

TARNÓCZI, L.: Ruhm und Fall der Übersetzungsmaschine

Verfasser analysiert die Ursachen des Aufschwungs und des Rückfalls der Forschungen zur Mechanisierung der Übersetzungsarbeit. Die maschinelle Übersetzung hält er von sprachlicher Sicht für unzureichend und auch im Fall ihres wider Erwarten eintretenden Erfolges für unwirtschaftlich. Er hielt es für zweckmäßiger, statt die fehlenden Maschinenoperatoren, Programm-Mathematiker, Logisten, theoretischen Philologen und terminologischen Experten auszubilden, die Übersetzungen sowohl von quantitativer als auch qualitativer Sicht so zu fördern, dass neue Übersetzer in organisierter Form ausgebildet und in Arbeit gestellt werden.



MI AZ IGAZSÁG A GÉPI FORDÍTÁS KÖRÜL
Varga Dénes

TARNÓCZI Lóránt a maga területén - mint nagy tapasztalattal rendelkező "hagyományos" fordító - igen hasznos elméleti és gyakorlati tevékenységet fejt ki /1/, s bizonyára sok értékes megfigyeléséről tud a továbbiakban is számot adni. Már kevésbé fogadható el az, hogy cikkében /2/ idézőjelbe tett "nyelvészeknek" tituláljon olyan szakembereket, akiknek tevékenységét meg sem próbálja tárgyilagosan értékelni, bár ő maga hangsúlyozottan a tárgyilagosság igényével lép fel. Még kevésbé indokolt, hogy a matematikai alapon megkonstruált modern számítógépek egyik megteremtőjének, NEUMANN Jánosnak* szemrehányást

* A cikkben eltorzítva J.R. NEWMAN szerepel, ilyen nevű jelentős matematikusról azonban szakmai körökben semmit sem tudnak.

tegyen, hogy azt a tévhitet keltette az emberekben: megtalálták a bölcsek követ. Hasonló sorsra jut Z.S. HARRIS, a kiváló nyelvész is, aki egyébként akkoriban /a 40-es években és az 50-es évek elején/ még nem is foglalkozott matematikával.

A szerző módszere egyébként is több szempontból erősen kifogásolható. Vitapartnereit természetesen mindenki maga választja meg - kisé különös azonban 1973-ban fő bizonyítási alapul egy 1959-ben megjelent népszerűsítő könyvecskét választani /3/, a számítógépek gyermekkorából származó véleményekkel és számadatokkal érvelni, s komoly véleményként idézni olyan abszolút laikus megállapításokat, mint Edmond CARY kijelentései az 50 ezres szakemberigényről.

Maga a szenvedélyesség még megbocsátható volna, ez sem menti azonban az olvasók befolyásolására törekvő túlzásokat, tényként közölt igazolatlan állításokat.

Első példa:

Állítás: /itt még feltételezés formájában/: DELAVENAY könyvét sugalmazhatták ... DELAVENAY-t csak az ipar "inspirációja" lelkesíthette ...

Bizonyítás: a könyvet franciából lefordították angolra is.

Előrevetett következtetés: itt "bunda" folyik, a "fejek" és szakemberek megvásárlása.

Második példa:

burkolt állításokkal igyekszik befolyásolni a tájékozódást kereső olvasókat:

"nálunk a gépi fordítás erőszakolását - az elleplezett egyéni érdekeken túlmenően - ... főként azzal indokolták ..."

Magyarországon nemcsak a gépi fordítás, hanem talán az egész számítógépes nyelvészet fejlesztésére összesen nem fordítottak annyit - az más kérdés, hogy helyesen-e - amennyit egy-két "maszek" műszaki fordító 5 év alatt keres. Jónéhány kutató tudományos érdeklődésből, társadalmi munkában végzett a Tudományos Akadémia égisze alatt nemzetközi viszonylatban is jelentős kutatásokat - és nemcsak az én véleményem szerint - ma lényegesen előrébb tarthatnánk a dokumentáció gépesítésében, ha az ilyen irányú kutatások fontosságát szélesebb körben hamarabb felismerik. /Vö. például a Szovjetunióban kibontakozott kutatásokkal /4./

Harmadik példa:

helyenként a hatásosság kedvéért erős túlzásokra ragadtatja magát.

DELAVENAY azt írta, hogy miért ne lehetne költészetet is géppel fordítani.

TARNÓCZI: - a gépi fordítás első /komolyan sem gondolt/ célkitűzése tehát a költészet gépi fordítása volt, először azt ejtették el;

- azután a szépirodalmi prózát;
- majd a társadalom- és embertudományi szövegeket;
- végül maradtak a szaktudományi és műszaki szövegek, ott is csak az olyanok, amelyekben sok a képlet és a nemzetközi vagy nemzetközi jellegű szó.

Ez a lefelé licitálás erős tulzás. A legelső gépi fordítási kísérletek és célkitűzések is tudományos-műszaki szövegek fordítására vonatkoztak. Mint elméleti lehetőség, természetesen foglalkoztatott több kutatót is /igy többek között több szovjet kutatót is/ az a kérdés, hogy kiterjeszhető-e, és ha igen, meddig a gépi fordítás hatóköre /5/. Másrészt semmiképpen sem könnyíti meg a gépi fordítást, ha a szövegben sok képlet és nemzetközi vagy nemzetközi jellegű szó van. Ha egy szótárban keresünk, teljesen mindegy, hogy hasonlít-e a szó a másik nyelvbéli megfelelőjéhez. A képletek, matematikai jelölések grammatikai funkciójának megállapítását éppen nehezíti az, hogy a grammatikai végződés ezeknél írásban gyakran elmarad.

