

DOKTORI ÉRTEKEZÉS TÉZISEI

FORDÍTOTT SZÖVEGEK SZÁMÍTÓGÉPES ÖSSZEVETÉSE

Autentikus magyar szövegek és fordítás eredményeként
létrejött célnyelvi magyar szövegek lexikai kohéziós
mintázatának összehasonlító elemzése

SEIDL-PÉCH OLÍVIA

Eötvös Loránd Tudományegyetem
Bölcsészettudományi Kar
Nyelvtudományi Doktori Iskola
Fordítástudományi Doktori Program

2011
Budapest

1 A kutatás tárgya, az értekezés célja

Az értekezés alkalmazott nyelvészeti megközelítésből foglalkozik a fordításokkal, fókuszában a célnyelvi szövegek szövegépítő sajátosságainak feltárása áll, amit a szövegnyelvészet és a korpusznyelvészet eszközeivel vizsgál. Az értekezés központi témájául választott célnyelvi szövegprodukciónak leírását azért tartom kiemelkedően fontos feladatnak, mert a „fordítási szöveg nyelvi és nem nyelvi (kognitív, illetve a társadalmi kontextus által meghatározott) sajátosságainak leírása [...] elengedhetetlen részét képezi minden, a fordítások minőségének értékelését, a fordítók képzését és a fordítói kompetencia fejlesztését célul kitűző vállalkozásnak” (Károly2007: 13).

Az értekezés csatlakozik ahhoz a Petőfi S. (2004) által szorgalmazott szövegnyelvészeti megközelítéshez, amely a szövegek komplexebb vizsgálatát tűzi célul maga elé. A korpusznyelvészeti módszerrel végzett kutatás egyrészt számszerű adatokkal bizonyítja a fordított szövegekre jellemző kohéziós eltolódásjelenséget, másrészt az elméleti fejezetekben részletezi az eredményközpontú megközelítés mellett a forrásnyelvi szöveg hatására bekövetkező célnyelvi szövegsajátosságok és eltolódások funkció- és folyamatorientált jellemzőit.

A kutatás során alkalmazott korpusznyelvészeti technika általánosításokat tesz lehetővé a variabilitást befolyásoló szövegkörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek feltárásán keresztül. A vizsgálat kiemelkedően fontos területe az írott nyelv szövegépítő sajátosságain belül a fordított szöveget az autentikus szövegprodukciónak megkülönböztető fordítási univerzálék feltárása (vö. Baker 1993, 1996). Az elemzés főbb területei az autentikus és célnyelvi szövegek összevetése, a nyelvpár specifikus fordítói magatartás feltárása, a forrásnyelvi és célnyelvi magyar szövegek szövegkohéziós eltolódásainak feltárása (vö. Péch 2009a, 2009b, 2010). Az értekezés három tudományterület – a fordításkutatás, a szövegnyelvészet és a korpusznyelvészet – kutatási szférájának metszéspontjában helyezkedik el, eredményeivel ezek elméleti és gyakorlati tudásbázisát gazdagítja.

Az összehasonlító elemzést nagyobb mennyiségű nyelvi adat kontrasztív egybevetése alapján végeztem: többféle műfaji és szövegtipológiai csoportba sorolható magyar nyelven keletkezett autentikus és magyarra fordított célnyelvi

szöveg szövegépítő sajátosságait vettem egybe. A szövegek lexikai kohéziós mintázatának leírása a mondathatárokon túlmutató nagyobb nyelvi egységek nyelvészeti elemzésének eredménye, azaz „a mondatokból létrejövő szöveg alkotását modellálja, annak szabályait és a nyelvi rendszerben meghatározó törvényszerűségeit kutatja” (Kocsány 2002: 146).

2 A kutatási hipotézisek

Az autentikus magyar szövegek és a célnyelvi magyar szövegek lexikai kohéziós mintázatának leírását célzó kutatásom során a következő hipotéziseket fogalmaztam meg:

1. A fordított szövegekre jellemző kohéziós eszközhasználat nem egyezik az autentikus szövegprodukcióna jellemző kohéziós eszközhasználattal.
2. A célnyelvi szövegek kohéziós eszközhasználatát a forrásnyelvi szöveg kohéziós eszközhasználata befolyásolja.
3. A célnyelvi szövegek kohéziós eszközhasználatának eltolódása azonos szövegtípus esetén nyelvpár-függő.
4. A célnyelvi szövegek kohéziós eszközhasználatának eltolódása fordítási tendencia.

3 Az értekezés felépítése

Az értekezés részei: *Bevezetés, Elméleti háttér, A kutatás leírása és az Összefoglalás*. Az elméleti háttér négy alfejezetre tagolódik, melyek (1) A fordított szöveg mint a tudományos vizsgálat tárgya, (2) A kutatás szövegnyelvészeti beágyazottsága, (3) A lexikai kohézió vizsgálata fordított szövegekben és (4) A korpusznyelvészeti módszer előnyei a fordításkutatásban. A kutatás leírásának főbb részei (1) A vizsgálat tárgya, (2) A kutatás főbb szakaszai, (3) A korpusz bemutatása, (4) Az elemzés bemutatása és (5) A Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian) eredményei. Az értekezés végén található a magyar nyelvű

Összefoglaló, a Táblázatok és ábrák jegyzéke, az Irodalomjegyzék és a Publikációs lista.

