

A GÉPI FORDÍTÁS Néhány kérdése

Hell György

A gépi fordítás a maga újdonságával, módszereinek és eljárásának látszólagos rejtélyességével, sokszor eltulzott eredmények közléssel az általános érdeklődésnek szinte középpontjába került. A gépi fordítás története, a gépi fordítással kapcsolatos kérdéseknek akár csak áttekintő jellegű ismerete is azt mutatják, hogy felesleges ezzel a kérdéssel kapcsolatban titokzatosságról vagy tulzottan bizonyult eljárásról beszélni.

A korszerű elektronikus számológépeken végezhető matematikai és logikai műveletek a negyvenes évek vége felé adták azt a gondolatot, hogy ezekkel a gépekkel fordítani is lehetne. Az angol Warren WEAVER és A.D. BOOTH 1946-47-ben abból az elgondolásból indultak ki, hogy a gép helyettesíthetné a szótárat, és az ember helyett elvégezhetné a szókeresést. Ha a gépet megfelelő szótárral látják el /"töltik fel"/, a gép bizonyos jelrendszer alkalmazásával a fordítandó szó alapjelentését adja ki a kívánt nyelven. Mondattani elemeknek alkalmazásáról itt még nem volt szó, ennek gondolatát R.H. RICHENS vetette fel, és dolgozta ki BOOTH-tal 1948-ban.

RICHENS és BOOTH a tulajdonképpeni mondattannak a lényegtelen-ségéből indultak ki. Cikkük mottójának is SCHLEGEL egyik mondatát választották: "Liser, liser; jetez vos grammaires au feu!" /Olvasni, olvasni; tűzre a nyelvtanaitokat!/. A műszaki szövegek elemzése alapján arra az álláspontra helyezkedtek, hogy a fordítás akkor is érthető, ha csak a szavak egymásutánjára ügyelnek. Feltételezésük az volt, hogy a mondattani szabályok minden nyelvben egyformák. Ezt ugyan nem tekintették kizárólagosnak, de úgy gondolták, hogy az esetek legnagyobb részére érvényes. Elgondolásuk igazolására husz nyelvből fordítottak le egy-egy mondatot angolra csak az említett szabályok figyelembevételével. RICHENS és BOOTH tervezett gépi szótára már nemcsak szótövet tartalmaz, hanem végződéseket is. A fordításban azonban nem kapjuk vissza teljes mértékben az eredeti nyelv

szavainak nyelvtani alakját, hanem csak nyelvtani utalásokat találunk, amelyek a szerzők szerint elegendőek a szóról szóra fordított szöveg megértéséhez még akkor is, ha történetesen japán nyelvről van a szöveg lefordítva. Példaszövegeiket növény származásból vették, és olyan mondatokat választottak, amelyekben sem személynév, sem számnév nem volt. A fordítás során néhány szófajt jelentéktelennek, elhanyagolhatónak vettek /pl. a névelőt/, nem jelölték az egyesszámot, a jelenidőt és a harmadik személyt. Betűjelzéssel 16 nyelvtani utalást fűztek a fordításba, az eredeti szövegben pedig egyöntetű írásjellel választották el az összetett szavak egyes tagjait egymástól és a szótöveket a ragoktól. A fordított szövegben szó után tett betűvel jelezték pl. a többesszámot, részes esetet, infinitívuszt, múlt időt, kihagyott szót stb., így a kapott szósor összefüggése világosabbá vált. Elgondolásuk szemléltetésére álljon itt két olyan mondat, melyet RICHENS és BOOTH fordítottak le említett módszerükkel.

Wenn in ein-em gröss-eren Gebiet zwei Form-en nebeneinander leb-en, ohne sich zu vermisch-en, so gehö-r-en sie verschie-den-en Form-en-kreis-en an.

fordítása:

if in a/one d large /more/ area two form m beside one another live without self to/too mix, so belong/hear p different m form m circle m at.

+

Az apró-bogyó-ju fajták, úgy termés-mennyiség-ben, mint száraz-anyaghozam-ban, felülmul-ják a nagy-bogyó-ju fajták-at.

fordítása:

small berry variety m so crop/fruit quantity in as dry matter yield in surpass m great berry variety m.

d = részes eset; m = többes szám.

1950-ben REIFLER cikke jelentett újat a gépi fordítás fejlődésében. A fordítógép munkáját két személy munkájával akarta teljessé tenni. Az egyik személy lenne az előszerkesztő, a másik az utószerkesztő. Az első az eredeti szöveg minden belső kétértelműségét ki-szöbölné ki, /azaz megszüntetné a lexikai és grammatikai homonimiákat/, a másik szerkesztőnek az lenne a feladata, hogy a gép fordította szöveget egyértelművé, olvasmányossá, kinyomtathatóvá tegye.

A gépi fordítás megalapozottságát és megbízhatóságát mi sem bizonyítja jobban, mint az, hogy REIFLER irányzatának követői hamarosan maguk is főlöszlegesen tartották a szöveg előzetes előkészítését.

