



INFORMATIKA A REGIONÁLIS FELSŐOKTATÁSBAN
KONFERENCIA

A konferencia fővédnöke:

Dr. Lőrincz Péter, főiskolai tanár intézetigazgató

2001. november 19.

A konferencia helyszíne:

8000. Székesfehérvár, Budai út 45. K. ép.

Pedagógiai és interdiszciplináris szekció

Elnök: Dr. Tóth Mihály kandidátus, főiskolai tanár

Titkár: Horváth Árpád főiskolai adjunktus

- Adatrejtés és az elektronikus vízjel (Dr. Tóth Mihály főiskolai tanár BMF KVK SZGTI)
- Oktatásmenedzsment és oktatásszervezés a kreditrendszer kialakításának folyamatában (Stáczér Livia főelőadó - Kodolányi János Főiskola)
- Elégedettségvizsgálat a nyugdíjreformról (Roncsekné Kontó Gizella főiskolai adjunktus Kodolányi János Főiskola)
- Minőségbiztosítás a felsőoktatásban (Monostori Endréné okl. villamos mérnök, szervező szakmérnök – Kodolányi János Főiskola)
- A tudományos terminológiarendszer emberi fordításának gépi támogatása (Dr. Virághalmy Lea főiskolai adjunktus BMF KGK Szervezési és Vezetési Intézet)
- A siker, mint tudatos tevékenység (Nagy Enikő főiskolai docens Dunaújvárosi Főiskola)
- Folyamatos fejlődés – fejlődési folyamat: a globalizáció (Gyulaffy Béláné dr. főiskolai docens Dunaújvárosi Főiskola Közgazdasági Intézet)
- Az új gazdaság értelmezésének problémái a regionális konvergencia szempontjából (Hervainé dr. Szabó Gyöngyvér főiskolai tanár Kodolányi János Főiskola)

A tudományos terminológiarendszer emberi fordításának gépi támogatása

B. Virághalmy Lea

főiskolai adjunktus

Absztrakt:

A tudományok rendszerezése, szervezése során elkerülhetetlenül terminológiai problémákba ütközünk. A tudományos lexika az adott nyelv szókészletén belül körülhatárolt, sajátos alrendszert képez, legfőbb követelménye, hogy logikus, rendszeres és egyértelmű legyen. A számítógép a fordításban részt vehet teljesen automatizált formában (FAMT), emberi segítséggel (HAMT), vagy fordítva: a gép támogathatja az emberi fordítást (MAHT). A fordítás alapvető eszköze a szótár, amelynek számítógépen rögzített változata lehet facsimile, szöveghű, kódolt vagy adatbázisos.

Kulcsszavak:

terminológia, fordítás, szótár, SGML, terminológiai adatbázis

1. A terminológia rendszere

A tudományos, gazdasági s általában mindennapi életünk globalizációs fejlődése szükségszerűen megköveteli a terminológia tervszerű, módszeres finomítását. A tudományok rendszerezése, szervezése során elkerülhetetlenül terminológiai problémákba ütközünk, s rádöbbenünk arra, hogy ezek megoldása napjainkban a nyelvészet egyik elsőrendű feladata. A tudományos lexika az adott nyelv szókészletén belül körülhatárolt, sajátos alrendszert képez, amelynek normái szigorúbbak az általános lexika nyelvérzékünk diktálta szabályainál.

Milyen egy „ideális” műszaki, gazdasági vagy informatikai terminológia? Legfőbb követelménye, hogy logikus, rendszeres legyen – azaz minden eleme szervesen illeszkedjék az adott szakterület terminológiájába. Egy-egy terminus technicus nem önmagában, hanem más terminusokhoz viszonyítva, vagy éppen azokkal szembeállítva fejezi ki valódi „mondanivalóját” – a terminus technicus szerepét tehát semmi esetre sem korlátozhatjuk csupán a megnevezésre, mivel a teljes fogalmi rendszert kell tükröznie. Sok pontatlanság éppen abból ered, hogy az ismert folyamatok, tárgyak, tulajdonságok közti valóságos kapcsolatokat nem képes az adott kifejezés elég világosan megragadni. Lehetséges tényező az is, hogy egyszerűen nem tisztázott a fogalom helye a diszciplína fogalmi rendszerében.

