

A fordítástechnológia és az alkalmazott nyelvtudomány

Doktori értekezés tézisei

Kis Balázs

**Pécsi Tudományegyetem, Bölcsészettudományi Kar,
Alkalmazott Nyelvészeti Doktori Iskola**

2008. február 3.

Bevezetés

A fordítás a közelmúltban jelentős paradigmaváltáson esett át. Míg korábban leginkább egyszemélyes alkotómunkának tekintették, ma már nemigen van olyan szakmai fordítási feladat, amelyet egyetlen fordító el tudna végezni. A fordítást – mint annyi más kreatív tevékenységet – csapatok végzik, s mivel a határidők is szűkebbek lettek, számos szervezési és számítógépes eszközt kellett bevezetni ahhoz, hogy időben elvégezhető legyenek.

A fordítás és a hozzá kapcsolódó technológia gazdasági jelentőségét mutatja, hogy a szakmai közösség számos konferenciát szervez. Ilyen a Localization World, Magyarországon az MFE által szervezett Szent Jeromos-napi találkozások vagy a legnagyobb nemzetközi fordítói szervezet, a Proz.com által szervezett összejövetel. A fordítástechnológiáról ugyanakkor nem született tudományos igényű irodalom. Részterületeken, a gépi fordításban, a korpusznyelvészetben, a fordítástudományban, a fordításoktatásban és a nyelvpolitikában azonban számos kutatás folyik: az irodalomjegyzékben alapvető és friss munkákat egyaránt feltüntettem. A fordítói munka műszaki, technológiai vonatkozásairól azonban nincsenek ilyen írások. A terület legfontosabbnak tekinthető forrásmunkái (Esselink 2000, Auster Mühl 2001) leginkább műszaki ismertetőnek, nem pedig rendszerező monográfiának tekinthetők.

Az értekezés célja és tartalma

Ez a doktori értekezés összefoglaló jellegű. Fő célja a fordítástechnológia definiálása az alkalmazott nyelvtudomány önálló szakterületeként. A fordítástechnológia nyilvánvalóan a fordítástudománnyal, a nyelvtechnológiával és a korpusznyelvészettel áll szoros kapcsolatban, de rendkívül fontos a szociolingvisztikai, nyelvpolitikai vonatkozása is, mivel épp a fordítástechnológia jóvoltából lehet elvégezni a mai, megnövekedett terjedelmű és abszurd határidőkkel kiadott fordítási munkákat.

Az értekezés célját a fordítástechnológiai kutatások leírásával, demonstrálásával kívánja elérni, ezért rendszerszerűen, az alkalmazott nyelvtudomány különböző területeihez igazítva mutatja be kutatásokat és azok eredményeit. Az első fejezetben *definiálok a fordítástechnológiát* mint szakterületet; a második fejezetben a *nyelvpolitikai jelentőségéről* ejtek szót (Szépe 2001; Szabari 1996; Horváth 2002). Mivel úgy vélem, az alkalmazott nyelvtudomány egyetlen területétől sem választható el az oktatási tevékenység, a nyelvpolitikai fejezetben a fordítástechnológia elemeinek oktatásával is foglalkozom (Kis B. 2004, Drugan 2004). A harma-

dik fejezet feladata látszólag könnyű: a *fordítástudománnyal* kell kapcsolatba hoznom. A negyedik fejezetben a korpusznyelvészeté és a *nyelvtechnológiáé* a főszerep, míg az ötödik fejezet a terminológiai vonatkozásokat tárgyalja.

1. A fordítástechnológia meghatározása

1.1. A fordítástechnológia mint szakterület

A fordítás mint gazdasági tevékenység rendeltetése a célnyelvi szöveg előállítás a forrásszöveg alapján. Ezt egyre ritkábban végzik egyéni fordítók: a fordítás mindinkább csapatmunka, méghozzá technikai eszközökkel segített csapatmunka lesz – ez következik az átlagos fordítási feladat nagyságából és a rendelkezésre álló időből. Arról, hogy a célnyelvi szöveg mikor tekinthető a forrásnyelvi szöveg fordításának, van közmegegyezés. A csapatmunka körülményei és az említett közmegegyezés együttesen határozzák meg azokat a szabályokat, amelyek alapján a fordítással foglalkozó embercsoportok előállítják a célnyelvi szöveget.

A fordítás így műszaki – gyártási – tevékenységnek tekinthető, amelynek során meghatározott eszközök segítségével, meghatározott eljárások és szabályok követésével terméket állítanak elő. Ez pedig nem más, mint technológia. A fordítás olyannyira műszaki tevékenység, hogy szabványok is vonatkoznak rá, legalábbis egyes részterületeire: UNI 10574 (olasz), Önorm D 1200 és D 1201 (osztrák), DIN 2345 (német), Taalmerk (holland), ISO 12616 (nemzetközi), EN-15038 (európai). (vö. Arevalillo 2007)

Ha a fordítás folyamatát a fordítástechnológia szemszögéből vizsgáljuk, azt mondhatjuk, hogy a fordítást végző személyek mikro- és makrostratégiákat alkalmaznak. A fordítás mikrostratégiája az elemi művelethez, egy szegmentum – korlátozott nyelvi szerkezet, a legtöbbször egy mondat – lefordításához kapcsolódik. Azt határozza meg, hogy a fordító – az erőforrásoktól kapott segítséget is figyelembe véve – hogyan jut el a forrásnyelvi szegmentumtól a fordításhoz.

A fordítás makrostratégiája az elemi műveleteket jól definiált módszerre összefogó folyamatot határozza meg. Ennek során megtörténik a munka előkészítése; az anyag dokumentumokra, a dokumentumok elemi szegmentumokra bontása. Ezután az elemi szegmentumok fordításából összeáll az egyes dokumentumok fordítása; megtörténik a minőségellenőrzés; végül a dokumentumok fordításából is összeáll a több dokumentumból álló célnyelvi anyag.

