

Началото на рогната компютърна лингвистика

Щастливият брак на съвременните комуникационни технологии с езика



Елена Паскалева – доайен на българската компютърна лингвистика, обича да казва, че е влязла в науката, както Рада Госпожина в национално-освободителната борба – по пътя на любовта. В последния гимназиален клас неин сърдечен приятел радиотехник възприема иронично литературните ѝ интереси, като я поучава, че и машините могат да превеждат. По това време в социалера пада идеологическото табу от кибернетиката, излизат първите руски популярни преводни издания по темата. Елена е привлечена от научнофантастичното звучене на тези трудове. Щастлива среща в научния ѝ живот е тази с Александър Людсканов, пионера на машинния превод у нас, който още в първата си лекция пред студентите филолози споменава машините-преводчици.

Тези електронноизчислителни машини (още не компютри) у нас са от съветската серия „Минск“

– 2, 22, 32 и т. н., прототипи на сериите IBM. В Института по математика с изчислителен център на БАН започва серия от експерименти – демонстрации на възможностите на тези машини в хуманитарни области. Още като студентка в трети курс руска филология Елена е включена в новосъздадената група по машинен превод, ръководена от Людсканов, където машината Минск-2 превежда 1 страница руски текст на български. Експериментите продължават със съставяне на речници за двата езика.

Причината тези изследвания да започнат в точните науки и изчислителната техника, а не в лингвистиката, е главно идеологическа. Докато математиците отрано са се запознавали с формалните езици и граматиките на Ноам Чомски, при езиковедите същият е дамгосван като буржоазен идеолог.

Изследванията в тази област се прекъсват с преждевременната смърт на Людсканов и закриването на групата. Завалят запитвания от чужбина за съдбата на тези изследвания, популяризирани от книгата му „Превеждат човекът и машината“, претърпяла 2 български, 2 немски, 1 полско и 1 френско издание. Този международен интерес обуславя създаването през 1978 г. на ново самостоятелно звено в Института по математика – секцията по математическа лингвистика.

„Компютърната лингвистика е истинска интердисциплинарна наука – пояснява Елена Паскалева. – В нея ролите на поръчители и обслужващи се разменят между лингвисти и програмисти – или лингвистите доставят за изчислителните модели надеждни езикови данни, или програмистите създават компютърни средства за обработка на лингвистичното знание. Истинското интердисциплинарно звено

по компютърна лингвистика с равноправно участие на информатици, логици и лингвисти бе създадено през 1987 г.“

През 1987 г. в системата на БАН има Координационен център по информатика и изчислителна техника, покровителстван на високо държавно и партийно ниво. Тройка ентусиасти (логик, лингвист-теоретик и компютърен лингвист), водени от талантия български логик Георги Гаргов, създават Лабораторията по лингвистично моделиране.

„След ранната смърт на Гаргов аз ръководих секцията 12 години във всичките административни метаморфози на въпросния Координационен център. На-

пуснах я по време на голямата чистка на феодалните старци, без да съм от най-феодалните. Но съм горда, че днес тя определя насоките на авангардните езикови технологии не само в Института за информационни и комуникационни технологии на БАН, а и в национален и международен план“, разказва Паскалева.

След раздялата с БАН Елена Паскалева се посвещава на автоматичната обработка на медицински текстове – задача, привлекателна и като специфично езиково знание, и като компютърно приложение в една бъдеща електронизация на здравните услуги.

Компютърната лингвистика не е просто машинно моделиране на човешка дейност. Тя е рожба на щастлив брак между две комуникационни технологии: най-модерната – изчислителната, и най-древната – езиковата („в началото бе Словото“), твърди Елена Паскалева. Рожбите на този брак са компютърните модели на двата основни аспекта на езика – комуникацията и знанието, в двете задачи – междуезиковата комуникация и извличането на знание от текст.

Първата гама

„Първата гама“ на компютърната лингвистика е машинният превод, подчертава Елена Паскалева. Той е най-ранната и ярка звезда на нейния небосклон. От възторзите след първия експеримент през 1949 г., през разочарованията и надеждите, идеята е претърпяла много промени. Първоначалната утопична идея за машинния превод е, че той е осъществим на универсално смислово равнище, ако за всеки език имаме компютърен модел на прехода от текста до смисъла и обратно (анализ и синтез). Тази утопия се сблъсква с реалната трудност, дори невъзможност на този пълен преход (но пък опитите за компютърен езиков анализ и синтез дават начало на всички модерни техники за компютризация на езика). Този подход, основан на знанието, днес е допълнен от подхода, основан на данните. Той се използва в т.нар. статистически машинен превод, който борави с получени от статистическата обработка на големи обеми текстове двузични съответствия от различни езикови равнища.

Друга езикова дейност, привлекателна за компютърно моделиране, е създаването на електронни речници, и то не само за компютърни езикови справки.

Елена Паскалева споделя една шеговита случка по този повод. На международна конференция в Любляна тя докладва софтуер, който позволява да се извлича ин-

формация от подробно структурирана речникова статия на тълковен речник. В дискусиата по доклада се включват две лингвистични светила – Джон Синклер и Чарлз Филмор. Обсъжда се и практическият аспект на използването на тази информация в различни езикови разрези и обеми, без да се нарушава авторското право. След 2 години колега на Паскалева среща Филмор на друг форум и при споменаването на София последният възкликва: „Познавам оттам Елена Паскалева, тя докладва софтуер за крадене на речници“. След още 3 години при лична среща с Филмор Елена споделя: „Поласкана съм, че ме помните, но ми е мъчно, че ме помните по този начин“. Той се засмива и отговаря: „Ех, Елена, ако знаете мен колко хора по какви начини ме помнят...“.

„Образно казано, аз съм еничарче, рано потурчено – от III курс филология попаднах сред цвезата на българската математика и останах там докрай – шегува се Елена Паскалева. – Никога не съм противопоставяла физиците на лириците, хуманитарното на математическото, за мен те са едно цяло. Служебната ми кариера е в номенклатурата на



Елена Паскалева и бъдещият професор по математика и главен редактор на сп. „Математика и информатика“ на Национално издателство „Аз Буки“ Сава Гроздев

математическите науки, но хуманитарното ми образование и интереси са ме свързали с цвезата на преводаческата гилдия не само в Съюза на преводачите. Както всяко еничарче, в края на пътя си се връщам към корените. Последните години съм посветила на просветителска дейност – да запознавам лингвистите с лингвистичното знание, обслужващо предния фронт на езиковите технологии“.