REIFLER az előszerkesztő alkalmazásával tulajdonképpen olyan problémát akart megoldani, amely a gépi fordítás kezdeti szakaszában még legyőzhetetlen akadálnak látszott. Ekkor még nem tudtak a

gépnek olyan működési menetet, programot előírni, amellyel a sokféle mondat szerkezetet és a bonyolult nyelvtani formákat le lehetett volna fordítani. A szintakszis és a morfológia leegyszerűsítését akarja elérni S.O. DODD is 1952-ben kiadott cikkével /Model English/. Azt ajánlja, hogy már a szerzők írják cikkeiket egy egyszerűbb és a gép számára sokkal könnyebben hozzáférhető és lefordítható nyelven. Az érdekesség kedvéért érdemes megemlíteni DODD három legfontosabb szabályát:

1. legyen egységes szórend;
2. legyen egyetlen jelentése minden szónak;
3. legyen egyetlen alakja minden szónak.

Cikkét DODD már ilyen "angol" nyelven írta, és szerinte kis gyakorlattal kötetlen angol szöveget ilyen angolba lehet gépelni.

A gépi fordítás kérdéseiről 1954-ig csak cikkekből és előadásokon szólt. A nagy változást tulajdonképpen 1954 hozta, amikor január 7-én Georgetownban egy IBM gépen lefordították az első szöveget.

Mielőtt ennek a kísérletnek a részleteire rátérnénk, célszerűnek látszik előbb a fordítógépek felépítéséről és működéséről néhány szót szólni. Legfontosabb tulajdonságuk talán az, hogy képesek adatokat befogadni, azt megtartani, azaz képesek emlékezni, és az adatokkal az adott utasítás-sorozatnak megfelelően képesek műveleteket elvégezni. Az utasításokat vagy a gép emlékező tárolóiba adagolják, vagy egy különleges kódolásnak megfelelő lyukasztott lapon írják elő. Az adható utasításoknak egyik lényeges vonása, hogy adott esetben a gép a lehetséges két vagy több további utasítás-sorozatból a már elért eredmények vagy a meglévő adatok alapján választani képes. Az adatokat a gépek kettes számrendszerben tárolják, és ebben a számrendszerben is dolgoznak. A gép által elvégezhető műveletek /összeadás, kivonás, szorzás stb./ közül a fordítás szempontjából az a legfontosabb, hogy a gép összehasonlítani is tud, és éppen az összehasonlítás eredményeként választ a lehetséges lépések közül. Így történik pl. a szókeresés a szótárból, egy adott szó környezetének felderítése stb. A fordításhoz a gépet természetesen elő kell készíteni. Ez abból áll, hogy a forrásnyelv és a cél nyelv szükséges morfémait /szóelemeit/ a gép emlékező egységeiben tároljuk, és az elvégzendő műveleteket vagy lyukkártyán, vagy magában a gépben rögzítjük. A fordítandó szöveg beadagolása után a gép most már automatikusan fog fordítani.

Ezek után térjünk rá az első gépi fordítás ismertetésére, amelynek a forrásnyelve az orosz, cél nyelve az angol volt. Egyes vélemények szerint azért választották az orosz nyelvet, mert gazdag alakzata könnyítést jelent a gépi fordítás számára, mások szerint gazdasági és politikai okok játszottak közre. Az irodalom mindenestre ezt az utóbbi véleményt támasztja alá. A georgetowni gép szótára 250 egy-, legfeljebb kétjelentésű szót tartalmazott. A teljes szóalakok mellett esetvégződéses és igeragok is szerepeltek, ami azt jelenti, hogy a szótár legnagyobb része szótövekből állt. A gép hat

alapvető nyelvtani szabály alkalmazásával végezte el a fordításhoz szükséges műveleteket.

Az egész munkát nagy apparátussal készítették elő, a nyelvi részt, a kódolást /a szöveg és nyelvtan 2-es számrendszerrel való visszaadását/ és a programozást külön személyek végezték.

Érdekes néhány szóban kitérni ennek az első fordításnak a részleteire a nyelvi részt előkészítő Leon DOSTERT ismertetése alapján. A szerző már cikkének elején leszögezi, hogy a gépi fordítás első-sorban nyelvészeti probléma, bár az itt szükséges szempontok eltérnek a nyelvészet hagyományos vizsgálódási szempontjaitól. A hagyományos nyelvészet inkább a tartalmi elemet nézi, és a szavak értelemszerinti összefüggését keresi, itt viszont a formai elem a lényegesebb. Ezért is történt, hogy kevés nyelvész foglalkozott a gépi fordítás problémáival, és éppen a nyelvészek közül voltak sokan, akik nem sokat vártak a gépi fordítástól. DOSTERT olyan nyelvi elemzést vett programjának alapjául, amelyből gépi programot tudott összeállítani. Munkájából kezdettől fogva kiküszöbölte az elő- és utószervező működését, és az eredeti szövegből közvetlenül jó fordításra törekedett. Programját a gépi kipróbálás előtt olyan személyeken próbálta ki, akik oroszul nem tudnak, és akiknek a fordítást a megadott utasítások alapján kellett elvégezniük. Ez az eljárás azóta is gyakorlatban van. Legfontosabb fordítási szabályai a következők voltak:

1. a szó egy az egyhez fordítható;
2. szórendszerét kell alkalmazni;
3. a szó jelentését a később jövő egység alapján meg kell változtatni;
4. a szó jelentését a megelőző egység alapján kell megváltoztatni;
5. a fordításban szükségtelen szó kihagyása;
6. a fordításhoz szükséges szó betoldása.

A georgetowni gép még lassan működött: percenként 72 számjegyet /betűt/ adagoltak be, a fordítást 5-9 másodperc múlva adta ki. Az átlagos szóhosszuságot 7-8 betűnek véve, ez kb. 10 szót jelent percenként.