A fordítástechnológia jelentős kölcsönhatásban van az alkalmazott nyelvtudomány különböző területeivel, amellett hogy maga is a fordítástudomány egyik kutatási területe lehet.

Kutatási területként és erőforrásként a fordítástechnológia rendszerszerű kapcsolatban áll az alkalmazott nyelvtudomány több elemével:

- (1) a nyelvpolitikával (a szociolingvisztikával), mivel a fordítások (megfelelő) elvégzése és léte a nyelvi jogok kérdése, sok esetben pedig jogszabály írja elő. A fordítás iránti jelenlegi kereslet, illetve a fenti körbe eső fordítási feladatok nagy volumene miatt ez csak a fordítástechnológia eszközeivel és folyamataival lehetséges, így a fordítástechnológia nyelvtervezési (korpusz- és státusztervezési) prioritást kap;
- (2) a fordítástudománnyal, több oldalról is: a fordítástechnológia befolyásolja a fordítás folyamatát, egyben pedig megkönnyíti a fordítás egyes aspektusainak kutatását, egyfelől azáltal, hogy a munka során párhuzamos korpuszok jönnek létre, másfelől – a jól definiált folyamatok révén – megfigyelhető a minőségbiztosítás folyamata, és ez által további ekvivalenciamodellek állíthatók fel;
- (3) a korpusznyelvészettel és azon keresztül a számítógépes nyelvészettel: a létrejövő fordítómemóriák és terminológiai adatbázisok alapanyagul szolgálnak a nyelvi elemzéssel és a gépi fordítással kapcsolatos kutatásokhoz, a lektorálás előtti és utáni szövegek összevetése pedig a fordításjavítás automatizálásának fejlesztését segíti;
- (4) a terminológiatannal¹, mert a szakfordításnak fontos eleme a helyes és konzisztens terminológiahasználat. A fordítástechnológia alkalmazása szinte kizárólag a szakfordításra irányul, így fontos eleme a terminológia előkészítése, alkalmazása és ellenőrzése; a terminusok egy adott célnyelvben nagyon gyakran valamilyen fordítás által jönnek létre. A fordítástechnológiai folyamat ezért a legtöbbször egyfajta terminológiai munkafolyamatot is magában foglal.

Mivel a fordítástechnológiát az előzőekben műszaki területként is definiáltuk, interdiszciplinaként rendszerszerű kapcsolatban áll a műszaki tudományok több elemével is:

- az informatikával, azon belül a nyelvtechnológiával, mivel amellett, hogy a párhuzamos korpuszok és a gépi fordítás integrációja népszerű kutatási terület, a fordítástechnológiai rendszerek nagy mennyiségű

nyelvi adat tárolását és nagy teljesítményű feldolgozását, és a meglévő párhuzamos korpuszok hatékony kihasználását igénylik. Emiatt az informatikai fejlesztés során nemtriviális adatmodellek és keresési algoritmusok kidolgozására van szükség;

- a folyamatirányítással és a projekttervezéssel, mivel a fordítástechnológiai rendszer jól definiált munkafolyamatot igényel. Azok a fordítási feladatok, amelyeket a fordítással foglalkozó szervezetek napjainkban kapnak, általában összetett projekt létrehozását igénylik.

1.2. A fordítástechnológia és a gépi fordítás

A gépi fordítás és a géppel támogatott emberi fordítás (más szemszögből: számítógépes fordítástámogatás) rendeltetése is különbözik. A két paradigma közötti „munkamegosztás” eredete az a szemlélet, amely szerint a gépi fordítás jórészt haszontalan, és amelyet a kezdeti eufória után az ALPAC-jelentés ültetett el a kutatókban és a társadalomban (Pierce, Carroll et al. 1966:32) Martin Kay (1980) pedig már a 80-as években definitív munkát írt az ember és a gép fordításban elfoglalt helyéről.

A munkamegosztás alapja a következő különbség: míg a gépi fordítás gyors és automatikus, rossz nyelvi minőségű és gyakran csak nagyjából érthető fordítást hoz létre, addig a géppel támogatott emberi fordítás a lényegét tekintve emberi fordítás, ezért minősége potenciálisan a lehető legjobb emberi fordításét is elérheti. Létrehozása nagyságrendekkel lassabb és több munkát igényel, mint az automatikus gépi fordítás kimenetéé, azonban lényegesen gyorsabb, mint a gépi támogatás nélküli emberi fordítás.

Az automatikus gépi fordítás és a géppel támogatott emberi fordítás minőségkülönbsége azt eredményezte, hogy felhasználásuk a kommunikációs lánc különböző részeihez kapcsolódik: az automatikus gépi fordítás a szöveg befogadását, míg a géppel támogatott emberi fordítás a szöveg előállítását segíti.

Minden szöveg-előállítással és fordítással kapcsolatos számítógépes rendszer feltételezi, hogy az emberi intelligencia összehasonlíthatatlanul magasabb rendű, s ezért az ember által létrehozott kimenet mindig elsőbbséget élvez a gépi kimenettel szemben. A mai számítógépes rendszerek tervezési filozófiája kimondatlanul is az, hogy az emberi kimenetet mindenféle vizsgálat nélkül is jobb minőségűnek kell tekinteni a gépi kimenetnél. [Holott Melby (1995) szerint: „[...] bad human translation is interesting because it was most likely done by a human yet in a manner similar to the way computers translate”².]