4 A kutatás módszerei

A fordított és nem fordított szövegek kvantitatív vizsgálatát csak az eredményközpontú leíró fordítástudományi módszer (Holmes 1972) teszi lehetővé, mivel csak a célnyelvi szöveg alapján kaphatunk pontos képet a fordító végső döntéseiről (vö. Hatim-Mason 1990: 3). A kutatott téma érdekes és megragadható feladatot kínál a számítógépes nyelvészet számára és a nagyobb szövegtulajdonságok különbségeinek kvantitatív módszerekkel mérhető relatív eloszlásának kimutatását teszi lehetővé.

A kontrasztív elemzés eredményeinek általánosíthatósága érdekében különböző szövegtípusú szövegeket vizsgáltam, amelyek a szövegprodukciónban részt vevők száma és egymáshoz való viszonya alapján mind monológikus, míg kommunikációs funkciójuk szerint elbeszélő, leíró és érvelő szövegek. A célnyelvi magyar szövegek esetében további fontos szempont volt, hogy eredetijeik különböző forrásnyelveket reprezentáljanak. A szövegtípusok kommunikációs színterei szerint a vizsgált szövegek közéleti (hivatalos), egyházi, tudományos és szépirodalmi szövegek, a hordozó közeg vagy csatorna szerint pedig mind átgondolt, tudatos felépítésű, igényes, választékos mondatfűzésű írásbeli megnyilatkozás. Az írásbeli szövegprodukción esetén a szövegalkotónak lehetősége nyílik a gondolatok körültekintő kifejtésére és a gondolatokat legpontosabban kifejező lexikai elemek és szerkezetek választékos, a kommunikációs szándéknak és a szövegtípusnak legmegfelelőbb alkalmazására.

Az elemzett és alkorpuszokba rendezett szövegeket kézi bevitel során láttam el header annotációval, ami az adott elektronikus fájl bibliográfiai leírását tartalmazza. A Magyar Lexikai Kohéziós Projektben a kohéziós elemek identifikációját a projekt céljainak megfelelően és érdekében a PPKE ITK két hallgatója, Orosz György és Laki László által összekapcsolt nyelvtechnológiai eszközök láncá segítségével (Laki 2011, Novák, Orosz, Indig 2011). A kiindulási Word-dokumentumokból XML-

dokumentumok készültek az OpenOffice program segítségével, majd az elemző azonosította a mondathatárokat és a tokeneket, a beépített Humor program (Prószéky 1995) elvégezte a morfológiai annotációt és egyértelműsítést; és végül az elemző az egyes bekezdéseken belül feltárta a főnevek más mondatokban történő ismétléses előfordulásait, illetve a magyar WordNet Program (Prószéky, Mihaltz 2008) élei mentén azonosította a mondathatárokon túlmutató szemantikai – szinonímia-, antonímia-, meronímia-, hiponímia- vagy domén reláción alapuló – kapcsolatokat.

5 A korpusz összetétele

A korpusz építésének legfontosabb kritériuma az autentikus és célnyelvi magyar szövegprodukciónak reprezentálása volt. Az alkorpuszokba gyűjtött szövegek esetében ezért nagyon elsődleges szempontnak tekintetem, hogy a szövegek keletkezésével kapcsolatos információk egyértelművé tegyék, hogy a szöveg eredetileg magyar nyelven keletkezett-e vagy fordítás eredményeként jött létre, illetve nyomon lehessen követni a szöveg keletkezésének eredeti nyelvét. A szövegek gyűjtésénél további fontos szempont volt, hogy a szövegek elektronikus fájlok formájában álljanak rendelkezésre. A dokumentumok gyűjtésének másik meghatározó kritériuma a nyilvános elérhetőség volt a szerzői jogi problémák kiküszöbölése érdekében.

Az autentikus magyar és a fordított magyar szövegeket a szövegtípusok kommunikációs színterei szerint közéleti, irodalmi, vallásos és tudományos alkorpuszokba rendeztem. A szövegek szövegépítő tulajdonságainak egybevetethetősége érdekében az autentikus és a fordított magyar szövegprodukciónak reprezentáló párhuzamos alkorpuszokba igyekeztem a szövegek keletkezésének körülményei szempontjából hasonló kritériumok alapján keletkezett szövegeket választani. A közéleti alkorpuszokba az Európai Unióval kapcsolatos szövegeket töltöttem le (2011. évi magyar EU-elnökség / az Európai Unió hivatalos weboldala), az szépirodalmi alkorpuszba regényeket és novellákat válogattam magyar, angol, német, francia és olasz nyelven. A vallásos szövegek esetében a hitbeli kérdésekkel

foglalkozó egyházi szövegek közül válogattam (püspöki konferenciák körlevelei – a Magyar Katolikus Egyház honlapja / pápai enciklikák - a Vatikán honlapja). Ezt a szövegtípust azért tartottam a kutatás szempontjából érdekesnek integrálni, mivel a katolikus egyház kétezer éves működésének folyamatos kísérőjelensége a Biblia szövegének, a pápai megnyilatkozásoknak és egyéb hivatalos egyházi dokumentumoknak több nyelvre történő lefordítása. A tudományos szövegek alkorpuszába történelmi, filozófiai, nyelvészeti és pszichológiai témájú tudományos szövegeket archiváltam (Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár / Educatio Társadalmi Szolgáltató Nonprofit Kft.).

6 A korpusz mérete

A korpusz jelenlegi méretét a digitális formátumban elérhető dokumentumok száma határozta meg. Meggyőződésem szerint a korpuszban archivált anyagok összetétele és mennyisége alapján a hipotézisek igazolására a jelen kutatás alapjául szolgáló korpuszállomány megnyugtató méretű és összetételű és a korpuszból kinyert eredmények megnyugtatóan alátámasztják a tézisek állításait.

