

kedhetett. Az állomány 44%-át játékfilmek, 15%-át gyermekprogramok tették ki.

Egy-egy kazettát évente átlagosan 21 alkalommal kölcsönöztek, ami lényegesen magasabb kölcsönzési gyakoriság, mint a nyomtatott dokumentumoké. 19 könyvtárban összesen 51 férőhelyet alakítottak ki a videofelvételek helybeni megtekintésére. Számos könyvtár szervezett rendszeresen előadással, szakmai vitával egybekötött videobemutatókat.

A könyvtárak számára nem gazdaságos, ha önállóan állítanak elő videoprogramokat. Célszerű azonban, ha együttműködést alakítanak ki a videostúdiókkal, filmgyártókkal és a pedagógusokkal.

A kölcsönzés a könyvtárak többségében (kétharmadában) ingyenes, kisebb részükben térítésköteles (3 DEM). A díjszedés számottevően csökkenti a forgalmat. A kölcsönzési időtartam rendszerint néhány nap, egy hétnél sehol nem több. Az állományról a könyvtárak szórólapszerű, ingyenes, elvihető listákat és/vagy tájékoztató füzeteket állítanak össze.

/LOEST, K.-G.: Videokassetten in Öffentlichen Bibliotheken der Bundesrepublik Deutschland. = ARBIDO-R, 5. köt. 1. sz. 1990. p. 3–7./

(Nagy Gábor Ádám)

Gépi fordítás és géppel segített fordítás – remények, gondok, kompromisszumok

A kezdetek

A gépi fordítás és a géppel segített fordítás igénye a második világháború idején merült föl. A gépi fordítás (machine translation = MT) a fordítási folyamat teljes gépesítését jelenti. A géppel segített fordítás (machine-aided translation = MAT) könnyebben megvalósítható: interaktív folyamat, azaz emberi beavatkozást igényel a gépi műveletek előtt, között és/vagy után – a fordítandó szöveg előkészítését, kétértelműségeinek feloldását, a legmegfelelőbb célnyelvi szó vagy kifejezés kiválasztását, a fordítás utószervezését stb.

A gyakorlati megvalósítást célzó kutatómunka az 1950-es években indult meg az Egyesült Államokban, a katonai szervek és a hírszerzés anyagi támogatásával. Az első kísérletek a *Georgetowni Egyetemen* folytak, 49 mondat oroszról angolra fordításával, 250 szavas szótár és hat nyelvtani szabály alapján.

Az *Amerikai Információtudományi Társaság* (American Society of Information Science = ASIS) már első szakcsoportjai (Special Interest Group = SIG) közül néhányat ebben a témakörben hozott létre. Felvette a témát kutatási tervébe az *Amerikai Dokumentációs Intézet* (American Documentation Institute = ADI) kutatás-fejlesztési bizottsága is.

A Szovjetunióban, a VINITI-ben az ötvenes évek végén indult meg a munka, ahol soknyelvű szótárakat állítottak össze a gépi fordítás céljaira, és egy közvetítő nyelv kifejlesztésével foglalkoztak.

Az Egyesült Államokban az 1960-as években hanyatlást hozott egy kongresszusi jelentés, amely kifogásolta a kutatás magas költségeit és csekély eredményét. Ez a jelentés az anyagi támogatás komoly csökkenését eredményezte. Egyetlen kivétel volt: Daytonban, Ohióban a légielő (U. S. Air Force) támaszpontján megkezdődött az azóta világszerte elterjedt *SYSTRAN* rendszer kifejlesztése.

Folyt azonban a munka ezekben az években Nyugat-Európában, Kanadában és Japánban.

A közelmúlt és a jelen

Az 1970-es években Nyugat-Európában indult meg nagy erővel a géppel segített fordítás fejlesztése. Franciaországban *Bernard Vauquois* és a *Grenoble-i Egyetem* munkája vonta magára a figyelmet. Ez lett később a kiindulópontja az *Eurotra* névre keresztelt EGK-programnak, amelynek azután nagy lendületet adott az 1992-re, az egységes EGK-piac megvalósítására történő felkészülés.

