

Н. В. Котова, М. Янакиев

ГЛОТОМЕТРИЯТА ЕКСПЛИЦИРА ОСНОВИТЕ НА СЪПОСТАВИТЕЛНАТА ЛИНГВИСТИКА*

В тази книжка на списанието се публикуват статии с резултати от такива изследвания, които от десетина години насам се извършват под название *глотометрически* (срв. Котова, Янакиев 1968, с. 58–77; Котова, Янакиев 1973, с. 238–260; Найденов и др. 1976, с. 65–70; Янакиев 1977). В тези изследвания се съпоставят количества (честоти, фреквентности) на появи на морфеми в текстове, принадлежащи към различни стилове на българския език, от една страна, и към подобни стилове на други езици.

Основната отличителна черта на тези изследвания, черта, която дава основание те да се характеризират със специално название, за да се обособят в многообразието от все по-широко разпространяващи се лингвистични изследвания, опиращи се на математико-статистически методи, е в това, че се съпоставят фреквентности на появи на типове факти в „равноразмерни“ проби от определени разновидности на езиковата практика.

Чрез понятието „равноразмерност на пробите“ в глотометрията се осигурява тази теоретическа предпоставка, която е необходима, за да се внесе в материята на езика метрика, т. е. измеримост. Измеримост може да се внесе в езиковата материя, като приемем под размер (обем, дължина, дименсия) на съобщение на даден език да разбираме числото на появите на факти от някакъв тип, често появяващи се в съобщенията на този език или в съобщенията на езици от някаква определена група.

В зависимост от избора на типа факти, появата на един от които приемаме за край на съобщение с размер „единица“, числото, характеризиращо размера на съобщението, получава определено название, става „именувано число“, а за размера казват, че е размер „в определени (назовани с определено название) мерни единици“. В глотометрическите изследвания като публикуваните в тази книжка на списанието се използват две вече доказали практическите си удобства мерни единици: *монолексата* (съкратено *лкс* или *lx*) и *монофоната* (съкратено *фн* или *phn*).

В съответствие с тези две мерни единици се съпоставят като статистически варианти числата, които означават честотите (фрек-

* В сп. *Съпоставително езикознание*, III, 1978, № 3, с. 3–15.

вентностите) на появите на определени морфемни или класове морфемни в два типа равноразмерни проби: *хектолекси* (съкратено *хлкс* или *hlx*) и *килофони* (съкратено *кфн* или *кphn*). Хектолексата е откъс от съобщение, съдържащ сто текстови и непременно записани с букви думи. Килофоната е откъс от съобщение (записано по правилата на определена, разяснена по-долу тук фонетична транскрипция), съдържащ 1000 букви.

Лексиметризация

Измерването „в монолекси“ (лексиметризацията) на един текст се осъществява сравнително лесно: номерират се (реално или мислено) в текста всички думи, пълнозначни и служебни, и номерът на последната е численият израз на (моно)лексовия размер на текста. Нужно е да се дадат съвсем малко допълнителни разяснения, свързани с посочването, че се броят непременно „записани с букви думи“. Записаните с цифри числителни имена при лексиметризацията трябва да си представим преписани „словом“ и да ги броим за толкова монолекси, от колкото „словом“ записани думи се състоят. Например българското числително „21“ се смята за „дълго“ три монолекси (3 *lx*), защото „словом“ се записва като „три думи“: *двайсет*₁ *и*₂ *един*₃. Съответното му руско числително е „дълго“ две монолекси, защото словом се записва *двадцать*₁ *один*₂. А словашкото му съответствие се смята за една монолекса – записва се словом и в двата си варианта като една дума: *dvadsat'jeden*, *jedenadvadsat'*. Като една дума се записва словом и немското *einundzwanzig*, та поради това се брои също като една монолекса.

Във връзка с лексиметризацията на цифровите комплекси в текстовете на различни езици разясненото по-горе положение понякога дава повод за тревоги. Така, ако в чешки приемем да запишем „словом“ „21“ в съответствие с разговорната практика, т. е. *jedenadvacet*, ще трябва да го броим за една монолекса (въпреки че в числителното добре личат „три думи“: *jeden*₁, *a*₂, *dvacet*₃). А ако го запишем с „пунктирното“ му четене, т. е. *dvacet jeden*, ще имаме работа с две „правописно обосновани“ монолекси.

Но, както показват съответни проверки, при съпоставяне на проби от различни стилове на един и същ език тези „правописни шумове“ не пречат да се улавя полезната информация за варирането на честотите на интересуващите ни факти в пробите. По-неприятни за съпоставителната глотометрия са различията в записването на някои типове изрази, съдържащи глаголни форми. Тези типове изрази са

много по-чести в езиковата практика и правописните им особености в различните езици могат да повлияят съществено върху съпоставянето на глотометрически данни от различни езици.

