

Гинка Димитрова

ИНФОРМАТИКАТА И ЛИНГВИСТИКАТА ВЪВ ВЗАИМОДЕЙСТВИЕ

(Информационните тезауруси като обект на лингвистиката)

Така формулираното заглавие предполага разглеждане на следните основни проблеми:

1. Информационните процеси в обществото като фактор за промяна на съвременната езикова ситуация. Разкриване на връзките между естествения език (ЕЕ), езика за специални цели (ЕСЦ) и изкуствените езици (ИЕ).

2. Изграждане и използване на информационните езици. Тезаурусът като речник за нуждите на информационното търсене и като средство за представяне и систематизиране на научното знание.

3. Информационният тезаурус (ИТ) и проблемът за унифициране на знанията (унификация и стандартизация на научната терминология).

4. ИТ и общата лексикография. Тезаурусна лексикография.

Японският учен Ю. Масуда определи бъдещата индустриална ера като ера на информационното общество, в което потреблението на знания ще представлява по-голяма ценност от натрупването на материални блага. Това е бъдещето, а засега ние навсякъде слушаме за информационната лавина, която затрупва човечеството. Термините „информационен взрив“, „информационна криза“, „информационно общество“, „компютърна революция“, „информационна технология“ и т. н. представляват отражение в езика на различни аспекти на извършващата се автоматизация на информационните процеси, внедрение на изчислителната техника във всички сфери на живота, осъзнаването на все по-голямата роля на информацията, на „информационните ресурси“ в живота на съвременното общество.

Автоматизацията на информационните процеси е изграждането на нов тип комуникации в обществото — човеко-машинни комуникации¹. Но особеностите на автоматизацията на информационните процеси предявяват свои изисквания и към езиковото развитие, към езиковите средства

на обществото, използвани за фиксацията, предаването, съхраняването, търсенето, преработката на цялата социално значима информация.

Следователно под въздействие на потребностите на информационното общество се изменя и ЕЕ, конструират се и ИЕ, усилват се процесите на интерференция между естествените и изкуствените езикови средства. Това може нагледно да бъде показано чрез разкриване на съотношението между ЕЕ, ЕСЦ и ИЕ. Елементите на тази верига са взаимно свързани, а не противопоставени. Т. е. между единиците на ЕЕ, ЕСЦ и ИЕ съществуват не отношения на контрарност, а отношения на логическа производност. Например термините като единици на ЕСЦ заедно с общоезиковите се характеризират и с други, специфични само за тях признаци (еднозначност или тенденция към еднозначност, емоционална неутралност). Лексическият състав на ЕСЦ, както сочат някои терминологии (В. Лейчик, Р. Котов, С. Никитина), все повече се променя и разграничава от състава на ЕЕ. Както семантичните, така и формалните различия ще доведат до приоритет на изкуствените елементи в ЕСЦ над естествените. Понастоящем под въздействието на информационната сфера се извършва фактическо срастване на ЕЕ с изкуствените знакови системи. Наред с думите, словосъчетанията, фрагментите на текста широко се използват цифрови и азбучно-цифрови схеми, индекси, условни знаци — еквиваленти на наименованията от ЕЕ. Процесът на изместване на средствата на ЕЕ от изкуствени се разпространява и върху граматичния строй. Морфологичната система и синтаксисът се ограничават в рамките на определен вид словосъчетания. Следователно езикът на деловата документация е хибридна естественоезикова система за обслужване на специални комуникации в сферата на управлението. Но най-ярко силното въздействие на обществото върху езиковото развитие се проявява в активното използване на ИЕ за масова употреба — езиците за програмиране и информационните езици. Създавани на базата на ЕЕ, те придобиват все по-голяма самостоятелност.

От казаното можем да направим извод, че езиковата ситуация в нашето съвремие е коренно променена. Тя изисква от езикознанието и на първо място от приложната лингвистика отчитане на установилото се естествено-изкуствено двуезичие и реална оценка на традиционния обект на изучаване.

И така разработката на лингвистичните основания за нормиране на специалните езици, създаването на изкуствени и естествено-изкуствени езици са най-актуалните проблеми на приложната лингвистика.