Próbáljunk azonban most eltekinteni TARNÓCZI erősen kifogásolható vitamódszerétől, a szubjektív burokból kihámozni és vizsgálat alá venni a szerző koncepciójának egyes fogyatékoságait.

1. A matematika

Alapvető tévedés a számítógépet kizárólag műszaki ujitásnak tekinteni, a műszakiakat, az elektromérnököket /sőt a műszerészeket/ vonni felelősségre - még hozzá teljesen indokolatlanul. /"Hogyan juthattunk idáig, elektromérnökök, műszerészek és formalista nyelvészek jóvoltából ..."/ A tárolt programu univerzális számítógép megteremtése a matematika hatalmas eredménye volt - TURING elméletének gyakorlati igazolása - ahhoz hasonlóan, ahogy az atomenergia felszabadítása EINSTEIN sokak által kételkedéssel fogadott elméleti fizikai megállapításait realizálta. A megvalósítás számos műszaki nehézségét és eredményét senki sem akarja kétségbe vonni, az elvi korlátok megállapításakor azonban nem a műszaki fejlettség pillanatnyi állapota a döntő. Az atomenergia-felszabadítás első, laboratóriumi sikerei is bizonyára roppant gazdaságtalanok voltak - sőt még ma is csak meghatározott feltételek mellett érdemes atomerőművet építeni.

Az univerzális számítógépek létrehozásának lehetőségét - akkoriban még csak matematikusok számára felfogható módon - TURING igazolta 1936-ban /6/: mindazok a számítási eljárások, amelyek bármilyen bonyolult tennivalók formájában pontosan leírhatók, megvalósíthatók egyetlen univerzális automatával is. /A "számítás" itt a szó legtágabb értelmében értendő, tehát a logikai műveleteket, sőt valóságos fizikai folyamatok vezérlését is magában foglalja./

A matematikusok számára ez sokkal többet mond, mint a laikusoknak. A matematika platóni koncepciója /hogy ti. a matematika egyszerűen "a számok tudománya"/ a XX. században - a modern matematika kifejlődésével - végképp tarthatatlanná vált. Talán legközelebb jár az igazsághoz az N. BOURBAKI néven publikáló híres francia szerzőcsoport megállapítása: "a matematika lényege, ez a nehezen megfogható fogalom ... úgy jelenik meg, mint az objektumok közötti relációk tanulmányozása, amelyeket most már /tudatosan/ néhány sajátosságukból kiindulva vizsgálunk és írunk le..." /7/.

A számítógépek megjelenése kézzelfogható valósággá tette azt a lehetőséget, hogy a matematika módszereivel modelláljuk a legbonyolultabb folyamatokat is a tudomány, a termelés és a társadalmi élet legkülönbözőbb területein, ha több-kevesebb pontossággal ismerjük az adott jelenség belső összefüggéseit /az "objektumok" közötti "relációkat"/, működésének törvényszerűségeit.

Nagy előnyben volt e tekintetben más tudományokkal szemben a fizika, amelynek "matematizálódása" már évszázadokkal előbb lezajlott. Maga a matematika is kialakította a fizikai jelenségek ügyes leírásához szükséges formális apparátust /ha úgy tetszik, "nyelvet"/ a differenciál- és integrálszámítás, valamint a komplex számok elmélete, a mátrixszámítás stb. formájában. Össze sem hasonlítható ilyen szempontból a műszaki tudományok felkészültsége a számítógépek közvetlen alkalmazására pl. a közgazdaságtani, biológiai vagy nyelvészeti területekével. A számítógép adta lehetőség persze egyszerűen roppant mértékben meggyorsította e területek alapvető törvényszerűségeinek kutatását. A matematika és a számítógépek gyors térhódítását ezeken a területeken ugyanúgy naiv dolog lenne - még kapitalista viszonyok között is - a számítógépes ipar üzleti érdekeivel magyarázni, mint, mondjuk, a villanyvilágítás elterjedését az izzólámpagyárak manipulációira vezetni vissza. Hasonlóképpen naiv dolog lenne pl. a közgazdaságtanban a folyamatok bonyolultságára hivatkozva tiltakozni a számítógépek alkalmazása ellen, mondván, hogy a gazdasági élet "történelmileg kialakult, s ezért régiségeket, hagyományokat hordozó, állandóan átalakuló rendszer". A matematika lényegét nem érti az, aki a számítógépekben nem lát mást, csak adatok tárolására, kezelésére, csoportosítására és konkrét számítások elvégzésére szolgáló eszközt.

2. "fordítógép"

Ez olyan terminus, ami nincs. Még akkor sem, ha utána angolul ott áll zárójelben /translation machine/. Köztudomásu ugyanis, hogy fordításra nem célgépeket készítenek, hanem az univerzális számítógépeket használják - sok egyéb mellett erre a célra is. TARNÓCZI ezt rossz néven is veszi, s felhánytorgatja, hogy miért a nyelvet akarták a géphez alkalmazni ahelyett, hogy a nyelv adottságainak megfelelő számítógép megszerkesztésével foglalkoztak volna - holott, mint láttuk, az univerzális számítógép mindazt el tudja végezni, amit egy speciális célgép, legfeljebb az gyorsabban, ügyesebben végzi.

Kimondottan helytelen szemléletet tükröz a szerző által használt gépi nyelvtan /machine-grammar/ kifejezés. Egyedül arról van szó, hogy a gép számára precízebben kell megfogalmaznunk a szabályokat, mint az értelemmel bíró ember számára, akit a pontatlanságok kevésbé zavarnak.

