

működtető könyvtárak, valamint a mikrofilmtechnikai, ill. COM-berendezéseket előállító cégek képviselőit.

A berendezések tartóssága

A könyvtárakban az olvasóberendezések jelentősebb igénybevételnek vannak kitéve, mint a bankokban, irodákban.

Sok könyvtár állapította meg, hogy a fényforrásként szolgáló izzók élettartama alacsony, s mind a szellőző, mind a képet élesre állító szerkezetek hamar elromlanak. Problémákat okozott a szellőző szerkezet rossz elhelyezése, a készülékek használatát segítő magyarázó feliratok hiánya, a külső lakkebevonat lepattogzása stb.

A gyártó cégek képviselői kifejtették, hogy a könyvtári piac rendkívül kicsi, ezért a könyvtári használók igényeit az olvasóberendezések kialakításakor csak kevés figyelembe venni. Csupán hosszabb élettartamú izzók forgalmazására tettek ígéretet. Az is világossá vált, hogy sem a COM, sem az olvasóberendezéseket gyártó cégek nem foglalkoznak a tekercsfilm használatával, fejlesztési tevékenységük elsősorban mikrofilmlapra irányul. Az egyik érdekelt cég képviselője úgy vélte, hogy ma már csak a könyvtárak alkalmaznak tekercsfilmet, forgalmuk 95%-át filmlappal működő berendezések teszik ki.

Ebből következően felmerült a mikrofilmlapos katalógus kifejlesztésére irányuló kísérletek szükségessége. Jelenleg már rendelkezésre állnak ilyen katalógusok is (néhány az olvasók, néhány a könyvtár munkatársai számára).

A kazetták tárolási lehetőségei

Súlyos kifogásként hangzott el a megbeszélésen, hogy a gyártó cégek úgy fejlesztik ki az olvasóberendezéseket, hogy nem veszik figyelembe a teljes rendszert. Így pl. a

CIL olvasóberendezésekhez nem lehet megfelelően csatlakoztatni a kazetta-tartókat. A kazetta-tartó ugyan ráragasztható a berendezés oldalára, de a kazetta a súlya miatt így sokszor leválik.

A Planet olvasóknál tíz kazetta elhelyezésére van lehetőség, de ez is oly módon, hogy a berendezés előtt ülő nem látja a kazetták feliratait. Leginkább a Microscot szabadon elhelyezhető, 14 kazetta tárolására alkalmas tartója vált be, amely az olvasóberendezés tetejére erősíthető.

Kazetták

Több könyvtárban tapasztaltak problémákat a kazettákkal kapcsolatban is, pl. gyors tekerceslésnél a film megcsúszik és hamar kopik. A konstrukció javítás azonban jelentősen növelné a kazetták árát.

A COM-ügynökség szerint egy kazettában 100 láb (30,48 m) hosszúságú film (1800 kép) helyezhető el. A kazetta gyártója ugyanakkor úgy vélte, hogy csak 50 láb hosszú film alkalmazható. Kétségtelen, hogy kevesebb film esetén a tekerceslési problémák egy része megszűnne.

Nehézségek merültek fel a kazetták feliratozását illetően is (munkaerőhiány).

Film-minőség

A COM-ügynökségek folyamatos ellenőrzése ellenére is több könyvtár emelt a film minőségére vonatkozó kifogást (kettős kép, képkimaradás, torzulás, fakultság stb.). Nagyon költséges valamennyi kazettás filmet a könyvtárban is ellenőrizni, ugyanakkor mégis szükségesnek tűnik.

/HADLOW, J. F.: Some problems with COM catalogues = Microdoc, 16. köt. 3. sz. 1977. p. 75–82./

(Sárdy Péter)



GÉPI FORDÍTÁS

Ténylegesen működő gépi fordítási rendszerek kialakítása

A fordítási folyamat automatizálásának problémája nem hogy vesztett volna fontosságából, hanem ellenkezőleg: a lefordítandó információáradat világszerte tapasztalható növekedése miatt különösen aktuális.

Így például a Szovjetunióban a Szovjetunió Minisztertanácsa mellett működő Tudományos és Műszaki Állami Bizottságnak és a Szovjetunió Tudományos Akadémiájának felügyelete alatt működő Tudományos-műszaki For-

ditások Össz-szövetségi Központjának (Vszeszojuznij Centr perevodov naučno-tehnicseszkoj literaturü i dokumentacii, VCPNTLD) adatai szerint csak a saját fordításuk volumene néhány év alatt a tízszeresére nőtt, és 1976-ban mintegy 40 ezer szerzői ívnyi fordítás készült. A fordítási megrendelések száma azonban a központ lehetőségeit is meghaladja, és az igények még tovább fokozódnak.