Az amerikai kísérlet után egy évvel a Szovjetunióban is elvégezték az első gépi fordítást angol nyelvről oroszra. A rendelkezésre álló irodalom alapján megállapítható, hogy az előkészítő munkák, amelyek ezt a fordítást megelőzték, szintén évekkal előbb kezdődtek. Az a tény, hogy egy évvel az amerikai kísérlet után a Szovjetunióban is végeztek gépi fordítást, méghozzá a gépi fordítás szempontjából sokkal több nehézséget jelentő angol nyelvről, a szovjet nyelvtudomány igen komoly teljesítménye. Az első angol-orosz fordítást egy BESZM számológép végezte, amelyet azóta is fordítási célokra használnak. A kísérletekről D.Ju. PANOV akadémikus ismertetése alapján a következőket lehet röviden elmondani:

A fordítás szótári alapja 952 angol és 1073 /!/ orosz szó volt, a szavakat különleges matematikai szövegből vették. Az első kísérlet után a gép szókészletét és nyelvtanát rendszeresen bővítették, és jelenleg már kb. 6000-es szókincsét tartanak. A fordítások meglepően jók, s ez annál is inkább figyelmet érdemel, mert az angol nyelv meglehetősen szegényes alaktani jegyeket tartalmazó mondatait kellett az alaktanilag gazdag orosz nyelvben visszaadni.

A fordítási program általános jellegének ellenőrzésére cikkeket fordítottak a Times-ből és részleteket DICKENS Copperfield David c. regényéből, jó eredménnyel. Egy mondat az angol-orosz fordításból:

Even in cases where explicit or implicit solutions are known, it is sometimes easier to obtain a numerical solution than attempt to calculate numerical values from the known solution.

A gépi fordítás:

Dazse v szlucsa jah, gde javnue ili nejavnue resenija izvesztuju, inogda bolee legko polucsit' csiszlennoe resenie, csem putat'szja vucsiszlit' csiszlennoe znacsenie iz izvesztynogo resenija.

Ugyancsak a gyakorlatban megvalósított fordításról számoltak be 1956 februárjában a SzU Tudományos Akadémiáján O.C. KULAGINA és I.A. MELCSUK nyelvészek. Ez a fordítás franciáról oroszra történt, és a szövegeket szintén matematikai tárgyú könyvekből vették. Mind a két fordítási algoritmus jónak és használhatónak bizonyult.

1955 óta a SzU-ban a gépi fordítás, szinte azt mondhatnám, közkedvelté lett a nyelvészek között. Több kutatóintézet mellett megalakult a Gépi Fordítók Társasága, amelynek tagjai rendszeres összejöveteleken vitatják meg a gépi fordítással összefüggő nyelvészeti problémákat, előadásokat tartanak, és kiadványokat jelentetnek meg. A "Voproszi Jazikoznanija" hasábjain 1955 óta minden évben több cikk jelenik meg a gépi fordítás kérdéseiről. A szovjet nyelvészeket nemcsak a nagy nyugati nyelvek érdeklik, de foglalkoznak a keleti nyelvekkel, a kínai nyelvvel /oroszlól kínaira már fordítottak/, és ami minket különösképpen kell, hogy érdekeljen, a magyar nyelvvel is. I.A. MELCSUK 1957-ben elsőnek készítette el egy magyar-orosz gépi fordítás programot, és néhány mondatot ezzel a programmal próbaképpen le is fordított. MELCSUK azért választotta éppen a magyar nyelvet forrásnyelvül, mert ez szerinte szerkezetét tekintve ragozó nyelv, mint a török nyelvcsalád nyelvei, sok összetett szava és elvált ige-kötője van, mint a germán nyelveknek, az előljáró szerepét névutók töltik be, és ezért a speciálisan "magyar" problémák megoldása előkészíti a talajt olyan más nyelvek nehézségeinek megoldására, amelyek a magyar nyelvvel bizonyos rokonvonalat mutatnak fel. /6/ Második szempontként a fordítás megvalósításához szükséges rendkívül sok és bonyolult szórendi változtatás kidolgozását is hasznosnak látta. Eredményeit a "Problémi Kibernetiki" kiadványaiban tette közzé.

Azonkívül, hogy a SzU-ban sok nyelvre terjesztették ki a gépi fordítással kapcsolatos kutatásokat, a szovjet nyelvészek munkáját

még az is jellemzi, hogy intenzíven fáradoznak egy közvetítő jelle-
gű gépi nyelv kialakításán és szövegek gépi kivonatolásán.

Jelenleg már sok államban végeznek eredményes gépi fordítást,
és jónéhány országban folynak kutató munkák, így nálunk is.

A gépi fordítás lehetőségének gondolatát nem nyelvészek vetet-
ték fel, és azok a megoldási módok, amelyekkel eleinte a fordítást
megvalósítani akarták, annyira a nyelvészet általános szemlélete el-
len voltak, hogy hivatásos nyelvészek részéről csak elutasító nyil-
latkozatokat lehetett várni. Elismert szaktekintélyek eleve tiltá-
koztak az ellen, hogy a fordítást, ezt az igen magasrendű szellemi
tevékenységet gépies munkává próbálja valaki lesüllyeszteni. Ennek
az ellentétbornak nem is volt nehéz dolga, hisz garmadával lehetne
felhozni olyan eseteket, amelyekben a gépi fordítás tehetetlen. Az
idiomatikus kifejezések fordításának nehézsége, a többértelmű sza-
vak szövegű fordításának bonyolultsága /és sokszor sikertelensége/
mellett valóban megoldhatatlan a szó hangulati velejárójának is a
visszaadása, a széppróza minden árnyalatának kifejezése, nem is szól-
va a versfordításról.