Така в чешки и словашки проклитичната отрицателна частица *ne* се пише слято с личните глаголни форми, а в другите славянски езици тя се записва по правило като отделна дума и тъй като имаме работа с една от най-честите морфемни форми в езиковата практика, чешката и словашката дума се оказват, средно взето, уловимо „по-дълги“, ако дължината им измерваме в букви, от думите в съответните стилове на другите славянски езици. Компенсира се (но частично) това удължаване в полския език, където пък континуантите на спомагателния глагол *съм* в така наречения „славянски перфект“ (например *-em, -eś* в *myślątem, myśląteś*) се пишат слято с еловата глаголна форма, в източнославянските литературни езици, където възвратните местоименни енклитики *ся, съ* (в белоруски и *ца*) се пишат слято с глаголните форми въобще, и в още по-голяма степен в украинския език, където спомагателните глаголни форми в единия тип „бъдеще време“ *иму, имеш* и т. н. се сливат графически с инфинитива (например *писатиму, писатимеш*).

Необходимо е да се отбележи изрично, че като една монолекса се отчита и всяка абривиатура – инициална (акроним), например *КПСС*, или силабична, например *агитбригада* или *U-Boot*.

Брахиграмите обаче, които изискват „развързано многословно четене“, се броят за толкова монолекси, от колкото се състои пълният им буквен запис. Например българското *и т. н.* се смята за „дълго“ три монолекси, защото пълно се записва „*и₁ така₂ нататък₃*“, немското *usw* също е „дълго“ 3 lx („*und₁ so₂ weiter₃*“). Английското и френското (в същност латинско) *etc* обаче е „дълго“ 2 lx, защото пълно се записва „*et₁ caetera₂*“.

Приема се при лексиметризацията да не се смятат за разделящи едно буквосъчетание на две монолекси дефисът (съединителната чертица, т. е. -) и апострофът (т. е. ’). Например българското *най-добре* е 1 lx, както и руското *наилуше*. Френското *j’aime* и английското *master’s* са също смятани за „дълги“ по 1 lx.

При определяне обема на текстовете, от които се извличат данни за честотите на думите, когато се съставят честотни (фреквентни) речници, обикновено използват като мерна единица монолексата, макар че я наричат „текстова дума“ и по този начин дават повод за редица недоразумения. Тези недоразумения са свързани с практиката някои словосъчетания да се характеризират като „сложни (съставни,

перифрастични) форми на изменяеми думи“ (най-често глаголи). Така например словосъчетанията *ще пиша* (в български), *буду писатъ* (в руски), *буду писати* (в украински), *będę pisać* или *będę pisał* (в полски), *werde schreiben* (в немски), *shall write* (в английски), *vais écrire* (във френски) се смятат за сложни глаголни форми и това дава „основание“ на някои да не ги броят за по две монолекси.

Крайният резултат от такъв начин на мислене е опитът на един френски (Гиро 1954) и един съветски лингвист (Головин 1971) да измерват обема на текстовите проби и текстовете не „в монолекси“, а „в пълнозначни думи (mots-forts)“. Тъй като понятието „пълнозначна дума“ не е достатъчно ясно, мерната единица на Гиро-Головин трудно може да бъде въведена в глотометрията: дискусиите относно непълнозначността на много думи и даже класове думи (частици, спомагателни глаголи и енклитични местоимения) няма да имат край!

С въвеждането на термина (и понятието) „монолекса“ тези дискусии стават безпредметни. Монолексата е просто мерна единица. Нейната практическа приложимост в глотометрията се опира на довежданата до практическа безизключителност правописна норма във всеки език. Ето защо в рамките на всеки език, както показва вече значителният опит в областта на глотометрическите изследвания, използването на равни по (моно)лексова дължина проби (най-често хектолекси, по-рядко килолекси) дава глотометрически резултати, които лесно получават лингвистична интерпретация.

Лексиметрически изравнените проби може да се използват за получаване на лингвистично интерпретируеми глотометрически характеристики не само на „думи“ (лексеми) и класове „думи“ (например глаголни лексеми), но и на по-малки от „дума“ езикови факти (например „звукове“ или „звукосъчетания“ и дори „фонетични диференциални елементи“). Това е така, защото чрез броя на монолексите в един текст се определя само обемът (размерът) на текста. А след като сме изравнили текстовете в някакъв набор по размер, т. е. след като сме отрязали от всеки текст равни проби (хектолекси) можем във всяка от пробите да определяме честотата на най-различни типове езикови факти.

На пръв поглед може да се стори, че изравнените по брой на монолексите проби е недопустимо да се използват за изследване честотите на езикови факти, по-големи от една дума (монолекса), защото последната дума на пробата може да не бъде последна дума на словосъчетание от изследвания тип. Ако се интересуваме например от честотата на така наречените „съставни собствени имена“, като

Стара планина, Стара Загора, Софийско поле, и последната дума на някоя от пробите ни се окаже *Софийско*, докато следващата дума, вече въвн от пробата, е *поле*, за лингвист без глотометрическа подготовка такъв факт може да компрометира лексиметризацията изобщо.

А затруднението може да се преодолее по съвсем прост начин: приема се, че интересуваният ни обект трябва да бъде регистриран с поява в пробата само ако в пробата е налице целият, т. е. ако пробата завършва със *Софийско*, а следващата дума в текста е *поле*, не бива да се приема, че в пробата е налице словосъчетанието *Софийско поле*.