Hogy valóban alapvető dolgok félreértéséről van szó, az kiderül a folytatásból, amikor arra hivatkozik, hogy "a gép csupán két alapelehetőséggel operálva tud értelmezni és szerkeszteni: igen-nem". Semmivel nem csorbitjuk a gép lehetőségeit akkor, amikor kettes számrendszerben fejezzük ki a lehetőségeket. Ha egy bit nem elég, használhatunk 2, 3, 10 vagy akárhány bitet a lehetőségek megkülönböztetésére /2 bittel már 4, 3 bittel 8, 10 bittel 1024 féle lehetőséget tudunk megkülönböztetni/. Az, hogy a legtöbb számítógép kettes számrendszerben számol, nem a gép korlátoltságát jelenti, - egyszerűen annak a matematikailag triviális ténynek a megnyilvánulási formája, hogy amit bármilyen más számrendszerben ki tudunk fejezni és számítani, azt a kettes számrendszerben is ki tudjuk. Kettes számrendszert használ a gép, mert az is elég - és műszakilag könnyű realizálni.

Egyébként a logikai lehetőségek ügyes kihasználásával* roppant mértékben le lehet rövidíteni a vizsgálatokat, s például igen gyorsan működő morfológiai és szintaktikai analízist lehet végezni, sőt a szemantikai lehetőségek kiszűrésének is ez a legmegfelelőbb módja. Valahol az 50-es évek derekán kellett megállnia az események követésében annak, aki azt hiszi, hogy "a gép csak a "szókörnyezetet" /az éppen elemzett egység és előtte vagy utána két további egység/ képes - és azt is csupán alakilag - elemezni".

Az 1957 előtti éveket tekinthetjük a gépi fordítás "előtörténetének". Ekkor jelent meg CHOMSKY korszakalkotó műve, a "Syntactic structures" /8/, amely először adott matematikai értelemben elfogadható definíciókat a "nyelv" és a "grammatika" fogalmára. A CHOMSKY által definiált "context free" nyelvek jelentették az alapját az 1960-ra létrehozott Algol-60 algoritmusos nyelvnek, amely először szabványosította a számítógépek nyelvét. Fordítóprogramok tucatjai születtek az Algol-nyelvű mondatok szintaktikai elemzésére. /Nem véletlen egyébként, hogy az Algol-60 nyelv létrehozói között ott volt B. VAUQUOIS, a kiváló francia nyelvész-matematikos, aki később - munkatársaival együtt - az orosz nyelvre egészen magas szintű elemző és fordító rendszert dolgozott ki./

TARNÓCZI a szintaktikai szintig már nem követi a gépi fordítás fejlődését. Számára ott megszakad a lánc, amikor az idiómák felmerülnek /isten hozzád, béke veled, isten veled/**, Ezek idiómaszótárral már a szintaktikai elemzés előtt viszonylag könnyen kiszűrhetők - ez nem jelent különös problémát.

A jelenleg működő /!/ gépi fordítórendszerek szintaktikai szintű fordítórendszerek, néhol kisebb szemantikai javításokkal. Infor-

* Az un. logikai vektorokon végzett műveletekkel.

** Ahol természetesen a "hozzád" és "veled" szavak nem határozószók.

mációs célra már ezek a rendszerek is jól használhatók mindenféle elő- és utószerkesztő nélkül. Fő előnyük a nagy sebesség, s az a tény, hogy mindig az adott szakterületnek megfelelően biztosítható az elfogadott terminológia. /Az igények állandó változása miatt a hagyományos fordítók között sokkal nehezebben megoldható problémát jelent a szabad kapacitással rendelkező és az adott szakterület terminológiájában járatos fordítók biztosítása! / Minőség dolgában a gépek még nem versenyezhetnek az emberi forritóval, s lényeges minőségi javulás nem is várható addig, míg kellő részletességű szemantikai szótárak nem állnak rendelkezésünkre. Ez a munka azonban folyik, és az eddigi eredmények alapján minden okunk meg lehet rá, hogy optimisták legyünk /9/.

Egyedül a gépi fordítás számára valóban meggondolandó volna, hogy ilyen óriási volumenű munkát, ami a szókinccs szemantikai tényezőkre bontásán és tipikus összefüggéseik feltárásán alapul, érdemes-e elvégezni. /Hasonló dilemma előtt állhat a szerves vegyten kutatója, amikor elébe tárul a lehetőségek végtelenné tűnő gazdagsága. / Közben azonban egy újabb igénylő is jelentkezett, az informatika, amelynek nagy szüksége van jó gépi szövegelemző rendszer kidolgozására az automatikus indexelés, osztályozás, annotálás, tartalmi visszakeresés problémáinak megoldásához. A gépi feldolgozásnak ugyanis egyik legnagyobb előnye az, hogy egyetlen "input"-ból nagyon sokféle "output"-ot tud készíteni: sok különböző célra felhasználhatók a gépi szövegelemzés eredményei. Információs célra már eddig is készítettek speciális szemantikai szótárakat, ún. tezaurusokat /10/, amelyek viszonylag egészen egyszerű strukturával is már minőségi ugrást okoztak a dokumentáció gépesítésében. Egy magasabb szintű szakmai tezaurus-rendszer kidolgozása alapját jelenthetné a szakmai terminológia felülbírálásának, kiegészítésének, tisztításának és egységesítésének is.

Igaz, ezek a célok már messzebb vezetnek a gépi fordítástól. De vajon szemrehányás illeti-e a geometriát azért, mert nem csak földméréssel foglalkozik?

Állandóan visszatérő motívum TARNÓCZI érvelésében az, hogy két adott nyelv között egy meghatározott irányban /"A-B viszonylatban"/ kidolgozott fordítóprogram más nyelvekre nem alkalmazható, sőt meg sem fordítható. Ez persze emberi fordítók esetében is így van - ott sem tud mindenki két adott nyelv között sem ide és oda fordítani.