Más színben tűnnek fel ezek a felhozott nehézségek, ha arra
gondolunk, hogy a gépi fordítás már kezdetől fogva csak bizonyos
szűkebb területről vett írott szöveg fordítását tekintette feladatá-
nak. A választás a szakmai jellegű szövegekre, ezen belül főképpen a
matematikai szövegekre esett, amelyek a "gépies", "betűhű" fordítás
számára a legalkalmasabbak. A szakmai szövegek jellegüknél fogva ke-
rültek a kétértelműséget mind szavakban, mind kifejezés-módban, és vi-
szonylag kis szókincset használnak. 1000 általános és 1000 speciális
szóból álló szótárral az irodalom adatai szerint már jó eredménnyel
lehet fordítani. A szakmai szövegnek szinte minden mondata definíció,
és így ezek már jellegüknél fogva is megkívánják, hogy kínos /gépi-
es/ pontossággal fordítsuk őket. Az ilyen szöveg szórendje állandóbb
jellegű, mondatainak szerkezete a bonyolultság mellett is egyöntetű.
A szakmai szövegeken végzett statisztikai jellegű kutatások pl. a kö-
vetezőket mutatták:

1528 mondat közül 1309 mondat tartalmazott személyragos igét,
ebből a mondatok 81 %-ában az alany az ige előtt állt, és csak 19 %
esetében az ige után. A fennmaradt 219 mondatból 195 mondatban volt
személytelen, ill. általános alany, és az 1528 mondatból csak 24 mon-
datnak volt névszó állítmánya. Egy másik számlálás adatai szerint
1129 mondatból a mondatok 88 %-ában a tárgy az ige után állt, és
csak 12 %-ban az ige előtt. 1484 esetből 1468 esetben állt a mellék-
név a főnév előtt, 1 esetben utána. /A többi esetben a melléknév ál-
lított, ill. középfoku volt./ Az igealakok száma a műszaki szöve-
gekben kevesebb, mint az általános jellegű szövegekben: a jelenide-
jű alakok közül az egyesszám 3. személy és a többesszám 1. és 3. sze-
mély fordulnak elő. A főnevekre két jellemző adat: 617 esetből két
egymásután előljáró nélkül következő főnév 93,4 %-ban birtokos kap-
csolatot adott. Ezek mind azt bizonyítják, hogy a szakmai szövegek
elsősorban alkalmasak a gépi fordításra.

A gépi fordítás elleni érvek alapja sokszor az a gondolat volt,

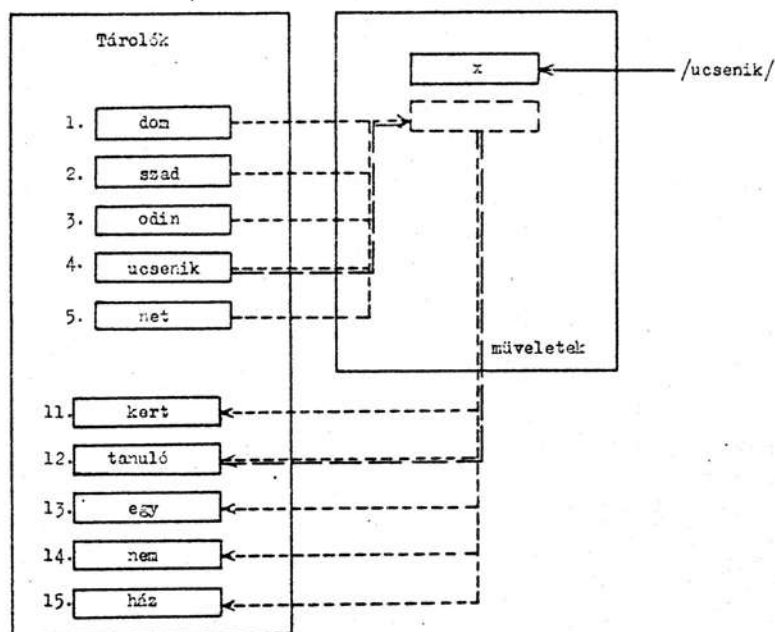
hogy nemcsak a fordítás, de a nyelv maga is alkotó jellegű tevékenység eredménye, és nem közelíthető meg gépies szabályokkal. Nem szabad azonban elfelejtenünk, hogy a nyelv a társadalmi érintkezés eszköze, mint ilyen jelrendszer, és a nyelv vizsgálatában is érvényesíthetjük a jelrendszerekre általában érvényes formális vizsgálati szempontokat. A gépi fordításban a fordítás megvalósítása főképpen a formális tulajdonságok elemzése alapján történik.

A gépi fordítás tulajdonképpen két nagy műveletcsoportból áll: az analízis és szintézis műveleteiből. Az analízis során összegyűjtjük a forrásnyelv szavain fellelhető alaktani vonásokat, hogy a szókereséssel megkapott célnyelvi szavakon a megfelelő alaktani változásokat létrehozassuk. Az utóbbi a szintézis. Mivel a nyelvtani szerkezet minden nyelvben más és más, a fordítási utasítások pedig ettől függenek, minden gépi fordítás egyszerre csak két nyelv között képzelhető el.

