

AKADEMIA E SHKENCAVE DHE E ARTEVE E KOSOVËS

AKADEMIA E SHKENCAVE E SHQIPËRISË

SHQIPJA DHE GJUHËT E BALLKANIT

ALBANIAN AND BALKAN LANGUAGES

Konferencë shkencore e mbajtur më 10-11 nëntor 2011 në Prishtinë
Scientific Conference held on 10-11 November 2011 in Prishtina



PRISHTINË
2012

KOSOVA ACADEMY OF SCIENCES AND ARTS

ACADEMY OF SCIENCES OF ALBANIA

ALBANIAN AND BALKAN LANGUAGES

Scientific Conference held on 10-11 November 2011 in Prishtina

Editor

Academician Rexhep Ismajli



PRISHTINA
2012

AKADEMIA E SHKENCAVE DHE E ARTEVE E KOSOVËS

AKADEMIA E SHKENCAVE E SHQIPËRISË

SHQIPJA DHE GJUHËT E BALLKANIT

Konferencë shkencore e mbajtur më 10-11 nëntor 2011 në Prishtinë

Redaktor
Akademik Rexhep Ismajli



PRISHTINË
2012

**Programi i punës i Konferencës shkencore “Shqipja dhe gjuhët e Ballkanit”,
Prishtinë 10-11 nëntor 2011**

E enjte, 10 nëntor 2011

Ora 9:00

Seancën e drejtojnë: Besim **BOKSHI**, Irena **SAWICKA**, Kolec **TOPALLI**

1. Rexhep **ISMAJLI**, Prishtinë:

Shqipja dhe gjuhët e Ballkanit – fjala e hapjes

2. Idriz **AJETI**, Prishtinë:

Shqipja dhe gjuhët fqinje

3. Besim **BOKSHI**, Prishtinë:

Format e futurit të shqipes në raport me gjuhët e Ballkanit

4. Bardhyl **DEMIRAJ**, Muenchen:

Velle futurum si dukuri dialektore dhe ballkanike në gjuhën shqipe

5. Seit **MANSAKU**, Tiranë:

Tipologji romane dhe ballkanike në strukturën gramatikore të shqipes

6. Francesco **ALTIMARI**, Cosenza:

Disa ndikime të greqishtes në fjalëformimin foljor të arbërishtes së Italisë

Ora 11:30

Seancën e drejtojnë: Bahri **BECI**, Heiner **EICHNER**, Francesco **ALTIMARI**

1. Heiner **EICHNER**, WIEN:

Protoshqipja në mesin e gjuhëve të lashta të Ballkanit

2. Wilfried **FIEDLER**, Meissen:

Disa veçori të urdhërores në gjuhët ballkanike

3. Victor **FRIEDMAN**, Chicago:

Problemi i urdhërores treguese në gjuhën shqipe dhe në gjuhët e tjera të Ballkanit

4. Shaban **DEMIRAJ**, Tiranë:

Ç’ dëshmon mënjanimi i paskajores në gjuhët ballkanike?

5. Irena **SAWICKA**, Warszawa:

Lidhjet fonetike të shqipes në kontekst ballkanik

6. Bahri **BECI**, Paris:

Konsiderata të përgjithshme për rolin e shqipes në formimin e bashkësisë gjuhësore ballkanike

7. Ethem **LIKA**, Tiranë:

Karakteristika tipologjike të shqipes

Ora 15:30

Seancën e drejtojnë: Victor **FRIEDMAN**, Shefki **SEJDIU**, Bardhyl **DEMIRAJ**

1. Bardh **RUGOVA**, Prishtinë:

Pozita dhe statusi i shqipes kundruall gjuhëve të tjera të Ballkanit

2. Kolec **TOPALLI**, Tiranë:

Shtresimi i huazimeve nga gjuhët ballkanike në një fjalor etimologjik

3. Rexhep **ISMAJLI**, Prishtinë:

Shqipja dhe rumanishtja

4. Marcus **GABINSCHI**, Chişinău:
Saktësime rreth fondit leksikor të quajtur „rumuno-shqiptar”
5. Shefki **SEJDIU**, Prishtinë:
Gjurmë interferimesh emërtimi helen(iko)-greke, ilir(iko)-albane dhe ital(iko)-romane në fushë të etnonimisë
6. Teuta **ABRASHI**, Prishtinë:
Nyja e prapavendosur në gjuhët e Ballkanit – ngjashmëri dhe dallime morfosintaksore
7. Aristotel **SPIRO**, Tiranë:
Përsëritja e nyjës para elementeve të sintagmës në shqipe dhe greqishte dhe problemi i nyjës shquese në shqipe

