն ունենար, կկազմեր մոտ 20-21 MiB (ազգային լեզվական կորպուսների և լեզվական շտեմարանների մասին ավելի մանրամասն տե՛ս<sup>7</sup>): Հավելենք, որ հազեցումով պրոցեսներին հաստուկ օրինաչափությամբ ադրբեջաներեն և վրացերեն հոդվածներից հայալեզու *Wiki*-հոդվածների թվի գրեթե կրկնակի գերազանցումը (100%) ողջ *Wikipedia*-ի կտրվածքով ապահովում է ընդամենը 18-20 կարգային տեղի առաջխաղացում (7-8%): Պարետոյի օրենքի<sup>8</sup> «ջանքի 80%-ը կպահանջվի արդյունքի վերջին 20%-ին հասնելու համար» տրամաբանությունը հուշում է, որ հետագա առաջխաղացման համար (օրինակ՝ տեղ զբաղեցնել առաջին 30 կամ 20-յակում) անհրաժեշտ կլինի որակական նոր թռիչք:

Ընդգծենք հետևյալը: 2013թ. վճռական թռիչքը հնարավոր եղավ իրականացնել այն բանից հետո միայն, երբ թվայնացվեց և համացանցում հանրայնացվեց 1974-87թթ. հրատարակված, շուրջ 50 հազ. հոդված ունեցող Հայկական սովետական հանրագիտարանը (ՀՍՀ)<sup>9</sup>, ինչը լուրջ ներդրում էր հայերեն լեզվական դաշտում համակարգված, որակյալ, թվայնացված կորպուսի ստեղծման գործում: Խորհրդային հանրապետությունների մեջ բարձր վարկանիշ ունեցող ՀՍՀ-ն այսպիսով նպաստեց Հայաստանի 3-րդ հանրապե-

<sup>7</sup> Մարջանյան Ս.Հ., Բառերի տիեզերք. Հաճախակիության բառարանները և Google N-grams գործիքը, «21-րդ ԴԱԸ», N4 (74), 2017:

<sup>8</sup> 2016-ի դեկտեմբերի դրությամբ հայալեզու Wiki-կորպուսում բացակայում էին «Պարետոյի օրենք», ինչպես նաև «Ցիպֆի օրենք», լեզվահաճախականային վերլուծություն, կորպուսային լեզվաբանություն և այլ հարակից հոդվածները:

<sup>9</sup> Հայաստանի Հանրային ռադիո. Հայկական սովետական հանրագիտարանն այսուհետ հասանելի է նաև Վիքիպեդիայում: 18:20, 02 օգո. 2013. Հայկական հանրագիտարան ստեղծելու 1932-36թթ. առաջին փորձը ձախողվեց:

տության «թվայնացվածության» կայացմանը: Խնդիրն այստեղ հետևյալն է. արդյո՞ք ունակ կլինի Հայաստանի 3-րդ հանրապետությունը գեներացնել իմաստային առումով՝ խոր, կառուցվածքային առումով՝ լավ ստրուկտուրացված և փոխկապակցված, իմաստների (տվյալների, գիտելիքների) այնպիսի բազմություն, որը ոչ միայն համեմատելի լիներ ՀՄՀ-ի հետ, այլև գերազանցեր այն թե՛ քանակապես, թե՛ որակապես: Քանզի գիտելիքների և տվյալների (*data*) արդի, վերին աստիճանի մրցունակ ու դաժան աշխարհում նախկին քանակա-որակական ցուցանիշներն այլևս բավարար չեն զարգացման համար:

### **Վիքիպեդիա. լեզվաքանակական և որակական ասպեկտ**

Համաձայն հեղինակավոր *Ethnologue* ամսագրի<sup>10</sup>, այսօր աշխարհում գոյություն ունի 7099 կենդանի լեզու: Մյուս կողմից, ներկայում աշխարհում առկա է 195 ճանաչված պետություն, որից 193-ը՝ ՄԱԿ անդամ: Այսպիսով, կենդանի լեզու ունեցող մարդկային հանրությունների կողմից պետականությունն իրացված է միայն մոտ 2.75% կողմից: Գծապատկեր 2-ում ամբողջացրել ենք 50 մլն մարդ և ավելի կրող (մայրենի՝ L1, 2-րդ լեզու՝ L2) ունեցող 26 լեզուների մասին *Ethnologue*, *SIL Int.* տվյալները: Գումարային առումով դրանք կազմում են 6.6 մլրդ մարդ՝ մոլորակի այսօրվա բնակչության (7.4 մլրդ) 89%-ը: Լուսամուտում բերվում է նաև կրողների կառուցվածքը (մլն մարդ և %) ըստ լեզվական ընտանիքների:

Ինչպես տեսնում ենք, կրողների մոտ 41%-ը պատկանում է հնդեվրոպական լեզվաընտանիքին, ևս 20%-ը՝ հնդարիական, ներառյալ՝ դրավիդյան լեզուները: Հետևաբար, բնակչության շուրջ 61%-ը պատկանում է հնդեվրոպա-արիական գերընտանիքին, այդ թվում L1 կրողների 58%-ը (մոտ 2.73 մլրդ, որից 1.69 մլրդ՝ հնդեվրոպական), L2 կրողների՝ 65%-ը (մոտ 1.27 մլրդ, որից մոտ 1.02 մլրդ՝ հնդեվրոպական)<sup>11</sup>: Իսկ Սինոտիբեթյան լեզվաընտանիքին է պատկանում միայն 18%-ը, կամ L1 կրողների

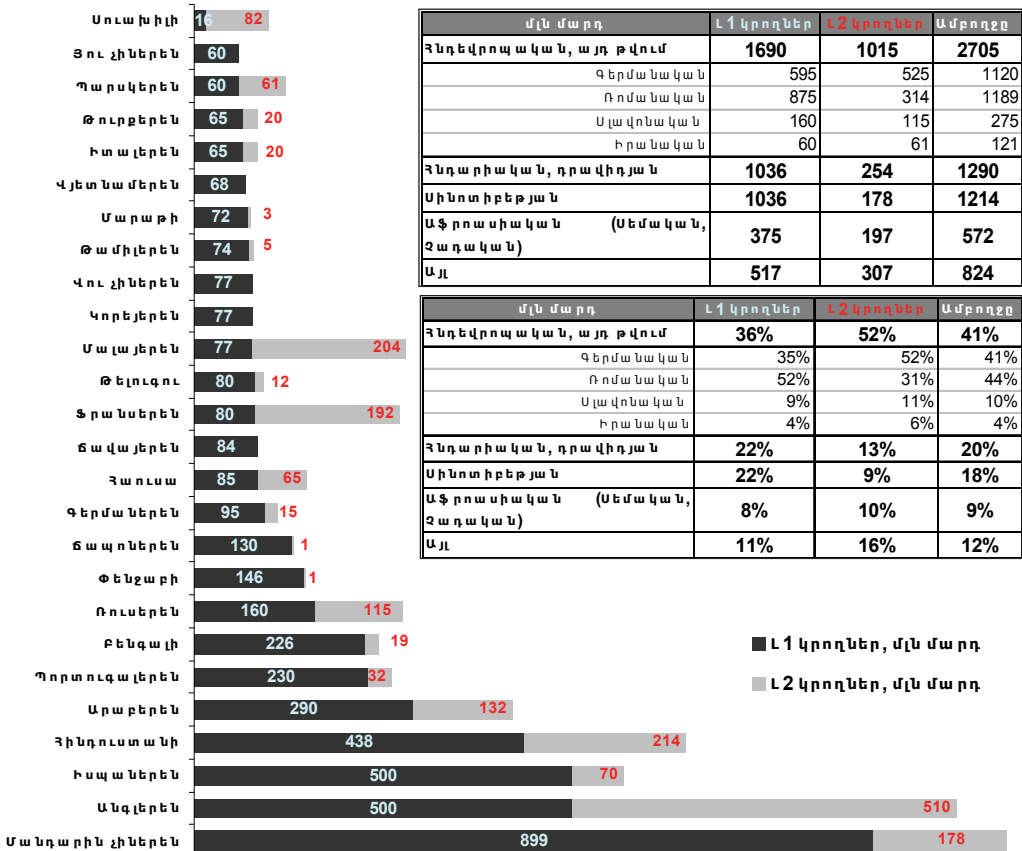
<sup>10</sup> *Ethnologue: Languages of the World*. 20-th edition, Simons & Fennig, 2017, <https://www.ethnologue.com/about>