A Magyar Lexikai Kohéziós Projekt teljes szövegállományának méretét az autentikus és a célnyelvi magyar alkorpuszokra vonatkozó szövegtípusonkénti adatok szemléltetik.

AUTENTIKUS	TOKENSZÁM
EU	249 315
IRODALOM	255 699
TUDOMÁNY	257 442
VALLÁSI	245 009
ÖSSZESEN	1 007 465

1. TÁBLÁZAT : Az autentikus magyar szövegeket tartalmazó alkorpusz összetétele

FORDÍTOTT	TOKENSZÁM
EU	249 420
IRODALOM	1 070 126
TUDOMÁNY	853 869
VALLÁSI	873 597
ÖSSZESEN	3 047 012

2. TÁBLÁZAT : A célnyelvi magyar szövegeket tartalmazó alkorpusz összetétele

Az autentikus magyar szövegeket tartalmazó alkorpusz közel 1.000.000 tokenszámú szöveget tartalmaz (1. TÁBLÁZAT), míg a célnyelvi magyar korpusz közel 3.050.000 tokenszámú szöveget tartalmaz (2. TÁBLÁZAT). Összességében az autentikus és a célnyelvi szövegprodukciónak a szövegépítő sajátosságait feltáró Magyar Lexikai Kohéziós Projekt a jelen szakaszban közel 4.000.000 tokenszámú szöveg elemzéseinek eredményein keresztül szemlélteti a lexikai kohéziós kapcsolatok realizációját.

AUTENTIKUS	TOKENSZÁM	FORDÍTOTT	TOKENSZÁM
KÖZÉLETI (EU)	249 315	KÖZÉLETI (EU)	249 420
		ANGOL	249 420
SZÉPIRODALMI	255 699	SZÉPIRODALMI	1 070 126
		ANGOL	278 909
		FRANCIA	257 731
		NÉMET	260 803
		OLASZ	272 683
TUDOMÁNYOS	257 442	TUDOMÁNYOS	853 869
		ANGOL	187 334
		FRANCIA	280 259
		NÉMET	257 079
		OLASZ	129 197
VALLÁSI	245 009	VALLÁSI	873 597
		ANGOL	169 363
		NÉMET	259 514
		LATIN	269 713
		GÖRG	175 007
ÖSSZESEN	1 007 465	ÖSSZESEN	3 047 012

3. TÁBLÁZAT : A Magyar Lexikai Kohéziós Projekt vizsgálati korpuszának összetétele

A 3. TÁBLÁZAT jól szemlélteti, hogy a vizsgált és elemzett szövegállomány mind a szövegtípusok, mind a nyelvpárok diverzitásának tekintetében, illetve a korpusz méretének tekintetében is egyedülálló és úttörő jellegű vállalkozás a hazai fordítástudományi kutatások között.

7 A kutatás újszerűsége

A jelen kutatás alapvető újszerűsége a kutatott témának, a szöveg lexikai-szemantikai tulajdonságának, a lexikai kohézió vizsgálatának nyelvtechnológiai alkalmazások segítségével történő feltárásán alapszik. A számítógépes nyelvészeti eszközök alkalmazásának köszönhetően a következő területeken lehet a vizsgálat újszerűségét tetten érni:

1. nyelvtechnológiai eszközök alkalmazása az autentikus és a fordított magyar szövegek lexikai kohéziós relációinak feltárásában,
2. a nyelvtechnológiai eszközök segítségével megvalósított nemzetközi kohéziókutatás gyakorlatától eltérően a szövegkohézió vizsgálata nem a kollokációk és együttes előfordulások gépi lekérdezésére támaszkodik, hanem a mondathatárokon túlmutató szemantikai kapcsolatok feltárásán nyugszik,
3. a WordNet formalizmus felhasználása a mondathatárokon túlnyúló lexikai kohéziós kapcsolatok automatikus gépi eszközökkel történő feltárásában,
4. a hagyományos kohéziókutatás gyakorlatából ismert korpuszoknál lényegesen kisebb időráfordítás mellett lényegesen nagyobb szövegtörzsen végzett vizsgálat,
5. két párhuzamos, a keletkezésük körülményeiben eltérő, korpuzon végzett összehasonlító vizsgálat,
6. az autentikus és a fordított magyar szövegek kohéziós mintázatának leírása a HuWordNet géppel olvasható lexikális adatbázis ontológiai kategóriáinak segítségével,
7. a HuWordNet géppel olvasható lexikális adatbázis ontológiai kategóriáinak alkalmazása miatt a kohéziós kapcsolatok bejelölésének következetessége a viszonylag nagyméretű szövegtörzsen ellenére is.

8 Az automatikus gépi eszközökkel azonosított lexikai kohéziós kapcsolatok a Magyar Lexikai Kohéziós Projektben

A közel 4.000.000 szövegszavas korpusz szövegeinek gépi elemzése és a korpuszban a lexikai kohéziós kapcsolatok automatikus azonosítása a vizsgált autentikus és célnyelvi magyar közéleti, szépirodalmi, tudományos és vallásos szövegek esetében eddig példa nélküli mennyiségben teszi lehetővé az egyes szövegtípusokra jellemző lexikai kohéziós kapcsolatok konkrét nyelvi realizációjának vizsgálatát. A kohéziós kötések típusai szerinti nyelvi példák lekérdezése az annotált szövegegységekben hozzájárulhat a szövegek szövegszemantikai szerveződéséről alkotott elképzelésünk kiteljesítéséhez és pontosításához. A Magyar Lexikai Kohéziós Projekt során azonosított lexikai kohéziós kapcsolatok kiinduló pontjai lehetnek a jelen kutatás folytatásaként egy a kohéziós kötések szövegtípusonkénti nyelvi realizációját feltáró szövegnyelvészeti vizsgálatnak.