A gép esetében azonban - éppen a problémák általánosabb megfogalmazása következtében - sokkal több a felhasználható, egy-egy adott nyelv vagy szakterület konkrét sajátosságaitól független elem, mint az ember nyelvtanulásában. /Bár ott is lényegesen könnyebb az újabb nyelvek elsajátítása, mint az első idegen nyelv. / A matematikailag képzett gépi fordítók olyan megoldásokat alkalmaztak, amelyekben teljesen külön van választva például a szintaktikai elemzésre használt algoritmus /ennek kidolgozása - természetesen a nyelvek formai követelményeinek eleget tevő vagy azt jól megközelítő modell alapján - a matematikus feladata/ és az algoritmus "paramétereit", táblázatait jelentő nyelvi szabályrendszer /ennek elkészítése - a nyelvészet feladata/. Magam is készítettem ilyen szintaktikai elemző algoritmust, amely a megfelelő szintaktikai szabályok beolvasása alapján az adott szabályrendszernek eleget tevő természetes vagy mesterséges nyelv tetszőleges elemzendő mondatában megállapítja a szavak mondat-

tá fűzésének valamennyi szintaktikailag lehetséges változatát.

E szabályrendszernek természetesen sokkal differenciáltabb szó-faji kategóriákból kell felépülnie, mint az a hagyományos nyelvészetben szokásos. A gépi fordítás kétségtelenül az eddigieknél nagyobb és szigorubb követelményeket támaszt a nyelvészettel szemben már a szintaktikai szinten is: az egész nyelv formális apparátusának egységes szemléletű leírását. Olcsó megoldás a feladatok elől az idiomákra vagy a szemantika nehézségeire való hivatkozással kitérni. Igazán tudománnyá a nyelvtudomány akkor válik, ha urrá lesz ezeken a nehézségeken, nem akkor, ha kitér előlük. Ahogyan a kémia is nekilátott a vizsgált jelenségek pontos leírásának, törvényszerűségeik felderítésének, pedig ott is vannak a szemantikával analóg "jelentéstani" problémák: a különböző összetételű vegyületek fizikai, gyógyászati stb. viselkedése.

TARNÓCZI számára azért tűnhet mindez "humorosnak", mert ő mind-ebben csak műszaki és programozói problémát lát, s el sem tudja képzelni, hogy valaki tudományos érdeklődésből foglalkozzon egy újonnan felvetődő /vagy új módon felvetődő/ problémával. A feltűnő éppen az, hogy milyen sok élvonalbeli nyelvésznek és matematikusnak a pályája kapcsolódott össze ilyen vagy olyan módon a gépi fordítással CHOMSKY-tól MELCSUK-ig, illetve BAR-HILLEL-től VAUQUOIS-ig. /Hogy csak néhány további nevet említsünk: APRESZJAN, BORSCSEV, CEJTYIN, DÖMÖLKI, FITYIALOV, FRUMKINA, GARVIN, GLADKIJ, HAYS, IVANOV, JOSHI, KALMÁR, KALUZYNYIN, KAY, KULAGINA, KUNO, LECERF, LEES, MÁRCUS, MÁRTYENJANOV, MATTHEWS, OETTINGER, PADUCSEVA, PAPP, PLATH, REFORMATSZKIJ, REVZIN, SGALL, SREJDER, USZPENSZKIJ, ZALIZNYAK, ZSOLKOVSKIJ és sorolhatnám még sokáig./ Vagy azt gondolja talán TARNÓCZI, hogy mindezeket a "fejeket" megvásárolták vagy félrevezették? /ll/

3. A gazdaságosság

A gazdaságosság kétségtelenül döntő kérdés egy új szervezeti forma vagy műszaki ujtás bevezetésekor. TARNÓCZI azonban tulságosan szűken értelmezi a gazdaságosság kérdését, amikor

- a/ a dokumentáció gépesítésének keretéből kiszakítja a gépi fordítást;
- b/ fejlődéséből kiszakítja a problémát;
- c/ a kapitalista viszonyokat, szervezeti formákat veszi alapul.

ad a/ A fordítás gépesítése csak egyik részletkérdése a dokumentáció rohamos növekedéséből adódó feladatoknak, amely szorosan összefonódik a szöveges információk gépi feldolgozásának egyéb területeivel. "A gyorsan fejlődő és változó ismereteket ma már nem lehet a régi módon közvetíteni, hiszen pl. a velük foglalkozó szakirodalom mennyisége is egy-egy szakterületen olyan rohamosan nő, hogy még válogatva is nehezen tekinthető át. A dokumentáció különböző gépesített

tárolási, visszakeresési módszereinek bevezetése - e problémákon van hivatva segíteni." /12/

E problémát nem oldják meg az olyan elvi kijelentések, hogy "ezt a munkát egy alkalmas kétnyelvű személy az idegen nyelvű szöveg gyors átlapozása, illetve átfutása alapján ... elvégzi", főleg nem igaz az, hogy "a gép teljesítményének a tört része alatt"! A probléma az, hogy gyakran nincsenek ilyen "alkalmas kétnyelvű személyek", s igen ritkán esik egybe az, hogy valaki a nyelvet is tökéletesen bírja és a szakmában is teljesen otthonosan mozogjon. A fordító ritkán tudja érdemben eldönteni, hogy ad-e a szakembernek új információkat az általa lefordított anyag. A nyersfordításnak igenis megvan a maga funkciója, s szó sincs arról, hogy minden fordítást tökéletesre kellene csiszolgatni - erre sem lehetőség, sem szükség nincsen. Egy idejében kapott gyöngébb minőségű fordítás adott esetben ezerszer többet érhet egy elkészített műalkotásnál. S mint láttuk, nemcsak a teljes anyag lefordításáról lehet szó, hanem például a cikkek címének gépi elemzése /tartalmi elemzése/ segítségével a dokumentumok automatikus osztályozásáról, mutatók készítéséről, tartalmi visszakeresésről stb. E módszer jelentőségét nem korlátozza lényegesen az, hogy az így készült elemzés nem mindig egyértelmű és nem mindig hibátlan. /A hibák okának emberi elemzésével számuk természetesen csökkenthető, de teljesen soha nem küszöbölhető ki./

ad b/ A közölt számítások kiinduló adatai már a maguk idejében is teljesen laikus forrásból származtak, az azóta végbement fejlődés pedig pontosan az ellenkezőjét bizonyítja annak, amit a szerző igazolni akar.