<sup>11</sup> Մենք չենք անդրադառնում L2/L1 հարաբերությանը, որը չափազանց էական ցուցիչ է «մոլորակի տարածաշրջանների» (*Regions of the World*) ձևավորման և զարգացման առումով: Մենք դրանք կոչում ենք «Աշխարհի աշխարհներ» (*Марджанян А.А., Арутюнян, Г.А., Союзники на фоне научно-технологического пейзажа. Тр. межд. конференции «Демократия и сетевые технологии в современном социально-политическом процессе», 18 мая, 2017. М. ФИП/РГГУ, 2017.*): Տարածում է ստացել նաև դրանց «աշխարհ-համակարգ» (*мир-система, world-system*) անվանումը:

22%-ը (մոտ 1.04 մլրդ) և L2 կրողների 9%-ը (178 մլն): Առանձին լեզուների տեսանկյունից ամենաբազմաթիվը Մանդարին չինարենն է (ներառյալ դասական չինարենը)՝ 899 մլն L1, և մոտ 178 մլն L2 կրողներ, որին հաջորդում են անգլերենն ու իսպաներենը (500 մլն L1 կրողներ յուրաքանչյուրը), և հինդուստանին (հինդի և ուրդու, 438 մլն L1 կրողներ): Առաջին 6-յակը եզրափակում են արաբերենն ու պորտուգալերենը (տես՝ Գծապատկեր 2):

Գծապատկեր 2

Լեզվի (լեզվական ընտանիքների) կրողներ, մլն մարդ.  
L1 կրողներ՝ մայրենի լեզու, L2 կրողներ՝ որպես 2-րդ լեզու:  
Ըստ Ethnologue 2015, SIL Int.



Միևնույն ժամանակ, 2017թ. մայիսի դրությամբ *Wikipedia*-ն առկա էր 263 լեզուներով: Այսինքն՝ համաշխարհային հանրագիտարանում կենդանի լեզու ունեցող մարդկային հանրությունների ներկայացվածության

