

НА ФАНЕТЫЧНЫХ СКРЫЖАВАННЯХ

У Аб'яднаным інстытуце праблем інфарматыкі НАН Беларусі (АІПІ) адкрытыу лекцыю «Корпусная лінгвістыка: кампіляцыя корпуса вуснага маўлення» прачытала маладая даследчыца камп'ютарнай лінгвістыкі Наталля Карлава-Бурбонус з універсітэта імя Юстуса Лібіха (г. Гісэн, Германія). Яна таксама правяла вучэбныя майстар-класы па практичным транскрыбіраванні вусных тэкстаў з выкарыстаннем спецыяльнай праграмы EXMARaLDA.

Пропанаваная тэма даклада сабрала каля паўсотні зацікаўленых слухачоў з акадэмічных інстытутаў, навуковадаследчай лабараторыі фонаграфічных даследаванняў Міністэрства юстыцыі, сталічных універсітэтаў і іншых арганізацый, якія займаюцца камп'ютарнай лінгвістыкай, распрацоўкай тэарэтычных і практичных пытанняў фанетыкі, аналізам гукавога ладу мовы, фарміраваннем разнастайных баз звестак з выкарыстаннем лінгвістичных рэурсаў, стварэннем прымесловых сінтэзатаруў маўлення. Аўдыторию аднадумцаў аб'яднала разуменне, што для далейшага паспяховага развіцця як лінгвістичных, так і тэхнічных науک, якія займаюцца, з аднаго боку, распрацоўкай пытанняў узорнага літаратурнага маўлення, а з другога, прымяненнем праграм для агучаных чалавека-машынных узаемадносін, неабходна правядзенне сумесных міждысцыплінарных даследаванняў, звязаных з вуснай формай літаратурных моў. Тому з асаблівай прыхільнасцю былі сустрэты прадстаўленыя Н. Карлавай-Бурбонус інавацыйныя распрацоўкі замежных вучоных у галіне камп'ютарнай лінгвістыкі: сучасныя магчымасці аналізу фанетыка-арфаэпічных з'яў, дэмантрацыя дасягненняў славісту Балгарыі і Чэхіі ў фарміраванні нацыянальных карпусоў вуснага маўлення.

Дакладчыца азнаёміла слухачоў з праграмнымі прадуктамі, распрацаванымі спецыялістамі Гамбургскага ўніверсітэта для працы з відэа- і аўдыяфайламі, – EXMARaLDA (Extensible Markup Language for Discourse Annotation), які ўжо атрымаў шырокую вядомасць у навуковых колах на ўсіх кантынентах планеты. Папулярнасць праграмы забяспечана, па-першае, сваім заснаваннем на базавым стандарце UNICODE (шматмоўнасць), а па-другое, магчымасцю бясплатнага і законнага спампоўвання яе

любым зацікаўленым карыстальнікам, па-трэцяе, даступнай, выдатнай дакументацыяй, якая суправаджаеца навучальнымі ўрокамі на англійскай і нямецкай мовах.

Як аказалася, сапраўды, гэтыя тры палажэнні з'яўляюцца неабходнымі і дастатковымі для паспяховай кампіляцыі любога па сваіх задачах і аўёме корпуса вуснага маўлення, а таксама для далейшай працы з ім. Удзельнікі презентацыі праграмы EXMARaLDA пераканаліся ў гэтym самі, паспрабаваўшы ўласныя сілы ў набытых за два дні правядзення мерапрыемства навыках.