Ora 17:30

Seancën e drejtojnë: Rexhep **ISMAJLI**, Leonardo **SAVOIA**, Seit **MANSAKU**

1. Matteo **MANDALÀ**, Palermo:
Mbi turqizmat në të folmet arbëreshe
2. Gjorgji **BUFLI**, Trieste:
Kontaktet ndërgjuhësore ballkanike në rastin e turqizrave
3. Lindita **LATIFI**, Tiranë:
Për një fjalor ballkanik të turqizrave
4. Skender **GASHI**, Wien:
Gjurmë të latinitetit e të romanitetit ballkanik në mikrotoponiminë e Kosovës
5. Evalda **PACI**, Tiranë:
Elemente latine në shqipen dhe rumanishten, veshtrim semantik dhe tipologjik
6. Qemal **MURATI**, Prishtinë:
Shqipja dhe maqedoarumanishtja nga aspekti i kontakteve midis tyre
7. Artur **KARASIŃSKI**, Torunj:
Një projekt për fjalëformimin ballkanik

E premtë, 11 nëntor 2011**Ora 9:00**

Seancën e drejtojnë: Murat **BLAKU**, Matteo **MANDALÀ**, Ethem **LIKA**

1. Remzi **PËRNASKA**, Paris:
Bashkësia e “trajtave të shkurtra të përemrit vetor” dhe zgjedhimi subjektor-objektor sui generis në gjuhën shqipe dhe në gjuhët ballkanike
2. Rita **MANZINI** & Leonardo **SAVOIA**, Firenze:
Case as Nominal Property: the Interaction of DOM and Discourse-linking in Shkodër Pronouns
3. Giovanni **BELLUSCIO**, Cosenza:
Si shqiptohet shqipja në Ballkan?
4. Blertë **ISMAJLI**, Prishtinë:
Sa ballkanike janë mënyrat e foljeve të shqipes? Kushtorja dhe habitoret
5. Julie **KOLGJINI**, Prishtinë:
Një strukturë alternative foljore në gjuhën shqipe: ndikimi prej një gjuhe të arealit ballkanik

6. Dieter **NEHRING**, Berlin:

Shqipja dhe gjuhët ballkanike në arsimin shkollor tek shqiptarët dhe fqinjët e tyre

Ora 11:30

Seancën e drejtojnë: Imri **BADALLAJ**, Anila **OMARI**, Skender **GASHI**

1. Murat **BLAKU**, Prishtinë:

Interferenca të shqipes në gjuhët sllave të jugut

2. Xhelal **YLLI**, Heidelberg:

Rreth vlerësimit të huazimeve sllave në gjuhën shqipe

3. Marjan **MARKOVIĆ**, Shkup:

Balkan-Slavic lexical parallels

4. Imri **BADALLAJ**, Prishtinë:

Infinitivi në gjuhën shqipe dhe në gjuhën serbe

5. Anila **OMARI**, Tiranë:

Shtresëzimit kronologjik dhe shtrirja dialektore e huazimeve sllave të shqipes së shek. XVII

6. Shkumbin **MUNISHI**, Prishtinë:

Probleme të zbatimit të dygjuhësisë shqip/serbisht në Kosovë

7. Haxhi **SHABANI**, Ulqin:

Ndikimi i gjuhës malazeze në emrat dhe mbiemrat e njerëzve

Ora 15:00

Seancën e drejtojnë: Bardh **RUGOVA**, Maria **MOROZOVA**, Remzi **PËRNASKA**

1. Martin **SUROVČÁK**, Brno:

Mallkimi si çështje e frazeologjisë ballkanike

2. Monica **GENESIN** & Domenico **MARTUCCI**, Bari:

Perceptimi i gjuhës dhe i kulturës shqipe nëpërmjet vëllimeve: “Le terre albanesi redenti: Cossovo” (1942) dhe “Le terre albanesi redenti: Ciamuria” (i pabotuar)