Букви и „звукове“

И така лексиметрически изравнените проби са удобни за глотометрическо изследване на всякакви по размер типове лингвистични обекти, но в рамките на един език. Лексиметризацията на текстовете лесно се осъществява „на ръка“, не изисква нито много време, нито особена лингвистична ерудиция. Затова лексиметрически изравнените текстови проби постепенно се превръщат в основа на всякакви глотометрически съпоставки между различни типове текст (стилове – индивидуални, функционални) на един език.

Когато обаче се съпоставят в лексиметрически изравнени проби от различни езици честотите на някакъв клас факти, за резултатите, които се получават, трудно се намира ясна лингвистична интерпретация. Причината на трудността е едно отдавна съществуващо у лингвистите практически необоримо убеждение – убеждението, че в различните езици текстовите думи (монолексите) са, средно взето, „различно дълги“. Едва ли ще се намери например лингвист, който да възрази срещу твърдението, че немската дума (монолекса) е значително по-дълга от английската.

Ако експлицираме глотометричната основа на това убеждение, ще открием, че то се опира на интуитивно използване на друга единица за измерване дължината на текста – буквата. По данни на Л. В. Малаховски (Андреев 1965, с. 139) в английски текстове от областта на радиоелектрониката средната дължина на монолексата е около 5,1 букви, а по данни на И. Б. Фитиалова (Андреев 1965, с. 158–159) в текстове на немски език също от областта на радиоелектрониката средната дължина на монолексата е около 6,6 букви.

Просто пресмятане показва, че английски текст от областта на радиоелектрониката с размер хиляда думи (една килолекса) ще съдържа $5,1 \cdot 10^3$, т. е. около 5100 букви, а подобен текст на немски език с

размер пак една килолекса, т. е. хиляда думи, но хиляда немски, т. е. по-дълги думи, ще съдържа приблизително 6600 букви. Другояче казано, английски текст с размер една килолекса почти ще се събере на 102 машинописни реда (при 60 удара на ред), а немски текст със същия лексов размер ще съдържа цели 25 машинописни реда в повече.

Лингвистите отдават различията между измерваните „в букви“ дължини на думите в разните езици само на условностите на различните правописни системи. Приема се без възражения, че един и същ (поне по времетраене) говорен звук в едни правописни системи се означава с монолитерема, т. е. с една буква (например в кирилицата *ш*, в чешката, словашката, хърватската и словенската латиница *š*), в други – с дилитерема, т. е. със съчетание от две букви (например в полската латиница с *sz*, в английската и албанската с *sh*, във френската с *ch*), в трети – с трилитерема, т. е. със съчетание от три букви (в немската с *sch*). Французите в някои случаи записват с тетралитерема, т. е. със съчетание от четири букви, говорния звук, който в други езици записват с монолитерема (например в *Bordeaux* с тетралитеремата *eaux* се записва един гласен звук, поне по времетраене равен на гласния звук, записван в други езици с монолитеремата *o*).

Има и, така да се каже, обратни примери: на руската монолитерема *ш* по общо признание отговаря звуков комплекс, по трайност равен на два консонанта (куриозен факт е, че в немска транслитерация руската монолитерема *ш* се превръща в *schtsch*, в хепталитерема, т. е. в съчетание от седем букви); на чешката, словашката и унгарската монолитерема *á* във финската и естонската латиница отговаря (поне по времетраене) гласен звук, записван с дилитерема – *aa*.

В действителност различията в средния размер на думата (монолексата) в различните езици, когато го определяме чрез броя на буквите, не се дължи само на различията в правописните системи. Това лесно се доказва, като съпоставим средните размери на монолексата, когато ги определяме чрез средния брой на буквите в нея, за различни стилове на един и същ език.

Например по данни на Т. А. Селютина (Андреев 1965, с. 150–152), получени по ексцерпт от романа на Ч. Дикенз „Hard Times“, монолексата в английската художествена литература едва ли ще бъде средно по-дълга от 4,5 букви, а видяхме, че в английските текстове от областта на радиоелектрониката монолексата е дълга средно 5,1 букви. И може да се очаква, че аналогично ще бъде положението в много езици. Например в руските текстове по радиоелектроника монолексата е „по-дълга“ (Андреев 1965, с. 77). И в българската беле-

тристика монолексата, измерена в букви, е средно по-дълга, отколкото в научните и техническите български текстове (Андреев 1965, с. 327 – данни на Н. И. Федулова и Г. Я. Мартиненко).

Но даже ако различията в средния размер на монолексата, измерена в букви, се дължаха само на различията в правописните системи на разните езици, пак бихме имали основание да преразгледаме „лексовото“ решение на проблемата за мерната единица в глотометрията.

Фонометризация

Появата на новата мерна единица, която в глотометрията се нарича монофона, е резултат от именно такова преразглеждане на лексовото решение на проблемата. Теоретически най-основателно е да се опита да изразим големината на езиковото съобщение с единиците, с които измерваме продължителността на процесите, т. е. с количеството на секундите (евентуално милисекундите, минутите, часовете), необходими за говорната реализация на съобщението. На практика обаче измерването на големината на писмените съобщения „в секунди“ се оказва много трудно осъществимо. Кратък текст (от две-три машинописни страници) един четец може да прочете за три пъти по-късо време, отколкото друг четец, без четенето нито на единия, нито на другия да звучи неестествено.