2. A fordítástechnológia nyelvpolitikai szerepe és hatása

Ebben a fejezetben a fordítástechnológia nyelvpolitikai vonatkozásait foglalom össze. Ez felöleli a terület kialakulásának nyelvpolitikai indítékait, a fordítótársadalomban bekövetkező változásokat, illetve a fordítástechnológia nyelvtervezési hatásait.

Ugyane fejezetben foglalkozom a gépi fordítástámogatáshoz és a fordítástechnológiával kapcsolódó oktatással is. Véleményem szerint ez szerves kapcsolatban áll a nyelvtervezéssel, mert az oktatás e tekintetben nem más, mint a társadalmi/gazdasági változások tudatos átvitele és érvényesítése.

2.1. A fordítástechnológia szükségessége

Ebben a fejezetben megállapítom, hogy a különböző csoportok kommunikációs igénye globális. A globális kommunikáció lokális kommunikációt, vagyis fordítást igényel. A fordítás erőforrásai végesek, ugyanakkor egyre inkább szükség van az azonnali vagy egyidejű kommunikációra, vagyis az ennek megfelelő sebességű fordításra. Mivel megfelelő minőségű fordítást egyelőre csak az embertől várhatunk, az emberi fordítás hatékonyságát kell megnövelnünk. Ennek része a gépesítés, amelynek létjogosultságát nem lehet elvitatni, azonban jelenlegi fejlettsége mellett kételkedhetünk a valódi hasznosságában.

2.2. A technológizált fordítás társadalmi-gazdasági vonatkozásai

A fordítói munka új körülményei szükségessé teszik a csoportos fordítás alkalmazását, amelynek a gépi támogatás mellett a szervezési kérdései is fontosak, és részletes tanulmányozást igényelnek.

Ha a lefordítandó szöveg terjedelme nagy, a határidő pedig szűkös, kézenfekvő a munka párhuzamosítása: ha a fordítást egyetlen ember nem képes a megadott határidőre elvégezni, fordítócsoporthoz kell alkalmazni. A szakmai szövegek fordításának párhuzamosítása esetén nem kerülhető el a gépi fordítástámogatás alkalmazása. Ezt a konzisztencia igényével indokolhatjuk.

A fordítási munka egyre technológiaszerűbbé változik: a relatíve rendszeretlen, illetve egyéni beosztás szerint végzett fordítás a csoportos fordításban legalább három, jól elkülöníthető szakaszra osztható, és minden szakaszban különböző, jól definiált feladatot kell elvégezni.

A fordítók tehát egyre inkább csoportmunkát végeznek, ami új eszközöket igényel, és új készségeket vár el tőlük. A csoportos fordítás centruma olyan szervezet, amelynek az a rendeltetése, hogy elvégezze a munkaszervezést, előállítsa

és rendelkezésre bocsássa a fordítási erőforrásokat. A csoportmunkában dolgozó fordító mindig hálózatban dolgozik, és folyamatosan kommunikál a többiekkel.

2.3. A fordítástechnológia szerepe a státusztervezésben

A fordítástechnológiát az Einar Haugen (1983) által bevezetett taxonómia szerint két szempontból, a státusztervezés és a korpusztervezés szempontjából vizsgálom.

A státusztervezés felvetésére igen egyszerű válaszolni: mindaddig, amíg egy tárgykörben vagy nemzetközi szervezetben biztosítják a fordítást egy adott nyelvre vagy dialektusra, adott kultúrának megfelelően, addig fenntartható a kérdéses nyelv és nyelvközösség egyenrangú státusza az adott tárgykör vonatkozásában vagy az adott nemzetközi közösségben.

A fordítás szempontjából a terminológiaalkotást lokális vagy decentralizált státusztervezésnek tekintem. Ezt Kis Ádám inkább korpusztervezési folyamatnak tekinti (Kis Á.-Kis B. 2004), azonban a státusztervezésnek a korpusztervezés nagyon sokszor eszköze. Magyarországon ezt felismerve jött létre a Magyar Nyelv Terminológiai Tanácsa (MATT), különböző kormányzati és nem kormányzati szervezetek, illetve magánszemélyek részvételével, a Magyar UNESCO Bizottság támogatásával.

2.4. A fordítástechnológia szerepe a korpusztervezésben

A fordítástechnológia alkalmazása korpusztervezési kérdés, mivel a forrásnyelven leírt fogalmi rendszer gyakran kizárólag a fordításon keresztül kerül a célnyelvi kultúrába. Ezen keresztül a fordítástechnológiához kötődő minőségbiztosítás, illetve a terminológiatervezés kap jelentőséget.

A fordítás által érintett tárgykör nyelvhasználatára elsődleges hatással van a fordítási terminológia és a fordítások minősége is. Ezért a fordítással és az adott tárgykörrel foglalkozók előtt a következő feladatok állnak:

- Tudatos és szervezett terminológiaalkotás a fordítási munkától függetlenül
- Tudatos és szisztematikus terminológiaalkotás és terminológiahasználat a fordításban
- Konzisztens és jó nyelvi minőségű (jól olvasható) fordítások előállítása

Összefoglalva a fentieket: a státusztervezés és a korpusztervezés kettős feladata egyfelől a megfelelő fordítási kapacitás, másfelől a terminológiai bázis és a fordítások konzisztenciájának biztosítása. E feladatok – ha nagy a forrásnyelvi doku-

mentumok keletkezésének intenzitása, és szűkös a fordítási kapacitás – nem oldhatók meg megfelelő számítógépes segítség nélkül. A jelenleg hozzáférhető technikai eszközök erre elégtelenek bizonyulhatnak, azonban léteznek olyan, kutatási fázisban levő eszközök, amelyek további segítséget nyújthatnak.