3. Dhori **QIRJAZI**, Selanik:

“Mjaltë” nga hojet dhe trojet e shqipes: Fjalori dorëshkrim shqip-greqisht i Th. Mitkos dhe suaza e tij ballkanike

4. Besim **KABASHI**, Erlangen:

Korpuse gjuhësore për shqipen

5. Timofej **ARKHANGELSKIJ** (Moskva), Mikhail **DANIEL** (Moskva), Maria **MOROZOVA** (S.-Peterburg), Aleksandër **RUSAKOV** (S.-Peterburg):

Korpusi i gjuhës shqipe: drejtimet kryesore të punës

6. Nebi **CAKA**, Ali **CAKA**, Prishtinë:

Korpusi i gjuhës shqipe – rezultatet e para, problemet dhe detyrat

Ora 17:00

Mbyllja e Konferencës – Kolec **TOPALLI**

Ora 17:30

Vizitë në Muzeun e Kosovës

shtunë, 12 nëntor 2011 Ora 8:30

Nisja për Gjakovë, Prizren para Grand Hotel Prishtina

SHQIPJA DHE GJUHËT E BALLKANIT
ALBANIAN AND BALKAN LANGUAGES
Konferencë shkencore - Scientific Conference

AKADEMIA E SHKENCAVE DHE E ARTEVE E KOSOVËS
KOSOVA ACADEMY OF SCIENCES AND ARTS

AKADEMIA E SHKENCAVE E SHQIPËRISË
THE ACADEMY OF SCIENCES OF ALBANIA

Prishtinë, 10-11 nëntor/ November 2011

Këshilli organizues i Konferencës
Organizing Board of the Conference

Akademik Rexhep Ismajli
Akademik Besim Bokshi
Akademik Idriz Ajeti
Akademik Kolec Topalli
Prof. dr. Seit Mansaku

Besim KABASHI, Universität Erlangen-Nürnberg

KORPUSE GJUHËSORE PËR SHQIPEN

Abstrakt

Korpuset gjuhësore janë përmbledhje tekstesh, në kuptimin përmbledhje e gjuhës së shkruar, asaj të folur apo e të dy formave së bashku, të cilat si të tilla ose veç e veç, të organizuara në një mënyrë të caktuar përbëjnë një njësi të quajtur korpus, me qëllim për të lehtësuar, përshpejtuar studimet gjuhësore e letrare, ngritjen e cilësisë së tyre, si dhe për të dokumentuar gjuhën në ato lëmi, kohë, shtrirje gjeografike e sipas veçorive të tjera, të cilat ai ka për qëllim që t'i mbulojë. Bërja e një korpusi varet shumë nga qëllimi i përdorimit të tij. Një korpus nuk mjafton të jetë vetëm përmbledhje e thjeshtë tekstesh apo/dhe e gjuhës së folur. Tekstet apo pjesët e korpusit duhen të jenë të kategorizuara, të përshkruara dhe të analizuara në mënyrë që të mund të përdoren më mirë si dije gjuhësore – aq më tepër kur sot për përdorim të një korpusi duhen vegla dhe metoda nga teknologjia e përpunimit të gjuhëve natyrore, të cilat bazohen, ndër të tjera, në të tilla përshkrime. Në kumtesë do të diskutohet rëndësia e bërjes së një korpusi gjuhësor të shqipes sa më parë që është e mundur, pastaj tipi i korpusit që do të duhej të bëhej së pari, struktura e tij dhe gjëra të tjera të cilat lidhen me bërjen e korpuseve.

Kryefjalë të këtij shkrimi janë njohuritë thelbësore për korpuset, roli i korpuseve në studime gjuhësore, llojet më të përhapura të korpuseve, rëndësia e ndërtimit të një korpusi gjuhësor gjithëpërfshirës të shqipes dhe gjëra të tjera, të cilat ndërlidhen me këto çështje – parë si punë promovuese.