Az egyértelműség követelménye gyakran azért nem valósulhat meg, mert sokszor még ugyanazon szakterületi terminológián belül is többféle fogalom jelölésére ugyanazt a terminust használják (homonímia). Kevesebb bonyodalmat okoz, de feltétlenül hátráltat, ha egy és ugyanazon fogalomra eltérő műszavakat alkalmaznak a különböző tudományágakban; erre napjainkban döbbennek rá a kutatók. Évtizedes viták gyakran egy kifejezés eltérő, vagy nem megfelelő használatából születtek. A homonímák és szinonimák a tudományban nem gazdagítják, hanem szükségtelenül

terhelik a nyelvet, időről időre szükség van tehát egy-egy tudományágban a kialakult terminológia revíziójára.

Gyakori jelenség a többjelentésű terminusok használatából eredő terminológiai pontatlanság, mivel azok nem mindig fedik egymást egészen. A homonímia különleges esetének tekinthető az is, amikor a fogalmi körben nem egyértelmű a kifejezés, néha egy szerző művein belül sem. Nem sorolható ide viszont az a probléma, amikor a terminus bár többjelentésű, de az adott szakterületen egyértelmű: pl. tűzfal, egér.

A szinonímiából néha egész terminológiai dialektusok fejlődnek ki, ha nem egységesítik időben a szakszókincset. Ez kétféleképpen mehet végbe: területi, földrajzi alapon (vagyis elképzelhető, hogy a mai és a történelmi Magyarország elcsatolt területein lévő fakultásokon vagy azok különböző tanszékein „más nyelvet beszélnek”), illetve szakmai szint által meghatározott módon (azaz más kifejezéseket használnak a gimnáziumi informatika órán vagy a műszaki menedzsereknek tartott előadáson). Ide sorolhatók a kiejtés „logikátlanságai” is, pl. Partium [parcium], de Pentium [pentium].

Érdekes vadhajtás, hogy néha olyan terminusokat használunk, amelyek nyelvérzékünk szerint el-lentmondanak az általuk jelölt fogalomnak. Ilyen például a „táblázatkezelő” program, amely táblázatok helyett pénzügyi, gazdasági nyilvántartások kezelését segíti.

Az idegen szavak használata nem egyértelműen helyteleníthető. Megszokottá vált már a karakter, generátor, modem, modul és processzor kifejezés, de pl. a „fájl”, „interfész” írásbeli használata igen elgondolkodtató.

2. Számítógéppel támogatott emberi fordítás

Már az 1970-es évek legelejétől kisgyerekként édesapámmal, Virághalmy Ferencsel többször megcsodálhattam a MIKI-ben, a KFKI-ban és az OMKDK-ban működő számítógépeket. Tanúja voltam, hogyan építették ki a különböző műszaki területeken a teauruszra alapozott lyukkártyás válogató rendszert, és ekkor találkoztam a permutációs indexelés kérdéskörével is. Ez a kontextusban szerepeltetett kulcsszavak betűrendes jegyzékének előállításával végzett tartalmi feltárás már akkor sem volt új gondolat, mivel a XIII. századtól használták a Páduai Szent Antalnak tulajdonított, a Vulgáthoz írott anonim konkordációt, majd a későbbi betűrendes tárgyszókatalógusok is erre az elvre épültek.