2.5. A fordítástechnológia oktatása

A fordítástechnológiához kötődő nyelvtervezés elengedhetetlen eleme a fordítástechnológia makro- és mikrofolyamatainak (makro- és mikrostratégiáinak) oktatása, mert a fordítástechnológiát nem lehet alkalmazni a szükséges eszközök kezelésére készségszinten képes fordítók és más közreműködők nélkül. Ez egyaránt jelenti a fordításhoz felhasznált műszaki eszközök alkalmazásának (a mikrostratégiának) és a fordításszervezésnek, illetve a fordítás technikai kiszolgálásának (a makrostratégiának) az oktatását is.

A fejezetben megállapítom, hogy a fordítóknak az alábbi műszaki készségekre van szükségük:

- **termelési készségek:** nagyon fontos, hogy fordítási teljesítményük és rendelkezésre állásuk versenyképes maradjon (Austerermühl 2001);
- **gépi kommunikációs készségek:** ezzel kiléphetnek viszonylagos elszigeteltségükből, és tudnak (virtuális) csapatban dolgozni;
- **műszaki-karbantartási készségek:** fontos, hogy saját számítógépes infrastruktúrájuk működését fenn tudják tartani akkor is, amikor nem kapnak külső segítséget.

Az elmúlt évek során fokozatosan kidolgozott kurzust három szempont szerint ismertetem: (1) a kiindulási pont, vagyis a hallgatók meglévő tudása; (2) a tananyag alapprioritásai; (3) a módszertan.

A tanmenet a hangsúlyt a hallgatók önállóságára helyezi. Ez azt jelenti, hogy a tantermi foglalkozás során a hallgatók mindvégig önállóan gyakorolnak, a tanári jelenlétre ahhoz van szükség, hogy a hallgatók előszörre is könnyen hozzáférjenek a gyakorlatok anyagához. A gyakorlatok az interneten érhetők el, és olyan módon vannak megszerkesztve, hogy önálló, tanári jelenlét nélküli tanuláshoz is használhatók legyenek.

3. A fordítástechnológia és a fordítástudomány

A fordítástechnológia mind a kutatási terület kiterjesztéseként, mind pedig kutatási erőforrásként hasznos a fordítástudomány számára. A következőkben ezt mutatom meg.

Ha a fordítástechnológiát kapcsolatba akarjuk hozni a fordítástudománnyal, meg kell fogalmaznunk a fordítástudomány azon kérdéseit, amelyekkel kapcsolatban a két terület kölcsönhatásba kerülhet. Ezeket a következőképpen előlegezhetjük meg:

- Melyek a fordítás kognitív folyamatai?
- Mit jelent a fordítási ekvivalencia, és hogyan vizsgálható?
- Melyek a fordításszövegek (grammatikai, szemantikai, szövegnyelvészeti) jellemzői? Milyen kapcsolatban állnak ezek a forrásszöveg jellemzőivel?

Erről a következő állításokat tehetjük:

1. fordítástechnológia a fordítástudomány vizsgálatának tárgya. Mivel a fordítástudomány egyaránt vizsgálja a fordítás kognitív folyamatait és a célnyelvi szövegek nyelvi megformálását, szükséges vizsgálnia a hatást is, amelyet a fordítástechnológia alkalmazása gyakorol ezekre. Ilyen jellegű kutatással a jelen értekezés szerzője nem foglalkozott, a kérdést tehát az értekezés nem válaszolja meg.

2. A fordítástechnológia alkalmazása során létrehozott erőforrások lehetővé teszik a fordítási folyamat, a fordítási ekvivalencia, illetve a forrásszöveg fordításnyelvre gyakorolt hatásának vizsgálatát.

3. A fordítástechnológia alkalmazásával új ekvivalenciamodell jött létre. Ez a modell azt a fordítást tekinti ekvivalensnek, amelyet valamely, meghatározott technológiai fegyelmet megtartó szerkesztőség közlésre elfogad. A technológiai fegyelem ez esetben a fordítás javítás előtti és javítás utáni változatának együttes megőrzését jelenti.

3.1. Ekvivalencia és minőség

Mivel a teljes (mentális, szemantikai) ekvivalencia nem vizsgálható *per se*, a kutatók az ekvivalenciát különböző szinteken próbálják megragadni (pl. Catford 1965, Komisszarov 1990). A gépi fordítástámogatás szempontjából az a fontos, hogy amennyiben az ekvivalencia nem vizsgálható könnyen a közlés egészére nézve, meg kell keresni azokat a kisebb vagy nagyobb egységeket, amelyek esetében már formálisan is beszélhetünk ekvivalenciáról. A formálisan ábrázolt ekvi-

valencia azt jelenti, hogy forrásnyelvi és célnyelvi egységek között formális megfeleltetést állítunk fel, nem vizsgálva a megfeleltetés eredetét és természetét. Ez tulajdonképpen a performanciaalapú ekvivalenciamodell, szemben a fordítástudomány hagyományos, kompetenciaalapú ekvivalenciaelméleteivel: a literális, funkcionális és szintagmatikus ekvivalenciával, a totális fordítással (Catford 1965, Jakobson 1959), illetve a formális és dinamikus ekvivalenciával (Nida 1964) – és sok tekintetben Komisszarov (1990) ekvivalenciaszint-elméletével.

A triviális formális ekvivalenciamodell a párhuzamos korpusz: az ember által párhuzamos korpuszokba rendezett közléseket a gépi módszerek kutatói *a priori* ekvivalensnek tekintik. Pedig a korpusz közvetlen felhasználása esetén a fordítási folyamat utánzása csak a felszínen, primitív formában lehetséges: a gép nem az ember átváltási műveleteit utánozza, hanem a forrásnyelvi szövegeket helyettesíti célnyelvi szövegekkel. A fordítás folyamata tehát itt is fekete doboz marad.