"Edmond CARY megvizsgálta a fordítógéppel kapcsolatos munkaigény és kihasználhatóság kérdését" - írja TARNÓCZI egy 1956-os adataira hivatkozva. CARY - aki egyébként a kézi fordítók egyik vezető személyisége volt - "kiszámította" egy akkoriban használt első generációs IBM 701 fordító-számítógép munkaerőszükségletét. E szerint a gép teljes kapacitásának kihasználásához 12 000 kódolóra van szükség. Minthogy a programozó szakértők /a lexikográfiai és terminológiai szakértőkkel, valamint a kiszolgáló és karbantartó műszakiakkal/ további 14 000 fővel /!/ szerepelnek, itt nyilván a kézi kódolásról, vagyis a lyukasztásról van szó. Aki ismeri a lyukasztás szokásos normáit, az tudja, hogy személyenként átlag 10 000 leütés számítható óránként. 2000 n-nel számolva tehát 5 gépelt oldalnak megfelelő mennyiség. Valószínű, hogy CARY életében nem látott számítógépet, ha elhitte egy akkori első generációs számítógépről, hogy óránként 60 000 oldalt le tud fordítani, a beolvasásról és a kinyomtatásról nem is beszélve. A számítástechnikai fejlesztési program korában még a szakmailag tájékozatlan olvasók számára is tulzott ijesztgetésnek tűnnek az ilyen kijelentések: "A háttér tárolók föltöltése ... az elemzési és szerkesztési programok jelkombinációinak a rögzítése a mágneses adattárolón maga is kényes és hosszadalmas művelet. Még bonyolultabb a lefordítandó szöveg kódolása, hiszen a teljes alapszöveget betűnként és írás- vagy hiányjelenként lyukszalagon ... kell rögzíteni." /Kiemelés tőlem. V.D./

Tény, hogy az adatok géprevitelének jelenlegi módja nehézkes, s

a jövő az optikai beolvasó berendezéseké.* Igaz, egyelőre nagyon magasan tartják az árát, de hát valamikor a szemüveg is olyan drága volt, hogy csak császárok viselték ...

A költségekkel való másik büvészkedés kiindulópontja az, hogy evidensnek tartja: a fordítás céljára szolgáló gép = fordítógép, amelynek semmi más rendeltetése nem lehet. Nem kell tehát mást tenni, csak kiválasztani az egyik legnagyobb számítógépet és annak a teljes költségeit venni alapul.

Az idézett - egyébként még az előbbinél is régebbi - adatok szerint egy "fordításra alkalmas" számítógép ára az 50-es évek közepén 800 000 dollár volt. A számítógépek árai azóta abszolút értékben sem igen emelkedtek, ha pedig a teljesítményt nézzük, akkor jelentős olcsóbbodásnak lehetünk tanui. Egy akkori első generációs gép sebessége jó, ha elérte a 10 000 műveletet másodpercenként. A TARNOCZI számítáiban szereplő ICL 4-72 gép minimálisan számítva 100-szoros teljesítményű, az ára viszont csak 1,5-szerese amannak, még ha a dollár értékcsökkenését figyelmen kívül hagyjuk is. Ugyanaz a teljesítmény tehát kevesebb, mint 1/60-ad olyan költséggel állítható elő.

Az persze általánosságban kijelentve nem igaz, hogy "a tranzisztorteknika fejlődése által lehetővé vált olyan műszaki előreugrásról van szó, amely semmiféle konkrét igénynek nem felelt meg, illetve nem felel meg". Először is, a tranzisztorteknika valóban lényeges fejlődést jelentett az első generációs gépekhez képest, de a 3. és 4. generációs gépek idején anakronizmus ezt emlegetni. Másodsor, nem azért fejlesztik a szocialista országok számítástechnikai hálózatukat, mert nincs rá igény. Az igény, a szükség megvan hozzá, csak megfelelő számítástechnikai kultúrának kell kialakulnia, hogy az igényeket ki is tudjuk elégíteni. Az igaz, hogy az igénybevétel soha nem lesz teljesen egyenletes - ez azonban éppen azt jelenti, hogy külön számítógép beszerzése nélkül, vagyis jelentős beruházások nélkül biztosítható lenne a dokumentáció gépesítését célzó kísérletekhez a szükséges gépidő. Feltéve természetesen, hogy újra lehet találni olyan gárdát, amely szívügyének tekinti az elkezdett munka folytatását.

ad c/ A történeti események erősen szimplifikált beállítása az alábbi megállapítás: "e negatív szakvélemények /t.i. TAUBE és BAR-HILLEL/ alapján s a kutatások eredménytelensége láttán az Egyesült Államok kormánya megvonta a költségvetési támogatást a kutatóintézetektől és kutatócsoportoktól, s ezzel a nagy port kavart ügy tulajdonképpen le is került a napirendről..."