AUTENTIKUS	TOKENEK	ÖSSZES LEXIKAI KOHÉZIÓS KAPCSOLAT
KÖZÉLETI (EU)	249 315	27 142
IRODALMI	255 699	30 998
TUDOMÁNYOS	257 442	58 414
VALLÁSI	245 009	53 348
ÖSSZES	1 007 465	169 902

4. TÁBLÁZAT : Az autentikus magyar szövegeket tartalmazó alkorpuszban feltárt lexikai kohéziós kapcsolatok száma

FORDÍTOTT	TOKENEK	ÖSSZES LEXIKAI KOHÉZIÓS KAPCSOLAT
KÖZÉLETI (EU)	249 420	26 107
IRODALMI	1 070 126	111 305
TUDOMÁNYOS	853 869	181 762
VALLÁSI	873 597	172 197
ÖSSZES	3 047 012	491 371

5. TÁBLÁZAT: A fordított magyar szövegeket tartalmazó alkorpuszban feltárt lexikai kohéziós kapcsolatok száma

A 4. TÁBLÁZAT és 5. TÁBLÁZAT az autentikus és a célnyelvi alkorpuszok esetében tartalmazzák a szövegtípusonkénti (közéleti, szépirodalmi, tudományos és vallási) tokenszámot (TOKENEK) és a szövegtípusban előforduló összes kohéziós kapcsolat számát (ÖSSZES LEXIKAI KOHÉZIÓS KAPCSOLAT).

Egy szöveg minta segítségével szeretném a következőkben érzékeltetni azt a folyamatot, ahogy a nyers internetről vagy egyéb forrásokból letöltött szövegeket az

elemző feldolgozta, és ahogy a kohéziós kapcsolatokat azonosította. Természetesen a jelen értekezés nem fogja tükrözni az elemző kimeneti dokumentumaként kapott színes annotációkat tartalmazó xml fájlok szemléletes struktúráját. A színes annotáció megkönnyíti az elemzett szövegekben az eredeti szövegrészek és a taggek között megjelenő metainformációk elkülönítését, illetve a nyelvészeti elemzés számára releváns adatok értelmezését.

Példaként szolgál erre *Az unió kapujában* (Molnár: 2000) című közéleti szövegtípusba tartozó szöveg eredeti Word-dokumentumának szövegrészlete (1. *ÁBRA*).

A második lépésben azonosított kohéziós kapcsolatokat szemléltetésére a kapcsolódásban szerepet játszó lexikai elemek félkövér betűtípussal jelennek meg. Az elemzés első lépésében történik a szöveg szegmentálása és morfológiai elemzése, amit az 1. *ÁBRA* által bemutatott bekezdés 2. mondatának elemzésén keresztül szemléltet a 2. *ÁBRA*.

- **Erre a kérdésre nem lehet fekete-fehér, igen vagy nem a válasz.** Az illúziók elkerülése végett fontos tudnunk, hogy az **alvilág** mindig törekszik **érdekei** szerinti befolyás szerzésére. A **politikába**, az államigazgatásba, a **közpénzek felhasználásának** eldöntésébe, a döntéshozásba próbálja betenni a lábát. A **politika** meg a **közigazgatás** pedig jól felfogott **érdeke** szerint védekezik. Tehát **ez is örök háború**. Egy maffiának szüksége van arra, ha külföldi, hogy útlevelet, betelepelési, beutazási, munkavállalási engedélyt kapjon. Ha valamit csempész, legyen fegyver-, kábítószer- vagy éppen embercsempész, akkor arra törekszik, hogy a határon beépüljön a hatóságba, de szüksége van arra is, hogy a **közpénzek** elosztásakor gazdasági előnyre tehessen szert, hozzájusson **közpénzekhez** is. Az **alvilág** részéről rendszeresen érzékelhető, kitapintható **ez** a támadás. Hogy hol állunk **ebben a küzdelemben**, **ez** az igazi kérdés. Vannak olyan pozíciók, például a **közigazgatásban**, amit aktívan és nem eredményte-lenül támadnak a **maffiák**. Nagyon sokszor előfordult a **közpénzek felhasználásakor**, az állami megrendelésekkor, hogy valószínűsíthető volt: a **maffiák** vitték el **ezeket a pénzeket**.

1. ÁBRA : Részlet Az unió kapujában című dokumentumból

A következő lépésben a bekezdésen belüli lexikai kohéziós kapcsolatokat azonosítja a program. Az értekezés fekete-fehér és lineáris ábrázolás módja korántsem tükrözi az elemző kimeneti dokumentumaként kapott színes annotációkat tartalmazó xml fájlok szemléletes struktúráját. A 3. ÁBRA lineárisan szemlélteti a példaként hozott közéleti szövegtípushoz tartozó szövegrészletben realizálódó kohéziós kapcsolatokat.

```
<sentence><origin>A politikába , az államigazgatásba , a közpénzek felhasználásának
eldöntésébe , a döntéshozásba próbálja betenni a lábát ./origin>
<token stem="a" tag="[DET]">A</token>
<token stem="politika" tag="[FN][ILL]">politikába</token>
<token stem="," tag="[PUNCT]">,</token>
<token stem="az" tag="[DET]">az</token>
<token stem="államigazgatás" tag="[FN][ILL]">államigazgatásba</token>
<token stem="," tag="[PUNCT]">,</token>
<token stem="a" tag="[DET]">a</token>
<token stem="közpénz" tag="[FN][PL][NOM]">közpénzek</token>
<token stem="felhasználás" tag="[FN][PSe3][DAT]">felhasználásának</token>
<token stem="eldöntés" tag="[FN][PSe3][ILL]">eldöntésébe</token>
<token stem="," tag="[PUNCT]">,</token>
<token stem="a" tag="[DET]">a</token>
<token stem="döntéshozás" tag="[FN][ILL]">döntéshozásba</token>
<token stem="próbál" tag="[IGE][Te3]">próbálja</token>
<token stem="betesz" tag="[IGE][INF]">betenni</token>
<token stem="a" tag="[DET]">a</token>
<token stem="láb" tag="[FN][PSe3][ACC]">lábát</token>
<token stem="." tag="[PUNCT]">.</token></sentence>
```