A fordítási ekvivalencia új modellje. A fordítások minőségbiztosításából világosan látszik, hogy a közmegegyezés szerint ugyanazon FNy szöveg különböző CNY fordításai között vannak „ekvivalensebbek” és „kevésbé ekvivalensek”. Ennek az egyszerűsítő szemléltetése a fordítási kontinuum, amely a szövegtranszformációk (fordítás és javítási fázisok) eredményét egy egyenes szakaszon helyezi el. A szakaszon balról jobbra haladás azt jelképezi, hogy a CNY szöveg halad a CNY norma felé.

Az első CNY szöveg és a publikált CNY szöveg összevetéséből jön létre a javításkorpusz. Ez úgy válik ekvivalenciamodellé, hogy láthatóvá teszi a javítás folyamatát, és összeveti a fordító és a lektor mentális ekvivalenciaképét.

3.2. A fordítás új körülményei – a fordítástechnológia keletkezése

Ebben a részben a fordítástechnológiát mint a fordítástudomány kutatásának tárgyát mutatom be. Ennek során kimutatom a fordítástechnológia mikro- és makrostratégiájának hatását a fordítás folyamatára, különös tekintettel azokra a fordítástechnológiai folyamatokra, amelyek a fordítás megváltozott körülményeinek ellensúlyozására jöttek létre.

A megváltozott körülmények lényege az, hogy a fordítási feladatok mennyisége és nagysága megnőtt, az elvégzésük határideje pedig csökkent. Ennek folytán igazolható, hogy a fordításnak a piacon elérhető minősége lényegesen romlik: a fordítástechnológia kialakulása erre a helyzetre válasz.

3.3. A fordítás mikrostratégiája

A fordítástechnológia mikrostratégiája annyit jelent, hogy a fordító a CNY szöveget számítógépes fordítási környezetben szerkeszti meg. Ebben a részben elsősorban a fordítómémória-használat és a fordítástudomány kölcsönhatásával foglalkozom, ezen belül pedig a következő három kérdésre keresem a választ:

- (1) A fordítómémória-használat mennyiben tekinthető az ekvivalenciaszintek elméletében leírt átváltási műveletek modelljének?
- (2) Hogyan befolyásolja a fordítási folyamatot a fordítómémória-használat?
- (3) Hogyan – milyen nyelvtechnológiai eljárásokkal – csökkenthető a fordítómémória-használat negatív hatásai?

1. A fordítómémória-használat nem tekinthető az átváltási műveletek modelljének. Az értekezés írásakor létező eszközök pusztán a transzformáció bemenetét és kimenetét jegyzik fel, illetve egyetlen manipulatív műveletre, a helyettesítésre alkalmasak, a terminus és a szegmentum (nagyjából a mondat) szintjén.

Kitérő: a gépi fordítás módszereit bizonyos szempontból sokkal inkább lehet az átváltási műveletek modelljének tekinteni, mint a gépi fordítástámogatás műveleteit. Az előbbi módszerei között pedig van olyan, amely közvetlenül is alkalmas az átváltási műveletek produktív modellezésére. A gépi fordítástámogatás alkalmazásakor azonban megvan az ember lehetősége az átváltások tényleges végrehajtására, az előbbiben azonban a döntést az automatára ruházzuk. Lényeges különbség még, hogy a gépi fordítástámogatás eszközei az emberi fordító átváltási döntéseit játsszák vissza, míg a gépifordító-rendszerek sajátos algoritmikus döntéseik alapján kombinálják az egyes műveleteket – ezért tűnnek alkalmasabbak rosszabb nyelvi minőségűnek, illetve a célnyelv struktúráitól és lexikájától távolibbnak a géppel automatikusan előállított fordítások.

2. A fordítómémória-használat az átváltási műveletek szempontjából azt jelenti, hogy a számítógép visszajátsszik egyes tárolt átváltásiművelet-sorozatokat, vagyis felkínálja a korábban már tárolt forrásszegmentumhoz tartozó egyik lehetséges fordítást, amelyet a fordítónak pedig ki kell igazítania. Feltételezhetjük, hogy ez némiképp másfajta kompetenciát igényel, mint a „tiszta” fordítás, mivel itt egyes átváltási műveleteket vissza is kell vonni. Ez egyfajta lektorálás, ám a fordításjavítás jóindulatú előfeltevése, miszerint a kapott fordítás az aktuális forrásnyelvi szegmentum ekvivalense, nem tartható.

3. A fordítómémóriák továbbfejlesztésével kapcsolatos kutatásnak két iránya van:

3. A fordítástechnológia és a fordítástudomány

- Hatékonyságnövelés: a fordítómemória-találatok arányának és gyakoriságának növelése új hasonlósági keresési módszerek bevezetése által.
- Utószerkesztés: algoritmusok kifejlesztése az adatbázisból kiemelt CNY fordítás módosítására úgy, hogy az az aktuális FNY szegmentum ekvivalensévé váljon (Kis et al 2004).

3.4. A fordítástechnológia makrostratégiája

A fordítástechnológia makrostratégiája a fordítás mikrostratégiai műveleteit rendezi jól definiált folyamatba. Ennek megfelelően három rendeltetése van:

- Az új típusú (megnövekedett volumenű és szorosabb határidejű) fordítási feladatok elvégzésének biztosítása szervezéssel, munkafolyamat felállításával; a projekt költségvetésének meghatározása és ellenőrzése;
- A fordítás műszaki szinergiájának biztosítása. Az összetett fordítási projektek magukban foglalják a FNY szöveg műszaki előkészítését és a kész CNY szöveg műszaki előállítását is;
- Minőségbiztosítás: a volumen/idő nyomás negatív hatásainak enyhítése. Ez az utólagos ellenőrzésen túl a fordítást megelőző és a fordítás közben érvényesülő minőségbiztosítási intézkedéseket is jelent.