A szófordítás lényege abban van, hogy a gép a fordítandó szöveg minden egyes szavát megkeresi a szótárban, és a megtalált szó alapján kiadja a neki megfelelő célnyelvi jelentést.

Az eljárás lényegének bemutatására tegyük fel például, hogy a géppel az ismeretlen "ucsenik" szót akarjuk lefordítani. A fordítást most a korlátozott terjedelmű gépi szótárral hajtjuk végre, melynek orosz részében a "dom", "szad", "odin", "ucsenik", "net" szavak, a magyar részben: "kert", "tanuló", "egy", "nem", "ház" szavak vannak. Mint látjuk, a szavak "párhuzamos" elhelyezése nem fontos, viszont lényeges, hogy a megfelelő orosz és magyar szavak között olyan "kapcsolatot" teremtsék, amellyel az ismeretlen fordítandó szónak bármely orosz szóval történő eredményes összehasonlítása után a gép "adja ki" /gépeli ki/ a megfelelő magyar szót mint jelentést. Egy ilyen "kapcsolat" elektronikus gépeken megoldható, és a fordítás most már abból áll, hogy a valóban ismeretlen szót egymás után minden forrásnyelvi szótári szóval összehasonlítom olyan feltétel mellett, hogy pozitív válasz esetén érvénybe lép az említett utasításszerű kapcsolat.

Mindjárt az első pillanatban felvetődik az a kérdés, hogy milyen alakban tároljuk a gép emlékezetében a szótári egységeket. A korszerű elektronikus számológépek elkepesztő teljesítményére jellemző, hogy vannak olyan fordítógépek, amelyek szótárukban egy-egy szó minden lehetséges szóalakját tartalmazzák. Az ilyen gépek szókészlete kb. 1 000 000 szóegységet tesz ki. Ez természetesen csak különleges, un. igen gyors memóriájú gépekkel lehetséges, amelyek még ilyen hatalmas szókészletből is óránként kb. 6 oldalnyi szöveg szavait keresik ki, és fordítják le. A gyors memóriájú gépek ma nem jelentenek különlegességet, az ilyen szótározási mód alapján mégis csak egy-két kutatócsoport dolgozik. A fordítógépek szótárában általában nem teljes szavakat, hanem csak szótöveket helyeznek el, toldalékok nélkül, sőt ujabban képzők nélküli szóalakokkal is működnek gépek /ebben az esetben a képzők értelmezésére külön szabályokat kell összeállítani/. A gépi szótár szótöveit azonban nem a hagyományos nyelvészetben megismert szótövek, hanem mechanikus, un. grafikus tövek, amelyek a szó minden írott alakjában változatlanul meg-



maradnak. Míg a nyelvtani alapokat ismerő ember számára elegendő egy szó egyetlen szótővének megadása, és a tanult hangtani változások alapján pl. a "szizz" szótőben felismeri a "szidet" ige "szid" tövének hangváltozós alakját, addig a gép erre nem képes, és minden tövvariánst külön kell elhelyezni. A gépi szókeresés során az eredeti szöveg toldalékos szóalakjáról a toldalékot el kell távolítani ahhoz, hogy a szótő jelentését megkaphassuk. Ez a gépi fordítás gyakorlatában többféle módon történhet: A legegyszerűbb eljárás az, amikor a gép a szótárban meg nem talált szó végéről egyenként betűket vág le, és minden betűlevágás után összehasonlítást végez, nem talál-e a megmaradt szóalakokkal egyező tövet a szótárban. Egy másik módszer szerint a gép a hosszabb tövektől a rövidebbek felé elrendezett szótáron végigviszi a toldalékos szóalakot, és a legnagyobb egyezést mutató szótári szó lesz a keresett tö, a megmaradó rész pedig a rag. /Ilyen módszerrel dolgozik pl. a francia-orsz fordítás programja./ Az orosz-magyar fordításban célszerűbbnek bizonyul A.G. OETTINGER megoldását választani, amely szerint a szóról előbb eltávolítjuk a teljes ragot, ha van, és a megmaradt szótövet a rag elemzése után abba a szótárrészbe küldjük, ahol a rag alapján a tö várható. Ezzel a szókeresésre fordított idő erősen lecsökken.

A gépi szótárral összefüggő másik lényeges kérdés, milyen jellegű és milyen terjedelmű legyen a szótár. Mivel a fordítás szakmai szövegekkel történik, célszerű a szavakat a szakterület szótárából

venni. Ebben a szótárban nagy számban lesznek általános jellegű szavak, de lesznek a szakterülethez tartozó különleges jelentésű szavak is, és az egész szókészletet az előfordulási gyakoriság és fontosság alapján kell összeállítani. Vigyázni kell arra, hogy a szó gyakorisága nem jelenti a szavaknak a szöveg értelmezése szempontjából számított fontosságát is. Ennek szemléltetésére egy igen jellemző kísérletet végeztek.