2. ÁBRA : Példa az előző bekezdés egyik mondatának morfológiai annotálására

Jól látható, hogy a program a kohézió kutatásból ismert és a lexikai kohéziós vizsgálatoknál alkalmazott módszer szerint először a bekezdés első lexikai egységének az összes többi tagmondatban megvalósuló ismétléses kapcsolatát listázza ki a kapcsolat fajtájának megjelölésével. Az első lexikai egység kohéziós

kapcsolatainak megtalálása után az elemző tovább halad és megvizsgálja a második, harmadik, stb. lexikai egység ismétléses kapcsolatait, míg elérkezik az utolsó tagmondat utolsó lexikai eleméhez. A megtalált kohéziós kapcsolatokat minden esetben a kapcsolat típusa szerint is azonosítja az elemző.

Egyes lexikai elemek esetében az azonos elem többféle kohéziós kapcsolatát is azonosítja az adott bekezdésen belül az elemző. Így például a *kérdésre* lexikai elemnek megtaláljuk az ismétléses (*kérdés*) és az antonima kapcsolatát (*válasz*) is a szövegrészletben, vagy a *közpénzek* lexikai elemnek megtaláljuk az ismétléses (*közpénzekhez*) és az hiponima kapcsolatát (*pénzeket*) is.

"repetition" :	"Erre"	>	"ez"
"repetition" :	"Erre"	>	"ebben"
"repetition" :	"Erre"	>	"ezeket"
"antonym" :	"kérdésre"	>	"válasz"
"repetition" :	"kérdésre"	>	"kérdés"
"antonym" :	"válasz"	>	"kérdés"
"repetition" :	"alvilág"	>	"alvilág"
"repetition" :	"érdekei"	>	"érdeke"
"repetition" :	"politikába"	>	"politika"
"repetition" :	"közpénzek"	>	"közpénzek"
"repetition" :	"közpénzek"	>	"közpénzekhez"
"repetition" :	"közpénzek"	>	"közpénzek"
"hyponym" :	"közpénzek"	>	"pénzeket"
"repetition" :	"felhasználásának"	>	"felhasználásakor"
"repetition" :	"közigazgatás"	>	"közigazgatásban"
"repetition" :	"ez"	>	"ez"
"repetition" :	"ez"	>	"ebben"
"repetition" :	"ez"	>	"ezeket"
"synonyme" :	"háború"	>	"küzdelemben"
"repetition" :	"maffiának"	>	"maffiák"
"repetition" :	"maffiának"	>	"maffiák"
"repetition" :	"szüksége"	>	"szüksége"
"repetition" :	"arra"	>	"arra"
"repetition" :	"közpénzek"	>	"közpénzekhez"
"repetition" :	"közpénzek"	>	"közpénzek"
"hyponym" :	"közpénzek"	>	"pénzeket"
"repetition" :	"közpénzekhez"	>	"közpénzek"
"hyponym" :	"közpénzekhez"	>	"pénzeket"
"repetition" :	"ez"	>	"ebben"
"repetition" :	"ez"	>	"ez"
"repetition" :	"ez"	>	"ezeket"
"repetition" :	"ebben"	>	"ez"
"repetition" :	"ebben"	>	"ezeket"
"repetition" :	"ez"	>	"ezeket"
"repetition" :	"maffiák"	>	"maffiák"
"hyponym" :	"közpénzek"	>	"pénzeket"

3. ÁBRA : Az 1. ÁBRA szövegrészletében található kohéziós kapcsolatok

9 A kutatás eredményei

Az Orosz-Laki-féle elemző (Laki 2011, Novák, Orosz, Indig 2011) segítségével feltártam az autentikus és a célnyelvi szövegprodukciónak reprezentáló, nyelvpár specifikus alkorpuszokba rendezett, egyes dokumentumokra jellemző lexikai kohéziós eszközhasználatot. Az eredményeket az egyes szövegtípusok és az egyes forrásnyelvek függvényében külön-külön összesítettem (ld. III. 5.2) és az eltolódásjelenségek kvalitatív módon történő szemléltetését és bizonyítását a négy vizsgált szövegtípus esetében forrásnyelvenkénti bontásban leírtam.