Само че различието се дължи почти изключително на количеството и продължителността на паузите, с които разчленяват говора си хората. „Звучащата част“ от едно съобщение, изричана от най-различни четци и при най-различни темпове на четене, е с почти еднаква продължителност. Относително малките, но все пак съществуващи отклонения имат за причина нерегламентираното от правоговора „разтягане“ на някои гласни при четене, което обаче добре се осъзнава като неестествено, та и съответният начин на четене се оценява от всички лингвисти като „нееталонен“ и при буквено записване се отразява само във вид на словесни характеристики на четеца от типа: „говори провлечено“, „разтегля гласните“, „напява“.

Доколкото така наречената фонетична транскрипция (с изключение, за които ще стане дума по-долу) литеризира (отразява буквено) тъкмо и само „звучащата част“ от съобщението, оказва се, че броят на буквите във фонетичната транскрипция на едно съобщение на кой да е език може да служи като надеждна числова оценка на размера на съобщението. Нужно е да се направят две уговорки. Едната е за т. нар. „дълги гласни“ и „дифтонгите“. В някои системи за фонетично транскрибиране те се означават с единични букви (монолитереми),

над които стоят диакритики, означаващи дължина или дифтонгичност. Дългите гласни и дифтонгите при измерване на съобщението „в монофони“ (съкратено *fn* или *phn*) се смятат като дифони, т. е. като „дълги“ две монофони.

Другата уговорка е за т. нар. африкати и т. нар. аспиrowани (придихателни) консонанти. Всеки африкат и всеки аспиrowан консонант се брой за една монофона, въпреки че в някои фонетични транскрипции може да се отбелязва със съчетания от две букви. Например *ч* (*č*) в една от фонетичните транскрипции на английския език се отбелязва с дилитеремата *tʃ* (съчетание от латинско *t* и математическия символ „интеграл“). Широко разпространеното в Азия аспиrowано *д* (*d*) също редовно се означава с дилитерема – *dh*. Обратно, означаваните в някои фонетични транскрипции с монолитеремни геминирани (двойни) консонанти трябва да се броят като дифони, защото за разлика от африкатите и аспиrowаните консонанти „траят“ колкото две монофони.

Повече нищо не е нужно да се знае, за да се превръща броят на буквите от фонетичната транскрипция на едно съобщение във фонетрична дължина на съобщението, т. е. в размер на съобщението, определян чрез броя на монофоните. Тази *фонометризация* на съобщението представлява процедура, твърде проста, ако разполагаме с негова фонетична транскрипция, каквато използват диалектолозите и лексикографите.

Няколко примера. Английското географско название *Worcester* в съответствие с фонетичната му транскрипция се фонетризира просто: $w_1 u_2 s_3 t_4 ə_{5:6}$ (двоеточието е знак за дължина на *ə*). Следователно то е „дълго“ 6 *phn*. Чешкото *říká* се оказва „дълго“ също 6 монофони, защото *í* и *á* са „дълги гласни“, т. е. дифони, а *ř* е африкат, т. е. монофона. Полското *posiedzenie* пък е „дълго“ 8 *phn*, защото *i* след *s* и *n* пред *e* „не се брой“ (във фонетичната транскрипция не съответствува на отделна буква), а с дилитеремата *dz* се означава африкат, т. е. монофона.

Примерите подсказват, че ако се познават правилата за трансформиране на едно записано според официалния правопис съобщение в съобщение, записано във фонетична транскрипция, няма да бъде трудно да се фонетризира то, като се „прескочи“ фонетичното му транскрибиране. Нещо повече – за много езици (например за всички славянски с изключение на сърбохърватския, защото сърбите и хърватите не отбелязват задължително по различен начин дългите и кратките гласни) фонетризацията на правописно запи-

саните съобщения може лесно да се автоматизира. Несложността на такава автоматизация се обяснява с това, че за да се фонетризира едно писмено съобщение, необходимо е да се вземат под внимание не всички особености на фонетичната транскрипция, а само тези, които имат отношение към размера (дължината) на звучащата част от съобщението. А тези особености за щастие се поддават леко на формализация.

Банка килофони

Но доколкото днес все още автоматичното фонетризиране на писмените съобщения е само възможност, а глотометрическите изследвания изискват значителен брой фонетрично изравнени проби – килофони – на разни езици, списание „Съпоставително езикознание“ ще започне да информира библиографски за използваните вече в публикуваните глотометрически изследвания килофони от разни езици, като ще ги номерира и ще посочва точно границите им (началото и края им).

Другояче казано, предстои да се образува банка от килофони, която ще позволи да се използват от разни хора на разни места за глотометрически изследвания едни и същи килофони. И това не само ще спестява на мнозина значителните усилия по отброяването на килофоните. Банката от килофони ще осигурява една и съща текстова основа за всякакви глотометрически изследвания. А, както ще се постарам да покажем, глотометрията експлицира количествената същност на всички съпоставителни лингвистични изследвания.