A gépi fordítás szovjetunióbeli helyzetét 1973-ban a Tudományos és Műszaki Állami Bizottság által szervezett bizottság tanulmányozta.

A bizottság következtetései szerint az országban megvannak azok a társadalmi, tudományos-műszaki és szervezeti előfeltételek, amelyek lehetővé teszik az üzemszerűen működő gépi fordítási rendszerek kialakítását. E rendszerek műszaki bázisaként rendelkezésre áll az ESZR számítógéppark. Szükség van azonban a koordinálatlan törekvések összefogására. Az erre kidolgozott javaslatot a Tudományos és Műszaki Állami Bizottság elfogadta, és az idegen nyelvekről orosz nyelvre történő gépi fordítási rendszerek létrehozásával, valamint a gépi fordítási törekvések terén kialakítandó nemzetközi együttműködéssel kapcsolatos feladatokat a VCPNTLD-re bízták.

A további javaslatok megvalósítása azonban számos tudományos-műszaki és szervezési kérdést vetett fel.

A szemléletesség és az áttekintés céljából ezek a kérdések az alábbiak szerint csoportosíthatók:

- a) az elmélet kiválasztása;
- b) a már kidolgozott elvek hasznosítása;
- c) a matematikai alapok és a számítógépi programok kimunkálása;
- d) szervezeti kérdések.

Mint minden más esetben, a gépi fordítási rendszerek realizálását is célszerű a megfelelő elmélet kiválasztásával (vagy kimunkálásával) kezdeni.

A gépi fordításhoz szükséges a nyelvi adatok teljes formalizálása, különben nem lehet azokkal egyértelműen operálni. E feladatot a nyelvészetnek, mindenekelőtt a strukturális nyelvészetnek kellett megoldania, ugyanis a strukturális nyelvészet formális-logikai szerkezeteit lehet a legkönnyebben matematikai jelekkel kifejezni.

A nyelv formalizálásával és algoritmizálásával kapcsolatos kísérletek új eredményekre vezettek és lehetővé tették a nyelvi jelenségek tanulmányozását elősegítő új módszerek kialakítását is. Új kutatási irányok keletkeztek, amelyek a hagyományos nyelvészetben eddig teljesen ismeretlenek voltak. Ilyen például a *nyelvi modellek szerkesztése, a nyelv és a beszéd statisztikai és elméleti-információs mérése* és több más irányzat. A vizsgálatba nemcsak nyelvészeket kapcsoltak be, hanem matematikusokat is. Együttműködésük meghozta az eredményt, többek között a matematikai diszciplinaként elkönyvelt matematikai nyelvészet formájában. Ennek eredményeképpen a gépi fordítás égisze alatt a nyelv formalizálását és több érintkező problémakört átfogó kutatás keretében az eredeti kísérleti alapfeladat – nevezetesen a fordítás automatizálása –, mindinkább közvetett céllá vált. A tényleges fordítási algoritmusok létrehozását a nyelvészet szempontjából triviális és másodrendű feladatnak kezdték tekinteni.

Az elmélet fejlődése napjainkban két irányú. Az első irányzatot *mérnöki irányzatnak* nevezhetjük. A nyelv és a beszéd statisztikai és elméleti-információs jellemzőinek tanulmányozását itt úgy értelmezik, mint a gépen belüli nyelvi jel speciális elméletének létrehozását. Ez az

irányzat a gépi fordítást a számítógépi lehetőségek fokozatos kihasználásával kívánja megvalósítani. A számítógép szóformák és szóösszetételek felismerésével a szöveget olyan értelmes szegmentumokra osztja, amelyek lehetővé teszik a nagyobb szövegrészek, mondatok elemzésére való áttérést.

A második irányzatot *szemantikainak* nevezhetjük. E megközelítés az ember nyelvhasználatának szemantikai modelljét alkotja meg. Míg az előbbi irányzatban a nyelvészeti leírás és a gépi algoritmus nem határolódik el egymástól, addig ez utóbbiban az elhatárolás elvi alapot képez. Feltételezik ui., hogy a természetes nyelv átalakításával (fordításával) kapcsolatos műveletek sikeres elvégzéséhez szükség van a nyelvi tények (a nyelvtan) nyelvészek által elvégzett és eléggé árnyalt formális leírásának, valamint az e tényekkel és leírással (mechanizmussal) végzendő operációk algoritmusának külön-külön, önálló létezésére.