Az értekezés téziseinek és bizonyításainak összegzése előtt szeretném még egyszer hangsúlyozottan kiemelni, hogy az értekezésben bemutatott módszer – amelyet a kohéziós kapcsolatok feltárására és a magyar autentikus és célnyelvi szövegek szövegépítő sajátosságainak leírására alkalmaztam – nem kacérkodott sohasem azzal az elképzeléssel, hogy egy nyelv szövegeire jellemző kohéziós mintázat pusztán kvantitatív adatokkal pontosan megragadható. Elképzelésem szerint egy adott nyelv kohéziós eszközhasználata nem hasonlítható azon mérnöki létesítményekhez, amelyeknél egy elemet csak egy vele teljesen kompatibilis elemmel lehet kiváltani. Ahhoz, hogy a szövegalkotó egy adott szövegtípusban, egy adott nyelven, egy adott gondolatmenet kohézívítását biztosítsa több lehetséges megoldás közül is választhat a nyelvi és kognitív kompetenciáknak köszönhetően. A szövegalkotó ezen választásait természetesen befolyásolja maga a szövegalkotó személyisége, a kommunikációs helyzet és cél, a szövegtípus, a szöveg keletkezésének körülményei, a szöveg keletkezésének nyelve, stb., ahogy ezt a II. 1.6. fejezet bővebben tárgyalja. A jelen kutatás a lexikai kohéziót befolyásoló összes tényező közül elsősorban a szöveg kohéziós mintázatának kialakulása és a szöveg keletkezését befolyásoló nyelvi tényezők közötti összefüggések megragadására törekedett.

Az értekezés során bemutatott eljárás a kohéziós elemek előfordulásának pontos adatolására soha nem törekedett. Feltételezésem szerint a jelen kutatás a vizsgált szöveganyag nagyságának köszönhetően arra bizonyosan alkalmas, hogy segítsen megragadni azokat a különbségeket, amelyek az adott szövegtípuson belüli, ám a keletkezésük körülményét tekintve éppen a különböző forrásnyelvi szövegek

alapján készített fordításokat különítik el egymástól. A jelen vizsgálat elsődleges célkitűzése tehát az volt, hogy megfelelő szöveg minta alapján jelezze azokat az egyes forrásnyelvi hatásoknak tulajdonítható eltolódás jelenségeket, amelyek a célnyelvi magyar szövegek kohéziós eszközhasználatában a fordítás következtében, az egyes forrásnyelvek függvényében bekövetkeznek.

Az a tény, hogy a magyar nyelven keletkezett szövegek lexikai kohéziós kapcsolatainak bejelölése a WordNet technológiának köszönhetően automatizálhatóvá vált, lehetővé tette nagyobb, szövegtípusonként több forrásnyelvű alkorpuszt tartalmazó korpusz vizsgálatát. A kohéziós kapcsolatok gépi azonosítása és négy szövegtípuson belül különböző nyelvpárokra jellemző disztribúciójának leírása tette lehetővé az autentikus magyar szövegekre és a célnyelvi magyar szövegekre jellemző lexikai kohéziós mintázatok egybevetését, valamint a szövegtípusok és a forrásnyelvek függvényében jellemző kohéziós mintázatok leírását. Az értekezés III. 5. fejezetében bemutatott elemzési eredmények alapján lehetséges tehát a célnyelvi magyar szövegek keletkezését befolyásoló forrásnyelvi hatások megragadása a szövegek lexikai kohéziós mintázatának tekintetében és ezáltal az értekezés bevezetőjében megfogalmazott hipotézisek (ld. I. 5.2) kvalitatív érvelésen túli (ld. II. fejezet) kvantitatív adatokkal történő alátámasztása.

1. TÉZIS

Az adatok bizonyítják, hogy a fordított szövegekre jellemző kohéziós eszközhasználat nem egyezik az autentikus szövegprodukciónak jellemző kohéziós eszközhasználattal.

Az értekezés III. 5.2 fejezetében bemutattam, hogy az autentikus magyar és a célnyelvi magyar szövegeknél a vizsgált négy szövegtípus esetében az autentikus magyar szövegprodukciónak jellemző kohéziós eszközhasználattól eltérő kohéziós mintázat jellemző a célnyelvi magyar szövegekre minden forrásnyelv – az angol, francia, német, olasz, latin és görög – viszonylatában. A tézis állítását a négy szövegtípus és a hat forrásnyelv viszonylatában összesen tizennégy féle relációban bizonyítottam.

2. TÉZIS

Az adatok bizonyítják, hogy a célnyelvi szövegek kohéziós eszközhasználatát a forrásnyelvi szöveg kohéziós eszközhasználata befolyásolja.

Az értekezés III. 5.2 fejezetében bemutatott eredmények rámutattak arra a tényre, hogy a kohéziós eszközök használatának gyakoriságában és eloszlásában bekövetkező változások a forrásnyelvek hatására különböznek egymástól. A célnyelvi szövegek lexikai kohéziós mintázata az egyes forrásnyelvek által előnyben részesített lexikai kohéziós eszközhasználat befolyásának hatására tér el az adott célnyelv esetében az autentikus szövegprodukcióna jellemző lexikai kohéziós mintázattól. A tézis állítását hat forrásnyelv viszonylatában bizonyítottam.

3. TÉZIS

Az adatok bizonyítják, hogy a célnyelvi szövegek kohéziós eszközhasználatának eltolódása azonos szövegtípus esetén nyelvpár-függő.

Az értekezés III. 5.2 fejezetében bemutatott szövegtípusonkénti összevetések arra a tényre is rámutattak, hogy a kohéziós eszközök használatának gyakoriságában és eloszlásában bekövetkező változások adott szövegtípuson belül a forrásnyelvek függvényében különböznek egymástól. A vizsgálat nem hozott az egyik szövegtípus esetében sem olyan eredményt, ahol a különböző forrásnyelvekről fordított célnyelvi szövegek esetében a kohéziós eszközök disztribúciójában bekövetkezett változások egymással megegyező eltolódásbeli tendenciákat mutattak volna, mivel a különböző forrásnyelvekről fordított célnyelvi szövegek esetében a kohéziós eszközök disztribúciója az azonos szövegtípus viszonylatában forrásnyelvenként mindig különböző. A tézis állítását három szövegtípus és a hat forrásnyelv viszonylatában összesen kilenc féle relációban bizonyítottam.