В банката всяка килофона ще бъде снабдена с номер, състоящ се от буквена и цифрова част. Буквената част регистрира езика, а цифровата – мястото на килофоната в банката (двете първи цифри – годината, третата – книжката на списание „Съпоставително езикознание“, в която е регистрирана килофоната, а следващите цифри – номера на килофоната в списъка на регистрираните в тази книжка на списанието килофони). Например „килофона R 7835“ означава „килофоната руски текст (R), регистрирана с номер 5 в трета книжка от (19)78 година на списание „Съпоставително езикознание“.

В настоящата трета книжка на списанието като начало са „вложени“ в създаваната банка 100 руски (R 7831 – R 783100) и 100 български (Bg 7831 – Bg 783100) килофони. И „ефектът“ на банката килофони е вече налице – в три от публикуваните статии са използвани отчасти едни и същи килофони. Предвижда се в бъдеще списанието да посредничи между вложителите в „банката килофони“, за да си

осигуряват те размяна и на копия от целите килофони, регистрирани в банката.

Всяко глотометрическо изследване е съпоставително лингвистично изследване

Откриването на банката килофони и особено постепенното нарастване на „влоговете“ в нея ще осигурят постепенно интензивирание на глотометрическите изследвания, които се отличават с това, че са винаги експлицитно, явно съпоставителни. Глотометрията изисква непременно съпоставяне на числа, всяко от които представлява честота (брой на появи) на някакъв тип лингвистични обекти в определена по размер текстова проба. За да има глотометрически обект, трябва да са налице поне две, а най-добре пет или даже десет проби, поне хектолексни, а най-добре килофонни. И глотометрическото изследване започва със съпоставяне на честотите на изследвания тип факти в пробите, а завършва с числа, характеризиращи степента на близост между честотите на типа факти.

Методиката за получаване на тези числа може да се разбере и даже да се усвои, като се проучи внимателно как е използвана в публикуваните в тази книжка на списанието глотометрически изследвания. От тези публикации може да се направи и важен извод: глотометрическите изследвания не са развитие само на статистико-лингвистичните изследвания, които доскоро си поставяха задачи, нестоящи пред „неколичествената (а в същност „неексплицитно количествената“) лингвистика“, а представляват усъвършенствувание на всички лингвистични изследвания.

Статистико-лингвистичните изследвания не се осъзнаваха от огромното мнозинство лингвисти като „истинска лингвистика“, защото не търсеха решения на проблемите, с които живееше, живее и днес „редовият лингвист“, а проблемите, които се опитваха да решат, изглеждаха на редовия лингвист твърде периферни и често даже непонятни. На числата, с които се изразяваха резултатите от тези изследвания, той гледаше или с неоснователно доверие, или с основателно недоверие (когато трябваше да бъдат интерпретирани по начин, противоречащ на интуицията му).

Така например натрупаните през последните десетилетия данни за „количеството информация (негентропията)“ на буква от писмената езикова практика и на свързаната с тази негентропия редуваност на „езика“ често, нещо повече – почти винаги се оказват статистически твърде ненадеждни, а им вярват и ги цитират, макар да не

знаят за какво са им (ако изключим свръхобщия извод, че естествените човешки езици съвсем не са „икономични“, извод, при това, философски незащитим).

От друга страна, заниманията на статистиците с „богатството на речника“ на отделни писатели или на цели езици се посрещат в средите на трезво мислещите лингвисти твърде скептично, защото на тези лингвисти е добре известно колко силно е подложено на субективни вариации причисляването на една или друга словоформа към определена речникова единица (лексема), да не говорим за неяснотата на понятието „речник“.

Глотометричните изследвания търсят решения не на проблеми, чужди или периферни за неглотометрическата (или по-точно предглотометрическата, или най-точно неексплицитно глотометрическата) лингвистика. Глотометрията е методика за решаване на същите проблеми, които стоят в центъра на вниманието на лингвистите и които те напразно са се опитвали да решават досега без нея, а доколкото са ги решавали, решавали са ги, като са се опирали на неявна (имплицитна) глотометрия, на каквато непременно се опира интуицията на добре работещия лингвист.

Всяко добро лингвистично изследване има в основата си имплицитна глотометрия

Ако се опитаме да сведем проблемите, които стоят в центъра на интересите на лингвистите днес, до възможно най-малко на брой пределно общи класове, ще стигнем до три такива класа проблеми: 1) проблеми, отнасящи се до извличането на носената от съобщението информация (можем да ги наречем „проблеми на лингвистичния анализ“); 2) проблеми, свързани с начините, по които някаква информация се превръща в съобщение („проблеми на лингвистичния синтез“); 3) проблеми, които имат отношение към изграждането на устройства, способни да анализират и синтезират съобщения („проблеми на лингвистичната инженерия“).