A szemantikai modellezés gyakorlati célja szótárak készítése, amelyek a különböző szintek (morfológia, szintaxis, szemantika és lexika) vonatkozásában széles körű információkat tartalmaznak, s ezek az információk elvileg elégségesek egy-egy adott lexikális egység valamennyi kontextusának a leírására. Ezen irányzat értelmezésében a fordítás folyamata – a gépi fordítás folyamatát is beleértve – nem más, mint különféle színvonalú, egyfelől felületi; másfelől mélységi (értelmi) nyelvi struktúrák egymásutánjának az átalakítása.

A *gépi fordítás elméletének fogalmát* az ismertetett irányzatok más és más módon értelmezik. A mérnöki irányzat szerint a gépi fordítás a szöveg egymás után következő, a kis egységektől a nagyobb egységek felé haladó gépi leírása, amelynek minden lépését gépi program ellenőrzi, illetve amelyben a gép állapítja meg, hogy a be- és a kimeneti nyelv egységei megfelelnek-e egymásnak. A szemantikai elmélet a szintek és a nyelvek közötti, egyetlen meghatározott jelentés kifejezését célzó átalakítások elmélete.

A ténylegesen működő gépi fordítási rendszerek azonban mind a Szovjetunióban, mind pedig más országokban egyaránt többé-kevésbé *vegyes vagy átmeneti elméleti változatot* hasznosítanak ilyen vagy olyan módon. Így a *grenoble-i rendszer* lehetővé teszi a nyelvi közvetítő tiszta ideájától való eltérést, és meglehetősen nagy szerepet játszó algoritmikus tökéletesítések sorozatába torkollik. *O. Sz. Kulagina kísérleti rendszere* korlátozott számú kiegészítő nyelvvél és kisméretű szótárral rendelkezik, és sok időt emészt fel a változatok kiadásában. Emellett a kísérleti üzem algoritmikus korlátai lényegesen befolyásolják az itt alkalmazott szűrőmódszer eredményességét. Átmenetinek minősíthetjük a *ciklikus elemzés módszereit* is, amelyek alkalmasak lévén a gépi fordításra, számos rendszerben kerültek bevezetésre. Ezek a módszerek egyfelől elégséges árnyaltsággal és teljességgel modellezik az ember nyelvi magatartását a

nyelvi analízis során, másfelől lehetővé teszik a beszéd statisztikai és mennyiségi karakterisztikáinak felhasználását, illetve ténylegesen fel is használják őket. Ennek ellenére, mivel nem élnek a szemantikával olyan mértékben, ahogy azt a szemantikai irányzat ajánlatosnak tartja, bonyolultabb elemzésekben ezek a módszerek elégtelennek mutatkoznak.

Napjainkban még nem mondható el, hogy már megszületett volna az az elmélet, amelynek közvetlen realizálásával minőségi gépi fordítás készíthető üzemszerűen.

A különböző változatok realizálásánál azonban célszerű egyetlen, leginkább esélyesnek tűnő stratégiát követni. E tekintetben totális és szelektív stratégiákat különböztethetünk meg. *A totális stratégia* a gépi fordítást előzetesen és előfeltételként – struktúráját és működési mechanizmusát illetően – teljes körű ismeretekre törekedve óhajtja megoldani. *A szelektív stratégia* az ismertől az ismeretlen, illetve az egyszerűtől a bonyolult felé halad. Ennek célszerűségét az a megfontolás valószínűsíti, hogy az elmélyült problémákat csak abban az esetben oldjuk meg, amikor a felület szintjén nem lehetséges a megoldás.

Ezért leszögezhető, hogy a gyakorlatilag működő rendszerek realizálása közben kényszerűen több elméletre és variánsra kell támaszkodni, ám ez mindig egyetlen stratégia alapján történjék. Felmerül a kérdés, hogy ebben a helyzetben miként hasznosíthatók a már elért eredmények.

A gépi fordítási rendszernek az alábbi négy részből kell állnia:

- a) a be- és a kimeneti nyelvek megfelelő mennyiségű információval rendelkező szótárai;
- b) a bemeneti szöveg elemzésének algoritmusai;
- c) a kimeneti szöveg szintézisének algoritmusai;
- d) programcsomagok.