4. TÉZIS

Az adatok bizonyítják, hogy a célnyelvi szövegek kohéziós eszközhasználatának eltolódása fordítási tendencia.

A vizsgálat a kutatásban szereplő mind a négy szövegtípus és mind a hat forrásnyelv relációjában, tehát összesen tizennégy vizsgált kategóriában bizonyította a fordítás és a forrásnyelv hatására bekövetkező eltolódást a célnyelvi szövegek kohéziós eszközhasználatának gyakoriságában és disztribúciójában. A kutatás másrészt a négy szövegtípus és a hat forrásnyelv relációjában vizsgált tizennégy kategória egyikében sem talált a fordítás és a forrásnyelv kohéziós eszközhasználati preferenciájának hatásától mentes olyan célnyelvi alkorpuszt, amely esetében a kohéziós eszközhasználat gyakorisága és disztribúciója megegyezne az autentikus szövegalkotásra jellemző kohéziós eszközhasználat szövegszervező tulajdonságaival.

5. TÉZIS

Megmutattam, hogy a szövegszerveződés lexikai-szemantika szerkezetére jellemző lexikai kohéziós mintázat feltárása automatizálható a WordNet adatbázis synset relációi alapján történő kohéziós kapcsolatok bejelölésével.

A több mint 4.000.000 szövegszavas korpusz elemzése alapján egyértelműen bizonyítottam, hogy a WordNet adatbázis relációi alapján a szövegekben feltárt kohéziós kapcsolatok hozzájárulhatnak az adott szövegtípusra jellemző kohéziós mintázat leírásához.

10 A tézisekhez felhasznált irodalom

- Hatim, B., Mason, I. 1990. *Discourse and the Translator*. London: Longman.
- Holmes, J. 1972. The Name and Nature of Translation Studies. In: Holmes, J. 1988. *Translated!*. Amsterdam: Rodopi. 67–80.
- Károly K. 2007. *Szövegtan és fordítás*. Budapest: Akadémia Kiadó.
- Kocsány P. 2002. Szöveg, szövegtípus, jelentés: a mondás mint szövegtípus. Budapest: Akadémia Kiadó.
- Laki L. J. 2011. Statisztikai gépi fordítási módszereken alapuló egynyelvű szövegelemző rendszer és szótövesítő. In: Tanács A., Vincze V. (szerk.) 2011. *VIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi tudományegyetem Informatikai Tanszékcsoport. 12–23.
- Molnár P. (szerk.) 2000. *Az unió kapujában*. Budapest: Mundus Magyar Egyetemi Kiadó. Elérhető: mek.oszk.hu/01900/01989
- Novák A., Orosz Gy., Indig B. 2011. Javában Taggelünk. In: Tanács A., Vincze V. (szerk.) 2011. *VIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi tudományegyetem Informatikai Tanszékcsoport. 336–340.
- Péché O. 2009a. Kimutatható-e eltolódás a fordítások kohéziós eszközhasználatában? In: Nádor O. (szerk.) 2009. *A magyar mint európai és világnyelv. A XVIII. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai*. Budapest: Balassi Intézet. 635–640.
- Péché O. 2009b. Analysing lexical cohesion in translation research using corpus linguistic methods. In: Váradí T. (szerk.) 2009. *Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból*. Budapest: MTA Nyelvtudományi Intézet. 103–116.
- Péché O. 2010. Szempontok a fordított szövegek sajátosságainak leírásához. In: Károly K., Fóris Á. (szerk.) *Nyelvek találkozásá a fordításban. Doktori kutatások Klaudy Kinga tiszteletére*. Budapest: ELTE Eötvös Kiadó. 73–84.
- Petőfi S. J. 2004. *A szöveg mint komplex jel*. Budapest: Akadémia Kiadó.
- Prószéký G. 1995. Humor: a Morphological System for Corpus Analysis. In: Rettig, H., Pajzs J., Kiss G. (eds.) *Language Resources for Language Technology*. Tihany: TELRI 149–158.
- Prószéký G., Miháltz M. 2008. Magyar WordNet: az első magyar lexikális szemantikai adatbázis. *Magyar Terminológia*. Vol. 1. No. 1. 43–57.