Като търсят решения на проблеми и от трите класа, лингвистите винаги имат работа с огромен масив от буквени писмени съобщения (текстове). Без посредничеството на буквени текстове днес никой не се опитва да прави лингвистични изследвания. Съзнавайки или не съзнавайки точно какво правят, лингвистите извличат (подбират си) от този текстов масив проби, съпоставят наблюденията си върху пробите и изводите от наблюденията си екстраполират върху целия масив или върху определени (но по правило големи) части от него.

Естествено е да се извършват предимно наблюдения върху факти, появяващи се в поне една от пробите. Значи в подбора на обектите за изследване има имплицитна глотометрия. Имплицитността ѝ обаче позволява по принцип значителни отклонения от научната истина: говори се не винаги основателно за „чести“ в някаква проба факти, без да е ясно спрямо какви еталонни проби фактите се обявяват за „чести“; обявяват се за „важни“ не особено чести факти по съображения, които можем да характеризираме като конюнктурни; правят се изводи за еднаквост в поведението на един и същ тип факти въз основа на единични появи на фактите от изследвания тип, без да се забелязва, че в част от пробите изследваните факти се появяват значително по-често. И, което е още по-неприятно, имплицитността на глотометрията, лежаща в основата на лингвистичните изследвания, не позволява на лингвистите да стигат до еднакви, убедителни за всички резултати при изследване на едни и същи текстове.

А ако е вярно, че признак за зрелост на една наука е получаването на едни и същи резултати при изследване на едни и същи набори от обекти, то предглотометрическият (имплицитно глотометрическият) етап от развитието на лингвистиката не може да се смята етап на зрелост. Без експлицитна глотометрическа методика научното в лингвистиката не може да се отличи от ненаучното и общността на лингвистите, твърде нараснала числено през последните няколко десетилетия, не може да се самоочиства радикално от носителите на квазинаучно и даже псевдонаучно фразьорство, което съществено понижава ефективността на лингвистическия труд.

Но това не означава, че на този етап не са получени и не се получават трайни, съответстващи на научната истина резултати. Трудно се различават истинските научни постижения от лъжестиженията в областта на т. нар. обща или теоретическа лингвистика, защото „общите лингвистични теории“ не е ясно по правило от колко голям текстов масив са абстрахирани, та не подлежат на глотометризиране. За голям брой лингвистични изследвания, които внасят в обектния фонд на лингвистиката нови, подредени факти, резултат от преки наблюдения на по-рано неописвана или грубо описвана езикова практика (диалекти – териториални и професионални; езици, използвани от малък брой хора; текстове от стари епохи), може със сигурност да се твърди, че ще запазят стойността си и след глотометризацията на лингвистиката, особено ако са съпроводени с публикации на проби от изследваните съобщения, въпреки че към такива

изследвания „теоретиците“ прикачват пренебрежителните етикети „фактология“ или „фактография“.

Даже когато не съдържат изрични данни за големината на наблюдавания текстов масив, фактоложките (фактографските) публикации позволяват, макар и с грубо приближение, да се оцени математико-статистически големината на този текстов масив, та поради това може да се очаква, че с натрупването на все повече и повече експлицитно глотометрически наблюдения върху езиковата практика съображенията на фактолозите ще стават все по-лесно глотометризуеми.

Основания да се насочваме към глотометрическа експлицитност в съпоставителната лингвистика

Широко разпространяващият се напоследък сред лингвистите интерес към т. нар. съпоставителни (контрастивни, конфронтативни) изследвания, интерес, „рожба“ на който е и списание „Съпоставително езикознание“, също даде в резултат и „теоретически“, и „фактографски“ публикации. Но, за разлика от „общотеоретическите“ лингвистични опуси, даже и „най-общотеоретично“ замислените и по различни начини изпълнени, но все съпоставителни по същество лингвистични изследвания, като френско-немската „сравнителна стилистика“ (Малблан 1961), немско-английската „контрастивна лингвистика“ (Бургшмит, Гьоц 1974) или френско-руската „сравнителна типология“ (Гак 1977), позволяват да се експлицира глотометрическата им основа, тъй като по заглавията им, макар и неексплицитно, можем да получим представа за текстовите масиви, на които се опират, чрез назоваването на съпоставяните езици.

Само че посочването на цели езици като текстови масиви за привържениците на експлицитната глотометрия (намираща се още „ин статус nascendi“ и поради това още твърде агресивно отрицателски настроена към неексплицитно глотометрическите опити за широки лингвистични обобщения) изглежда и недостатъчно ясно, и същевременно ненаучно претенциозно.

Глотометрията по принцип е в състояние да внесе ясност в разбирането на термини като „английски“, „френски“, „немски“, „руски“, „български“ език, тъй като са известни начините да се получава математико-статистически надеждна оценка за големината на текстовия масив, който има (или трябва да има) пред вид един лингвист, когато назовава „цял един език“. Няма днес разумен човек, който през живота си да стига до усещането, че напълно владее родния си език. Ето защо, за да се разбере какво точно означава твърдението „еди-кой си

владее даден език като роден“, трябва добре да се планира и осъществи социологичен експеримент, който да установи колко голям текст (чужд и свой) е реципирал средно един човек докъм 25-ата си година, и ще имаме нужната ни оценка за големината на текстовия масив, „маскиран“ под названието „роден език“.