Roli i korpuseve në studime gjuhësore

Para rreth 50 vitesh, në fillim të viteve '60 të shekullit të kaluar, u krijua *Korpusi Brown*, në origjinal *Brown Corpus* (emërtimi nga Brown University), ndërsa para rreth 40 vitesh, pra në fillim të viteve '70 të shekullit të kaluar, gjermanët, sapo kishin parë rëndësinë e një korpusi, pa u vonuar filluan punën dhe pak vite më vonë, ndërtuan korpusin gjuhësor të quajtur *Limas* (emër i përfituar nga fjalët “*Linguistik und maschinelle Sprachverarbeitung*”).¹

Nga ajo kohë, nga ajo ide, nga ato prodhime të para, sot është zhvilluar një degë shkencore që në anglishte quhet *Copus Linguistics*,

¹ Khs. [FRANCIS/KUČERA 1964] dhe [GLAS 1975].

në shqipe le të quhet *gjuhësi e korpuseve*, apo *gjuhësi e bazuar në korpuse dhe metoda statistikore*. Për një pjesë studiuesish kjo mbetet një metodë pune, pa arritur mëvetësinë si shkencë. Sidoqoftë, ajo ka nxitur studimet gjuhësore me metodat dhe rezultatet e saja, ka përsheptuar dhe njëkohësisht ka lehtësuar punën e gjuhëtarëve, ka ndihmuar që ata të kalojnë nga introspekcioni dhe nga mbledhja e shembujve dhe e argumenteve gjuhësore në mënyrën tradicionale, d. m. th. përmes leximit të shkrimeve përkatëse, në një mënyrë, e cila mbështetet në kërkime të shpejta në tekste elektronike. Mund të përmendet këtu si shembull kërkimi i një fjale, fjalëforme, një shenje apo një vargu shenjash në kuptimin e shprehjeve të rregullta (angl. *Regular Expressions*), si dhe nxjerrja e ndonjë statistike të fjalëve apo fjalive nga një tekst.²

Metodat e kësaj fushe kanë ndihmuar që të rritet cilësia e studimeve, duke u bazuar në të dhëna konkrete, d. m. th. empirike, dhe në rastin optimal edhe në të dhëna të shumta. Si rrjedhojë, në shumë raste janë nxjerrë dije të reja, sidomos nga puna me të dhëna të sasive të mëdha. Një standard madhësie që është arritur para më shumë se një dekade vitesh, është 100 milionë fjalë teksti, hap i bërë së pari nga *British National Corpus* (shkurt *BNC*).³

Kujt i shërben korpusi? Një përgjigje e thjeshtë do të ishte : leksikografisë. Jo më pak ai do t'i shërbente edhe fushave të tjera, si drejtshkrimi (ortografisë), fonetikës, morfologjisë, sintaksës, semantikës dhe fushave tjera, prej nga mund të nxirreshin shembuj të shumtë, kuptohet edhe dije tjetër, p. sh. në formë rregullash apo përjashtimesh.

Korpuset gjuhësore janë përmbledhje tekstesh, në kuptimin përmbledhje e gjuhës së shkruar, asaj të folur apo e të dy formave së bashku, të cilat janë të organizuara në një mënyrë të caktuar, me qëllim për t'u përdorur për hulumtime shkencore e didaktike, si të tilla për të lehtësuar e përsheptuar studimet gjuhësore e letrare, për të ngritur cilësinë e tyre, si dhe për të dokumentuar gjuhën sipas veçorive të

² Khs. [KENNEDY 1998] dhe [LEMNITZER/ZINSMEISTER 2010].

³ Khs. [BNC-XML 2007]. Sasia e fjalëve që përmbajnë korpuset sot shkon shumëfish përtej 100 milionë fjalësh teksti. Në këtë kuptim, një shembull është korpusi i projektit DWDS (*Das digitale Wörterbuch der deutschen Gegenwartssprache des 20 Jh*, Berlin-Brandenburgische Akademie der Wissenschaften), shih. <http://www.dwds.de/> (21.10.2011).

ndryshme – sipas fushës, kohës, shtrirjes gjeografike e sipas veçorive tjera, të cilat ai ka për qëllim që t'i mbulojë.

Llojet e korpuseve dhe veçoritë kryesore të tyre

Varësisht se nga cila pikëpamje trajtohen veçoritë e teksteve, përkatësisht korpuseve, aq lloje korpusesh mund të identifikohen. Nuk ka një kufi të paracaktuar. Ato më të përhapura janë *korpuset e baraspeshuara* dhe *korpuset paralele*.

– *Korpuset e baraspeshuara* apo ndryshe të quajtura edhe *korpuse reference* (angl. *Reference Corpus*). Këto korpuse mbulojnë në formë gjithëpërfshirëse gjuhën dhe përdoren në rend të parë për studime të ndryshme gjuhësore.

