

GÉPI FORDÍTÁS A VILÁG TUDOMÁNYOS KÖZPONTJAIBAN

Kiefer Ferenc

A gépi fordítás helyét a modern tudomány ágazatai között ma még nehéz lenne meghatározni. Sem kimondottan elméleti, sem kimondottan gyakorlati tudomány: magában egyesíti az elméleti és gyakorlati tudományok számos tulajdonságát. A gépi fordítás nem tekinthető az alkalmazott nyelvészet egyik ágának, hiszen a nyelvészetten kívül - és legalább olyan mértékben - támaszkodik bizonyos matematikai módszerekre is. Tulajdonképpen a kibernetika körébe sem tartozik, mert azon elsősorban a matematikai gépek elméletét /"gondolkodó gépek"/ értik.

Elméleti vonalon a gépi fordítás elsősorban a modern nyelvészet-re /strukturális, matematikai nyelvészet/ és a matematikai gépek elméletére hat, gyakorlatilag pedig elsődleges célján, vagyis a fordításon kívül elsősorban a tudományos dokumentáció /gépi kivonatolás és adatfeldolgozás/ fejlődését mozditja elő. E gyakorlati hatásoknak köszönheti komoly népgazdasági jelentőségét is.

Mind ezeknek a hatásoknak és kölcsönhatásoknak köszönhető, hogy a legutóbbi években egyre nő azoknak a nyelvészeknek, matematikusoknak, logikusoknak, számítógép-szakembereknek, dokumentációs kérdésekkel foglalkozóknak a száma, akik a gépi fordítás különböző szempontú problémáival foglalkoznak. 1959-ben az egész világon kb. 500 kutató dolgozott a gépi fordítás kérdésein, és kb. 3 millió dollárt költöttek erre a célra. Azóta - hozzávetőleges adatok szerint - ezek a számok megkétszereződtek.

Az orosz, angol, német és francia nyelven kívül a gépi fordítással foglalkozó intézményekben főleg a kínai nyelv iránt mutatkozik nagy érdeklődés /az USA több mint 10 intézete dolgozik a kínairól angolra történő fordítás kérdésein/, de a frekvenciált nyelvek közé tartozik a hindi, japán és arab is. Néhány intézetben egészen speciális nyelvet választottak, így például a pennsylvaniai egyetem a magyar, tamil és héber nyelv gépi fordításának problémáival foglalkozik.

A legtöbb intézetben elsősorban a forrásnyelv analízisével foglalkoztak és kevés figyelmet szenteltek a célnyelv szintézisének. A

forrásnyelv alakítani /morfológiai/ elemzése a legtöbb helyen már megoldottnak tekinthető. A mondattani elemzés általános megoldását még nem sikerült megtalálni. A gyakorlati fordításoknál eddig a szintaktikai elemzésre ad hoc megoldásokat alkalmaztak.

A gépi szótárak általában körülbelül 10 000 "szótövet" tartalmaznak, néhány intézetben ennél lényegesen nagyobb szótárral /100 000 szótó, pl. Kaliforniai egyetem/ dolgoznak. A szótári egységek növelésének ma már nincsenek gyakorlati akadályai, a gépi kapacitás megnövekedése miatt.

Az empirikus módszereket /próbálgatás, ad hoc megoldások stb./ a legtöbb intézményben az elméletibb kutatás szorította háttérbe. Ad hoc megoldások helyett általánosabb, elméletileg is indokolt megoldásokat keresnek. Ennek egyik módja megfelelő modellek konstruálása /Leningrádi egyetem, Cambridge-MIT, Texasi egyetem, Ohioi egyetem stb./. Világossá vált, hogy komoly elméleti előkészítés nélkül nincs sikeres gépi fordítás. Sok intézményben ezért a gyakorlati munka teljesen háttérbe szorult, másrészt olyan intézmények is bekapcsolódtak a gépi fordítási munkálatokba, amelyek jellegüknél fogva csak elméleti munkát tudnak végezni.