Нещо повече – някои разсъждения ни позволяват и сега да направим, разбира се, съвсем грубо занижена оценка за количеството езикова материя, която се реципира например от българина, докато той стигне до двайсет и петата си година. Ако приемем, че той „пропуска“ в съзнанието си за минута средно по стотина монолекси, когато разговаря, слуша други да говорят или четете, и че разговаря, слуша други да говорят или четете средно по осем часа в денонощие, за една година влезият в съзнанието му текстов масив се оказва не по-малък от 15 мегалекси, т. е. 15 милиона текстови думи. За двайсет и една-две години обемът на този масив достига поне до 300 мегалекси (т. е. до 300 милиона текстови думи).

Тази оценка е грубо занижена, защото, като изключим особено увличащите се в слушане на инструментална музика или в съзерцаване на творби на изобразителните изкуства, хората са езиково активни (в смисъл: слушат говорене, говорят или пишат, четат) много повече от осем часа на денонощие. Но и 300 мегалекси е достатъчно внушително число, за да направим веднага поне един ясен извод.

Да се разчита, че на съзнанието на някои хора може да се „подадат“ два масива от по толкова мегалекси за същото време, за което на съзнанието на други хора се „подава“ само един такъв масив, е невъзможно. А в последна сметка степента на овладяването на един език се определя от големината на текстовия масив, подаден на съзнанието на изучаващия езика, и то подаден в ранната му младост. А при това, когато на съзнанието на един човек е подаден вече достатъчно голям текстов масив (на родния език), другият, новият текстов масив не попада на празни, отредени за него места, а се свързва с вече наличните знания по родния език, усвоява се „на фона“ на тези знания, смесва се с тях.

За полиглотта, т. е. за този, който се ползува от услугите на повече от един език, единият език (родният) е все пак основното средство за общуване (разбира се, ако полиглотът не живее дълго сред хора, които не знаят и не желаят да учат родния му език). Останалите езици не може да са му даже приблизително толкова добре познати, колкото му е познат родният. Поради това именно в трудове, които имат претенцията, че съпоставят цели два езика, единият от двата езика

(по-малко познатият на автора) е непременно поставен в неравноправно положение спрямо другия.

Последиците от това неравноправие най-добре се виждат в двуезичните (даже големите, наричани „пълни“) речници: речниците, входният език на които е роден за съставителите, се различават значително от речниците, на които изходният (гълкуващият) език е роден за съставителите. Но не бива да се мисли, че ако се комплектува съставителски колектив „с равно участие“ на носители на двата езика, ще се получи в резултат „добре сбалансирани речник“: първо, не са известни начините за такова „балансиране“, и второ, по-важно, на тези, които ще използват речника, не е нужен „балансиран“ речник, защото те владеят или единия, или другия от двата езика по-добре.

По-малко личи „неравноправието“ на езиците в трудове, които имат претенции, че са посветени на съпоставянето на граматическите или стилистическите съставки на цели два езика, като посочените по-горе книги на Малблан и Гак, но не защото неексплицитно глотометрическите съставки на тези компоненти на езиците по-малко се влияят от неравноправието на съпоставяните езици, а защото при такива съставки още никак не е ясно какво се съпоставя.

Ясно е това неравноправие в друг вид трудове, в които се прави граматическо или стилистическо описание на цели езици (макар че в заглавията на тези трудове съпоставително-лингвистичният им характер не е обявен), когато описанието се прави за нуждите на хора, за които описваният език не е роден. Такива са граматиките на българския език за поляци (Славски 1954) и за руси (Маслов 1956), които са примери за по същество съпоставителни граматически (а в много случаи и стилистически) описания на българския език, писани за нуждите на хора, владеещи полски, съответно руски език като родни.

Но и на тези прекрасни, образцови съпоставително-лингвистични трудове са присъщи недостатъците, непреодолими в неопиращите се на експлицитна глотометрия изследвания: от една страна, се отделя внимание на много редки в българската езикова практика факти, а, от друга, не се дава информация за твърде често проявяващи се и поради това „по-опасни“ за чужденци особености на описвания език.

Например говори се за изчезнали днес даже в диалектите инфинитивни форми като *че, ре, пле, до*, а задължителната (съкратена) инфинитивна форма на спомагателния глагол *съм* в *може би* не се споменава, та и поляци, и руси напразно търсят в тези граматички обяснение защо е съвсем чужда за българското ухо калката *може да бъде* от полското *może być* и руското *может быть*.

А в едно опряно на експлицитна глотометрия съпоставително българско-полско или българско-руско граматическо описание такъв пропуск би бил изключен, защото ще бъдат ясни две неща: първо, най-честият инфинитив и в полския, и в руския език е инфинитивът на спомагателния глагол *съм* (*бу́с, быть*) – може да се очаква, че неговото съответствие ще бъде преди всичко нужно на човек, който иска да пише и (или) да говори български; второ, най-честите лексеми най-често имат „неправилни“ алолекси, рецидиви от по-стари състояния от историята на езика, като при това тези алолекси са най-вероятни в т. нар. „петрифицирани“ или „лексикализирани“ (т. е. много чести) съчетания с други думи – затова *би* е налице само в адвербиализацията *може би*, но даже на *nie może być* и на *не может быть* българското съответствие е единствено *не може да бъде*.