Ebben a fejezetben vázlatosan bemutatom a makrostratégia elemeit, majd két olyan makrostratégiai részfolyamatot, amely széttagolt projektekből, illetve kiélezett időbeosztás esetén is lehetővé teszi a fordítás minőségbiztosítását. Ez a két módszer a szimultán lektorálás és a fordítás javításának automatizálása.

4. A fordítástechnológia kapcsolata a korpusznyelvészettel és a nyelvtechnológiával

4.1. Általános megállapítások

A fordítástechnológiával fennálló kapcsolat meghatározásakor a nyelvtechnológiát a számítógépes nyelvészet alkalmazásának tekintem.

Mivel a számítógépes nyelvészet kutatásait kezdetben a gépi fordítás létrehozása motiválta, a fordítástechnológiával szerves kapcsolatban is állhatna. A fordítástechnológiai eszközök ugyanakkor – sem a mikro-, sem a makrostratégiában – nemigen használják a nyelvtechnológia által létrehozott eszközöket, holott ezt számos szerző (pl. Hodász G. et al. 2004, Callison-Burch et al. 2005) javasolja. A

4. A fordítástechnológia kapcsolata a korpusznyelvészettel és a nyelvtechnológiával

fordítástechnológia számos olyan szövegkeresési és szövegmanipulációs eszközt alkalmaz, amelynek olyan értelemben nincs köze a nyelvtechnológiához, hogy nem veszi figyelembe, nem kísérli meg feltárni és manipulálni a szöveg nyelvi struktúráját.

Ha az ellenkező irányt tekintjük, a modern számítógépes nyelvészet és nyelvtechnológia számos területen profitál a fordítástechnológiából, mégpedig azért, mert a fordítás spekulatív, kvázi-kompetenciamodelljeivel szemben a megvalósult fordítás performanciaalapú modellt kínál a fordítás tanulmányozásához. Ez ugyanaz a megközelítés, mint a korpusznyelvészeté a nyelv modellezésével kapcsolatban. A fordítástechnológia felügyelete alatt végzett fordítás során jelentős, szegmentumszinten szinkronizált párhuzamos korpuszhoz jutunk, amelyek különböző szempontok szerint „bányászhatók”. Már az ALPAC-jelentés (Pierce et al. 1966) és a XEROX (Kay 1980) is felvetette, hogy a spekulatív modellek helyett a megvalósult emberi fordítás újrahasznosításával kellene növelni a fordítás hatékonyságát. A kiterjedt, nagy tömegű párhuzamos korpuszok létrejötte mindenesetre elősegítette a statisztikai gépi fordítás létrejöttét (vö. pl. Callison-Burch et al. 2004).

A fejezet általános részében megvizsgálom, hogy a nyelvtechnológia egyes közismert eszközei hol jelennek meg a fordítástechnológiában. Így foglalkozom a párhuzamos korpuszokkal, a szövegszinkronizálással és a fordítómemóriák konkordanciafunkciójával.

4.2. A SZAK javításkorpusz

A fordítás javításának legjobb modellje a megvalósult javítások elemzése. A javítások elemzéséhez a fordító által előállított CNY szöveg és a publikált CNY szöveg közötti különbséget kell vizsgálni. Ehhez olyan párhuzamos korpuszt kell létrehozni, amelyben a FNY és a publikált CNY szövegek helyett a CNY szöveg első változata és a publikált CNY szöveg található.

A SZAK Kiadó műhelyében létrejött egy ilyen korpusz, amely informatikai szakkönyvek és weblapok szövegét tartalmazza. A korpusz valójában három komponensből áll: az FNY szövegből, az első CNY szövegből és a publikált CNY szövegből.

A 4.2. fejezetben először a SZAK javításkorpusz jellemzőit és kialakítását ismertetem, majd a javítási folyamat rekonstrukciójához alkalmazható módosított Levenstejn-algoritmust. (Levenstejn 1965).

4. A fordítástechnológia kapcsolata a korpusznyelvészettel és a nyelvtechnológiával

4.3. A fordítómemóriák értékelése és kihasználásuk javítása

A fordítómemória mint számítógépes eszköz rendeltetése a fordító vagy fordító-csoport korábbi fordításai között is előforduló szövegrészek gyors felismerése és megkeresése, illetve a korábban adott fordítás visszaadása valamely aktuális forrásszöveg fordítása közben. Ehhez a fordítómemóriának olyan adatbázist kell fenntartania, amelyben – megfelelő felbontással – tárolva vannak a korábbi forrásszövegek és fordításaik.

A fordítómemóriák a szegmentumokat fordítási egységekbe szervezik. A fordítási egység (*translation unit*, TU) egy forrásnyelvi szegmentum (a legtöbbször egy mondat) és CNy megfelelője által alkotott pár. A fordítómemória felügyeletével végzett fordítás során ilyen fordítási egységek jönnek létre.

A fejezetben igazolom, hogy a fordítómemória alkalmas a szakmai szövegek fordításának gyorsítására, mert azok az irodalmi szövegekhez képest lényegesen homogénabbak (több ismétlődést tartalmaznak).

A fordítómemóriák hatékonyságának mérése. A fordítási munka hatékonyságát általában nem mérjük, de ha mérnénk, valamilyen szöveg-egység lefordítására fordított időt értenénk alatta. Itt is szükségünk van egy jóhiszeműségi hipotézisre: feltételezzük, hogy a fordításra fordított idő az ekvivalens CNy szöveg előállításához szükséges, anélkül hogy az ekvivalencia fogalmát közelebbről meghatároznánk.