Az eddigi kutatások számos intézményben /elsősorban a Szovjetunióban és az USA-ban/ gyakorlatilag jól használható fordításhoz vezettek. A gépi uton nyert fordítás átnézésé utószekesztő feladata. Világossá vált ugyanis, hogy a teljesen automatizált elsőrendű minőségű fordítás /Fully automated high quality translation/ gépi uton elérhetetlen. Ennek felismerése két következménnyel járt: egyrészt a kutatások kizárólag bizonyos szakterületek szövegeinek vizsgálatára korlátozódtak, másrészt még szakaszok esetén sem törekedtek hibátlan fordításra és inkább utószekesztőre bízták a hibák kijavítását. A gépi fordítás éppen korlátainak felismerése révén vált gazdaságossá, a gép ugyanis kevesebb ideig "bajlódik" az elemzéssel és igen nagy gyorsasággal ad nyers fordítást /több gép már lehetővé teszi az 1 millió szó per másodperc fordítási-sebességet!/.

A következőkben rövid áttekintést adunk a gépi fordítással foglalkozó intézmények munkájáról és eredményeiről.

Amerikai Egyesült Államok

1. Computer Concepts Inc., Peter Toma

Különböző szakterületek glosszáriumát készítették el. A szótár tárolása lyukkártyákon történik. Fordítási algoritmusukat 35 000-es orvosi, 12 000-es atomenergiail műszótárral és 55 000-es általános szótárral próbálták ki. Az "egységesített szintézis" módszerét dolgozták ki, melynek segítségével az angolról történő fordítás minden lépését szimultán követte a következő nyelvekre történő fordítás: francia, német, olasz, portugál, spanyol. A tapasztalatok szerint sokkal nagyobb előny származik abból, hogy egyszerre fordítanak több nyelvre, mintha mindig csak egyre fordítanának. Ez főleg a rokon nyelvekre történő szimultán fordítás esetén igaz. Az elsősorban nem szakaszok esetén jelentkező többértelműség kiküszöbölésére "érte-

lem-átvivő" programot dolgoztak ki.

2. Georgetown University, L.Dostert

A szavakat lyukkártyán tárolják /kb. 10 millió szó/, melyek négy főbb szakterületet ölelnek fel: kémia, fizika, biológia, társadalomtudományok. Ezeken a szakterületeken jól használható fordítást sikerült elérni. Az angolról oroszra és oroszról angolra történő fordításon kívül az egyetem oroszról-franciára és angolról-törökre való fordításra dolgozott ki algoritmust.

3. Harvard University, A.G.Oettinger

A szintaktikai elemzéshez használt program az angol nyelvnek közel 3500 grammatikai szabályát tartalmazza és a mondatelemzést az eddigi módszereknél lényegesen gyorsabban végzi el, továbbá a szintaktikai többértelműséget is gyorsabban ismeri fel. A szótár tárolása mágnesszalagon történik, a tárolt egységek száma kb. 30 000 /matematikai és elektronikai szakszótár/, ami kb. 15 000 szónak, illetőleg 150 000 paradigmának felel meg. Minden egységhez grammatikai kódok tartoznak, amelyek közül egyesek a forrásnyelvre /oroszra/, mások a célnyelvre /angol/ vonatkoznak.

4. International Business Machines Corporation, G.W.King,
H.W.Chang, F.Y.Wang

Az orosz, francia és kínai nyelv gépi analízise és az angol nyelv gépi szintézise számára grammatikai modell készült. A kínai írás számára speciális dekódoló eljárást dolgoztak ki.

5. Massachusetts Institute of Technology, Victor H.Yngve

Az intézet a fordítás adta közvetlen problémákon kívül az emberi közlés folyamatának tanulmányozásával is foglalkozik, valamint azzal, hogy mikor lesz a fordítás hasznos és gazdaságos. Az angol, német, francia, arab és orosz nyelv esetére olyan programot dolgoztak ki, amely szótár és megfelelő grammatikai információk segítségével tetőzés szerinti mondatokat tud képezni. Különös gondot fordításnak szemantikai kérdések tanulmányozására. Külön programot dolgoztak ki a mondat-szerkezet /phrase structure/ felismerésére. A programok írására a Comit programozó nyelvet alkották meg, amely különösen alkalmas gépi fordítási feladatokra.