Коментарът на примера конкретно насочва към извода, че ако се вземат под внимание честотите на появите на езиковите факти, съпоставителните лингвистични изследвания ще бъдат предпазени от пропуски като описания. Но такъв извод все още не лансира необходимостта от експлицитна глотометрическа опора даже само за съпоставителните лингвистични изследвания, които имат за обекти „цели езици“. Още по-малко може да се отнесе той към съпоставките между части от два или повече от два езика.

И интуитивната (имплицитната) глотометрия може да направлява лингвиста съпоставител, когато пробира от езиковата материя фактите, заслужаващи да бъдат описани. Достатъчно е той да бъде предупреден, че трябва да пробира „по честотен принцип“.

Експлицитната глотометрия може да се осъзнае като необходима опора за съпоставителната лингвистика само от тези контрастивисти, които са склонни да приемат, че съпоставителното изучаване на езиците трябва да се превърне в ефективно управляван процес, които са близко до убеждението, че трябва да се изследва не какво да е, не „интересното лично за изследвача“, а интересното за цялата научна общественост, за цялото общество. Именно убеждението, че трябва да се дава приоритет на това, което е интересно за обществото, прави необходима експлицитната глотометрия в съпоставителната лингвистика.

Между интересното за даден лингвист и важното за обществото не може да се постави знак даже за приблизително равенство, но между интересното за обществото и важното за обществото има строго равенство. Нещо повече: за обществото като цялост е по-интересно (и по-важно) да се изследва това, което в езика е по-често, и

поради обществения характер на лингвистиката, която не само работи за обществото като цялост, но и ангажира големи човешки колективи: за да се разпределя разумно работата между много хора, нужно е да се използва някаква „по-тънка“ методика за ранжиране на обектите на лингвистиката въобще и на съпоставителната лингвистика в частност.

А известно е, че интуитивната (имплицитната) глотометрия ранжира езиковите явления по честотна скала, която е твърде крупно разграфена (Фрумкина, Василевич 1971, с. 7–28), докато експлицитната глотометрия може да йерархизира честотно „кандидатите“ за изучаване в много повече степени и да осигури по такъв начин и по-широк фронт, и по-дълбока перспектива за планиране на предстоящата работа.

Освен това експлицитната глотометрия чрез апарата на пробите и ексцерптите дава възможност да се осигури строго равнопоставяне между съпоставяните текстови масиви, а следователно и между езиците, на които принадлежат тези текстови масиви, като съпоставянето може да се организира така, че да се получат изводи за различия или подобия в поведението на определени типове факти от съпоставяните езици чрез обработване на минимални по обем текстове, изводи, чиято надеждност е не само висока, но и измерима.

БИБЛИОГРАФИЯ:

- Андреев 1965:** *Статистико-комбинаторное моделирование языков* (под редакцияй Н. Д. Андреева). Москва–Ленинград, 1965.
- Бургшмит, Гьоц 1974:** E. Burgschmidt, D. Götz. *Kontrastive Linguistik deutsch/englisch*. München, 1974.
- Гак 1977:** В. Г. Гак. *Сравнительная типология французского и русского языков*. Ленинград, 1977.
- Гиро 1954:** P. Guiraud. *Les caractères statistiques du vocabulaire. Essai de méthodologie*. Paris, 1954.
- Головин 1971:** Б. Н. Головин. *Язык и статистика*. Москва, 1971.
- Котова, Янакиев 1968:** Н. В. Котова, М. Янакиев. Сопоставление некоторых количественных характеристик славянских языков. „Славянская филология“. Сборник статей, выпуск седьмой. Москва, 1968.

- Котова, Янакиев 1973:** Н. В. Котова, М. Янакиев. Глоттометрический подход к вопросу о близости славянских языков. „Славянская филология“. Сборник статей, выпуск девятый. Москва, 1973.
- Малблан 1961:** A. Malblanc. *Stylistique comparée du français et de l'allemand*. Paris, 1961.
- Маслов 1956:** Ю. С. Маслов. *Очерк болгарской грамматики*. Москва, 1956.
- Найденов и др. 1976:** К. Найденов, М. Котаров, М. Янакиев, Н. В. Котова. О „силе“ межморфемной связи. *Использование математических моделей и электронных вычислительных машин в лингвистике. Сборник докладов Первой национальной конференции по использованию математических моделей и электронных вычислительных машин в лингвистике*. София, 1976.
- Славски 1954:** Fr. Sławski. *Gramatyka języka bułgarskiego*. Kraków, 1954.
- Фрумкина, Василевич 1971:** Р. М. Фрумкина, А. П. Василевич. Получение оценок вероятностей слов психометрическими методами. „Вероятностное прогнозирование в речи“. Москва, 1971.
- Янакиев 1977:** М. Янакиев. *Стилистиката и езиковото обучение*. София, 1977.