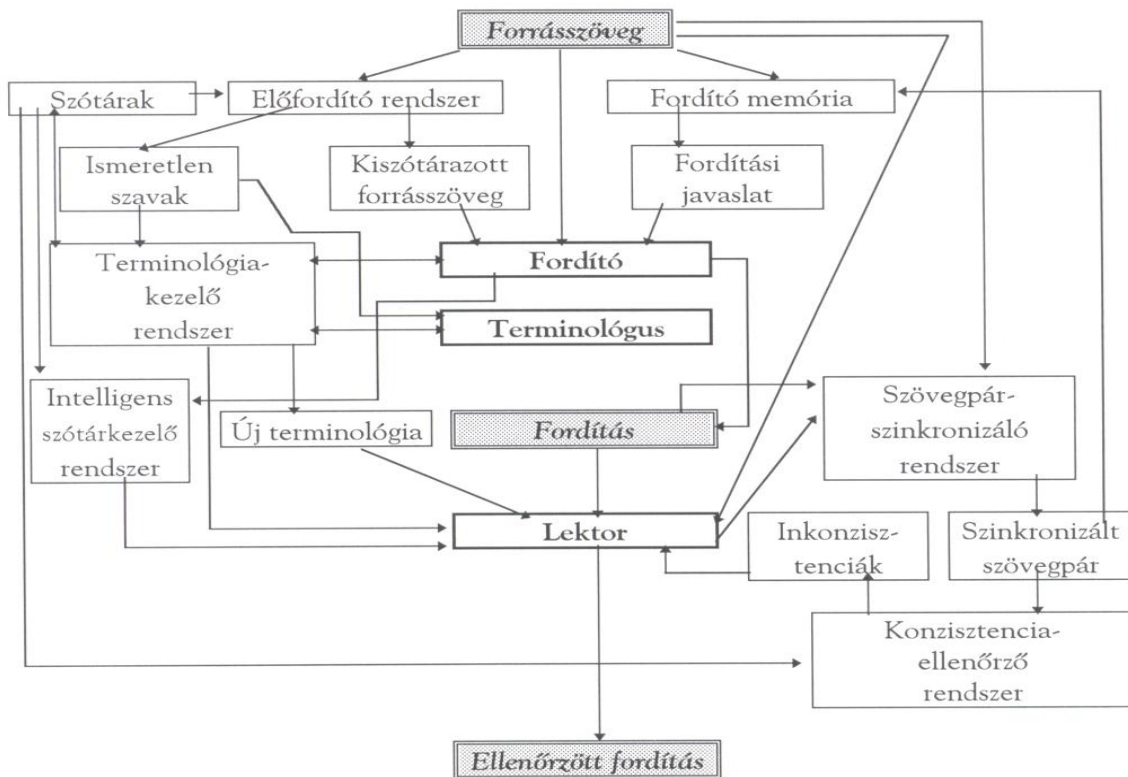
A céljának megfelelő tárgyszavazás – különösen a luhni permutációs indexelés gépi módszerével – csak akkor alkalmazható, ha egységesített a terminológia-rendszer. Ehhez a fogalmakat először rostálni, majd rendszerezni, osztályozni kell, s csak ezután lehet a megfelelő terminust kiválasztani vagy megszerkeszteni. Jelen témánkban az egységesítéshez, az információszaj csökkentésére használt szótárakat két nagy csoportra oszthatjuk: értelmező- és terminológiai szótárakra, illetve két- és többnyelvű szótárakra. Ezen kívül sokáig fontos szerep jutott az ún. teauruszoknak, vagyis egy meghatározott diszciplína tárgyszavait és azok alapvető relációit tartalmazó normatív

szótárnak. Ezen szigorúan egyértelmű tárgyszavak képezték aztán az információ-kereső rendszer mesterséges nyelvének alapszótári állományát.

Hogyan lehet egy terminust kiválasztani vagy megszerkeszteni? Napjaink felgyorsult integrációs, multinacionális világában az egyértelműség és a logika a tükörfordítást vagy a lényeg megragadását követeli. (Ilyen például az olaszban az „invio” = enter.) De mit tehetünk a már kialakult és használt szakszavakkal? Ez a legfontosabb kérdés, amit minden szakterületnek tisztázni kell; csak ezután léphet egy magasabb, az intelligens szövegkezelés szintjére.

A számítógép a fordításban többféleképpen vehet részt. Először teljesen automatizált formában, amikor a program hajtja végre a teljes fordítási folyamatot – ezt a szakirodalom FAMT-ként /fully automatic machine translation/ említi. Mindezt végezheti a gép emberi segítséggel is (ember támogatta gépi fordítás: HAMT / human aided machine translation/), vagy fordítva: a gép támogathatja az emberi fordítást (MAHT / machine aided human translation/). A FAMT lehetőségei még korlátozottak, legfeljebb szigorú, ún. kontrollált szakmai szövegek felszínes fordítására alkalmazhatók, mert „ökörként” túl következetesek egy környezetére reagáló, gondolkodó emberi agy számára. A HAMT és a MAHT eszközei fokozatosan egybeolvadnak, egyaránt igényelve az emberi interakciót.

Témánk szempontjából vizsgáljuk meg az emberi fordítás gépi támogatásának azon modelljét*, amikor csak egy-egy konkrét részfeladatot végeznek a számítógépes eszközök, és a feladat az emberre (esetünkben a fordítóra, a terminológusra és a lektorra) hárul.