Асаблівую вартасць праведзенае мерапрыемства набывае ў святле рэальнага стану распрацоўкі фанетыка-фоналагічных пытанняў мовы і культуры беларускага маўлення. Адзначу, што ў наш час планамерным даследаваннем фанетыкі беларускай літаратурнай мовы займаюцца толькі ў дзвюх-трох установах краіны. Па-першае, гэта фанетысты аддзела сучаснай беларускай мовы Цэнтра даследаванняў беларускай культуры, мовы і літаратуры НАН Беларусі, якія, працягваючы традыцыі заснавальнікаў беларускай фанетычнай школы, вывучаюць актыўныя моўныя працэсы, рэальный стан і дынаміку літаратурнай арфаэпічнай нормы, рыхтуюць адпаведныя даведнікі і дапаможнікі. Па-другое, нельга не адзначыць супрацоўнікаў лінгвістичных кафедр Мінскага дзяржаўнага лінгвістичнага ўніверсітэта, дзе таксама даследуюць беларускую фанетику, але ўжо ў асноўным з мэтай асвятлення пытанняў беларускай інтэрферэнцыі. Апошнім часам выказваюць зацікаўленасць у даследаваннях фанетычнага ладу беларускай мовы спецыялісты лабараторыі распознавання і сінтэзу маўлення АІПІ, якія вырашаюць інавацыйныя задачы, звязаныя з сістэмамі сінтэзаванага маўлення, у тым ліку і на беларускай мове. Перакрыжаванне інтарэсаў аkadэмічных вучоных філалагічнага і тэхнічнага профілю ў сферы камп'ютарнай лінгвістыкі ўяўляеца ўзаемадапаўняльным і перспектывным. Сумесная праца двух калектываў заклікана распрацаўваць праграмы прадукт для шматаспектнай апрацоўкі гукавых файлаў на аснове сучасных дасягненняў камп'ютарнай лінгвістыкі. Тым самым гэта будзе садзейніцаць захаванню і ўмацаванню норм беларускай літаратурнай мовы на ўсіх узроўнях яе сістэмы і дазволіць выкарыстоўваць

літаратурнае маўленне ў прыкладных распрацоўках. На аснове якаснага сінтэзатара маўлення могуць быць пабудаваны разнастайныя мультымедыйныя прадукты: гаворачы электронны інфарматар (агучванне транспартных прыпынкаў, электроннай чаргі, парушэння правілаў дарожнага руху), аўдыякнігі (паслядоўнага чытання, адукатыўнага дыялогу) і г.д.

Спецыяльнае падраздзяленне ЮНЕСКА, якое займаецца захоўваннем моў свету як жывых прадстаўнікоў культурнай спадчыны планеты, аднесла беларускую мову «к потенціяльно угражаемым языкам». Калі сучасная тэндэнцыя кардынальна не зменіцца, у найбліжэйшыя дзесяцігоддзі мы можам згубіць яшчэ адну жывую мову Зямлі – славянскую беларускую мову, якая, да таго ж – дзяржаўная, якія ніякай іншай ва ўказанай катэгорыі ёўрапейскіх моў. У святле нават гэтых палажэнняў значнасць аб'яднання дасягненняў вучоных двух аkadэмічных інстытутаў для стварэння аўтаматызаваных праграм канвертациі арфаграфічных записаў беларускага слова ў жывую форму ўзорнага маўлення цяжка пераацаніць.

Сумеснымі сіламі даследчыкаў плануецца стварыць электронны фанетычны корпус беларускай літаратурнай мовы, апрацаўваўшы з дапамогай спецыяльных камп'ютарных праграм жанрава-тэматычныя запісы літа-ратурнай мовы. Гэта дазволіць акумуляваць і захаваць для будучых пакаленняў не толькі пісьмовую форму мовы, але і вусную, фанетыку і фаналогію. У той жа час сучасныя карыстальнікам камп'ютараў і Інтэрнэту гэты праграмы прадукт зможа дапамагчы ў авалоданні і ўдасканаленні ведаў і навыкаў па беларускай мове, у тым ліку пры дыстанцыйным навучанні, а таксама забяспечыць магчымасць выкарыстоўваць яго ў разнастайных прыкладных сістэмах.

Навуковая сустрэча, якая прайшла ў АІПІ пад кіраўніцтвам загадчыка лабараторыі распознавання і сінтэзу маўлення Юрыя Гецэвіча, дазволіла не толькі азнаёміцца з новай для беларускай аўдыторыі праграмай аналізу вусных тэкстаў, але і дапамагла ўпэўніцца, што трэба больш паглыблена вывучыць пытанні варыятыўнасці маўлення, адаптацыйныя магчымасці маўленчай сістэмы ў сувязі з новай хвалій запазычанняў, а таксама зрабіць гэтыя веды даступнымі як мага больш шырокаму колу карыстальнікаў праз Інтэрнэт.

Валянціна РУСАК,
загадчык аддзела сучаснай беларускай мовы
Цэнтра даследаванняў беларускай культуры,
мовы і літаратуры НАН Беларусі