Először is, TAUBE ugyan derék könyvtáros, de magánvéleménye még az Egyesült Államok kormányának sem volt irányadó.

másodsor, a BAR-HILLEL-re való hivatkozásnál tudni kell, hogy ő és tanítványai nagyon sokat tettek éppen a matematikai nyelvészet és a gépi fordítás fejlesztéséért. Ő 1960-ban publikált nézeteivel

* Ilyen berendezésre egyébként magyar szabadalom is létezik.

a teljesen automatizált magasszintű fordítás lehetetlenségéről szinte ki akarta provokálni a szemantikai problémák komoly vizsgálatát. Hogy ez mennyire így van, igazolják például az alábbi, 1964-ben írt szavai: "Félek, hogy jelen tanulmány is, - csak úgy, mint jónéhány korábbi írás - destruktív jellegű /jóllehet azok a javaslatok, amelyek a számítógépekkel szembeni politika megváltoztatására vonatkoznak, építő jellegűek/ ... Lehet, hogy már a közeljövőben kiderül, hogy nincs igazam. De meg kell mondanom, hogy emiatt nem leszek nagyon szomorú. Oszintén szólva, ha magamba tekintek, meg kell vallanom, hogy igen nagy várakozás él bennem az elektronikus számítógépek használatát illetően, és nagyon örülnék, ha tudnám, hogy többre képesek, mint amennyit ez idő szerint fel tudok tétélezni róluk." /13/

1964-ig az Egyesült Államok kormánya viszonylag könnyen engedélyezett különböző kutatási támogatásokat, abból az elvből kiindulva, hogy "majd meglátjuk, melyikből lesz valami". Összehangolt kutatásokról természetesen - kapitalista viszonyok között - szó sem volt, mégis kialakult néhány komoly központ - a Harvard, az MIT és a RAND Corporation elsősorban - amelyek az elméleti kutatásokban is élen jártak. 1964-ben egy tanácsadó bizottságot alakítottak az elért eredmények vizsgálatára és a támogatás helyes szervezeti formáinak megállapítására. Ez volt az ún. ALPAC bizottság /Automatic Language Processing Advisory Committee/, amely kétéves vizsgálat alapján készítette el jelentését /14/.

E jelentésből a következők derülnek ki:

1. A National Science Foundation 1959-ben megállapodást kötött Izraellel és Lengyelországgal, majd 1960-ban Jugoszláviával, amelynek értelmében tudományos irodalmat és szabadalmakat fordítanak, kivonatokat, szemléket és bibliográfiákat készítenek, Lengyelország és Jugoszlávia a hazai, Izrael pedig főleg az orosz nyelvű tudományos irodalomból - természetesen az Egyesült Államok-beli áraknál lényegesen kedvezőbb áron /Public Law 480 Translations/. E megállapodás keretében 1959-től 1965-ig 250 000 oldal tudományos irodalom fordítása készült el /95 folyóirat-évfolyam, 374 könyv, 1004 válogatott cikk, 18 495 kivonat és 13 000 szabadalom/. /p.41-42./

2. A Bizottság megvizsgálta több működő gépi fordító és un. gépi segítséggel dolgozó /tkp. gépi szótárázást, szövegpreparációt végző/ rendszer működését, s a következő összehasonlító adatokat közli az 1965-ös állapotnak megfelelően /p.66./:

A fordítás költsége:

egy 11 millió orosz szóból álló szöveg lefordítása

hagyományos módon	440 ezer dollárba,
szerződés keretében	350 ezer dollárba,
nem hivatásos fordítókkal	240 ezer dollárba,
utószerkesztés nélküli gépi fordítással	80 ezer dollárba,

utószerkesztett gépi fordítással	400 ezer dollárba,
gépi segítséggel készült fordítással	310 ezer dollárba kerül.

A fordítás minősége

a hagyományos fordításnál	jó
a szerződéses fordításnál	jó közepes
a nem hivatásos fordítók esetében	közepes
az utószerkesztés nélküli gépi fordításnál	nem kielégítő
utószerkesztett gépi fordításnál	közepes
a gépi segítséggel készült fordításnál	kitűnő

E vizsgálatok alapján a Bizottság az adott helyzetben elsősorban a Luxemburgban és Nyugat-Németországban működő, gépi segítséggel történő fordítást, valamint a nem hivatásos fordítók által készített /bár csak közepes minőségű/ fordítást propagálja.

A gépi fordítással készült szövegek hibáinak elemzése alapján a következő eredményre jutottak /az adatok a lefordított szöveg százalékosában/:

lyukasztási hiba	0 - 1%
több jelentés megadása	6 - 11%
szórendi hiba	1 - 2%
különböző beszúrások és javítások /főleg névelőhiba/	3 - 7% /egy esetben 29%/

/Azóta lényeges változás elsősorban a "nyers" gépi fordítás minőségének javulásában történt. A lexikográfia gépesítésében elért eredmények különösen a több jelentésű szavak megfelelő jelentésének kiválasztásában voltak sikerrel alkalmazhatók. Megbízhatónak látszó értesülések szerint jelenleg az Atomenergia Bizottság /AEC/ az üzemszerű gépi fordítás legfőbb felhasználója az Egyesült Államokban, amely az 1966-os jelentés szerint még jórészt a nem hivatásos fordítók szolgálatait vette igénybe./

3. Tíz év alatt az Egyesült Államok kormánya kb. annyit költött a gépi fordítás és a belőle kinőtt gépi nyelvészet fejlesztésére, amennyi az államilag finanszírozott fordítások 1 évi költsége. A támogatás eredményét a Bizottság a következőkben látja /p.29-31./.

a/ Lényegesen hozzájárult ez a tevékenység a gép software programtechnika és programrendszerek/ fejlődéséhez.

b/ Olyan forradalmi hatással volt a nyelvészet fejlődésére, amely csak a részecske-fizika felfedezéseire hasonlítható - "a gépi nyelvészet eszközei azonban lényegesen olcsóbbak, mint a részecske-fizika több-billió voltos részecske-gyorsító". "Egy életnek a munkája egy generációval ezelőtt ... az első kis lépésnek felel meg ma, amit néhány hét alatt /a jövőben pár nap alatt/ meg tudunk tenni abból a sok 10 000 lépésből, amely a természetes nyelvnek mint az emberi kommunikáció eszközeinek megértéséhez vezet."