6. National Bureau of Standards, I.Rhodes

Orosz szövegek szintaktikai elemzésére olyan módszert dolgoztak ki, amely a mondat első szavánál kezd az elemzést, majd a másodiknál folytatja és így tovább. A program a gyakorlatban is jól működött. Speciális programot konstruáltak az angol mondatok struktúrájának felismerésére is, elsősorban a szórend-változtatás szempontjából.

7. Ohio State University, W.S-Y.Wang

Elsősorban a kínai és angol nyelv analízisével foglalkoznak. A fordítói algoritmus kidolgozásánál különös gondot fordítanak a két

nyelv grammatikáira, elsősorban a köztük lévő hasonlóságok megállapítására.

8. The Rand Corporation, D.G.Hays, Th.W.Zieve

Elsősorban a funkcionális grammatikai modell /dependency grammar/ segítségével sikerült a szintaktikai elemzés számára használható programot megalkotni. A szótári alakokhoz kapcsolódó kódok az illető szó összes funkciót tartalmazzák. A csoport több szakterület glosszáriumát is összeállította.

9. Thompson Ramo Wooldridge INC., H.P.Edmundson, P.L.Garvin, J.Mersel, C.A.Montgomery, G.Reitz, Steven B.Smith, D.R. Swanson

Kutatási programjukban az oroszról angolra és kínairól angolra történő fordítás szerepel. Előbbi esetén jól használható fordítást sikerült elérniük szakszövegek esetén. Olyan eljárás kidolgozásán fáradoznak, amely lehetővé teszi, hogy a gép maga találjon meg fordítási hibákat.

10. University of California, S.M.Lamb

Oroszról és kínairól kívánnak angolra fordítani. Olyan programmal is kísérleteznek, amely bármely nyelvről bármely nyelvre tudna fordítani a megfelelő információk cseréje mellett. A szintaktikai elemzésnél az empirikus utat követik, vagyis adott szövegek feldolgozása révén próbálnak általános következtetésre jutni.

11. University of Texas, W.P.Lehmann, E.D.Pendergraft

Automatikus programozással egyidejűleg több feladatot próbálnak megoldani: gépi fordítás, tudományos dokumentáció, gépi kivonatolás, információ tárolás és visszanyerés. A munkában három csoport vesz részt: nyelvészek, matematikusok és programozók. Az angol és német nyelv közel 120 000 konstrukciós szabályát sikerült a gép memóriájában /mágnesszalag/ rögzíteni. Ezek segítségével elvégezhető a szintaktikai analízis jelentős része.

12. University of Washington, E.Reifler

Kínai-angol algoritmust dolgoztak ki a már régebben elkészült orosz-angol algoritmus mintájára. A legfontosabb feladatnak a csoport a különböző szemantikai megfeleltetések kiderítését tekinti. A kínai nyelv esetén külön problémát jelentett a transzkripció és a megfelelő gépi szótár összeállítása.

13. Wayne State University, H.H.Josselson

A csoport oroszról angolra kíván fordítani. E problémakörön belül a következő kérdésekkel foglalkozott: orosz-angol glosszárium összeállítása matematikai szakszövegek alapján, nominális szerkezetek és külön a viszonyos szerkezetek vizsgálata, az orosz ragozási rendszer egyes osztályainak funkcionális elemzése.

Anglia

1. Cambridge Language Research Unit, M.Mastermann, D.S.Linney

Szemantikai analízis számára ugynevezett thesaurus-modellt készítettek. A modell nemcsak angol mondatok elemzésére alkalmas. Az elkészített szintaktikai elemző program legfontosabb részei: a függési viszonyban álló kapcsolatok felderítése, a szórend vizsgálata és az idők egyeztetése.

2. Birbeck College, London, A.D.Booth

A csoport legfontosabb kérdésének egy közvetítő nyelv megkonstruálását tekinti, amely redukálná az algoritmusok számát.