– *Korpuset paralele*, p. sh. një korpus paralel shqip–anglisht.⁴ Këto korpuse përdoren në rend të parë për studime gjuhësore krahasuese, për ndihmë gjatë përkthimeve apo edhe për leksikografi dy- apo shumë-gjuhëshe.

Nisur nga veçori të tjera, siç është koha e shkrimit të teksteve, korpuset mund të ndahen në *korpuse statike* dhe *korpuse dinamike*.

– *Korpuset statike (historike)*. Përmbajtja e teksteve të ky tip i korpuseve caktohet gjatë fazës së ndërtimit të tij dhe pastaj mbetet e pandryshueshme.

– *Korpuset dinamike*. Përmbajtja e korpuseve të këtij tipi ndryshohet pandërprerë apo kohë pas kohe, por duke mbajtur strukturën dhe disa veçori tjera themelore në bazë të një formule të caktuar.

Veçori tjera kryesore, duke theksuar të parën, janë edhe këto në vijim :

– *Gjuha*. E folur apo e shkruar, ajo ka shumë ngjyime, të cilat janë të ndërthurura (ortogonale) me njëra-tjetrën dhe në shumë raste nuk mund të ndahen veç e veç. Ato mund të jenë *regjionale, sociale, raporti në mes folësit përkatësisht shkruarit të tekstit dhe atij me të cilin flitet përkatësisht me atë, të cilit i është kushtuar teksti*, e shumë të tjera, të cilat mund të nënkuptohen. Pjesët e korpusit (angl. *Pieces*),

⁴ Në këtë kontekst, ndonëse qëllimi i shkrimit është tjetër, do të ishte me vlerë të theksohej këtu se një korpus paralel i llojit arvanitisht–shqip përkatësisht arbërisht–shqip do të ishte shumë i rëndësishëm për studime gjuhësore të shqipes.

përkatësisht *tekstet* mund të jenë *tekste “normale”*, d. m. th. *gjuhë e shkruar*, apo *tekste të përfituara nga gjuha e folur*.⁵

— *Madhësia*. Madhësia e korpusit është shumë e rëndësishme. Sa më i madh të jetë korpusi, aq më tepër ka gjasa, që kërkimet e bëra në korpus të përcillen me rezultate të kënaqshme, së pari për nga ana sasiore, dhe më pas edhe nga ana cilësore. Me rritjen e madhësisë së korpusit, rritet edhe gjasa që fjalët, strukturat, shprehjet apo diçka tjetër që kërkohet të gjenden në korpus. Qëllim i korpusit është dhënia e sa më shumë shembujve, të paktën disave ose tek e fundit edhe të vetëm një shembulli, për një kërkim të caktuar.

— *Fusha*. Varësisht se çfarë tekstesh e përbëjnë korpusin, ai mund të jetë *i përgjithshëm* apo *i veçantë* (special). Një *korpus i përgjithshëm* do të ishte ai, i cili ka tekste nga sa më shumë fusha të jetës materiale e shpirtërore, të dijeve shkencore dhe të artit e kulturës. Një *korpus i veçantë* do të përmbante tekste të një fushe të caktuar, b. fj. nga fusha e drejtësisë, e ekonomisë ose e mjekësisë. Në të dyja rastet me rëndësi thelbësore është hapësira që zë një tekst në krahasim me tekstet tjera.

— *Mediumi*. Sot nuk mund të diskutohet kjo çështje edhe pse korpuset kanë një histori, e cila fillon para “kohës së kompjuterëve” : korpuset e sotshme duhet të jenë në formë elektronike. Këtu bëhet fjalë për një zgjidhje të shkallës së dytë : *korpuset të rëndomta* apo *multimediale*. Këto të fundit bashkërendisin pamjen (video), zërin (audio), rrethanat ku flitet gjuha me përkohimin e tyre në formë teksti. Kuptohet, bëhet fjalë për korpus të gjuhës së folur.

Këto janë disa nga veçoritë më kryesore, jo të gjitha, dhe se përzgjedhja mund ishte edhe ndryshe. Në bazë të veçorive të këtyra mund të fillohej me punë dhe mund të ndërtohej një korpus, i cili do të ishte në gjendje të ashtuquajtur *lëndë e parë e klasifikuar*. Megjithatë një korpus i tillë do të ishte i përdorshëm dhe do t’u hynte padyshim në punë shumë gjuhëtarëve – por me shumë kufizime.