Ledolgozta hátrányát és az élre tört Japán, ahol gazdasági szükségszerűségnek tekintik a géppel segített fordítás megoldását. Ez a munka része az *ötödik generációs számítógépek* kifejlesztésére 1982-ben indult nagy állami programnak, amelynek az a célja, hogy az ország az 1990-es évekre az elektronikában és a digitális technikában a világ vezető hatalmává váljék. A kutatásra fordított nagy pénzek eredményeképpen már öt év alatt tíz japánról angolra fordító kommersziális rendszer született. Ipari kutatócsoportok is részt vesznek a munkában, például a *Fujitsu*, a *Hitachi*, a *Toshiba* és a *NEC* kutatói. Az Egyesült Államokból is igyekeztek bevonni a mesterséges intelligenciával foglalkozó kutatócsoportokat.

Kínai nyelvterületen a *Hongkongi Kínai Egyetem* fejlesztett ki 1968-ra egy sikeres rendszert. Ez a *CULT* nevű rendszer erősen interaktív folyamatban, jelentős előszerkesztési igénnyel fordítja a kínai matematikai és fizikai folyóiratokat angolra.

Kanadában a *Montreali Egyetem* fejlesztett ki sikeres, rutinszerűen működő rendszert a napi időjárás-jelentések angolról franciára fordítására. Ez a *TAUM-Meteo System* kevés nyelvtani szabályt és mindössze 1500 szótári egységet ismer. A lefordítandó időjárás-jelentésektől megköveteli, hogy kötött, szűk szóincset és szabványos nyelvtani szerkezeteket használjanak. Ezen az áron azonban emberi beavatkozás nélküli, tiszta gépi fordítást valósít meg.

A géppel segített fordításban kezdeményező szerepet játszó Egyesült Államok alaposan lemaradt energikus európai és japán versenytársai mögött.

Jelentős fejlesztő munka folyik a *Carnegie-Mellon Egyetemen*. Itt különbséget tesznek az egyszerűbb, géppel segített emberi fordítás (machine-aided human translation = MAHT) és a bonyolultabb, emberi segítséggel működő gépi fordítás (human-aided machine translation = HAMT) rendszerek között. Az utóbbi kategóriába tartozó kísérleti rendszerük egyszerű közvetítő nyelven keresztül IBM kézikönyveket fordít angolról japánra. Ez emberi segítséget nyújtó fordító igénybe veheti munkájához az Augmentor nevű nagyméretű tudásbázist. Az interaktív munka olyan fordítást eredményez, amely a szavakon túlmenően visszaadja a szöveg tényleges értelmét, belső kapcsolatrendszerét, sőt, retorikai jellemzőit is.

Említést érdemel a *Texasi Egyetemen* működő *Siemens* kutatócsoport *METAL* (machine evaluation and translation of natural language = természetes nyelv gépi elemzése és fordítása) nevű, fejlesztés alatt álló rendszere, valamint a *LOGOS* nevű rendszer, amelyet a légierő 1971-ben állított munkába a katonai berendezések karbantartási kézikönyveinek angolról vietnamira fordítására. Ez utóbbiaknak nyelvészeti alapelveit szinte üzleti titokként őrzik. A rendszert később a *Wang* cég németről angolra fordításra alkalmazta, és ebben a formában több más felhasználója is akadt, folyamatban van az angolról franciára és angolról németre fordító változat kifejlesztése.

Ismeretes több más működő rendszer is. Így a *Weidner Communications Corporation* angol–spanyol, spanyol–angol fordítója 22 helyen működik. A *Pánamerikai Egészségügyi Szervezet* (Pan American Health Organization) *SPANAM* nevű rendszere közvetítő nyelv nélkül fordít a szervezet hivatalos nyelvei (angol, francia, spanyol és portugál) között. A *Xerox*nál és még két helyen működő *ALPS* (automated language processing systems = automatikus nyelvfeldolgozó rendszerek) nevű rendszer angolból fordít franciára, németre, portugálra és spanyolra.