A ténylegesen működő rendszerek kialakításakor *célszerű igénybe venni a már létező szótárakat, algoritmusokat és programokat*. Mivel azonban a gépi fordítási rendszerek egyes részterületein dolgozó kollektívák nem nagyon törődnek a többiekkel, ahhoz, hogy a különböző kész részekből összeállhasson egy teljes rendszer, legalább *az alábbi kompatibilitásokat kell biztosítani:*

egy adott nyelvpár esetén a bemeneti nyelv szótárának kompatibilisnek kell lennie a bemeneti szöveg elemzésének algoritmusával, minthogy az algoritmus szótári információt hasznosít. A kimeneti nyelv szótárát viszont a kimeneti szöveg szintézisét szolgáló algoritmus szükségleteivel összhangban kell megteremteni, mivel ez az algoritmus ugyancsak szótári információkat hasznosít; az elemzés algoritmusának kompatibilisnek kell lennie a szintézis algoritmusával, minthogy az utóbbi az előző által rendelkezésre bocsátott információkra alapszik;

kompatibiliseknek kell lenniük egymással az algoritmusok és a szótárban található információk kódolására szolgáló rendszereknek;

tekintetbe kell venni, hogy a gépi fordítás korszerű szemlélete csupán ágazati, s nem pedig általános tematikájú szövegek fordítását kívánja megoldani, s ezért a bemeneti és a kimeneti nyelv szótárainak is kompatibiliseknek kell lenniük egymással az adott tudományos és műszaki szaknyelv jegyében.

Ha mindez fennáll, feltételezhető, hogy a rendszer egyes viszonylag független részei egymás által szigorúan meghatározott és egyértelmű eredményeket fognak produkálni. A gyakorlatban ez azonban igen ritkán következhet be, mivel minden következő részben számolni kell a megelőző szakaszban elkövetett hibák kijavításával.

Jelenleg a Szovjetunióban az idegen szövegek orosz nyelvre történő gépi fordításához a következő anyagok állnak rendelkezésre:

viszonylag kis terjedelmű *tudományos és műszaki területekről angol, német és francia nyelvű szakszótárak*. Ezek eléggé reprezentatív mintákból készültek, és mind általános, mind szaknyelvi lexikát tartalmaznak. E szótárakat az esetek többségében frekvencia-szótárak egészítik ki, illetve frekvencia-szótárak révén javították meg minőségüket. E téren továbbra is aktív tevékenységet fejt ki a nyelvstatisztikai munkacsoport számos kisebb kollektívája. A német szakemberek szintén érdekes eredményeket értek el az orosz nyelvű szakszótárak kifejlesztésében:

egyed-egy tudományos és műszaki területeken széles körű *szemantikai és szintaktikai információkkal kiegészített szakszótárak* léteznek. Ezek mélységi struktúrákkal operáló kísérleti rendszerek számára készültek. Ilyen szótárral dolgoznak pl. a VINITI-ben és a Szovjetunió Tudományos Akadémiáján;

rendelkezésre állnak *a német, angol és a francia nyelv morfológiai és szintaktikus elemzésére szolgáló algoritmusok*. Ezek részint a már működő kísérleti rendszerek részét képezik (pl. O. Sz. Kulagina algoritmus), részint önállóan léteznek, s így a szövegek automatikus átalakításával összefüggő feladatok megoldására is felhasználhatók (pl. a leningrádi Zsdanov egyetem számítóközpontjának szintaktikus elemzésre alkalmas algoritmus);

kísérleti gépi fordításhoz kidolgoztak ilyen rendszerek részét képező *szintetizáló algoritmusokat*. Kisebb mértékben készültek olyan szintetizáló algoritmusok is, amelyek önálló életet élnek;

kidolgoztak már *software-t* is teljes kísérleti gépi fordítási rendszerek, illetve egyes részeik számára. Ez utóbbiak közül a szótár alapján való keresés, a morfológiai elemzés algoritmusait, a szövegszegmentumok elkészítésére alkalmas algoritmusokat stb. említhetjük meg példaként. A programok a szovjet gyártmányú számítógépek különféle típusain hasznosíthatók. A gépi fordítás programozását szolgáló speciális nyelvet a Szovjetunióban ugyan nem fejlesztettek ki, de vannak olyan programnyelvek, amelyek e célra is felhasználhatók, így pl. a szabvány-operátorok nyelve.

A gépi fordítás nagy teljesítményű és gyorsaságú számítógépeket igényel, pl. az *ESZ-1040* és az *ESZ-1050* típust. Ezzel szemben a legtöbb gépi szótár és nyelvszerkesztési program régebbi, kisebb gépen (Minszk, BÉSZM-4) került kidolgozásra. A programok átírása viszont meglehetősen bonyolult.