4. Mindezek alapján a támogatás kereteinek megváltoztatását javasolja:

a/ Alaputatásokat kell folytatni a nyelvi adatok kezeléséhez a megfelelő gépi módszerek kidolgozása területén;

b/ Módszertani kutatásokat kell folytatni a komplex nyelvészeti elméletek /grammatikák, jelentésemleletek/ részletes kidolgozásának és kipróbálásának elősegítése érdekében.

A támogatás addigi és azutáni mértékére vonatkozólag a következő adatok találhatóak a jelentésben:

- az addigi átlagos évi 2 millió dollár helyett /p.107-112./ évi 2,5 - 3 millió dollár /1/ látszik reálisnak /a "Tudomány és Közérdek" Bizottság elnökének levele a Nemzeti Tudományos Akadémia elnökéhez a kötet elején/.

A válogatás nélküli és jórészt közvetlen gyakorlati célra fordított kutatómegosztásról tehát áttértek a hosszabb távú, jobban elméleti alapokon nyugvó alaputatások finanszírozására. Más szóval: a "gépi fordítás" név alatt futó, de gyakorlatilag már nagyon sokfelé differenciálódott kutatás támogatásáról a gépi fordítás érdekeit is szolgáló "gépi nyelvészet" /computational linguistics/ támogatására.

Mély igazság van abban a mondásban, hogy "a ma fizikája - a holnap technikája". De ma már ez a mondás nem csak a fizikára érvényes. A nyelvtudomány jelentőségének növekedését napjainkban éppen az okozta, hogy alkalmazásai révén beigazolódott: fejlődése egyre közvetlenebbül érinti a termelőerők fejlődését.

A szocialista tudománypolitika lehetőségei lényegesen kedvezőbbek a népgazdaság érdekeit szolgáló tudományágak tervszerű fejlesztésére. A távlati igényeknek a napi konjunkturától független objektív felmérése lehetővé teszi a rendelkezésre álló eszközök gazdaságosabb felhasználását, összehangoltabb szervezeti formák kialakítását.

Közvetlen aktualitást a nyelvészet gépi alkalmazásainak az egész népgazdaság érdekeit szolgáló információs rendszer kialakításának na-

* Összehasonlításul: ez az összeg kb. annyi, mint az Egyesült Államok egy napi hadikiadásának egy század része.

pirenden lévő feladata ad, ahol számításba kell vennünk nemcsak az adatszerű, hanem a szöveges információk tárolását, visszakeresését is. Az olyan agnosztikus kijelentések, hogy "az adott műszaki lehetőség ... a nyelv jellegétől idegen" stb. nem segítik elő a közvetlenül előttünk álló feladatok megoldását.

4. A kutatások

Néhány adalék a TARNOCZI állításai szerint a 60-as évek elején lassacskán kimúlt kutatásokról. Szerinte a Mechanical Translation c. folyóirat "1954-től 1959-ig egyre csökkenő érdeklődés mellett jelent meg s 1960-ban szűnt meg". Az 1968-ig /!/ megjelent számokhoz bárki hozzájuthat Budapesten is a Magyar Tudományos Akadémia könyvtárában. Több ízben emlegeti az 1961-es teddingtoni konferenciát, hol azért, mert "csak" 1970 résztvevője volt, hol mert a Szovjetunióból csak KULAGINA vett rajta részt. Nem szólva most arról, hogy elég veszélyes vállalkozás konferenciák, kongresszusok hatékonyságát a résztvevők számával mérni, szeretném figyelmébe ajánlani a Szovjetunióban Jerevánban rendezett össz-szövetségi és nemzetközi gépi fordítói konferenciákat, ahol viszont - a dolog természetéből kifolyólag - amerikai résztvevő volt lényegesen kevesebb. Szerinte 1964 után "az ügy tulajdonképpen le is került a napirendről". Az 1965 óta kétévenként rendszeresen 100-150 fővel megrendezett nemzetközi ICCL /International Conference on Computational Linguistics /konferenciák/ 1965. New York, 1967. Grenoble, 1969. Stockholm, 1971. Debrecen és legközelebb 1973. Pisa/ nem ezt mutatják. Hogy csak néhány számmal érzékeltessem a jelenleg folyó kutatásokat: a Referativnűj Zsurnal "Informatika" c. sorozata, amely külön rovatot tart fenn a tudományos fordítás témakörnek, e rovaton belül pl. 1972. augusztusában 3 cikkről referál a "Hagyományos fordítás", 32-ről a "Gépi fordítás" alrovatban. De hivatkozhatnk a Center for Applied Linguistics "Language and Automation" c. nemzetközi referáló kiadványára is, amely évente átlag 1400-1500 gépi nyelvészeti vonatkozású cikket ismertet.

Lehet, hogy nem csak a dolgok tehetetlensége mozgatja tovább a kutatásokat?