A jelenlegi helyzetről az Oisóban, Japánban 1989 áprilisában tartott konferencia adott jó áttekintést. Az ott elhangzottak szerint áttörésről nem beszélhetünk, de stabil, folyamatos fejlődés tanúi lehetünk.

A közeljövő céljai

A jelen és a közeljövő fejlesztő munkájában néhány jól körülhatárolt problémakör dominál. Ezek a következők:

- ▶ Árkérdések. Olcsóbb-e már a gépi vagy géppel segített fordítás a hagyományos emberi fordításnál?
- ▶ A telepítés kérdése. Személyi számítógépen, helyi nagygépen vagy bérelt vonalon elérhető központi nagygépen fusson-e a program?
- ▶ A gyorsaság és hatékonyság kérdése.
- ▶ Az elő- és utószerkesztő gárda betanításának kérdése.

- ▶ Soknyelvű rendszert ésszerűbb-e létrehozni, vagy nyelvpáronként külön-külön rendszert?
- ▶ A kulturális különbségek, jelentésárnyalatok, kétértelműségek lefordításának problémái.
- ▶ A témaorientált szótárak és a szakkifejezés-gyűjtemények összeállítása.
- ▶ Anyagi támogatás szerzése a kutatómunkához az állami költségvetésből és a magánszektortól.
- ▶ A működési alapelv kérdése: közvetítő nyelv vagy összekötő nyelvtan.*

Világszerte nagy erővel dolgoznak ezeken a kérdéseken, és gyorsan közeleg a soknyelvű számítógép ideje.

Fejlesztés az EGK keretében

Az *Európai Gazdasági Közösség* működése nagymértékben igényli a géppel segített fordítás megvalósítását, hiszen már az EGK hivatalos nyelvei is 72 lehetséges nyelvpárt adnak ki.

A munka 1976-ban a *SYSTRAN* átvételével indult, néhány nyelvpárral. Ma a szerződéses munkatársakkal megerősített 35 fős fejlesztő gárda 12 változaton dolgozik, ezek közül hat változat forrásnyelve az angol, négy változaté a francia, kettőé a német. Az angol és a latin nyelvek általában kevesebb nehézséget okoznak, mint a germán nyelvek. Így például az angolról franciára és a franciáról olaszra fordítás minősége sokkal jobb, mint a németről franciára vagy franciáról hollandra fordításé. Utószerkesztést azonban valamennyi változat igényel, bár a lényegét a nyersfordítás kielégítően visszaadja. A fordítás minősége jelentősen javítható specializált szakterületi szótárak, úgynevezett mikroszótárak összeállításával.

A *SYSTRAN* elsősorban az *Európai Bizottság* (European Commission = EC) belső céljait szolgálja. Itt a program egy IBM-kompatibilis Amdahl nagygépen fut MVS operációs rendszer alatt, óránként 500 000 szavas fordítási sebességgel. Ez a gép Unix kiszolgálóegységgel csatlakozik az EC hálózatához. A brüsszeli és luxemburgi felhasználók helyi mikro-gepeken, Unix vagy MS-DOS operációs rendszer alatt végzik az elő- és utószerkesztést.

Külső felhasználókat is kiszolgálhatnak, közülük jelentősek: brüsszeli NATO-központ, a karlsruhei nukleáris kutatóközpont és az NSZK-vasutak. Jelentős az a fejlesztés, amely az *Európai Szabadalmi Hivatal* (European Patent Office = EPO) kötendő szerződést szolgálja. A cél a szabadalmak fordítása az EPO négy hivatalos nyelve, az angol, a francia, a német és a spanyol között.

Jelentős új fejlesztési program indult 1982-ben az *Európai Közösség Tanácsa* (Council of the European Communities) égisze alatt *EUROTRA* néven. Ezt a hét

* Ez szorosan összefügg a kétnyelvű vagy soknyelvű rendszerek kérdésével. A soknyelvű rendszer gyakorlatilag csak közvetítő nyelvvél oldható meg, míg az egyetlen nyelvpárt kezelő rendszer általában egyszerűbben oldható meg összekötő nyelvtannal. – A ref.