A feladatok programozására alkalmas speciális gépi nyelv kidolgozása kétségtelenül aktuális feladat. Azzal azonban számolni kell, hogy iparszerű alkalmazás esetén egy ilyen nyelv használata meghosszabbítja a munkaidőt.

Új lehetőségeket kínál az ember és a számítógép közötti lehetséges együttműködési formák szélesedése. A párbeszédű üzemmód lehetővé teszi azoknak a nehéz eseteknek a megoldását is, amelyekre az automatikus elemzés nem ad lehetőséget. Az emberi közreműködés két változata lehetséges:

- a közvetlen beavatkozás,*
- a folyamat nyelvszerkesztési részébe való beavatkozás.*

Az első esetben az ember társszerkesztőként működik közre, a másodikban meg utószerkesztőként javítja ki a gépi fordítás hibáit. Mindkét esetben természetesen lehetőség van a szöveg előzetes szerkesztésére is.

A legnagyobb nehézséget jelenleg azoknak a szövegeknek az elemzése jelenti, amelyek sok formulát, általában nem betű formájú írásjeleket tartalmaznak. Ezek külön feldolgozást igényelnek. A fordításra kerülő anyagnak ugyanis meglehetősen egyneműnek kell lennie, és nem szabad meghaladnia a szótár lehetőségének határait.

A gépi fordítás vizsgált nyelvei elsősorban az angol, a német és a francia.

Az angol nyelvű szövegek orosz fordítására két szaknyelvi területen fejlesztettek ki rendszereket. Az *AMPAR-rendszer* (*Avtomatizirovannij masinnij perevod sz anglijszkogo na ruszskij = Automatizált gépi fordítás angolról oroszra*) a beszéd- és mondatrészek ciklikus elemzésének elveire épül a sok jelentésű angol szavak orosz nyelvre való lefordítását lehetővé tevő szótárral. Programozásához a szabvány-operátorok nyelvét veszik majd igénybe, működésében közbeeső és utólagos szerkesztéssel számolnak.

A másik angolról oroszra fordító rendszer a *SZPAR* (*Szisztema perevoda sz anglijszkogo na ruszskij = Fordítási rendszer angolról oroszra*). Ez a folyamatos szintaktikus elemzés elve szerint fog működni. Olyan összefüggés-grammatikát kívánnak benne felhasználni, amely képes az összes lehetséges helyes angol mondat-szerkezet értelmezésére. Az ilyen rendszer szótára megkívánja, hogy minden szócikkéhez terjedelmes szintaktikus és szemantikus információt rendeljenek hozzá. Jelenleg azt vizsgálják, hogy a rendszer szótáraként nem felel-e meg az a szótár, amelyet korábban a szemantikai szintézis céljából állítottak össze.

A németről fordító rendszer, a *NEPRA* (*Nemecko-ruszskij perevod avtomaticheskij = Automatikus német-orosz fordítás*) számítástechnikai szövegek fordítására létesül. Ciklikus elemzéssel fog működni. Szótára gyakorlatilag a szintaktikus szóosztályok szemantikai alosztályait hasznosítja. Külön figyelmet kell azonban fordítani a német szóösszetételek elemzésére, s e célból olyan kombinációs szótárat kell létrehozni, amely az összetett szóalakokat és az alapszavakat is tartalmazza. A rendszerhez nem kívánnak új programokat készíteni, hanem a programokkal rendelkező rendszerektől kívánják az alkalmasnak látszókat átkölcsönözni. A szótárakat várhatóan a segédészótárak és a konkordanciák széles köre fogja kiegészíteni.

A francia szövegek oroszra való fordításának rendszerét, a *FRAP-ot* (*Francuzszko-ruszskij avtomaticheskij perevod = Automatikus francia-orosz fordítás*), a BÉSZM-4 számítógépre kifejlesztett rendszerrel elért eredményeket hasznosítva kívánják megalkotni, a programot azonban teljesen át kell írni. Szótára, amely eredetileg matematikai szakszövegekhez készült, kicsi. Ezért új szótárakat kell készíteni, segédészótárakkal és konkordancia-jegyzékekkel is kiegészítve.

A távlati tervekben szerepel a különböző rendszerekhez egységesen használható orosz nyelvű szintetizáló módszer és program kialakítása.

/MARCSUK, Ju. N.: O posztroenii dejsztvujuscsh szisztem masinnogo perevoda = Naucsno-Tehniczeszkaja Informacija, 1. sor. 4. sz. 1977. p. 12-15./

(Futala Tibor)

