

A nyugati Európa Bizottság megbízásából...

Számítógépes szövegfordítás -eszperantó közvetítéssel

Lapunk 1983/6. számában közzeltük A. P. M. Witkam holland mérnök cikkét, amelynek főcíme: „Gépi fordítás teletex rendszerekben”, alcíme pedig: „A Közös Piac sok milliót áldoz az eszperantó számítógépes hasznosítására”. Ez a cikk a tavalyi budapesti eszperantó világkongresszuson elhangzott Witkam-előadás rövidített változata volt. Az előadó a holland Buro voor Systeemontwikkeling (BSO) nevű rendszerprogramozási vállalat munkatársa. 1982-ben a nyugat-európai Közös Piac megbízásából és finanszírozásával, a BSO keretében megalakult a gépi fordítás kutatásával foglalkozó munkacsoport, Witkam irányításával.

A z Európa Bizottság — az Európai Közösség vezető szerve — megállapította, hogy a különféle elektronikusinformáció-szolgálatok hamarosan belépnek a családi otthonokba. Az ilyen „telex”, és egyéb hasonló szolgáltatásoknak az Európai Közösség minden nyelvére rendelkezésre kell állniuk. Nagyon költséges lenne azonban az információkat különböző nyelveken megfogalmazni, tárolni és javítani, továbbá minden tagállam nyelvéről az összes többire egyenként lefordítani. Ezért az Európa Bizottság egy olyan tervezet elkészítésével bízta meg a holland BSO vállalatot, melynek célja annak a kérdésnek a megvizsgálása, hogy lenne-e lehetőség az információknak csupán egy nyelven történő tárolására, illetve továbbítására. Az otthonokban azután a szöveges információk vételére alkalmas számítógépes vevőkészülékek lefordítanak a kapott információt bármely nyelvre — a gépekezelő választásának megfelelően. (Egyszerű információkról, és nem irodalmi szövegekről van szó.) A pusztán nyelvi célokon túl, a bizottság ösztönözi kívánja a magas szintű technológia kifejlesztését Európában is, hogy túlszárnyalja az ilyen irányú fejlődésben már előbbre levő Egyesült Államokat és Japánt.

Mivel a tárolónyelvről a célnyelvre való fordítás nem egy nagy, központi számítógépen történne, hanem a családi ott-

Ez a csoport az eddigi munkásságával felismerte és bebizonyította, hogy a gépi fordítást a legcélszerűbben az eszperantó hídnnyelvi szerepére alapozva lehetne megoldani. E munkának néhány fővonásáról szolt Witkam kongresszusi előadása, illetve a lapunkban közölt cikke.

Most a Közös Piac e kutatást megrendelő illetékes bizottsága, és az érdekelt tudományos közvélemény részére A. P. M. Witkam elkészítette átfogó, tanulmány-értékű jelentését az eddig végzett munkájuk eredményeiről és a további teendőkről. A könyvterjedelmű, 370 oldalas, angol nyelvű jelentés címe: „Distributed Language Translation”.

Az Eszperantó Világszövetség „Eszperanto” című, Rotterdamban megjelenő folyóirata felkérte Bruce Arne Sherwood akadémikust (az USA-beli Illinois Egyetem Számítógépes Oktatási Kutató Laboratóriumának igazgatóhelyettesét, aki egyben a fizikának és a nyelvészeti tudományoknak is professzora), hogy ismertesse és értékelje Witkam tanulmányi jelentését. Az eszperantista amerikai professzor a folyóirat részére írt cikkében különös figyelmet szentel az eszperantó nyelv bizonyos „módosításainak”, amelyeket Witkam munkacsoportja, nyelvészek és eszperantológusok közreműködésével valósított meg. Épp ezért hangsúlyozza

honokban elhelyezett kis számítógépeken, az ilyen rendszereket a BSO „osztott nyelvi fordítási” rendszereknek nevezi. A. P. M. Witkam „Distributed Language Translation” című jelentése bizonyítja, hogy ilyen rendszer megvalósítható lenne, mégpedig több szakaszban. A nagy kérdés: melyik legyen az a közbülső, úgynevezett „tároló nyelv”, amelyből a fordításokat az összes többi nyelvre elvégeznék? A válasz: a módosított eszperantó.

Az eszperantóul beszélők örömmel hallják, hogy az általuk használt nyelv egy ilyen magas szintű technológiában is sikerrel alkalmazható. Ugyanakkor feltehetően helytelenítik is, hogy módosítják a nyelvhasználat során már kialakult nyelvet. Ezzel kapcsolatban szükséges megjegyeznünk, hogy a módosított eszperantóval nem lesz dolga a nagy közönségnek. Ezt a nyelvet csak a programozók, nyelvészek és más rendszerszakértők „látják”, hiszen csak a fordítórendszeren belüli belső munkaeszköz lenne. A külső megfigyelő csak a különböző nyelvi fordításokat látná, amelyeket a módosított eszperantó segítségével ültettek át más nyelvekből — köztük feltehetően — a köznyelvi eszperantóból is.