Belgium

Université Libre de Bruxelles, P .P.Gillis, Im.Faverge, L.S. Hirschberg

Fizikai tárgyú szöveg oroszról franciára történő fordítása az elsődleges feladat. A gyakorlati tevékenységen túl elméleti munkát is végez a csoport, ami elsősorban a francia nyelv vizsgálatára vonatkozik. A francia szintézis jó része elkészült. Együttműködnek Euratom Cetus csoportjával és a Georgetown University-vel.

Csehszlovákia

Karlova Universita, P.Novak, P.Sgall

Elektrotechnikai szövegek angolról csehre való fordítására készítettek algoritmust. A cseh szintézis részben szintén elkészült, a főnév szintézisét gépen is kipróbálták. A fordítást közvetítő nyelv segítségével tervezik a későbbiek folyamán megoldani. Egy ilyen közvetítő nyelv vázlatos terve szintén elkészült már.

Franciaország

1. Centre National de la Recherche Scientifique, A.Sestier, Paris

Szintaktikai elemzés céljaira a függőségen alapuló grammatikát használják fel; az orosz, francia és német nyelvre vonatkozóan behatóan vizsgálták az egyes szavak és szókapcsolatok funkcióit. Ugyanakkor a transzformációs grammatikát is igyekeznek felhasználni a gépi fordítás céljaira.

2. Centre National de la Recherche Scientifique, B.Vauquois, Grenoble

A csoport orosz-francia, német-francia és japán-francia fordítási algoritmusok készítésével foglalkozik. A szintaktikai elemzést a matematikai nyelvészet különböző modelljeinek /elsősorban Chomsky modelljeinek/ segítségével igyekeznek megoldani.

Japán

1. First Research Center, Defence Agency, I.Sakai

Univerzális fordítási algoritmus összeállításán fáradoznak. Ennek kipróbálását - egyelőre szűk nyelvi anyagon - meg is kezdték.

2. Kyushu University, T.Tanati, T.Kurihara, Y.Oono

Angol, német és japán szövegeket tanulmányoznak közös program konstruálása végett. A japán grammatikai szahályok tekintélyes részét sikerült már formalizálniok. A fordítás számára külön fordító-gépet konstruáltak, amely alkalmasabbnak látszik nyelvészeti vizsgálatok elvégzésére.

Német Demokratikus Köztársaság

- Humboldt Universitát, E.Agricola

Angolról németre való fordítás számára készítenek algoritmust. A német főnév-csoport strukturális analízise elkészült, az angol analízis egy része szintén.

Olaszország

1. Euratom, Cetus, P.Braffort, A.Leroy, Y.Lecerf

Holland, francia, német és olasz független analízise elkészült, a szintézisen dolgoznak jelenleg.

2. Università degli Studi di Milano, S.Ceccato

Első helyen az oroszról angolra történő fordítás algoritmusára áll a kutatási programban. Ezenkívül a következő feladatokkal foglalkoznak: angol analízis, német és olasz szintézis.

Románia

- Academia Republicii Populare Romine, Gr.Moisil

Az angol-román algoritmus elkészült. Az algoritmus tartalmazza a szórendre vonatkozó változtatásokat is és bonyolultabb mondatok elemzésére, illetőleg fordítására is alkalmas. A grammatikai többértelműség legtöbb esetét is sikerült az algoritmus segítségével kiküszöbölni.

Szovjetunió

1. Insztitút Élektroniki, Avtomatiki i Telemekhaniki Akademii Nauk Gruzinszkoj SzSszR, A.I.Éliasvili

Kizárólag a gruz nyelv gépi analízisével foglalkoznak. Ennek segítségével azonban olyan általános következtetésekre is igyekeznek jutni, amelyek érvényesek a forrásnyelvi analízisre általában.

2. Insztitut Jazükoznaniija Akademii Nauk SzSzsZsR, I.A.Mel'csuk

Először magyarról oroszra történő fordítás céljaira készült magyar analízis. Jelenleg orosz szintézis, illetőleg orosz független analízis készül.

3. Insztitut Matematiki Szibirszkogo Otdelenija Akademii Nauk SzSzsZsR, A.V.Gladkij, M.V.Rübakova, T.I.Sed'ko

Közvetítő nyelv segítségével szándékoznak fordítani. A közvetítő nyelv szintaxisa elkészült. Közvetítő nyelv segítségével először németről oroszra fordítanak. A német analízis egyes részei készen vannak.