100 000 szót tartalmazó szövegből kikeresték azt az 1000 szót, amely a teljes szókészlet 78 %-át tette. A kísérlet tulajdonképpen abban állt, hogy az így kapott 1000 szóból 750 szóval és a hiányzó 250 szóval kinyomtatták ugyanazt a szöveget. /A hiányzó szavak helyét üresen hagyták/. A két fordított szöveg közül a 250 szóval visszaadott szöveg érthetőbb volt, mint a 750 szóval visszaadott szöveg, jóllehet a 250 szó csak 3 % többletet jelentett a 750 szó %-os előfordulásához képest. /A kísérletet angol szövegen végezték./ Ez a kísérlet azt bizonyítja, hogy a gépi fordításban sem elégedhetünk meg csak a leggyakoribb szavak ismeretével és használatával.

A gépi fordítás szótározására jellemzőek a francia-orsz fordítás előkészítő munkái. Ennek során a kiválasztott matematikai szövegből átnéztek 20 500 szó terjedelmű szöveget, és ebben 2300 eltérő szót találtak. Ezekből kiválasztottak 1000 olyan szót, amelyek legalább négyszer fordultak elő a szövegben. Ehhez az 1000 szóhoz statisztikai válogatás nélkül szakszavakat és kifejezéseket tettek hozzá /kb. 50/, valamint az előljárókat, kötőszókat, névmásokat. Így egy 1100 szóból álló szótárt kaptak a fordításhoz.

A szótározás különleges problémáját és nehézségét jelenti, hogy egy-egy szónál még akkor sem elégedhetünk meg egy jelentéssel a megadásával, ha a lehető legáltalánosabb érvényű jelentést választjuk. Mindig maradnak két- és többjelentésű szavak, illetve morféma. A szótározásnak egyik nagy kérdése és teljesen még meg nem oldott problémája az, hogyan keresse ki a gép a szövegnek megfelelő jelentést. Több elgondolás létezik. Elsőnek a legegyszerűbb megoldást említem, amikor a gép a többjelentésű szó minden jelentését egymásután kiadja, és az utószekesztő feladata, hogy a megfelelőt kiválassza. A gépi fordítás első próbálkozásai ezzel a módszerrel indultak, és pl. a többjelentésű esetvégzódésekre, ragmorfémákra a fordításban 4-5 jelentést is feltüntetett a gép. Ma már különféle elemzési módokkal ki lehet küszöbölni az ilyen zavaró többértelműséget. Egy ilyen módszer a következő: a szó jelentéseit jelzésekkel látják el, annak a szakterületnek megfelelően, amelyben a szónak az adott jelentése van: A "szag" szó pl. a közbeszédben lépést jelent, a műszertechnikában "osztás"-t; a "luk" szó a mezőgazdasági szakcikkekben minden valószerűség szerint "hagymá"-t jelent, egy archaeológiai értesítésben inkább "íjj"-at. A különböző jelentések tehát a szakterületnek megfelelően kapják meg jelzéseiket, és hasonló jelzéssel látják el magát a szakcikket is. A fordítás során azt a jelentést veszi a gép, amelynek jelzése egyezik a szakcikk jelzőszámával.

Az említett esetben a gép a szó tágabb környezete alapján állapítja meg, mi a jelentése. Egyes esetekben a szó szűkebb környezetét is vizsgálhatjuk, és a szó előljárós és egyéb eseti, szófaji kapcsos-

latai alapján választhatjuk meg a helyes jelentést. /Kuda vü idete; otkuda vü idete./ Néha magának a szónak a nyelvtani alakja is eldönteti a jelentést. Pl. a mehanika és mehanik szavak a szótárban egy közös mehanik tövel fognak szerepelni. Lesznek esetek, pl.: mehanika, mehaniku, mehanike, mehaniki, amikor külön környezetvizsgálattal kell eldönteni, mit jelent a szó, de a mehanikov szóalak pl. nem lehet kétjelentésű.

Az ismertetett megoldási módok ellenére a szavak többértelműsége a gépi fordítás komoly problémája.

A szótározás kérdéséhez tartozik még az idiomatikus kifejezések fordítása. Érdekes módon ezeknek a megcsontosodott szóalakoknak a fordítására BAR-HILLEL azt a megoldást ajánlja, hogy szóról szóra fordítsuk őket, mert több nyelven végzett vizsgálatai azt mutatták, hogy a tükröszavakhoz hasonlóan igen elterjedtek a tükrö-idiómák is, /pl.: poterjat' golovu/, és szóról szóra történő fordításuk a legrosszabb esetben csak nehezen érthető.

Az orosz-magyar gépi fordításban nem ezt az elgondolást használjuk az idiomatikus kifejezések fordítására. Az orosz-magyar gépi szótár orosz részében külön jelzéssel látjuk el azokat a szavakat, amelyek idiomatikus kifejezéseket fejezhetnek be. Minden ilyen zárószó felbukkanásakor a további szótárazáshoz a gép azt az utasítást követi, hogy a fordítandó szöveg előző szavait a szótárban külön helyen tárolt kifejezésekkel hasonlítja össze. Ha a zárószó előtti rész megegyezik az idioma-gyűjteményben tárolt valamelyik kifejezéssel, akkor az ennek megfelelő jelentést kapjuk a fordításban. Ha a jelzett zárószó előtti szavak ilyen egyezést nem mutatnak, a szótározás az általános szabályok szerint fog tovább haladni.

A szótározással kapcsolatos kérdések szemléltetően mutatják, hogy a gép a szó jelentéstartalmáról nem tud, annak alapján a következő lépéssorozatot kiválasztani nem tudja. Egyedüli tájékoztatást a gépi fordítás számára a szónak formális kapcsolatai adnak.