А една от най-перспективните области на взаимодействие между теоретичната и приложната лингвистика е работата по създаването на автоматизирани системи за търсене на информация. С цел намиране на търсените документи в информационните системи текстовете на тези

документи трябва да бъдат представени във вид на един или друг семантичен запис (т. е. текст на един или друг информационен език), а това предполага отчитане както на парадигматичните, така и на синтагматичните връзки между думите. Конкретната задача — построяването на информационна система за определен отрасъл от науката и техниката, изисква не само решаването на най-сложните проблеми на семантиката, но и работа по непосредственото моделиране на смисловата страна на езика. С тези проблеми се занимава информационната лингвистика. Тя разглежда информационните езици като лингвистични обекти. Създаването на тази приложна лингвистична дисциплина е свързано и с факта, че на дадения етап развитието на теоретичната лингвистика не може да служи за основа на практическата дейност, която извършват информационните работници. Лингвистичните теории се намират в постоянно развитие и по своя характер са по-скоро абстрактни, отколкото утилитарни. Създаването на нова, интегрална по своята същност научна дисциплина се явява своеобразен компромис между строго теоретичните лингвистични изследвания и конкретните, практически нужди на информационното обслужване. Понастоящем с информационната лингвистика се свързва цял комплекс от мероприятия, свързан с разработването и приложението на определени лингвистични средства. Те включват собственоезикови средства и процедури (програми) за обработка на текстовата информация. Езиковите средства на автоматизираните информационни системи са естественият език и неговите производни — ограничен естествен език и формализирани езици. Процедурите — това са средства за манипулиране със смисловите единици на различните нива (с морфемите, думите, словосъчетанията, фразите, свръхфразови единства) и в частност средствата за техния семантико-синтактичен анализ и синтез.

Следователно обособяването на информационната лингвистика като раздел на приложната лингвистика е свързано с необходимостта от изучаването на езика в действие, в комуникация. Новата лингвистична дисциплина не е чужда на интегралните научни методи (логико-лингвистични, лингво-философски, лингво-математически и др.).

От друга страна, ние сме убедени, че лингвистите, решили да се занимават с проблемите на информационното обслужване, ще намерят огромно удовлетворение в това и ще имат възможност да погледнат на традиционните езикови проблеми по един нов начин. Бихме цитирали и думите на френския изследовател на информационни езици Жан Клод Гарден като потвърждение на това наше мнение: „Малко лингвисти знаят за онова богатство от наблюдения и открития, които ги очакват в тази област, и нежеланието да се занимават с лексиката на специалните области на знанието не може да бъде оправдание за отхвърлянето на теоретичните проблеми, възникващи тук“².

А отхвърлянето на проблемите е просто невъзможно, защото те са общи (макар и разгледани по специфичен начин) както за лингвистиката, терминологията, така и за информатиката. В този смисъл можем да посочим въпросите за систематизирането на терминологичната лексика, за съотношението между теоретичната и тезаурусната семантика, за унифициране на знанията и свързаните с това проблеми на културата на речта, на стандартизацията на научните термини, въпросите на съвременната лексикография за мястото на тезаурусите сред останалите видове речници.

Но да се спрем на всеки от поставените въпроси малко по-подробно.

Основен елемент от всяка информационна система е определен информационен език, с помощта на който се извършва кодирането на информацията. А семантичната основа на всички информационни езици е лексиката, систематизирана в ИТ. Неин източник е терминологията на различните науки. Основна единица в ИТ е дескрипторът.

В ИТ всичко е насочено към разкриване семантиката на подезика като цяло. Семантичната характеристика на термина е ограничена от дескрипторната речникова статия. Свообразието на този вид речници от идеографски вид е в това, че те съдържат значителна смислова информация, която е представена не само в дескрипторната статия, а и в рубриците и разделите като цяло. Основното предназначение на ИТ е максималното разкриване на семантичните отношения на всеки термин.

И така какво дава тезаурусът като „речник с концептуален вход и експлицирани семантични връзки“³? Нас ще ни интересуват структурните и функционалните възможности на информационния речник, създаден за нуждите на информационното търсене и съдържащ лексиката на информационния език. Но за нас са важни не неговите функции за търсене на информация (това е въпрос, който засяга информатиката), а възможностите му за представяне на научното знание и способността му да бъде средство за семантичен контрол в областта на науката.