IRODALOM

- /1/ TARNÓCZI Lóránt: Fordítókalauz. A szakirodalmi fordítás elmélete és gyakorlata. Budapest, 1966. Közgazdasági és Jogi Könyvkiadó. 529 p.
- /2/ TARNÓCZI Lóránt: A fordítógép tündöklése és bukása. Kézirat, 1972.
- /3/ DELAVENAY, E.: La machine à traduire. Paris. Presses Univ. de France, 1959. 126 p.
- /4/ Magyar nyelvű áttekintés:
PAPP Ferenc: Matematikai nyelvészet és gépi fordítás a Szovjetunióban. Budapest, 1964. Országos Műszaki Könyvtár és Dokumentációs Központ. 222 p. /O.Sz. KULAGINA és I.A. MELCSUK tanulmányával./
A legfrissebb publikációk a folyó munkákról a Naucno-Tehnicoszka Informacija, Szerija 2. folyóiratban találhatók,
a legrészletesebb elemzések a VINITI, a Szovjet Tudományos Akadémia Orosz Intézetének /Insztitut Ruszszkogo Jazúka/ és az Idegen Nyelvek Főiskolájának /MGPIIJa/ kiadványaiban.
- /5/ Mélyrehatóan elemzi a kérdést például IVANOV, V.V.:/Lingviszticeszkie voproszú sztihotvornogo perevoda. In: Masinnúj perevod. Moszkva, 1961. p.369-395.
- /6/ TURING, A.M.: On computable numbers with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, Ser.2. vol.42. /1936/, 230-265.p.
- /7/ BOURBAKI, N.: Théorie des ensembles, 3. Paris, Herman, 1958.
- /8/ CHOMSKY, N.: Syntactic structures, 's - Gravenhage, 1957, Mouton and Co.
- /9/ Néhány irodalmi utalás:
- a/ LEONT'EVA, N.N.: O szozdani informationnogo jazúka na baze szemanticesz-kogo analiza teksztva. M. Insztitut Ruszszkogo Jazúka AN SZSZSZR, 1970.
- b/ KULAGINA, O.F. - MEL'CSUK, I.A. - ÉRASZTOV, K.O.: Ob odnoj vozmozsnój szisz-teme masinnogo perevoda. M. Insztitut Ruszszkogo Jazúka AN SZSZSZR, 1971.
- c/ ZSOLKOVSKIJ, A.K. - SCSEGLÓV, Ju.K.: K opiszaniju szmúszla szvjaznogo tekszta. M. Insztitut Ruszszkogo Jazúka AN SZSZSZR, 1971.
- d/ MEL'CSUK, I.K.: Urovní predstavlénija vúszkazúvanij i obscsee sztiroenie modeli "szmúszl - tekszt". M. Insztitut Ruszszkogo Jazúka AN SZSZSZR 1972.

/10/ Lásd például:

VARGA Dénes: Információs tezauszok készítésének módszertana, Budapest, 1969. Országos Műszaki Könyvtár és Dokumentációs Központ.

/11/ Az 1954-es IBM kísérlet kimerítő leírása egyébként rendelkezésre áll a kísérlet tudományos végrehajtójának, Paul I. GARVIN-nak a tollából, s ez nem látszik igazolni TARNÓCZI állításait: The Georgetown - IBM Experiment of 1954: An evaluation in retrospect, in: Papers in linguistics in honor of Leon E. DOSTERT. - William M. AUSTIN ed. MOUTON Co., The Hague, Netherlands, 1967. p.46-56. - Ujranymtatva in: On Machine translation. Selected papers by Paul L. GARVIN. MOUTON Co., The Hague Netherland, 1972. /1/, p.51-64.

/12/ DUZS János: Az audio-vizuális technika országosan összehangolt fejlesztése. Audio-vizuális technikai és módszertani közlemények, 1.évf. 2.sz. p.6.

/13/ BAR-HILLEL, Y.: Theoretical aspects of the mechanization of literature searching. Language and Information, Jerusalem, 1964. 356-364.p.

/14/ Language and Machines, Computers in Translation and Linguistics. A Report by the Automatic Language Processing Advisory Committee, National Academy of Sciences, National Research Council. Publication 1416. Washington, D.C., 1966.

*

VARGA, D.: What is the truth about machine translation?

In comparison with the conclusions of L. TARNÓCZI concerning the totally hopeless perspectives of machine translation /MT/, practical results of research in the field of MT and perspectives of further projects are presented. The author explains the importance of modelling in recognition of complex processes, with special regard to the development of computational linguistics. The importance of mechanized text analysis for documentation is emphasized. Economic problems are discussed in details, determining the role and importance of current operating /substantially syntactical/ systems of MT and of long-range development plans.

*

VARGA, D.: Как обстоят в действительности дела автоматического перевода?

Автор сопоставляет действительные результаты исследований и перспективы дальнейших изысканий в области автоматического перевода с заявлениями Лоранта ТАРНОЦИ о полной бесперспективности автоматического перевода. Излагается роль моделирования в познании сложных процессов, в особенности в отношении развития машинной лингвистики. Подчеркивается важность машинного анализа текстов с точки зрения механизации информационных процессов. Подробно рассматриваются вопросы экономической эффективности и определяются роль и значение уже действующих (преимущественно синтаксических по уровню) систем машинного перевода и долгосрочных перспективных разработок (главным образом семантических по направлению), которые ведутся в настоящее время.

*

VARGA, D.: Was ist die Wahrheit über die maschinelle Übersetzung?

Verfasser stellt den Äusserungen über die totale Zukunftslosigkeit der maschinellen Übersetzung des Autors L. TARNÓCZI die tatsächlich erzielten Ergebnisse der einschlägigen Forschungen und die Perspektiven der weiteren Forschungsarbeit gegenüber. Er beschreibt die Rolle der Modellierung in der Erkenntnis der komplizierten Vorgänge, mit besonderer Rücksicht auf die Fortschritte in der Entwicklung der kybernetischen Sprachwissenschaft. Er betont die Wichtigkeit der Textanalyse vom Gesichtspunkt der Mechanisierung der Dokumentation. Die Aspekte der Wirtschaftlichkeit sind eingehend erörtert, wobei die gegenwärtig funktionierenden /im wesentlichen syntaktischen/ maschinellen Übersetzungssysteme und die bereits begonnenen, auf längere Sicht geltenden /hauptsächlich semantisch ausgerichteten/ Entwicklungspläne auf ihre Bedeutung hin geprüft werden.