évre tervezett programot az EGK finanszírozta, részben központi keretből, részben a tagországok közvetlen hozzájárulásából 27 millió ECU-val, amit később 4,5 millióval megtoldottak, de még további költségnövekedés várható. A program első két éve az előkészítést szolgálta: a szervezést és a feladat specifikálását. A következő három év középpontjában a nyelvészeti kutatás állt: az egyes nyelvek analitikus és szintetikus modellezése, valamint a nyelvek közötti átmenetek modellezése. Ezzel egyidejűleg elkészült a szoftver. A két befejező év a nyelvészeti modellek stabilizálását és az eredmények értékelését szolgálta.

Kísérletek Franciaországban

Franciaországban a gépi fordítás és géppel segített fordítás megvalósítását célzó kutatások központja a Tudományos-Műszaki Információs Intézet (L'Institut de l'Information Scientifique et Technique = INIST), a PASCAL és FRANCIS adatbázisok előállítója. Az első cél ezeknek az adatbázisoknak a többnyelvű indexelése volt. Ehhez 100 000 szavas kulcsszógyűjteményt állítottak össze. A francia kulcsszavakhoz automatikusan válogatták hozzá az angol és német megfelelőket.

Az 1970-es években magasabb célt tűztek ki: a tényleges többnyelvűséget, a címek és tartalmi kivonatok gépi lefordítását. Ehhez több számba jöhető fordító rendszert megvizsgáltak.

Az elsőnek kiválasztott TITUS II csalódást keltett merevsége és az adatbevitel kötöttségei miatt. 1988-ban megpróbálkoztak a fejlettebb TITUS IV változattal. A kohászat témakörében végeztek próba-fordításokat, ehhez 9000 szavas szakszótárt állítottak össze. A kísérlet tárgya ezer francia, valamint száz német és angol tartalmi kivonat lefordítása volt. A fordítások minőségét kielégítőnek értékelték, de túl időigényesnek bizonyult az előkészítés: egy 50–70 szavas kivonat előkészítéséhez fél óra kellett, így a fordítási teljesítmény mindössze 10 kivonat volt fejenként és naponta, ezért a TITUS rendszerről lemondtak.

Sikertelennek bizonyultak a SYSTRAN rendszerrel végzett kísérletek is. Ezeket 1977-ben kezdték meg a rendszer angol–francia változatával, majd 1982–83-ban folytatták az EGK-fejlesztésű új változattal. A próbafordításokat ezer tartalmi kivonaton a geológia, a fizika és a biotechnika témakörében végezték 120 000 szavas szótárral. A SYSTRAN komoly utószerveztést kíván. Bár a fejlettebb változattal ez már csak 8–15 percet jelent egy 70 szavas kivonathoz, de nagyon drága szövegszerkesztőre van szükség hozzá.

Az ALPS-TRANSMATIC rendszert 1982-ben vizsgálták a kohászat témakörében. A témakör 5000 szavas szakszótárával egészítették ki a rendszer alapszótárát. A kétértelműségek feloldása, az interaktív folyamat azonban így is hivatásos fordító munkáját igényelte, a végeredmény pedig ennek ellenére elmaradt a többi vizsgált rendszerétől.

1978-ban vizsgálták az energetika témakörében az ARIANE rendszert, az azonban nem bizonyult eléggé fejlettnak az INIST céljaihoz. Ugyanez derült ki a szintén 1978-ban vizsgált SYGMART rendszerről is.

Nem vizsgálták ténylegesen a WEIDNER rendszert, de a lehetőségeit tanulmányozva arra a következtetésre jutottak, hogy bizonyos további fejlesztésekkel reményt keltő lehet.

Továbbra sem tettek le arról, hogy a PASCAL és FRANCIS adatbázisok többnyelvűvé tételére a jövőben igénybe vegyék az automatikus fordítás eszközeit, de a jelenlegi rendszereket nem ítélték megfelelőnek az azonnali bevezetésre.