A szótározás, a szóazonosítás a gépi fordításnak tulajdonképpen részproblémája, amely terjedelmét, valamint bonyolultságát tekintve is kevesebb gondot jelent, mint az, hogyan tudjuk a forrásnyelv szerkezetével /morfológiájával és szintaksziséval/ kifejezett tartalmat a célnyelv szerkezetével /egy lényegében tartalmilag ismeretlen szón/ pontosan visszaadni. Ezeknek a kérdéseknek a válaszára célszerűnek látszik, ha a továbbiakban az orosz-magyar gépi fordítás néhány ilyen jellegű kérdését vizsgáljuk meg részletesebben.

Ahhoz, hogy az orosz-magyar gépi fordításhoz az un. fordítási algoritmust, azaz a gép által elvégzendő műveleteket speciális jelekkel leírassuk, ezt megelőzően egy összehasonlító szerkezettől hiányában a két nyelv formális kapcsolatait kell elemezni és párhuzamba állítani. Ehhez a munkához jól fel lehet használni a két nyelv meglévő szókapcsolat /szintagma/ tanát.

A fordítás kérdése a szókapcsolatok alapján így vetődik fel:

1. Fel lehet-e ismerni az orosz nyelv szókapcsolatait a formális elemek elemzése alapján, és
2. az orosz nyelv szókapcsolatai milyen magyar szókapcsolatoknak felelnek meg?

Ha arra gondolunk, hogy az orosz nyelvnek bő alaktana van, akkor az első kérdés megoldására kielégítő támaszpontot kapunk az orosz szöveg morfológiai elemzéséből. A második kérdés tulajdonképpen nem más, mint az, hogy lehet-e az orosz nyelvű szöveget szóról szóra haladva szavanként magyarra fordítani. Nos, az orosz-magyar fordításban már kezdeti fokon is le kellett mondani az irodalomban szavankénti fordításnak nevezett módszerről. Ennek lehetetlensége nem egyszerűen a két nyelv közötti szórendi különbségből adódik, hanem abból, hogy a ragozó nyelv típusba tartozó magyar nyelv több orosz szófajt toldalékkal fejez ki. A program ezeket a szavakat "toldalék értékű szók" címen külön utasításrendszer alapján dolgozza fel. Az orosz-magyar gépi fordításban toldalékértékű szónak számítanak az előljárók, a birtokos névmások, a lehetőséget és műveltetést kifejező segédigék és a feltételezést kifejező "bü" partikula. Az ezeknek megfelelő toldalékok az ige, ill. a főnév toldalékai közé kerülnek a magyar nyelvhelyesség sorrendjének megfelelően. A toldalék értékű szavaknál nehezebben megoldható feladatot jelentenek bizonyos jelzős szókapcsolatok. Gondolnunk kell arra, hogy

1. a magyar mondatban a jelzők közül csak a ragnélküli jelzőnek van kötött szórendi helyzete /melléknévi jelző, névmási jelző, ragnélküli birtokos jelző/, a ragos birtokos jelző már pl. távolabb kerülhet a jelzett szótól, bár akkor is rendszerint előtte áll;
2. a magyar mondatban a szókapcsolatilag a ragozott igehez tartozó szavak a mondatban bárhol elhelyezkedhetnek, és mindig az igehez fognak kapcsolódni. Ez a szórendi szabadság első sorban a különböző határozóragos főnevekre vonatkozik. A magyar mondatnak ez a tulajdonsága sajátos, hiszen az indoeurópai nyelvekhez hasonlóan az oroszban is pl. egy előljárós vagy esetragos főnév nem mindig az igei állítmánnyal alkot szókapcsolatot, hanem az őt megelőző főnévvel vagy a megelőző főnevek egyikével. Ugyanakkor az oroszban az ilyen hátravetett, nem egyeztetett jelzőt adó főnév alaktanilag semmiben sem különbözik az igehez tartozó határozótól, illetve közvetett tárgyi kiegészítőtől.

Az orosz-magyar fordításban a mondottakból kifolyólag két nehézséget kell megoldani:

1. a birtokos jelzőt szórendileg a jelzett szó elé kell helyezni;
2. a nem egyeztetett hátravetett jelzőket /határozói jelzőket/, és szükség esetén szintén a jelzett szó elé kell vinni őket jelzői determinánsként.

A birtokos jelző néha egyedül, néha bővítményekkel áll a jelzett szó után, amelynek szintén lehetnek saját bővítményei. Mindez az általában jól ismert szórendi változtatást teszi szükségessé. Amíg a birtokviszonyban álló főnevek jelzői bővítmény nélkül állnak egymás után, a szórendi változtatás aránylag egyszerű módon megoldható. Más a helyzet, ha a birtokszónak és a birtokos szónak vagy szavaknak kiterjedt egyeztetett jelzője van. Az ilyen jellegű szórendi változtatásokhoz az orosz-magyar fordításba külön meg kell alkotni a birtokszó, majd a birtokos szó mellékneves, névmásos, illetve számneves jelzős szókapcsolatát, hogy a szórendcserét a teljes szókapcsolattal lehessen elvégezni.