Много специалисти отбелязват връзката между създаването на системата от понятия в определена научна област и разработката на ИТ. И в единия, и в другия случай се използват общи типове класификации (преди всичко йерархични). Следователно ИТ се разглежда като „модел на съответната наука“⁴.

Основните семантични отношения в ИТ, които определят и семантичните възможности на информационните системи (ИС), са:

1. Синонимични (класове на условна еквивалентност).
2. Йерархични: род — вид, отношения на принадлежност, а за някои научни дисциплини като география и биология — част — цяло.
3. Асоциативни: обект — предназначение, обект — операция, обект — характеристика и др.

Трябва да отбележим, че съпоставянето на основните понятия на теоретичната и тезаурусната семантика води до извода за наличието на аналогия между всички техни понятия. Но тезаурусната семантика не дублира лингвистичната. Развивайки се в значителна степен достатъчно автономно и намирайки се под влияние на практическите потребности на информационното търсене, тезаурусната семантика започва да изработва и свои специфични идеи и свой особен език, който несъмнено ще окаже влияние върху теоретичната семантика.

Непосредствена е връзката между изграждането на ИТ и унификацията и стандартизацията на терминологията. В идеален вид тезаурусът трябва да представлява цялостно лингвистично описание на езика на науката, на дадения отрасъл от знание.

Тезаурусите позволяват да се реши и проблемът за установяване на съответствието между:

1. Авторската терминология.
2. Терминологията на системата.
3. Терминологията на потребителя на ИС.

Тезауруса определихме като речник с концептуален вход. Т. е. основен негов принцип е представянето на информацията в посока от смисъла към думата (както е при всички идеографски речници). За разлика от другите терминологични словници той включва по-широк кръг от терминологични и понятийни полета.

Тезаурусът не съдържа конкретни логически определения на термините. Терминът (дескрипторът или ключовата дума) се определя косвено чрез съвкупността от семантични отношения в дескрипторната статия. Общото в целите и методите на научно-техническата терминология и тезаурусите е в това, че се обработват едни и същи обекти — термините. Една от предварителните задачи в процеса на създаване на стандартизирана терминология и тезауруса е осъществяване на коректна класификация на понятията, което позволява от своя страна установяването на най-съществените смислови връзки между тях.

Използвайки информацията, получена от различните семантични отношения на термина дескриптор, биха могли да се създават по-точни определения в обикновените тълковни речници, защото традиционната процедура за дефиниране на термините не дава основания за оптимистични прогнози.

Посочените различия между ИТ и терминологичните речници поставят необходимостта от отделянето на нова област на лексикографията — тезаурусната лексикография. Тя трябва да систематизира всички сведения за създадените ИТ, трябва да предложи тяхна типология. Тезаурусната лексикография има широк кръг от общи и специални задачи, които могат да се окажат далечни за специалистите по обща лексикография.

И накрая, обобщавайки казаното и набелязвайки нови перспективи за изследването и разработката на тезаурусите, трябва да уточним, че тезаурусът не е само работен инструмент на информатиката, подобно на библиографските указатели и рубрикатори. Подобен подход би стимулирал изключително прагматичния подход при тяхното съставяне и избор на структурата им. Не трябва да се забравя, че тезаурусът е начин за описание на семантиката на езика. Следователно изучаването на ИТ трябва да се осъществява не само от гледна точка на удобството при тяхното използване, но и от теоретико-семантични позиции.

Много важни са съвременните широки изследвания по описание на теоретичния модел на тезауруса като форма за представяне на информация, знание.

Би било извънредно полезно да се проведе сравнителен анализ на вече съществуващите тезауруси, да се състави каталог за тях.

Ще завършим с това, че така представените проблеми не изчерпват всички аспекти на тезаурусната семантика и тезаурусната лексикография, а по-скоро само ги набелязват.

БЕЛЕЖКИ

¹ Естественый язык, искусственный язык и информационные процессы в обществе. М., Наука, 1988, с. 7.

² Цитира се по: В. А. Москович. Лексико-семантическая система естественного языка и информационные языки. Автореферат диссертации на соискание ученой степени доктора филологических наук. Л., 1971, с. 4.

³ С. Е. Никитина. Семантический анализ языка науки. М., Наука, 1987, с. 53.

⁴ Пак там.