* Ez a modell a PRÓSZÉKY – KIS pp.232-234 található 3 ábra átdolgozott variánsa.

Egy idealizált szituációban a fordító egy személyben terminológus is, de sajnos megváltozott világunkban a szerepek annyira elaprózódtak, a szaktudományok oly mértékben specializálódtak, hogy gondolnunk kell a modellezésnél a terminológusra külön is. Ő olyan kitüntetett jogokkal felruházott személy, aki meghatározhatja, módosíthatja, egyértelműsítheti a teauruszokat, mai elnevezéssel a terminológiai adatbázist. A lektor tisztje a minőségbiztosítás: a forrásszöveg egybevetése a fordítással. Ez a folyamat feltételezi számára a hozzáférést a fordító és a terminológus által használt összes szótárhoz, adatbázishoz, valamint a félautomata konzisztenciaellenőrző rendszerhez, amely a terminológiai következetesség figyelésén kívül észleli a fordításból hiányzó bekezdéseket, mondatokat, kifejezéseket is.

3. Szótárak és terminológiakezelő rendszerek

Miben különböznek a számítógépen rögzített szótárak a könyvben publikált elődjüktől? Tartalmukban nyilvánvalóan azonosak – de akkor miért jobbak? Rideg dolog a megszokott és szeretett lexikonoknak felállított könyvszekrény helyett egy CD-tár előtt dolgozni. Bevallom, számomra először kényszerű volt az áttérés az elektronikus eszközhasználatra, hiszen az 1912-es Kelemen-féle német szótár illatát érezni, vagy az 1858-as kis Finályval Horatiust meghallani valóban életre szóló élmény. „Non omnis moriar, multaue pars mei vitabit Libitinam.” (HORATIUS: Ad Melpomenen III.30.) Vallva ezt, mulandó barbárságnak tartom a grafikai képet tároló facsimile változatot, vagy a szöveghű verziót, amely minden visszakeresési lehetőséget szerephez juttat, mivel a gép szöveggént kezeli a szótárat.

Az igazi elektronikus szótár azonban nem kizárólag egyes lexikonok helykímélő funkcióval bíró CD-je, hanem egy minden, rokon és kevésbé rokon tudományág művelői által elérhető, releváns információt tartalmazó, központilag módosítható normatív szótárrendszer. Elérése fontos, hogy ne csak a tradicionális felhasználó, hanem a programok formális világa számára is biztosított legyen reprezentációs nyelve révén, amelynek szabványosítása már a 80-as évektől megkezdődött. Ez az általános dokumentum-leíró formalizmus az SGML (Standard Generalized Markup Language) és a szótárleíráshoz kifejlesztett TEI (Text Encoding Initiative) együttes „munkájából” született; a tartalmi szerkezetet speciális szabványos jegyek formájában jeleníti meg, a formátum-információ segítségével.

Ezzel a módszerrel egy biztos, valamennyire statikus lexikális készletet tudunk feldolgozni, ahol az elemekhez morfoszintaktikai, szemantikai, néha fonológiai leírás is kapcsolódik. A terminológiai szókincs viszont a már tisztázott okoknál fogva szinte állandó átalakulásban van, az izoláció által veszélyeztetve. Azonban, ha minden szótári egységet azonos alakra hozunk, akkor a szöveges alak alkalmassá válik adatbázis-kezelő programmal való feldolgozásra. Erre a munkára több projekt is specializálódott, ezek közül kettőt emelek ki: az ASCOT-ot, amely az amszterdami

egyetem Angol Tanszékén Akkerman, Masereeuw és Meijs kutatók nevéhez fűződik, valamint a kanadai Waterloo Egyetem munkatársai által konvertált nagy Oxford English Dictionary. Mindezekből világosan következik, hogy az Európai Unió csatlakozásunk küszöbén, egy többnyelvű közösség leendő egyenjogú tagjaként a magyar felsőoktatás kiemelt nyelvészeti, informatikai feladatai közé tartozik a logikus, egyértelmű, magyar anyanyelvünkbe szervesen beépülő, állandó frissítésre képes terminológia-rendszer kiépítése, megőrzése.