4. Leningradszkij Goszudarsztvennűj Universzitet imeni A.A.Zsdanova, N.D.Andreev

A legtöbb nyelvvel /közel 30/ foglalkozó csoport. Ezekről a nyelvekről egy mesterséges nyelvre kívánnak fordítani /közvetítő nyelv/ és erről oroszra, németre, indonézre. A közvetítő nyelv grammatikája részben elkészült. A közvetítő nyelv szótára jelenleg körülbelül 10 000 szót tartalmaz. Több kísérletet végeztek közvetítő nyelvre történő fordításra /Ural 2 géppel/, melyek biztató eredményt hoztak.

5. Insztitut Tocsnoj Mehaniki i Vücsiszlitel'noj Tehniki, I.Sz. Muhin, I.K.Belszkaja, T.M.Nikolaeva

Angolról oroszra való gépi fordítással foglalkoznak elsősorban. Megfelelő glosszáriumok készültek kínai, japán és német nyelvről. Erősen empirikus beállítottságú csoport. Az orosz nyelvet tekintik közvetítő nyelvnek. A gépi fordítás céljaira speciális gépet konstruáltak.

6. Matematiceszkij Insztitut imeni Szteklova AN SzSzsZsR, A.A. Logpunov, O.Sz. Kulagina, T.N.Molosnaja

Empirikus úton indult el ez a csoport is, amikor a franciáról oroszra történő fordításra dolgozott ki algoritmust. Ma elméleti munkásságuk révén a legjobb gépi fordítói csoportok közé emelkedtek. Új alapon készült el az angol-orosz illetőleg orosz-angol fordítási algoritmus. Itt dolgozták ki gépi fordítási célokra először a programozó programok elvét.

7. Laboratorija Élektromodelirovaniija VINITI, V.A.Uszenszkij

A különböző nyelvek /elsősorban angol, francia, német, kínai/ vizsgálata alapján olyan operátor-rendszert igyekeznek kidolgozni, amely alkalmas lenne bármely nyelvről bármely nyelvre történő fordításra.

8. Pervűj Moszkovszkij Goszudarsztvennűj Pedagogiceszkij Insztitut Inosztrannűh Jazükov, I.I.Revzin

Főleg a gépi fordítás elvi problémáival foglalkoznak. Magpedig az orosz forrásnyelvként, az angol, a francia és a spanyol pedig célnyelvként szerepel.

Forrásmunkák:

1. Current Research and Development in Scientific Documentation, No.11. National Science Foundation, Nov.1962
2. PAPP Ferenc, Gépi fordítás és matematikai nyelvészet a SzU-ban, kézirat
3. Y.BAR-HILLEL, The present status of automatic translation of languages, Advances in Computers, vol.1. Academic Press, N.Y.1960 pp.91-145.

A magyarországi munkára vonatkozóan pedig lásd:

HELL-KIEFER-SIPÓCZI-VARGA: Gépi fordítás; algoritmusok orosz szövegek elemzésére, Bp. 1963. OMKEK, Időszertű Műszaki Dokumentációs Kérdések 3-4, 1963.

:
:: ::

KIEFER, F.: Translation in the scientific centers of the world

The author deals with the importance of machine translation from a practical and theoretic point of view. He tries to find the place of this new science among linguistics, cybernatics and mathematics.

The author gives a short survey of the different languages having been dealt with as yet. Most of the institutes have focused their attention to problems concerning the /morphological, syntactical and semantical/ analysis of the source language and only a few of them deal with the synthesis of the target language.

Most of the investigations may be characterized by a theoretic approach to the problems of analysis or synthesis, respectively. But there are some empirical approaches, too. Obviously, empirical methods can only be used with great success in the first step in machine translation work. A general solution for the analysis can only be given by the help of models.

Fully automatized high quality translation is unattainable. It seems suitable therefore, to make use of posteditors in machine translation.

In the second part of his paper the author gives a short survey on the work and aims of the most active machine translation groups.

..... :
 :