A szövegelemzés kérdésének befejezéséül még egy kérdést szeretnék érinteni: meddig terjedhet a szövegelemzés, milyen határok között érvényes ez? A legtermészetesebb területnek már első pillantásra is a mondat látszik, s ez írásjelekkel formailag is jól felismerhetően körül van határolva. Egyértelműségről persze itt sincs szó, hisz pont állhat rövidítések után, arab számjegyek után, a képletekben szorzás jelölése, vagy faktoriális jelként pl. a felkiáltó jel. Mégis mindezek elemzése keresési szabályokkal aránylag könnyen megoldható.

A mondaton belüli egységek összetett mondat esetében a különböző alá- és mellérendelt mellékmondatok, közbevetések. Ha ezeknek elválasztása pontosvesszővel történik, úgy megbírható támaszpontot kapunk az elemzéshez, a vessző mint mondathatár azonban komoly elemzési feladatot jelent. Vesszőt használunk az orosz nyelvben pl. alárendelt és mellérendelt mondatok elválasztására, azonos mondatrészek felsorolásánál, értelmező előtt, közbevetések elkülönítésére, megszólítás után, igeneves szerkezetek, hasonlító szerkezetek, egyes kiterjedtebb szókapcsolatok elkülönítésére, ugyanakkor a vessző használatának szabályai bizonyos szubjektív alkalmazási lehetőséget hagynak az író számára, s ezek mind igen megnehezítik a gépi fordításban a vessző adott szerepének felderítését. Ezen a téren a gépi fordítás megkönnyítésére bizonyos javaslatokat is lehetne tenni az interpunkció egyértelműbbé tétele érdekében. Hogy a helyesírásnak egyébként milyen fontos szerepe van a gépi fordításban, azt talán nem is kell külön hangsúlyozni. Egységes helyesírás nélkül írott szöveg elemzése el sem képzelhető. Betűhibákat a gép igazán nem tud figyelembe venni. Egyébként ez igen figyelmes és precíz munkát kíván a gép fordításra való előkészítésekor, mert egy betűhibával beadott szó a fordítás számára elveszett szónak számít.

A gépi fordítás néhány kérdésének áttekintő jellegű ismertetése is azt mutatja, hogy a sikeres fordítás elsősorban a nyelvi kérdések eredményes megoldásán múlik. A gépi fordítás nem annyira a technika kérdése, ezen a téren inkább a nyelv formális elemzésének lehetősége fog eredményt hozni. A meglévő kutatási központokban lényegében még ma is kísérleti jelleggel folynak fordítások, gépi fordítással /elő/ készített kiadványok eddig még nem jelentek meg, bár az ezzel kapcsolatos technikai kérdések már napirenden vannak. A gazdaságossági számítások azt mutatják, hogy a gépi fordítás akkor lesz igazán kifizetődő, ha a fordítandó szöveg "bevitele" is gépi uton történik, ún. "olvasó gépekkel". Ilyen gépek építésére biztató eredménnyel folynak kísérletek.

Még néhány szót a magyarországi kutatásokról is. A gépi fordítás, elsősorban az orosz-magyar gépi fordítás nyelvészeti, műszaki és matematikai kérdéseivel nálunk is többen foglalkoznak; a kutatások jelenleg még az alapvető kérdések kidolgozására irányulnak. Ennek eredménye egy átfogó jellegű fordítási algoritmus lesz majd, amelynek "működése" gyakorlati eredményt is fog szolgáltatni. A jelenleg meglévő fordítási algoritmusok csak egy-két mondat fordítására szolgálnak, rendeltetésük pusztán demonstratív jellegű. Kielégítő eredményt kétségkívül csak alapos elméleti és gyakorlati munka hoz majd.

+

A legfontosabb irodalom:

1. W.N. LOCKE - A.D. BOOTH: Machine Translation of Languages, 1955.
2. D.Ju. PANOV: Avtomaticheskij perevod, M.1958 /angolul is/.
3. E.DELAVENDY: An Introduction to Machine Translation. London 1960.
4. A.G. OETTINGER: Automatic Language Translation, Man: 1960.
5. TARJÁN Rezső: A hazai kibernetikai kutatások M.T. 1959.
6. A "Voproszű jazükoznanija" és a "Problemü kibernetiki" folyóiratok cikkei.



HELL, Gy.: Einige Fragen der maschinellen Übersetzung

Der Gedanke, dass es möglich sei Übersetzungen maschinell zu verrichten, ist mit der Schaffung und Entwicklung der elektronischen Rechenmaschinen verbunden. Viele Linguisten lehnten die vollkommen mechanischen Übersetzung ab, obwohl die maschinellen Übersetzungen stets nur darauf abgezielt waren, geschriebene Texte je eines engeren Fachgebietes in Übersetzung vorzulegen.

Die maschinelle Übersetzung besteht aus der mechanischen Wortwahl /dessen Mechanismus auf der bedingten elektronischen Verbindung beruht/ sowie aus der aus der mittels logischer Operationen erfolgenden, vor allem auf der formalen Analyse des Textes aufgebauten Lösung der grammatischen und lexikalischen Mehrdeutigkeiten.

Die russisch-ungarische maschinelle Übersetzung bedeutet /im Grunde genommen/ das Erkennen der Wortverbindungen und deren Übersetzung von Wortart zu Wortart. Zufolge der typologischen Verschiedenheit der beiden Sprachen ist bei russisch-ungarischen Übersetzungen die verlässliche Lösung der Fragen der Wortfolge von erstrangiger Wichtigkeit.

□
□ □