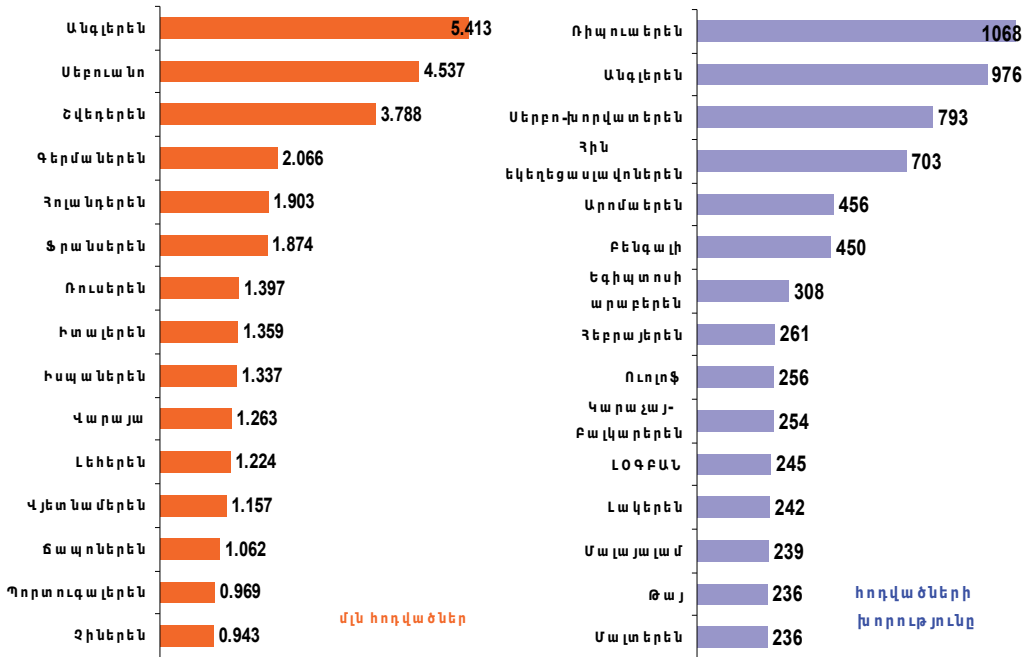
աստիճանը կազմում է մոտ 3.94% կամ 43%-ով ավելի, քան պետականության իրացման ցուցիչը: Ընդ որում, լեզվամշակութային դաշտի պահպանման և զարգացման առումով, հատկապես վտանգված լեզուների<sup>12</sup> և պետականություն չունեցող ազգային փոքրամասնությունների համար, ինչպես նաև աշխարհաքաղաքական գործընթացների «Ազգերի կերտման» (*Nation building*) տեսանկյունից, վերլուծաբանների համար մեծ նշանակություն ունեն *Wikipedia*-ի լեզվական կորպուսների քանակաորակական ցուցանիշները և դրանց դինամիկայի գնումը: Դրանցից երկուսը հանդիսանում են *Wiki* տվյալ լեզվական կորպուսում հոդվածների քանակը (A) և այսպես կոչված հոդվածների խմբագրության խորությունը կամ պարզապես խորությունը (D)<sup>13</sup>: Գծապատկեր 3-ում՝ ձախից, բերում ենք *Wiki*-հոդվածների թվով առաջատար 15 լեզվական կորպուսները: 2017թ. հունվարի դրությամբ բացարձակ առաջատարը շարունակում էր մնալ անգլերենը՝ մոտ 5.3 մլն հոդված (5.0 մլն շեմը այն հասել է 2015-ի սեպտեմբերին, 4.0 մլն-ը՝ 2012թ. հոկտեմբերին, 1.0 մլն-ը՝ 2006թ. ապրիլին): Առաջատարների ցանկում է նաև շվեդերենը՝ մոտ 3.7 մլն հոդված, գերմաներենը՝ մոտ 2.1 մլն հոդված (1.0 մլն շեմը այն հասել է 2009-ի օգոստոսին), հոլանդերենը՝ մոտ 1.9 մլն հոդված (1.0 մլն շեմը՝ 2012թ. հունիսին), ֆրանսերենը և ռուսերենը՝ մոտ 1.4 մլն հոդված (1.0 մլն շեմը հատվել է 2013թ. մայիսին): Ուշագրավ է, որ թուրքերեն *Wiki*-կորպուսը 2016թ. հոկտեմբերի դրությամբ ուներ մոտ 285 հազ. հոդված, ինչը հայերեն *Wiki*-կորպուսից ավելին է ընդամենը 35%-ով: Այնինչ, 2016-ին Թուրքիայի բնակչությունը գերազանցում էր Հայաստանի բնակչությանը մոտ 27 անգամ (2700%-ով):

<sup>12</sup> Լուրջ բացթողում պետք է համարել այն հանգամանքը, որ 2000-ականների սկզբին 3-րդ հանրապետությունը թերացավ գրանցել արևմտահայերենը, համշեներենը և այլ հայկական լեզուներ՝ որպես վտանգված լեզուներ: Աններելի է նաև արևմտահայերեն *Wiki*-կորպուսի բացակայությունը և արևմտահայերենի գրանցման բացակայությունը *Wikimedia*-յում: Հավելենք, որ UNESCO-ի Վտանգված լեզուների ատլասը (*UNESCO Atlas of the World's Languages in Danger*) ճանաչում է Հայաստանի համար՝ 2 լեզու (Pontic Greek, Suret), Վրաստանի համար՝ 11 լեզու (այդ թվում և համշեներենը), Ադրբեջանի համար՝ 11 լեզու (բուղուլ, ջուհուր, խինալուզ, կիլիթ, կրիզ, լեզգի, ռութուլ, թալիշ, թաթ, ծախուր և ուրի), <http://www.unesco.org/languages-atlas/index.php?hl=en&page=atlasmap>