A fordítómemóriák ismerete nélkül nem definiálhatnánk úgy a hatékonyságot, mint az adott dokumentumban ténylegesen lefordítandó szöveg tömegét. Ha a korpusznyelvészetben szokott módon szövegszavakban mérjük a terjedelmet, ez a következőt jelenti (jelölések tőlem):

$$\eta = 1 - \frac{w_f}{w_t}$$

ahol η a hatékonyság mértéke, w_f a lefordítandó szegmentumok együttes terjedelme szövegszóban, a w_t pedig a forrásszöveg teljes terjedelme ugyancsak szövegszóban. A hatékonyság mértéke 0, ha minden szövegszót le kell fordítani, és 0,2, ha a szövegszavak 20%-ára valamilyen automatikusan képzett javaslatot kapunk, így csak a fennmaradó részt kell lefordítani. A fordítómemóriák gyártói ezt a számot adják meg, amikor a legfeljebb 30%-os hatékonyságnövekedést említik.

A 4.2. fejezetben ezt a modellt finomítom tovább, meghatározandó annak módszerét, hogyan lehet számszerűsíteni a fordítómemória használatából eredő hatékonyságnövekedést.

4. A fordítástechnológia kapcsolata a korpusznyelvészettel és a nyelvtechnológiával

A karaktersorozat alapú fordítómemória korlátai. A kereskedelmi forgalomban ún. karaktersorozat alapú fordítómemóriák kaphatók, amelyek a szegmentumok hasonlóságát általában karakterhármak és szókettesek összehasonlításával – a számítógépes nyelvészet statisztikai módszereihez hasonló módon – vizsgálják. A karaktersorozat alapú fordítómemóriák (Navarro 2001, Navarro et al. 2001., Planas 2000) hiányossága, hogy csak teljes szegmentumok hasonlóságát vizsgálják, így a fordítás közben a tárolt szegmentumok igen nagy hányada rejtve marad – minél hosszabb az aktuális forrásszegmentum, annál kisebb a valószínűsége annak, hogy az adatbázisban lesz kellőképp hasonló tárolt szegmentum.

A fordítómemóriában így lehetnek olyan szegmentumok, amelyek egyik vagy másik része teljesen egyezik az aktuális forrásszegmentum egy részével (vagy éppen egészével), s lehetnek olyanok is, amelyek szókincsükben különböznek, de analóg szintaxist mutatnak.

Erre a problémára jelenleg két megoldás létezik, az egyik hozzáférhető kereskedelmi forgalomban, a másik egyelőre kísérleti fázisban van:

- Töredékkeresés karaktersorozat-alapon;
- Nyelvi dekompozíció: nyelvi támogatású fordítómemória.

A nyelvi támogatású fordítómemória. Korábbi kutatásom (Hodász et al. 2005, Kis B. et al. 2004) alapján olyan fordítómemória-kezelési eljárást ismertetek, amely a szegmentumok hasonlóságát nyelvi szerkezetük alapján határozza meg, és képes az egyes szerkezeti elemek dinamikus cseréjére. A módszer részletesen értékelem. Ennek alapján kijelenthetjük, hogy a nyelvi támogatású fordítómemória a karaktersorozat-alapú fordítómemóriához képest jelentős hatékonyságnövekedést jelent, mert

- (1) olyan esetekben is javasol tartalmilag adekvát fordítást, amikor a karaktersorozat-alapú fordítómemória javaslatának tartalma már nagyon távol esik az aktuális forrásszegmentumtól – vagy éppen nem is jelenik meg javaslat;
- (2) az utólagos javítás igényét azáltal is csökkenti, hogy egyrészt a főnévi csoportok ragozását a mondatsablonhoz igazítja, másrészt pedig megjelöli a potenciális javítási helyeket, így a javaslat annak előzetes végigolvasása és teljes értelmezése nélkül is kijavítható. Pszicholingvisztikai szempontból ez azt jelenti, hogy a javítást a fordító vagy a lektor „sekélyebb” nyelvi műveletekkel – kis részek fordításával és kiemelt nyelvtani hibák javításával – el tudja végezni.

4. A fordítástechnológia kapcsolata a korpusznyelvészettel és a nyelvtechnológiával

A fordítómemóriák értékelési szempontjai és módszerei. A nyelvtechnológia vizsgálja a nyelvi modellek jóságát. Ennek alapján felállítottam egy szempontrendszert a fordítómemóriák minőségének vizsgálatára. A három szempont a következő:

- (1) kihasználtság (r_m): a fordítómemória-adatbázis időben változó mennyiségi mutatója, amely egyfelől utal a fordítómemóriát lekérdező algoritmus hatékonyságára, másfelől pedig közvetlenül megmutatja, hogy a fordítómemóriában mekkora a „holt teher” – a tárolt, de fel nem használt szegmentumok tömege. Ennek hosszú távú, adott algoritmus mellett több fordítómemóriára kiterjedő mérése adja meg az algoritmus egyfajta jellemzését.
- (2) hatékonyság (η): a fordítás hatékonysága (hatékonyságnövekedése) adott fordítómemória, fordítómemória-algoritmus és forrásszöveg mellett, a fordítás elvégzéséhez szükséges tényleges munka alapján. Közvetlenül a gépi fordítástámogatás hasznosságát méri.
- (3) informativitás (i): a fordítómemória által a találatokhoz adott pontszámok (*scoring*) értékelése, annak meghatározása, hogy az adott találathoz adott pontszám milyen korrelációban állnak az adott szegmentum kijavításához szükséges munkával. Ez közvetlenül a fordítómemóriát lekérdező algoritmus minőségét méri.

Az értekezésben javasolok módszereket a fentiek mérésére és kiszámítására is.