11 Az értekezés témájához kapcsolódó saját publikációk

11.1 Tanulmányok

- Pécs O. 2007. A lexikai kohézió vizsgálata a fordításkutatásban – új távlatok a korpusznyelvészeti módszernek köszönhetően. *Fordítástudomány* IX. évf. 1. szám 79–96.
- Pécs O. 2007. Die Erforschung der lexikalischen Kohäsion mit korpuslinguistischen Mitteln. Schmitt, Peter A., Jüngst, Heike E. (Hrsg.): *Translationsqualität. Leipziger Studien zur angewandten Linguistik und Translatologie Vol. 5*. Peter Lang Publishing Group.
- Pécs O. 2007. A lexikai kohézió vizsgálata a fordításkutatásban egy korpusznyelvészeti eszközzel végzett kutatás tükrében. In: *I. Alkalmazott Nyelvészeti Doktorandusz Konferencia kötete* 114–127.
Elérhető: <http://www.nytud.hu/alknyelvdok/proceedings07/>
- Pécs O. 2007. Korpusznyelvészeti eszköz alkalmazása a lexikai kohézió vizsgálatánál. In: Gecső T., Sárdi Cs. (szerk.) *Nyelvelmélet – Nyelvhasználat*. Budapest: Tinta Könyvkiadó. 222–230.
- Pécs O. 2009. Miért érdekes a lexikai kohéziós eszközhasználat vizsgálata a fordításkutatás számára? In: Menyhárt J., Presinszky K. (szerk.) *Fordítás – kétnyelvűség (Nyelvészeti tanulmányok a fordítás elméletéről és gyakorlatáról)* Nyitra: Konstantin Filozófus Egyetem Közép-európai Tanulmányok Kara. 27–38.
- Pécs O. 2009. Analysing lexical cohesion in translation research using corpus linguistic methods. In: Váradai T. (szerk.) *Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból*. Budapest: MTA Nyelvtudományi Intézet. 103–116.
- Pécs O. 2009. Kimutatható-e eltolódás a fordítások kohéziós eszközhasználatában? Nádor O. (szerk.) 2009. *A magyar mint európai és világnyelv. A XVIII. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai*. Budapest: Balassi Intézet. 635–640.
- Pécs O. 2010. Szempontok a fordított szövegek sajátosságainak leírásához. In: Károly K., Fóris Á. (szerk.) *Nyelvek találkozása a fordításban*. Budapest: Eötvös Kiadó. 73–85.
- Pécs O. 2010. Korpusznyelvészeti módszer a fordításkutatásban. In: Zimányi Á. (szerk.) *A tudomány nyelve – A nyelv tudománya. Alkalmazott nyelvészeti kutatások a magyar nyelv évében*. Eger: Eszterházy Károly Főiskola. 208–216.
- Pécs O. 2011. Sprachliche Eigenschaften von Übersetzungen. Korpusbasierte Untersuchung für das Sprachenpaar Deutsch-Ungarisch. In: Schmitt, P. A., Herold, S., Weilandt, A. (Hrsg.) *Translationsforschung. Tagungsberichte der*

LICTRA IX. Leipzig International Conference on Translation & Interpretation Studies 19.-21.5.2010. Teil 1 & 2. . Leipziger Studien zur angewandten Linguistik und Translatologie Vol. 10. Peter Lang Publishing Group.

Seidl-Péché O. 2011. Warum wird die Verwendung lexikalischer Kohäsionselemente bei der translatorischen Tätigkeit beeinflusst? Neues zur Beschreibung sprachlicher Eigenschaften von Übersetzungen. In: Zybatow, L., Petrova, A., Ustaszewski, M. (eds.) 2012. *Translationswissenschaft interdisziplinär: Fragen der Theorie und der Didaktik*. Frankfurt am Main u.a.: Peter Lang. (megjelenés alatt)

11.2 Előadások

Péché O. 2006. A lexikai kohézió vizsgálata korpusznyelvészeti eszközökkel. Elhangzott: II. Fordítástudomány PhD Konferencia. Budapest: ELTE.

Péché O. 2006. Erforschung der lexikalischen Kohäsion mit korpuslinguistischen Mitteln. Elhangzott: VIII. Leipziger Internationale Konferenz zu Grundfragen der Translatologie. Universität Leipzig: Institut für Angewandte Linguistik und Translatologie.

Péché O. 2006. A lexikai kohézió vizsgálata korpusznyelvészeti módszerrel. Elhangzott: „Nyelvelmélet és nyelvhasználat” Budapest: ELTE-KJF.

Péché O. 2006. A lexikai kohézió vizsgálata a fordításkutatásban. Egy korpusznyelvészeti eszköz alkalmazásának előnyei. Elhangzott: Nyelvészeti Doktoranduszok 10. Országos Konferenciája. Szeged: SZTE.

Péché O. 2007. A lexikai kohézió vizsgálata a fordításkutatásban egy korpusznyelvészeti eszközzel végzett kutatás tükrében. Elhangzott: I. Alkalmazott Nyelvészeti Doktorandusz Konferencia. MTA Alkalmazott Nyelvészeti Munkabizottsága. Budapest: MTA Nyelvtudományi Intézet.

Péché O. 2008. Hozzájárulás a szövegnyelvészet interkulturális megközelítéséhez. A lexikai kohézió vizsgálata fordított szövegeknél. Elhangzott: Nyelv-Kultúra-Tudományok Konferencia. Pápa: Pápai Református Gimnázium.

Péché O. 2008. Kimutatható-e eltolódás a fordítások kohéziós eszközhasználatában? Elhangzott: XVIII. Magyar Alkalmazott Nyelvészeti Kongresszus. Budapest: Balassi Intézet.

Péché O. 2009. Korpusznyelvészeti módszer a fordításkutatásban. Elhangzott: „A tudomány nyelve – a nyelv tudománya” XIX. Magyar Alkalmazott Nyelvészeti Kongresszus. Eger: EKF.

Péché O. 2009. Szövegfeldolgozó eszközök és alkalmazásaik a fordításkutatásban. Elhangzott: „Új módszerek az alkalmazott nyelvészeti kutatásban” Budapest: ELTE-KJF.

- Pécs O. 2010. Sprachliche Eigenschaften von Übersetzungen. Eine korpusbasierte Untersuchung zum Sprachenpaar Deutsch–Ungarisch. Elhangzott: IX. Leipziger Internationale Konferenz zu Grundfragen der Translatologie. Universität Leipzig, Institut für Angewandte Linguistik und Translatologie.
- Pécs O. 2010. Szempontok a fordított szövegek sajátosságainak leírásához. Elhangzott: Nyelvek találkozás a fordításban. Doktori kutatások Klauy Kinga tiszteletére. Budapest: ELTE.
- Seidl-Pécs O. 2011. Warum wird die Verwendung lexikalischer Kohäsionselemente bei der translatorischen Tätigkeit beeinflusst? Elhangzott: 1.TRANSLATA-Konferenz. Institut für Translationswissenschaft der Leopold-Franzens Universität Innsbruck.