<sup>13</sup> Նկատենք, որ սա *Wiki*-կորպուսի որակական միակ չափելի ցուցանիշն է: Հաղվածների ակադեմիական որակի այլ գնահատականներն անխուսափելիորեն սուբյեկտիվ են: Ավելի մանրամասն տե՛ս [https://meta.wikimedia.org/wiki/Wikipedia\\_article\\_depth](https://meta.wikimedia.org/wiki/Wikipedia_article_depth)

Գծապատկեր 3

Wiki լեզվական կորպուսների առաջատար 15-յակը՝ ըստ հոդվածների և խորության, 2017թ. մայիսի 27-ի դրությամբ



Ուշագրավ է նաև այն, որ առաջատարների ցանկում են Ֆիլիպիններում տարածված սերբուհերեն (4.54 մլն հոդված, 2-րդ տեղ) և վարայա-վարայա (1.26 մլն հոդված, 10-րդ տեղ) լեզուները: Այնինչ, լեզու կրողների առումով (սերբուհերեն՝ 15.8 մլն L1 և L2 կրող, վարայա-վարայա՝ 2.56 մլն L1 և L2 կրող)<sup>14</sup>, այս լեզվական կորպուսները կարծես թե չպետք է ունենային հոդվածների այսքան քանակ: Բանն այստեղ այն է, որ այս քանակական արդյունքն ըստ էության միայն մեկ անհատի՝ շվեդերեն *Wikipedia*-ի ակտիվիստ և ծրագրավորող Ս.Յոհանսոնի (*Sverker Johansson*) գործունեության հետևանք է, որը ժամանակին մշակեց և 2012-ի վերջին թողարկեց *Lsjbot* համացանցային ռոբոտ ծրագիրը, սերբուհերեն լեզվով *Wiki*-հոդվածների ստեղծման համար<sup>15</sup>:

<sup>14</sup> <https://www.ethnologue.com/language/ceb> , <https://www.ethnologue.com/language/war>

<sup>15</sup> *Lsjbot*-ը ապահովել է սերբուհերեն *Wiki*-կորպուսի ծավալի 2/3-րդը և շվեդերեն *Wiki*-կորպուսի ավելի քան 1.0 մլն հոդվածների ստեղծումը, <https://www.quora.com/Why-are-there-so-many-articles-in-the-Cebuano-language-on-Wikipedia>



Այսպիսով, վերլուծությունների համար դեռ բավարար չէ հիմնվել միայն քանակական ցուցանիշների վրա: Ուստի, Գծապատկեր 3-ում՝ աջից, բերում ենք նաև առաջատար *Wiki*-կորպուսների 15-յակը՝ դասակարգված ըստ հոդվածների խորության: Ինչպես տեսնում ենք, ցանկն այստեղ միանգամայն այլ է՝ մեկ հատկանշական բացառությամբ: Անգլերեն *Wiki*-կորպուսն առաջատար է թե՛ հոդվածների քանակով, թե՛ հոդվածների խորությամբ: Այս հանգամանքը լավագույնս արտացոլում է «Անգլոսաքսոնական աշխարհի» ներքին աշխարհաքաղաքական, լեզվական ու քաղաքակրթական կոնսոլիդացիան<sup>16</sup>, անգլերեն լեզվի կրողների բաշխվածությունը (տես՝ Գծապատկեր 3) և լեզվի հաճախա-հակիրճային առանձնահատկությունները<sup>17</sup>: Մինչդեռ, օրինակ, ծավալով 2-րդ ու շատ իմաստներով՝ արևեստածին սերուանո *Wiki*-կորպուսը խորությամբ ընդամենը 252-րդն է 263-ից: Հավելենք, որ 2017թ. մայիսի դրությամբ հայերեն *Wiki*-կորպուսի խորությունը կազմում էր ընդամենը 35.41 (123-րդ կարգ 263-ից), այնինչ ծավալային առումով այն գրավում էր 36-րդ կարգը՝ 263-ից<sup>18</sup>:

Տվյալ լեզվական *Wiki*-կորպուսի քանակա-որակային և ժամանակագրական առանձնահատկությունները հանդիսանում են աշխարհաքաղաքական, մշակութային, քաղաքական և լեզվա-ազգայնական գործընթացների ինտեգրալ ցուցիչ: Այս առումով օգտակար է դիտարկել աշխարհի *Wiki*-կորպուսների «քարտեզը»՝ կառուցված հոդվածների թիվ (A)-հոդվածների խորություն (D) առանցքներով (լոգարիթմական մասշտաբ): Այդպիսի մի օրինակ բերում ենք Գծապատկեր 4-ում, ուր նշված են որոշ *Wiki*-կորպուսների դիրքերը 2017թ. մայիսի 27-ի դրությամբ: Հատուկ ընդգծենք այն հանգամանքը, որ *Wiki*-կորպուսների խորությունը (D) հանդիսանում է տվյալ լեզվի (լեզվական խմբի) ներսում և շուրջ, ինչպես նաև դրա սփռվածության տարածքում հանրային, հասարակական, քա-

<sup>16</sup> *Марджанян А.А., Арутюнян, Г.А.*, Союзники на фоне научно-технологического пейзажа. Тр. между. конференции «Демократия и сетевые технологии в современном социально-политическом процессе», 18 мая, 2017. М. ФИП/РГГУ, 2017.

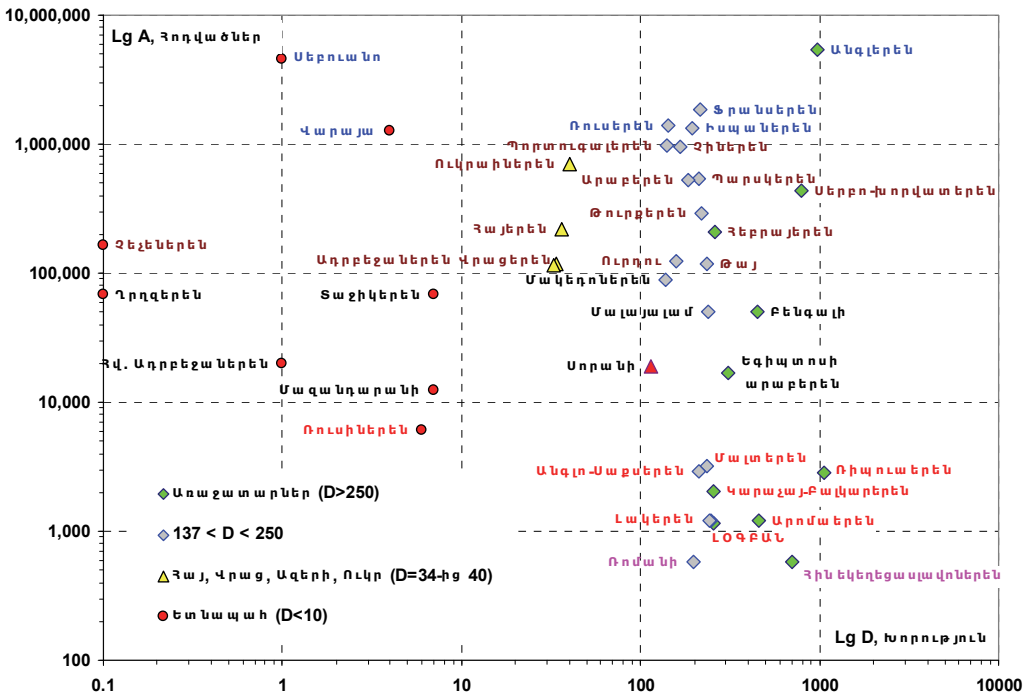
<sup>17</sup> *Մարջանյան Ա.Հ.*, Բառերի տիեզերք. Հաճախակիության բառարանները և Google N-grams գործիքը, *Նկար 1*, «21-րդ ԴԱԸ», N4 (74), 2017:

<sup>18</sup> 2017թ. մայիսի 27-ի դրությամբ թուրքերեն *Wiki*-կորպուսը գրավում էր 29-րդ կարգը՝ ըստ ծավալի և 17-րդը՝ ըստ խորության (D=221), ադրբեջաներենը՝ 56-րդն ըստ ծավալի և 130-րդը՝ ըստ խորության (D=34), վրացերենը՝ 58-րդն ըստ ծավալի և 133-րդը՝ ըստ խորության (D=33):

դաքական և աշխարհաքաղաքական ակտիվության ինտեգրալ ցուցանիշ: Իրապես օգտակար կլինեն այսպիսի քարտեզի պատկերումը ժամանակի մեջ, որի դեպքում յուրաքանչյուր լեզվական կորպուսի նիշն իր ուրույն հետազիծը կգծեր նշված առանցքներով հարթության վրա<sup>19</sup>: Այսպիսի դի- նամիկ պատկերը կհանդիսանար քաղաքակրթական և աշխարհաքաղա- քական լարվածությունների յուրօրինակ «ռադարային դիագրամ»:

Գծապատկեր 4

Որոշ Wiki-կորպուսները՝ ըստ հոդվածների քանակի (y-ի առանցք, lg մասշտաբ) և խորության (x-ի առանցք, lg մասշտաբ), 2017թ. մայիսի 27-ի դրությամբ



Եզրափակելով հոդվածը՝ նշենք, որ, օրինակ, կատալոներեն Wiki- կորպուսը համացանցում թողարկվեց 2001-ին, անգլերեն կորպուսի թո- դարկումից տասնացիորեն բուպեներ անց: Այսպիսով, այն առաջին ոչ անգ- լերեն Wiki-կորպուսն էր աշխարհում: Իսկ կենտրոնական քրդերենի՝ սո-

<sup>19</sup> Այս տրամաբանությամբ է կառուցվում, օրինակ, տվյալ երկրի տնտեսական-գիտատեխնո- լոգիական «հետազիծը» (Հարությունյան Գ.Ա., Մարջանյան Ա.Հ., Բազմակենտրոն աշխարհա- կարգի գիտատեխնոլոգիական բնապատկերը: "21-րդ ԴԱԸ", N3 (73), 2017, Նկար 6):

րանի կորպուսի խորությամբ 2017թ. մայիսի դրությամբ կազմում է  $D=113.87$ , ինչն ապահովում է դրա 47-րդ կարգային տեղը: Այնինչ, ըստ ծավալի, սորանի կորպուսն ընդամենը 112-րդն է 263-ից: Ավելին, սորանի կորպուսն էապես գերազանցում է քրդական մի այլ *Wiki*-կորպուսի՝ հս. քրդերենի կուրմանջի կորպուսին, որը ներկայում ունի  $D=29.95$  խորություն (144 կարգը 263-ից), թեև ունի մոտ 1.5 անգամ ավելի հոդված, քան սորանին: Ընդհանրապես, 2017թ. մայիսի դրությամբ, խորության ցուցանիշով սորանին առաջ էր, օրինակ, 92 մլն Լ1 և Լ2 լեզվական կրող ունեցող թելուգու կորպուսից (դրավիդյան լեզվաընտանիք) կամ մոտ 2.1 մլն հոդված ունեցող և ծավալով 6-րդ՝ գերմաներեն *Wiki*-կորպուսից:

Այս համատեքստում միանգամայն այլ լույսի ներքո է դիտվում այն հանգամանքը, որ 2017թ. սեպտեմբերին նախատեսվում է անցկացնել Իրաքյան Քրդստանի ինքնորոշման հանրաքվեն (տե՛ս Քարտեզ 1), իսկ հոկտեմբերին՝ կատալոնական ինքնորոշման հանրաքվեն Իսպանիայում:

Քարտեզ 1

*Քրդական լեզուների աշխարհագրական բաշխումը,  
Le Monde diplomatique, 2007*