Kísérlet a BIOSIS-nál

A BIOSIS, a legnagyobb biológiai adatbázis 1926 óta dolgoz fel nem angol nyelvű szakirodalmat. Kb. évi 19 000 címet és 30 000 tartalmi kivonatot fordítanak angolra 40 különböző nyelvről. Ehhez a hét belső fordító és lektor mellett a keleti és a ritka nyelvekhez külső fordítókat is igénybe vesznek.

A géppel segített fordítás megvalósítására a SYSTRAN rendszerrel kezdték meg a kísérleteket. Ehhez a SYSTRAN Translation Systems, Inc. szolgáltatóközpontot vették igénybe (La Jolla, Kalifornia, USA), amelyet távközlési vonalakon értek el helyi IBM PC/XT mikrogepeikről. A szövegek elő- és utószerveztését szövegszerkesztő programmal végezték, a szolgáltatóközponttal a kapcsolatot a mikrogepeken futó SYSTRAN EXPRESS távközlési program segítségével építették fel. A szolgáltatóközpont teljesen automatikus, még az elgépelések sem javíthatók, emberi beavatkozásra csak az elő- és utószerveztés során van mód.

Az első kísérletet 1988-ban végezték címek oroszról angolra fordításával (a szolgáltatóközpont ezt tartja a legmegbízhatóbb nyelvpárnak), az élettudományi szakszótár felhasználásával. Mellőztek minden előszerveztést, hogy a lefordított címek (a BIOSIS szerkesztéspolitikájának megfelelően) a lehető legközelebb álljanak az eredetihez. A SYSTRAN táblázata szerint transliterált címetek úgy, ahogy voltak, begépeltek és elküldték. Néhány percen belül megkapták és lemezre töltötték a fordítást. A transliterálás nagyon időigényesnek bizonyult, és sok hibával járt.*

A kapott eredményt a SYSTRAN rendszer minőség-ellenőrzési irányelvei szerint vizsgálták oroszul folyékonyan tudó belső munkatársak. Az eredmény nem érte el a várakozásokat. A SYSTRAN szolgáltatóközpont saját nyelvésze magasabb minőségi mutatót hozott ki, mint a BIOSIS-é, de az ő értékelése is jóval alatta maradt annak, amire az oroszról angolra fordítást a SYSTRAN általában értékeli. Ennek az egyik

* Ez a nagyon időigényes és súlyos hibaforrást jelentő "gyalog" transliterálás roppant egyszerű programmal gépesíthető akkor, ha van cirill betűkészlettel is rendelkező mikrogep. — A ref.

lehetséges oka az, hogy talán az élettudományi szakszótár gyengébb, mint a többi. A BIOSIS fordítási csoportjának vezetője és kutatási koordinátora közölhetetlennek ítélték a lefordított címeket, és úgy értékelték, hogy rendbetételük nagyobb erőfeszítést igényel, mint az eredetiből való lefordításuk. Így a SYSTRAN rendszert céljaik megvalósítására alkalmatlannak minősítették.

Kísérletek a DoD műszaki információs központjában

Az 1980-as években rohamosan növekedett a külföldi információ fontossága az *Egyesült Államok védelmi minisztériuma* (Department of Defense = DoD) számára, különösen a DoD kutatóintézetei számára. A minisztérium műszaki információs központja (Defense Technical Information Center = DTIC) által üzemeltetett online információs hálózat részben közvetlenül, részben közvetve kapcsolatot biztosít olyan online szolgáltatóközpontokhoz is, amelyekben az adatbázisok egy része nem angol nyelvű (STN International, ESA-IRS, Télésystemes Questel, Data-Star, SINORG). Hogy a nem angol nyelvű adatbázisokat a csak angolul tudó felhasználók is hasznosíthassák, a DTIC tervbe vette a gépi fordítás megvalósítását. A terv az, hogy a felhasználó az online keresés talátaiból azonnali fordítást kapjon olyan mennyiségben és minőségben, amiből megítélheti, hogy a teljes dokumentumról kérjen-e emberi fordítást.

