

Рецензия

върху дисертацията на Георги Емилов Илиев
«Езиково мотивирана оптимизация на машинния превод»
за получаване на научната и образователна степен «доктор»

Научен ръководител: проф. д-р Светла Коева,
Институт за български език – БАН

Рецензент: доц. д-р Иван Александров Держански,
Институт по математика и информатика – БАН

Дисертационният труд разглежда проблеми на машинния превод и по-точно на фразовия статистически машинен превод на български език. Както общата тема, така и конкретната методика понастоящем се радват на изключително внимание в глобален мащаб, но по разбираеми причини за българския език като целеви език на превода не може да се каже същото. Толкова по-интересни и важни са разглежданите въпроси както от теоретично, така и от практическо гледище.

Дисертацията съдържа 208 страници. Състои се от увод, три глави, заключение, приложение и библиография.

В **увода** кратко, но убедително се аргументира актуалността на изследването, определят се предметът и целта му, изброяват се основните теоретични и приложни приноси на отделните части на труда и се излага структурата му. Обсъжда се значението на избора на шведски като изходен език на превода и на български като целеви: и двата езика не са с висока плътност и не са били предмет на много разработки в областта на машинния превод, но пък са относително близки в типологично отношение, имат (особено шведският) близко

родствени езици, върху които биха могли да се пренесат непосредствено голяма част от резултатите, а и за тях все пак съществуват и са достъпни електронни ресурси, в това число паралелни двуезични корпуси.

В Глава 1, «Машинен превод — кратък исторически обзор», се проследява развитието на темата, преминало през поредица вълни на ентузиазъм и скептицизъм. Обособяват се четири етапа — до 1949 г., до 1966 г., до около 1990 г. и съвременният, — като за всеки се проследяват политическите и технологичните обстоятелства и доминиращите мотиви и методи. Отбелязват се особеностите на проблемите и резултатите в САЩ, СССР, Западна Европа и България, като се разглеждат голям брой системи за машинен превод (с позоваване на оригиналните публикации, където са описани те за пръв път). Разбира се, читателят се изкушава да попита защо не е спомената още и тази или онази система или проблемна група, но целта на кандидата не е да изброи всички научни центрове и техни проекти, а да покаже многообразието на методите. Историческият преглед завършва с твърде полезната Таблица 1.1, която съпоставя пет съвременни подхода към машинния превод с по една типична система — представител на всеки и с характерни особености, предимства и недостатъци.

Глава 2, «Построяване на система за статистически машинен превод с български като целеви език», провежда читателя през целия процес на изграждането на такава система, започвайки от създаването на паралелен двуезичен корпус, преминавайки през обосновка на математическия апарат, прилаган в тази област, и завършвайки с оценката на получената базова система. Изложението е илюстрирано с многобройни примери от български и шведски, както и с резултати от експерименти с машинен превод от шведски на английски и от пет германски езика (английски, шведски, датски, нидерландски и немски) на български, извършен с програмния пакет Moses. Резултатите показват, че морфологичната близост на двата езика и сложност на всеки поотделно е от голямо значение при машинния превод и че методите за

построяване на базова система за статистически машинен превод не работят еднакво добре за всички езици.

Глава 3, «Езиково зависими компоненти и методи», изследва някои похвати, които могат да бъдат приложени при изграждането, обучението или работата на системата за статистически машинен превод, за да се подобри качеството на превода на български език. Тук отново се започва от най-ранния етап, подравняването на паралелния корпус (този път с употреба на езиковозависими средства), и се стига до нормализирането на получения преведен текст, като се прави обосновка на всички предлагани методи. На много места се дават примери на шведски конструкции, чийто превод на български изисква особено внимание или нестандартен подход, както и на симптоматични грешки на машинния превод, и се сравняват резултатите от експерименти с привличане на различни езиковозависими методи и различни видове нормализация.

Заклучението накратко припомня изводите, резултатите и приносите на всяка глава и поставя изследването в глобалния контекст от разработки в областта на машинния превод. В **приложението** е посочено какво има във всяка директория на файловото пространство, съдържащо експерименталните корпуси и резултати, които съпровождат дисертацията (държа да отбележа, че предоставянето на тези материали е едно от най-големите достойнства на този труд). **Библиографията** включва 150 заглавия на източници на английски, руски, български, френски, шведски и датски език. В списъка публикацииите на кандидата са три (една самостоятелна и две в съавторство) — в сборници от доклади на международни конференции.

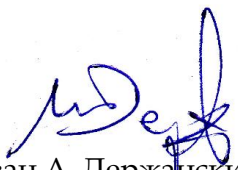
Темата на дисертацията е избрана отлично. Работата е извършена извънредно прецизно и със стремеж към изчерпателност. Особено приятно впечатление прави, че авторът си служи навсякъде с българска терминология, макар и да посочва, където е целесъобразно, английските съответствия на термините. Техническото оформяне също е на много

високо равнище. Малкото технически грешки (точката преди кавичката в «„китайския код.“» на стр. 15, небългарският словоред в «преводи на *пиджин* език» на стр. 24, главните букви в «езици като Есперанто, Интерлингва и др.» на стр. 44) могат да привлекат вниманието само на необикновено педантичен читател, макар че бездруго трябва да бъдат поправени при евентуалното публикуване на труда като монография.

Като имам пред вид всичко казано, препоръчвам без всякакво колебание да се присъди на кандидата Георги Емилов Илиев научната и образователна степен «доктор».

София, 4 ноември 2014 г.

Рецензент:


(доц. д-р Иван А. Держански)