(Folytatás a 6. oldalon)



A tanulmány címlapja

a cikkíró, hogy itt nem a beszélt eszperantó megreformálásáról van szó, hanem csupán olyan módosításokról, amelyek lehetővé teszik, hogy a nyelv alkalmassá váljon a számítógépes fordításban betöltött hídnnyelvi szerepére. Sherwood írását az alábbiakban, rövidítve közöljük.

persze nem tudom megmondani. Így kuksoltam Prizrenben november 2-án, november 25-én...

nap és három éjjel vesződött a titkos írás megfejtésével.



A. P. M. Witkam és egyik közvetlen munkatársa,
Ales Olde-Kalter nyelvész

Miért e módosítás? Miért nem használjuk a normál, beszélt esperantót? — teszik fel a kérdést a hozzá nem értők, akik gyakran azt állítják, hogy az esperantónak, mint közbülső nyelvnek az alkalmazása a gépi fordításokhoz rendkívül hasznos lehet. Ezzel kapcsolatban én mindig szkeptikus voltam, mivel véleményem szerint az esperantó is — mint minden emberi nyelv — túlságosan pontatlan ahhoz, hogy azt egy gép képes legyen elemezni. Egyszer azt írtam: azon kellene dolgozni, hogy létrehozzunk egy viszonylag pontos és egyértelmű esperantót; és hogy elvben kidolgozhatunk egy megfelelő, módosított nyelvi modellt bármelyik nemzeti nyelv alapján is.

Nos, részben igazam volt, részben tévedtem. Witkam ugyanis részletesen bebizonyította, hogy az emberi (beszélt) esperantó automatikusan nem alkalmas a számítógépes szövegfordítás céljaira. Például: az ember képes megérteni, hogy a „terhes asszonyok és gyerekek” kifejezésben csak az asszonyokra vonatkozik a „terhes” jelző. De a világról való ilyen jellegű ismeretek hiánya miatt egyszer egy számítógép a fenti kifejezést úgy fordította le angolról franciára, hogy a lefordított szövegben a gyerekeket is terheseknek lehetett volna értelmezni, a szórendnek a francia nyelv szabályai szerint kötelező megváltoztatása miatt. Witkam feltételezi, hogy néhány évtized múlva a fordítóprogramoknak már lesz olyan „mesterséges intelligenciájuk”, amely majd lehetővé teszi az ilyen kétértelműségek elkerülését. Jelenleg azonban még az esperantónál is precízebb nyelvre van szükség, hogy lehetővé váljon az automatikus fordítás abból az úgynevezett módosított esperantóból, bármilyen más nyelvre.

Witkam nem csupán körvonalazta e fentebb kifejtett követelményeket, hanem részletesen ki is dolgozta a saját céljainak megfelelő módosításokat az esperantón. Amikor annakidején feltételeztem, hogy senki sem lesz hajlandó végigcsinálni az ilyen kétértelműségekől mentes esperantó kifejlesztésével járó nagy munkát, el sem tudtam képzelni, hogy valaki három év alatt valóban elvégzi egy ilyen feladat előtervezését, majd utána megszerzi a szükséges állami, pénzügyi támogatást a fejlesztési terv megvalósításához.

Az Európa Bizottság számára készült kutatómunka egy évig tartott: ennek keretében a BSO két munkatársa foglalkozott a feladattal, bevonva a munkába tanácsadóként neves nyelvészeket, matematikusokat és esperantistákat is. A bevezetőben Witkam köszönetét fejezi ki egyebek között Heinz-Dieter Maas-nak, aki az általa kifejlesztett — és ugyancsak az esperantóra alapozott — SUSY fordítórendszer révén vált nemzetközileg ismertté; továbbá Victor Sadler, John Wells és Ivo Lapenna, neves esperantista akadémikusoknak.

Witkam, más szakemberekre is hivatkozva, rámutat arra, hogy a módosított nemzeti nyelvek használhatóságával összefüggésben több probléma merül fel, mint a módosított esperantóval kapcsolatban. Az esperantó előnye a következők: egyértelmű és egyszerű nyelvtani szabályok szerint, a szövegzödések alapján egyszerűen és gyorsan elvégezhető a mondattani elemzés, és a szógyökök között sokkal kevesebb a többjelentésű szó, mint a nemzeti nyelvek esetében. Mégis, még az esperantót is módosítani kell a gépi szövegfordítás lehetővé tétele érdekében. A módosított esperantó egyébként körülbelül csak 20 százalékban különbözik az élő, beszélt esperantótól. Egy megfelelően módosított nemzeti nyelv valószínűleg ennél sokkal nagyobb mértékben különbözne az eredetitől, tehát ezt a

nyelvi modellt sokkal nehezebb lenne kidolgozni. A beszélt esperantó ezért jó alapot nyújt a számítógép által értelmezhető nyelvhöz, bár közvetlenül ilyen célra mégsem használható.