Az első kísérleteket a SYSTRAN franciából, németből és transliterált orosz szövegből angolra fordító változataival végezték. Az online elérhető SYSTRAN szolgáltatóközpontot a Wright-Patterson légitámaszponton helyezték el Ohióban. Egyoldalnyi transliterált orosz szöveg angolra fordítása három percet vesz igénybe. Teljes szövegek fordításához ez a sebesség interaktív rendszerben nagyon alacsony volna, és a minőség sem éri el az emberi fordítást. Az ízelítőül adott kis fordítások azonban valószínűleg hatékonyan támogatják a felhasználók döntését. Ha a kísérlet beválik, akkor installálják a SYSTRAN több nyelvpár fordítására is képes üzemszerű változatát.

Különösen izgalmas feladat a nem latin írású nyelvek fordítása. A DTIC részt vesz abban a programban, amelyet a kereskedelmi minisztérium irányít, és

amelynek a célja a japánról angolra fordítás megoldása, hogy több japán műszaki-tudományos információhoz férjenek hozzá.

Az NTIS tervei

A *National Technical Information Service* (NTIS) a "szürke irodalmat" (kereskedelmi forgalomban nem kapható irodalmat) gyűjti. Az évi kb. 60 000 referált dokumentum mintegy 30 százaléka származik az USA határain kívülről, főleg Nyugat-Európából (elsősorban az NSZK-ból), Kanadából, a Szovjetunióból és Japánból. Az 1980 óta bevitt dokumentumok közül 52 534 a fordítás. Ez a szám nagyobb is lehetne, ha a fordítás olcsóbb volna.

A légierről működő SYSTRAN rendszert úgy ítélte meg az NTIS, hogy ott az oroszról angolra fordító változat már kielégíti az igényeit, megkezdhetik a használatát. Ugyanott működik egy japánról angolra fordító változat is, de azt nem tartották eléggé érettnek.

Sikeres vállalati alkalmazások

A gépi fordítás gazdaságos gyakorlati alkalmazására jó példa két iparvállalat: a *Xerox* és a *Caterpillar*. Világszerte árusított termékeik specifikációját, karbantartási kézikönyvét és más dokumentumait fordítják a vásárló ország nyelvére. Már az eredeti angol dokumentumban szigorúan korlátozott szókészletet és nyelvtani szerkezetet használnak, így érik el, hogy a fordítások minden elő- vagy utószerkesztés nélkül, minden emberi beavatkozás nélkül elfogadható színvonalúak.

/Interpreting machine translation. = NFAIS Newsletter, 31. köt. 3. sz. 1989. p. 45–47.

LEAMY, Ch. (szerk.): Machine translation. A view from abroad. = Uo. p. 47–49.

LOHNER, D.: Automatic translation trials at CNRS/INIST. = Uo. p. 49–50.

HODGE, G.: Publishing machine-aided translation: a BIOSIS experiment. = Uo. p. 50–51.

BILLINGSLEY, R.: Translation in the DoD research community. = Uo. p. 51–52.

FEINSTEIN, P. T. (szerk.: BYRNA, T.): New FOCUS: Translation overview from NTIS. = Uo. p. 52–53./

(Válasz György)

Szerzői jogvédelem a számítógépes programokra – Csehszlovákiában is

A számítástechnika fejlődésének következményei világszerte különös figyelmet keltettek maguk iránt az elméleti és gyakorlati jogászok körében. S nem véletlenül, ui. a számítástechnika alkalmazása felettébb gyorsan nyomul be az élet szinte valamennyi területére, aminek következtében mind termékeivel, mind alkalmazásával kapcsolatosan merőben új tár-

sadalmi kapcsolatok keletkeznek. Mindez egy sor jogi szabályozást is megkövetel.

E követelmény teljesítése alól Csehszlovákia sem vonhatta ki magát. A hardverre vonatkozó innovációk jogi védelmével az országban nem is fordult elő nehézség, mivel a felfedezésekről, találmányokról, újítási javaslatokról és ipari mintákról szóló 84/1972.