Analiza e lëndës së parë

Tekstet apo pjesët e korpusit duhen të jenë të kategorizuara, të

⁵ 10 % e teksteve, të cilat i përmban BNC, janë të përfituara nga gjuha e folur, shih. <http://www.natcorp.ox.ac.uk/corpus/index.xml> (21.10.2011).

përshkruara dhe të analizuar në mënyrë që të mund të përdoren më mirë si dije gjuhësore – aq më tepër kur sot për përdorim të një korpusi duhen vegla dhe metoda nga teknologjia e përpunimit të gjuhëve natyrore, të cilat bazohen, ndër të tjera, në të tilla përshkrime. Për këtë arsye është i nevojshëm një hap tjetër, një punë e mëtejme me “lëndën e parë”. Ajo është shumë e rëndësishme dhe po ashtu e kushtueshme. Është fjala për *analizën dhe përshkrimin e informacionit gjuhësor dhe jogjuhësor* të teksteve të korpusit nga ana e ekspertëve gjuhësorë të fushave të ndryshme, duke filluar nga drejtshkrimi, për të kaluar me përshkrimin statistikor të shumanshëm, përshkrimin fonetik, leksikografik, morfologjik⁶, sintaksor, semantik e kështu me radhë me përshkrime tjera. Këtu mund të përmendet edhe një ndarje, ajo në *korpuset të papërshkruara* (të paanalizuara, të paannotuara [lat. *annotare* : shënim, vërejtje, sqarim]) dhe në *korpuset të përshkruara*.⁷ Varësisht nga mundësia e përdorimit, një korpus mund të jetë *jopublik*, p. sh. të jetë për qëllime të caktuar vetëm në rrethet e një institucioni, grupi njerëzish, *publik i kufizuar* (për njerëz të regjistruar për përdorim të korpusit apo ata të cilët do të blinin licencën/atë), apo dhe *publik*, d. m. th. për të gjithë.

Para fillimit të punës duhet të jetë një kusht që duhet pasur pa ndërprerë në mendje : *përfitimi i të drejtës së autorit/autorëve për përdorim/përpunim pa kufizime të teksteve që do të ishin pjesë e korpusit*. Përrjashtim bën këtu rasti, nëse tekstet “nuk kanë të drejta”, b. fj. tekstet e vjetra, siç mund të jetë “Meshari” i Gjon Buzukut.

Korpuset gjuhësore për shqipen

Një rol të një korpusi të shqipes për një kohë e ka pasur teksti i romanit “Koncert në fund të dimrit” nga I. Kadare, botuar nga ECI/MCI më 1994 si pjesë e një korpusi përmbledhës, të thjeshtë, të gjuhëve evropiane.⁸ Me gjithë hapin e parë, dhe vlerën e padiskutueshme, aq më tepër për atë kohë, si tekst përfaqësues i shqipes, ai tekst, nuk mund të quhet korpus gjuhësor në kuptimin e një korpusi gjithë-

⁶ Kjo bëhet në pjesë të madhe ose tërësisht në mënyrë automatike me ndihmen e programeve (veglave) përkatëse, khs. p. sh. [KABASHI 2004].

⁷ Khs. [BNC-XML 2007].

⁸ Khs. këtu [KONCERT 1994] dhe [ECI/MCI 1994].

përfshirës. Që nga ajo kohë, në këtë fushë nuk ka ndonjë përparim, i cili mund të përmendej këtu, kur është fjala për gjuhën shqipe.⁹

Duke u nisur nga fakti se një korpus reference do t'u shërbente të gjithëve, ai përbën një pikënisje të mirë për të qenë korpusi i parë i shqipes së sotshme. Detyra e parë do të ishte *përzgjedhja e teksteve*, punë e cila do të bëhej duke u bazuar në baraspeshën e teksteve në bazë të shumë veçorive. Ato mund të jenë shumë të ndryshme nga njëra-tjetra si *gjuhë e folur dhe e shkruar, cilësi të ndryshme sociolinguistike dhe psikolinguistike të gjuhës, vendi ku ajo flitet përkatësisht shtrirja gjeografike e saj, cilësi organizative formale të saj dhe së fundi statusi i saj*.