Witkam megmagyarázza, hogy a rendszerszakembereknek gyakran ellenőrizniük kell a fordítórendszer működését. Nagy könnyebbség ilyen esetekben, ha az úgynevezett belső „kód” egy olyan olvasható emberi nyelv, mint amilyen a módosított esperantó (amiről külsőleg úgy tűnik, hogy lényegében azonos az eredeti esperantóval).

A módosított esperantó és a nemzeti nyelvek közötti szövegfordítás nem egyszerű feladat. Nem elég ugyanis csupán a szövegfeleltetések listáját elkészíteni. Például minden ígéről részletes kiegészítő információkkal kell rendelkeznie arra vonatkozóan, hogy az tárgyas, vagy nem tárgyas, tartozhat-e hozzá közvetlen tárgy stb. Ilyen típusú információhalmaz viszont általában nem található a szokásos szótárakban, ezért ezek összeállításához részletesebb leírásokra van szükségünk. Abból eredően, hogy az osztott fordítórendszerben a belső struktúra és az információ kezelése egynyelvű, nagy lehetőség van arra, hogy csökkentsük a fordítóprogram-rendszerhez elkészítendő lexikon kifejlesztéséhez szükséges emberi munkát.

A mai számítógépes fordítórendszerek nem automatikusan működnek, hanem csupán a fordító szakember munkáját könnyítik meg. Ezáltal a fordító egy óra alatt több szöveget képes lefordítani gépi támogatással, mint anélkül. Az embernek azonban, aki mind a forrás-, mind a célnyelvet ismeri, mindig ellenőriznie és javítania kell a fordítást. Az osztott fordítórendszer viszont egynyelvűek számára is használható. A módosított esperantó precizitásából következik ugyanis lehetővé válik az automatikus fordítás bármely nemzeti nyelvre. Sőt, még a bemenő információt is közölheti a rendszerrel a csak valamilyen nemzeti nyelvet beszélő, egynyelvű ember. Ime, Witkam zseniális ötlete: valaki angolul gépeli be a szöveget a rendszerbe, és az megpróbálja azt lefordítani a módosított esperantóra.

Természetesen vannak kétértelműségek is a szövegben. Például a rendszer nem tudja, hogy a „saw” azt jelenti, hogy „látta”, vagy azt, hogy „fűrészel”. Ilyenkor a fordítóprogram megkérdezi a szövegközlőt — a képernyőn keresztül, angol nyelven —, hogy a két jelentésváltozat közül melyikről van szó. A gépről erre egyszerűen válaszol (a legegyszerűbben például az 1 vagy 2 billentyűk leütésével). Egy ilyen nemzeti nyelvű dialógus után a rendszer elkészíti a módosított esperantóra történő pontos fordítást. Ezután egy „tisztítási” fázis következik, ami valamilyen próbarendszert jelent abból a célból, hogy bebizonyosodjon: a fordítás nem okoz majd semmilyen problémát a későbbi, eszerint történő további fordítások során. Végül így bizonyossá válik, hogy az információ a célnak teljesen megfelelő és egyértelmű formában van tárolva a további felhasználás érdekében.

Itt most nincs elég helyünk a módosított esperantónak nyelvészeti szempontok szerinti elemzésére. Célszerű lenne, ha az Esperantó Akadémia létrehozna egy bizottságot ennek a jelentős tervezetnek a tanulmányozására. Witkam néhány észrevétele ugyanis egészségesen befolyásolhatná a beszélt emberi esperantónak a fejlődését is.

Witkam azt mondja, hogy három független tanácsadó a rendszeréről szóló jelentést igen jónak minősítette, és ez segíteni fogja őt abban, hogy megszerezze a kutatások folytatásához szükséges anyagi eszközöket. E kutatásoknak az a közvetlen céljuk, hogy néhány év alatt kifejlesszenek egy próbarendszert. Mégis, attól kell tartani, hogy az Európai Közösségben jelenleg dúló pénzügyi válság fékezni fogja ennek az igen érdekes tervezetnek a kifejlesztését.

Két technikai, kereskedelmi problémát viszont már most meg kell említenünk. A családi otthonokban elhelyezhető, szöveges információk vételére alkalmas készülékek sokkal drágábbak lesznek, mint az egynyelvű vevőkészülékek. Witkam feltételezi, hogy minden családban meglesz majd a szükséges járulékokkal ellátott képmagnó, és ez akkor erre a célra is megfelel.

A másik probléma: hogy a lefordított szöveg viszonylag lassan jelenik meg a képernyőn. Ez a sebesség jelenleg 600 szó percenként; és vannak emberek, akik ennél gyorsabban olvasnak. E sebesség növelésére már jelenleg is léteznek különféle javaslatok, de ez a probléma még megoldatlan.

A jelentés világos, jó angolsággal van megfogalmazva (még akkor is, ha néhány jelentéktelen hiba előfordul a szövegben). Francia, német, holland és dán nyelven is készült összehasonlítás az anvagról, hogy a Közös Piac tagállamainak minél több szakembere megismerje a lényegét.

BRUCE ARNE SHERWOOD
Esperantóból fordította Márkus Gábor