5. Fordítástechnológia, terminológia és lexikográfia

5.1. Terminológiai folyamatok a fordításban

A fordítás mindennapi eleme a terminológiahasználat: a fordítók és a fordítással foglalkozó szervezetek szavakat és kifejezéseket gyűjtenek, terminológiai adatbázisokat építenek, és olykor szótárakat is kiadnak.

A terminológián irodalma ugyanakkor meglehetősen gyéren foglalkozik a terminológia fordítási vonatkozásaival. Sager (1990) áttételesen másodlagos terminusalkotásról (*secondary term formation*) beszél, amely többek között „történhet [...] másik nyelvi közösség felé irányuló tudásátadás során, terminusok létrehozásával.”³ Arntz (1993) észreveszi, hogy a terminológiai kutatás a fordításban problémát jelenthet: „Hirtelen felmerülő fordítási probléma megoldásához is szükség lehet az adott jelenség részletes tanulmányozására. Az ilyen vizsgálódás gyakran csak a szomszédos fogalmakat említi meg, további

részletezés nélkül, így csak az adott tárgykör vagy fogalmi rendszer egy részét kezeli.”⁴

Általánosságban azt mondhatjuk, hogy a leíró terminológiában elsősorban a terminusok (strukturális) nyelvészeti jellemzőivel foglalkozik, alkalmazott tudományként jellemzően azért, hogy valamiféle nyelvészeti modellt találjon a terminusok viselkedésére, amely aztán különféle alkalmazásokban – például a gépi terminuskivonatolásban – felhasználható. A terminológiához kapcsolódó szociolingvisztikai, onomasziológiai vizsgálódások elsősorban előíró jellegűek. Tipikus példa Pavel (1993) írása, amely elviekben a terminológia keletkezésével kapcsolatos neologizmusokat és frazeológiát vizsgálja, ám a folyamatok leíró vizsgálata helyett nyelvoktatási szempontokat ad – egy teljes fejezet foglalkozik az új terminusok helyességének nyelvi feltételeivel. Általában is elmondhatjuk, hogy a terminológiával foglalkozó kutatók jellemzően *n ó m e n k l á t o r o k* vagy *m e t a - n ó m e n k l á t o r o k*: elsősorban olyan módszereken dolgoznak, amelyekkel lehetséges a fogalmak pontos leírása, megnevezése, illetve amelyekkel ilyenek rendszerét lehet kialakítani.

A fentiekkel szemben az értekezésben leírom azokat az elveket és azt a módszertant, amely szerint a fordítás során a terminusok alkotása végbemegy.

A fordítástechnológia terminológiai folyamatait az alábbiak szerint vizsgálom:

- (1) A terminusalkotás folyamata
- (2) A terminusalkotás fordítási kérdései
- (3) A terminusalkotás lehetséges stratégiái a fordításban
- (4) A terminuskezelés lehetséges gépi eszközei a fordításban

5.2. Terminuskivonatolás

A terminuskivonatolással kiterjedt nemzetközi irodalom foglalkozik (pl. Jacquemin 2001), és számos alkalmazása is született. Ebben a fejezetben olyan fejlesztési projektumról és kísérletről számolok be (Kis B. 2005), amely arra irányult, hogy nagy fordítási feladatokhoz használható terminuskivonatoló eszköz jöjjön létre, amelynek segítségével

- leképezhető a FNy szöveg teljes terminológiai váza, és
- jelentősen rövidíthető a terminológiai előkészítésre fordított idő.

Az utóbbi kritérium konkrétan úgy fogalmazható meg, hogy a szövegspecifikus terminológiaalkotáshoz szükséges időnek lényegesen rövidebbnek kell lennie, mint amennyi idő alatt a FNy szöveget egy ember figyelmesen végig tudja olvasni.

Az 5.2. fejezetben közölt leírás az erre irányuló korpusznyelvészeti kutatást írja le, és megmutatja, hogy a létrehozott eszköz mindkét fenti követelménynek megfelel. A leírás elején ismét foglalkozom a terminológia definíciós problémáival, de ezúttal a gépi modellezés szemszögéből.

A kialakított terminuskivonatolási eljárás, illetve eszköz legfőbb jellemzője, hogy több – statisztikai és szabályalapú – algoritmus együttes alkalmazásával olyan terminusjelölt-listát állít össze, amelynek elemei pontozva vannak. A listában a terminusjelöltek a pontszámok alapján kategóriákba sorolhatók, amelyek alapján megjósolható, hogy az egyes kategóriákba eső terminusjelöltek milyen valószínűséggel lesznek helyesek. Két kísérletet ismertetek: a második esetben a kategóriákra vonatkozó hipotézist igazolom. Ennek alapján meghatározható olyan kategória, amelyen belül a terminuskeresés pontossága 90% körüli.

¹ A „terminológia” szó a „-lógia” utótagban elvileg már utal a „tan”-ra, ám a terminológia közkeletű értelmezése nem ez: adott szakterület terminusainak halmazát jelenti. Ezért fontosnak tartotam, hogy a kutatási területet megkülönböztessem a nyelvi eszköztől, így került a megnevezés végére a „-tan” utótag. Az ezzel kapcsolatos szóhasználatot részletesen az 5. fejezet tisztázza.

² „A rossz emberi fordítás azért érdekes, mert bár ember készítette őket, a gép által végzett fordításhoz hasonló módon.” (Saját fordításom)

³ „Secondary term formation [...] happens [...] as a result of knowledge transfer to another linguistic community which is carried out by means of term creation.” (Sager 1990:80) (Saját fordításom)

⁴ „A detailed study of an individual phenomenon is often necessary in order to solve an acute translation problem. Investigations of this kind will frequently mention the neighbouring concepts without going into more detail, so that only a part of the field or system of concepts is handled.” (Arntz 1993) (Saját fordításom)