Puna teknike-shkencore me korpus

Pasi që korpuset e sotshme që nga fillimi i ndërtimit të tyre janë të parapara për t'u përdorur me ndihmen e veglave elektronike, puna teknike-shkencore me korpus përbën në shumë raste një pjesë të pandashme apo përcjellëse të punës tjetër për korpusin. Ajo përfshin, ndër të tjera, çështjet në vijim : *kodifikimin teknik të teksteve, gjuhën përshkruese, gjuhët pyetëse, kompatibilitetin me korpuset e tjera dhe përshtatshmërinë e përdorimit të veglave punuese me standarde të përhapura dhe aktuale*. Disa nga këto veçori teknike-shkencore duhet pasur në mendje qysh nga fillimi i punës.¹⁰

Përmbledhje

Sot për sot, një korpus gjuhësor do të duhej të kishte të paktën 100 milionë fjalë teksti. Gjuhësisë së sotshme shqipe pa mëdyshje i duhet një korpus gjithëpërfshirës, i cili do të duhej të ishte i baraspeshuar.

Për shkak të natyrës së mbledhjes së teksteve, në rrethanat e tashme në hapësirën shqipfolëse, një korpus të mirëfilltë mund ta ndërtonin vetëm institucionet shkencore përkatëse të specializuara – të cilat sot për sot, në hapësirën shqipfolëse, janë ende vetëm një dëshirë.

⁹ Khs. këtu për çështje të resurseve të gjuhës shqipe shkrimin [KABASHI 2009].

¹⁰ Khs. këtu [WYNNE 2005].

Ndërtimi i një korpusi të gjithëpërfshirës është një ndërmarrje jo e vogël dhe se puna për ndërtimin e tij nuk do të duhej të nënvlerësohej. Në këtë rast, ai që fillon punën “lehtë”, e përfundon me vështirësi dhe çmim të lartë, për më tepër, në dëm të cilësisë së korpusit, dhe sidomos, në dëm të atyre që do ta përdornin atë. Një korpus i ndërtuar jo mirë, duke mos mbuluar të dhënat e duhura gjuhësore apo p. sh. duke i tjetërsuar raportet e tyre statistikore, do të ngadalësonte zhvillimin e shkencave gjuhësore, krahasuar me një korpus të ndërtuar mirë.

Referencat / Literatura

- [BNC-XML 2007] British National Corpus. Version 3. BNC XML Edition. Distributed under license by Oxford University Computing Services on behalf of the BNC Consortium, 2007.
- [ECI/MCI 1994] European Network of Excellence in Human Language Technologies (ELSNET): *European Corpus Initiative Multilingual Corpus I*. (CD-ROM). Utrecht : ELSNET, 1994.
- [FRANCIS/KUČERA 1964] Nelson W. Francis / Henry Kučera : Manual of Information to Accompany 'A Standard Corpus of Present-Day Edited American English'. Brown university : Dept. of Linguistics. Providence 1964.
- [GLAS 1975] Glas, Reinhold : „Das LIMAS-Korpus, ein Textkorpus für die deutsche Gegenwartssprache”. [In :] *Linguistische Berichte 40* (1975), 63–66.
- [KABASHI 2004] Kabashi, Besim : “Analiza automatike e fjalëformave të gjuhës shqipe”. 129–135. [Në :] (Materialet nga) *Seminari Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare, XXIII*. Libri 23/1. Prishtinë : Universiteti i Prishtinës, 2004.
- [KABASHI 2009] Besim Kabashi : “Resurset e gjuhës shqipe – një diskutim rreth gjendjes së tyre të tashme”. 97–102. [Në :] (Aktet nga) *Seminari III i Albanologjisë*. Tetovë, Universiteti Shtetëror i Tetovës, 2009.
- [KENNEDY 1998] Graeme Kennedy : *An Introduction to Corpus Linguistics*. Longman: London / New York, 1998.
- [KONCERT 1994] Romani *Koncert në fund të dimrit* nga I. Kadare, Tiranë, “Naim Frashëri”, 1988. *alb01*. [Në :] [ECI/MCI 1994].
- [LEMNITZER/ZINSMEISTER 2010] Lothar Lemnitzer/ Heike Zinsmeister : *Korpuslinguistik. Eine Einführung*. 2. Auflage. Tübingen: Narr Verlag, 2010.
- [WYNNE 2005] Wynne, Martin (Ed.) : *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford : Oxbow Books, 2005.