Felhasznált és ajánlott irodalom:

- AKKERMAN, E. – MASEREEUW, P. – MEIJS, W.: Defining a Computerized Lexicon for Linguistic Purposes. (In: ASCOT Report 1.) Amsterdam, 1985.
- BÁTORI S. István: Grundprobleme der Anwendungen in der Computerlinguistik (In: Computerlinguistik /Handbücher zur Sprach- und Kommunikationswissenschaft Bd.4./ Berlin – New York, 1989. pp.481-489)
- BIAL, Marcel: Beispiele zur Übersetzungsorientierten Terminologearbeit mit Hilfe des Softwarepakets Framework III.TM Saarbrücken, 1990, 165 p.
- BUDIN, Gerhard: Auszug aus „Empfehlungen für die Terminologearbeit“, Küwes 1990 (In: Theorie und Praxis der übersetzungsbezogenen Terminologearbeit /Version 1/ Wien, 1999. pp. I.1/1-6/5)
- BUDIN, Gerhard: GTW–Report „Guidelines for the Design and Implementation of Terminology Data Banks” (In: Theorie und Praxis der übersetzungsbezogenen Terminologearbeit /Version 1/ Wien, 1999. pp.III.1-35)
- BUSA, Roberto S.J.: Fondamenti di informatica linguistica. Milano, 1987. 412 p.
- CALZOLARI, Nicoletta: Computer Aided Lexicography: Dictionaries and Word Data Bases (In: Computerlinguistik /Handbücher zur Sprach- und Kommunikationswissenschaft Bd.4./ Berlin – New York, 1989. pp.510-519)
- CRYSTAL, David: A nyelv enciklopédiája. Budapest, 1998. 606 p.
- GALINSKI, Christian: Bibliographie der Terminologielehre (In: <http://starwww.uibk.ac.at/dolm/termlit.html>)
- GALINSKI, Christian: Infoterm. International Information Centre for Terminology (In: <http://www.infoterm.or.at>)
- HOHNHOLD, Ingo: Übersetzungsorientierte Terminologearbeit. Eine Grundlegung für Praktiker. Stuttgart, 1990. 223 p.
- JIMÉNEZ, Manuel Martín: Sprache, Computer und Übersetzen. Graz, 1995. 136 p.
- JONES, Randall L. – SONDRUP, Steven P.: Computer Aided Lexicography: Indexes and Concordances (In: Computerlinguistik /Handbücher zur Sprach- und Kommunikationswissenschaft Bd.4./ Berlin – New York, 1989. pp.490-509)
- MAYER, F.: Grundlagen der rechnergestützten Terminologearbeit. Modellversuch. Saarbrücken, 1992.

- MAYER, F.: Terminologieverwaltungssysteme für Übersetzer (In: Lebende Sprachen, 3/1990)
- NYOMÁRKAY István (szerk.) Európai terminológia (A Modern Filológiai Társaság és a Magyar Nyelvtudományi Társaság Modern Filológiai Szakosztályának nemzetközi konferenciája Budapesten, 1998. május hó 20-22-én)
- PAMMINGER, Heide Marie: Zur Entwicklung der maschinellen Übersetzung als Anwendungsgebiet der Computerlinguistik. Salzburg, 1993. 159 p.
- PRÓSZÉKY Gábor: Számítógépes nyelvészet. Budapest, 1989. 602 p.
- PRÓSZÉKY Gábor: Természetes nyelvek (In: Mesterséges intelligencia. Szerk. Futó Iván. Budapest, 1999. pp.756-814)
- PRÓSZÉKY Gábor – Kis Balázs: Számítógéppel – emberi nyelven. Intelligens szövegkezelés számítógéppel. Bicske, 1999. 338 p.
- SAGER, Juan C.: The Dictionary as an Aid in Terminology (In: Wörterbücher – Dictionaries – Dictionnaires. Ein internationales Handbuch zur Lexikographie. /Handbücher zur Sprach- und Kommunikationswissenschaft Bd.5.1./ Berlin – New York, 1989. pp.167-170)
- SCHMITZ, K. D.: Rechnergestützte Terminologieverwaltung am Übersetzerarbeitsplatz (In: Terminologie et traduction – Saarbrücken, 3/1990)
- SCHWANKE, M.: Maschinelle Übersetzung /Ein Überblick über Theorie und Praxis/ Berlin – Heidelberg – New York, 1991.
- TOMPA, F.: Database Design for a Dictionary of the Future. Preliminary report. Ontario, 1986.